EXPERTISE, UNDERUSE, AND OVERUSE IN HEALTHCARE[*]

Amitabh Chandra
Harvard and the NBER

Douglas O. Staiger
Dartmouth and the NBER

Version: March 27, 2011

Abstract

The reasons for variation in treatment rates across hospitals serving similar patient populations are not well understood. Differences in the use of a treatment across hospitals may be due to greater benefits of treatment in some hospitals (expertise), withholding of beneficial treatment in some hospitals (underuse), or providing harmful treatment in other hospitals (overuse). We develop an empirical model that can distinguish between these explanations, based on a behavioral model in which hospitals choose to treat patients if the benefit from treatment exceeds a hospital-specific threshold. Expertise, underuse, and overuse are identified based on differences across hospitals in both their treatment rates and the treatment effect on patient survival. Using data on heart attack treatments, we find that expertise varies considerably across hospitals, but a substantial amount of variation in treatment is due to overuse.

[*] This research was funded by the National Institute of Aging (NIA) P01 AG19783-02. We thank Jonathan Skinner and seminar participants at Dartmouth 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.

I. Introduction

An enormous body of literature in economics and medicine has documented variations in the use of medical and surgical treatments across hospitals and regions with observationally similar patient populations. Generally, higher treatment rates are not found to be associated with improved satisfaction, survival, or other patient outcomes. These facts have often been interpreted as evidence of overuse: high use hospitals or regions are choosing to treat marginal patients who receive no benefit from the treatment. However, a fundamental problem with this explanation is that it implies that higher treatment rates should be associated with lower benefits to treatment for the average patient getting treatment, while a number of studies have found the opposite (Beck et al., 2003; Pilote, McClellan, et al., 2003; Chandra and Staiger, 2007).

In this paper, we develop a simple empirical model that can distinguish between alternative explanations for variations in treatment rates, based on a behavioral model in which hospitals choose to treat patients if the benefit from treatment exceeds a hospital-specific threshold. In our model, differences in the use of a treatment across hospitals may be due to greater benefits of treatment in some hospitals (expertise), withholding of beneficial treatment in some hospitals (underuse), or providing harmful treatment in other hospitals (overuse). Expertise, underuse, and overuse are identified based on differences across hospitals in both their treatment rates and the treatment effect on patient survival. Using data on heart attack treatments, we find that expertise varies considerably across hospitals, but a substantial amount of variation in treatment is due to overuse.

In Section II we develop the theoretical model underlying our analysis. Section III discusses the etiology of heart attacks and their treatments, and introduces data from the Cooperative Cardiovascular Project (CCP). In Section IV we detail our estimation strategy, paying particular attention to how the theoretical model developed in Section II will be evaluated using the CCP data. Section V presents results and Section VI concludes.

II. Theory

A simple model of patient treatment choice guides our empirical work. We assume that treatment is provided to each patient whenever the expected benefit from the treatment exceeds a minimal threshold. Thus, 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. Within this framework, there are two ways in which a patient's hospital could affect treatment choice. First, the expected benefit of treatment for a given patient may vary across hospitals, reflecting each hospital's idiosyncratic level of expertise in providing the treatment. Second, the minimum threshold for receiving care may vary across hospitals, as determined by local

incentives and norms at each hospital. 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.

Setup of the Model

Let B_{ih} represent the expected benefit from treatment for patient i at hospital h. Benefit is the *gain* or *improvement* in survival relative to not receiving a treatment, not the level of survival. We focus on the health benefits of the treatment, which 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). In principal, the benefit could also incorporate the expected impact of the treatment on the patient's medical cost, and capture the health benefits net of costs. However, in our application we focus on survival for simplicity.

Suppose that the expected benefit from treatment depends on the hospital's expertise (α_h), observable patient characteristics (X_{ih}) such as age, medical history, and lab results, and other factors that are known to the medical care provider when making the treatment decision but unobserved by the econometrician (ε_{ih}):

(1) $$B_{ih} = \alpha_h + X_{ih}\beta + \varepsilon_{ih}$$

Each patient receives treatment if the expected benefit from treatment exceeds a minimal threshold (τ_h), where the threshold varies across hospitals due to local incentives or norms. This corresponds to a simple Roy model of treatment allocation, where a patient receives treatment if the gain from the treatment exceeds a threshold. Assuming that B_{ih} captures the total net benefit to the patient of providing treatment, then the optimal decision from the patient's perspective would let $\alpha_h = 0$, and provide treatment whenever the benefits to the patient exceed zero. There is underuse if $\alpha_h > 0$, since patients with positive benefits are under the threshold and do not receive treatment. There is overuse if $\alpha_h < 0$, since patients with negative benefits (who would do better without treatment) are above the threshold and receive treatment.

This decision process implies a very simple tobit structure that determines both the probability of treatment as well as the expected benefit conditional on being treated (the treatment-on-the-treated parameter). The probability of receiving treatment is just the probability that expected benefits exceed the minimum threshold:

(2) $$\Pr\left(Treatment_{ih} = 1\right) = \Pr\left(B_{ih} > \tau_h\right) = \Pr\left(-\varepsilon_{ih} < I_{ih}\right),$$

$$where\ I_{ih} = X_{ih}\beta + (\alpha_h - \tau_h)$$

Equation (2) highlights that differences in treatment rates across hospitals, holding all else equal, can be due to either differences in hospital expertise (α_h) or differences in the treatment threshold (τ_h) reflecting over or underuse. A hospital may be more likely to provide treatment because of greater expected benefits of treatment $(\alpha_h > 0)$ or because of a lower benefit threshold for providing care $(\tau_g < 0)$. Conversely, even if treatment rates were the same across hospitals, there could still be overuse or underuse if, say, hospital's with greater expected benefits of treatment set a correspondingly higher threshold for providing care $(\alpha_h = \tau_h > 0)$. Thus, because variation in treatment rates across hospitals masks variation in hospital expertise and hospital treatment thresholds, such variation cannot by itself say anything about overuse or underuse. In other words, while the conventional wisdom is to interpret variation in hospital treatment rates or spending as saying something about differences in their treatment threshold (as in flat of the curve models of provider behavior), this inference ignores the role of expertise in also explaining that variation.

However, overuse and underuse can be identified separately from hospital expertise if information on the treatment effect among the treated population is available. The treatment-on-the-treated parameter is defined as:

$$(3) \qquad E(B_{ih} \mid Treatment_{ih} = 1) = E(B_{ih} \mid -\varepsilon_{ih} < I_{ih}) = X_{ih}\beta + \alpha_h + E(\varepsilon_{ih} \mid -\varepsilon_{ih} < I_{ih})$$

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

$$(4) \qquad E(B_{ih} \mid Treatment_{ih} = 1) = \tau_h + g(I_{ih}),$$

$$where\ g(I_{ih}) = I_{ih} + E(\varepsilon_{ih} \mid -\varepsilon_{ih} < I_{ih})$$

Equation (4) states that in the absence of any difference across hospitals in the minimum threshold to receive care, two patients receiving treatment who have the same propensity to get the treatment (same *I*) will have the same expected benefit from the treatment. This relies on the assumption that *g(I)* depends only on the index, i.e. we must assume a single-index selectivity model, so that the truncated mean of the error in equation (4) depends only on the truncation point (*I*). This would not be the case if the distribution of the unobservable factors determining treatment (ε) differed across hospitals, which in turn would happen if some hospitals were better at measuring these unobservable factors. Examining equation (4) makes it clear that hospitals with larger unobservables will have larger treatment on the treated effects (which will look like having a higher treatment threshold) because of the higher values of conditional error term. Since there is no evidence to suggest that this is not the case, we will maintain the single index assumption for our analyses.

Identification Conditional on the Estimated Propensity

If the propensity to get the treatment (or equivalently the index *I*) can be estimated directly from Equation (2), we can identify differences in the minimum threshold from an estimate of the treatment-on-the-treated parameter: For patients with the same propensity, those treated at a hospital with a higher minimum threshold (a higher τ_h) will have a larger treatment effect. Moreover, because $g(I_{ih})$ goes to zero as the propensity to receive treatment goes to zero (or equivalently as I_{ih} gets increasingly negative), we can also identify overuse $(\tau_h < 0)$ or underuse $(\tau_h > 0)$: 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.

It is important to note that our model does not imply that the treatment-on-the-treated parameter is the same for all patients treated in hospitals with the same minimum threshold. In fact, Equation (4) implies that the treatment-on-the-treated effect will tend to be larger among patients that have a higher propensity to be treated (since $g(I_{ih})$ is increasing in I_{ih}). The treatment effect is the same only for patients with the same propensity to be treated *and* treated in hospitals with the same minimum threshold. Since hospital expertise (α_h) raises the propensity to be treated, it will also raise the treatment effect. The key difference, however, is that differences in hospital expertise have an impact on treatment effects by shifting the propensity to be treated (and, therefore, $g(I_{ih})$), while differences in the minimum threshold have an impact on treatment effects that is independent of the propensity.

The graphical intuition for our model can be seen in Figure 1. The expected benefit from treatment (B) is given on the vertical axis, while the propensity of being treated (which depends on *I*) is given on the horizontal axis. The top curve in Figure 1 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 $(\tau_{high} > 0)$, *i.e.* it represents $E(B_{ih} \mid B_{ih} > \tau_h) = \tau_h + g(I_{ih})$. The lower curve represents the same thing for a hospital with a low minimum threshold $(\tau_{low} < 0)$. Treatment-on-the-treated approaches the minimum threshold (τ_{high} or τ_{low}) for a patient with a low propensity of being treated (a very negative *I*), since no patient is ever treated with a benefit below this threshold. For a patient with a high propensity of being treated (a very positive *I*), 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 in the hospital with the higher threshold. In this figure, differences in hospital expertise would show up as a movement along the curves – changing the propensity of patients to be treated (and therefore the treatment-on-the-treated effect), but not affecting treatment effects conditional on propensity.

Identification Using Joint Distribution of Hospital Random Effects

The above approach identifies differences in the minimum treatment threshold by comparing the benefit from treatment once the propensity to receive treatment has been controlled for. But it is possible that our ability to control for this propensity is imperfect-- especially on the dimension of hospital effects in the propensity to receive treatment-- because of small numbers of patients at some hospitals. In some hospitals we have fewer than 20 patients, and the resulting imperfection would cause us to overstate the role of treatment threshold differences across hospitals when in fact there are actually large differences in hospital expertise.

Alternatively, hospital expertise (α_h) can be separately identified from the minimum treatment threshold (τ_h) using estimates of the joint distribution of the hospital effects in equations 2 and 4. This second approach requires us to constrain g(.) into a simpler form. To see this, we take a linear first-order taylor approximation of the function *g(.)*, and combine the hospital-level parameters into random effects to rewrite the equations as:

(2') $\Pr\left(Treatment_{ih}=1\right)=\Pr\left(-\varepsilon_{ih}<X_{ih}\beta+\theta_h\right),\ \ where\ \theta_h=\left(\alpha_h-\tau_h\right)$

(4') $E\left(B_{ih}\mid Treatment_{ih}=1\right)=\tau_h+\lambda_0+\lambda_1 I_{ih}=\lambda_0+\lambda_1\left(X_{ih}\beta\right)+\mu_h,\ \ where\ \mu_h=\tau_h+\lambda_1\theta_h$

Equation 2' defines the propensity to receive treatment as the sum of a linear combination of patient observables ($X_{ih}\beta$) and a hospital random effect (θ_h) which captures both the hospital treatment threshold and hospital expertise. Equation 4' defines the treatment effect parameter as proportional to the same linear combination of observables that determines treatment ($\lambda_1\left(X_{ih}\beta\right)$) plus a different hospital random effect (μ_h). We can write the original parameters of our model that captured hospital expertise and the minimum treatment threshold as functions of the parameters in this transformed model, with $\alpha_h=\mu_h+\left(1-\lambda_1\right)\theta_h$ and $\tau_h=\mu_h-\lambda_1\theta_h$. Thus, estimates of the parameters in this transformed model identify estimates of hospital expertise and the minimum treatment threshold. In particular, the joint distribution of expertise and the minimum threshold can be derived based on estimates of the joint distribution of the random effects (θ_h , μ_h) combined with an estimate of λ_1. We provide more details of this strategy in Section IV.

III. Heart-Attacks: Biology, Treatments, and Data

Heart-Attack Biology and 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. Timely thrombolytics, which are also known as fibrinolytics, are administered intravenously and break down blood clots by pharmacological means (these drugs include tissue plasminogen activators, streptokinanse and urokinase). Angioplasty (where a balloon on a catheter is inflated inside the blocked coronary artery to restore blood flow) and thrombolytics are two treatments that are used for immediate reperfusion (opening up the coronary artery). 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, as a consequence of what is known about heart attack treatments from randomized controlled trials, and more specifically for our setting, the benefits from reperfusion, 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. This would not be true if we focused on treatment variation for more chronic conditions such as diabetes, chronic obstructive pulmonary disease, or arthritis.

Our fourth reason for focusing on heart attacks is that it is an acute condition for which virtually all patients are hospitalized at a nearby hospital and receive some medical care. 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. The fact

that patients are generally taken to the nearest hospital for immediate treatment, makes it likely that a patient's hospital choice of hospital is exogenous and not driven by differences across hospitals in expertise or the treatment threshold. This feature of heart attack treatments would not be true, for example, of cancer therapies where two clinically identical patients may chose different providers based on their evaluation of the provider's expertise and standards for providing treatment.

Data

Because acute myocardial infarction is both common and serious, it has been the topic of intense scientific and clinical interest. One effort to incorporate evidence-based practice guidelines into the care of heart attack patients, begun in 1992, is the Health Care Financing Administration's Health Care Quality Improvement Initiative Cooperative Cardiovascular Project (CCP). Information about more than 200,000 patients admitted to hospitals for treatment of heart attacks in 1994/1995 was obtained from clinical records. The CCP is considerably superior to administrative/claims data (of the type used by McClellan et al. (1994)) as it collects chart data on the patients—detailed information is provided on laboratory tests, enzyme levels, the location of the myocardial infarction, and the condition of the patient at the time of admission. Detailed clinical data were abstracted from each patient's chart using a standard protocol. Further details about the CCP data are available in Marciniak et al. (1998), O'Connor et al. (1999), and in the appendix to this paper. The choice of sample and variables is identical to what we used and described in Barnato et al. (2005) and Chandra and Staiger (2007, 2010).

IV. Estimation

The Propensity to Receive Treatment

We use data on heart attack treatments to estimate the key components of our model, using receipt of reperfusion within 12 hours of the initial heart attack as our treatment. The propensity to receive treatment (I in the theoretical model) is estimated from a random effect logit model that regresses whether a patient received reperfusion within 12 hours of the heart attack on all the CCP risk-adjusters listed in the appendix (X_{ih}) and a random hospital effect (θ_h) that is assumed to be normally distributed:

$$(5) \qquad \Pr(reperfusion_{ih}=1) = F(X_{ih}\beta + \theta_h), \quad where\ \theta_h \sim N(0, \sigma_\theta^2)$$

Where the function F(.) represents the logistic function. Equation 5 is the empirical analog of Equation 2 in our model, where the hospital random effect is the difference between expertise and the threshold ($\theta_h = (\alpha_h - \tau_h)$) as in Equation 2'. Estimating equation 5 (using xtmelogit in Stata 11) yields maximum likelihood estimates of the coefficients $(\hat{\beta})$ and the standard deviation of the hospital random effect $(\hat{\sigma}_\theta)$,

and posterior estimates of the hospital random effects $\left(\hat{\theta}_h\right)$. We combine these to form an estimate of the propensity index for each patient $\left(\hat{I}_{ih} = X_{ih}\hat{\beta} + \hat{\theta}_h\right)$.

Estimation of Treatment Effects Conditional on the Estimated Propensity

We focus on patient survival as the outcome of interest which captures the key benefits of treatment. Survival is measured as a binary variable that measures survival to a certain date (e.g. survival to 30 days). Our model, and Equation 4 in particular, suggests that the effect of reperfusion on survival should be heterogeneous across patients and hospitals. This suggests estimating models of the following form (where F(.) represents the logistic function):

$$(6) \qquad \Pr\left(Survival_{ih} = 1\right) = F\left(reperfusion_{ih}\delta_{ih} + X_{ih}\gamma + \varphi_h\right),\ where\ \delta_{ih} = \tau_h + g\left(I_{ih}\right)$$

In equation 6, survival for all patients depends on patient risk adjusters (X_{ih}) and a hospital effect that captures general skill of the hospital staff (φ_h). The coefficient on reperfusion (δ_{ih}) is the empirical analog of Equation 4 in our model, and captures the survival benefit of reperfusion. Our model says the effect of reperfusion on survival depends only on a patient's propensity index (I_{ih}) and on the minimum treatment threshold at the hospital (τ_h).

Equation 6 suggests a simple test for whether the minimum threshold for treating a patient (τ_h) varies across hospitals. Conditional on a patient's propensity index (I_{ih}), variation across hospitals in the benefit of reperfusion is due to differences in the minimum threshold. This minimum threshold should be negatively correlated with the hospital random effect θ_h from the propensity equation (since $\theta_h = \left(\alpha_h - \tau_h\right)$). Intuitively, according to our model, hospitals with higher minimum thresholds for treatment should both treat fewer patients (conditional on *X*) and have higher benefits to treatment (conditional on *I*).

To implement this test, we use estimates of I_{ih} and θ_h from the propensity equation (5), and approximate the treatment effect with a linear function $\delta_{ih} = \lambda_0 + \lambda_1\hat{I}_{ih} + \lambda_2\hat{\theta}_h$, yielding an estimating equation:

$$(7) \qquad \Pr\left(Survival_{ih} = 1\right) = F\left(reperfusion_{ih}\lambda_0 + \left(reperfusion_{ih}\hat{I}_{ih}\right)\lambda_1 + \left(reperfusion_{ih}\hat{\theta}_h\right)\lambda_2 + X_{ih}\gamma + \varphi_h\right)$$

If hospitals vary in their expertise (α_h) but not in their minimum threshold for treating a patient (τ_h), then our model implies $\lambda_2 = 0$; controlling for patient propensity, the treatment effect is unrelated to the hospital effect in the propensity equation ($\hat{\theta}_h$). Alternatively, if hospitals vary in their minimum threshold but not in expertise, then we expect $\lambda_2 < 0$; controlling for patient propensity, the treatment effect is

smaller in hospitals with a high propensity to treat. If both expertise and the minimum threshold vary across hospitals, then we expect $\lambda_2 < 0$ unless there is a strong positive association between expertise and the minimum threshold (since the sign of λ_2 depends on cov(τ,θ) = cov(τ, α-τ) = cov(τ,α)-var(τ)).

Equation 7 provides two other tests of hospital behavior. First, if hospitals choose patients for treatment based on the benefit of the treatment, then the treatment effect should be increasing in the patient's propensity index ($\lambda_1 > 0$). Second, the treatment effect among patients with a low propensity index is an (upper bound) estimate of each hospital's minimum threshold, since $\delta_{ih} = \tau_h + g(I_{ih})$ and *g(.)* approaches zero as the propensity index falls. Therefore, if the treatment effect among low-propensity patients is negative in some hospitals, this is evidence of overuse ($\tau_h < 0$).

We explore a number of alternative specifications to Equation 7 in order to check the robustness of our results. First, in some specification we include hospital fixed effects (using fixed-effect logit models) in order to capture general differences in hospital skill affecting survival of all patients (φ_h). Second, in some specifications we estimate *g(I)* non-parametrically with 100 indicator variables for the percentiles of *I* interacted with reperfusion (and included directly in X), thus allowing the return to reperfusion to vary flexibly with a patient's propensity to receive reperfusion. Finally, we also use semi-parametric methods (described in more detail in a future appendix) to flexibly estimate how the effect of reperfusion on survival varies with $\hat{I}_{ih}$ and with $\hat{\theta}_h$.

Joint Estimation of Treatment Effect and Propensity Equations

One problem with Equation 7 is that the estimates and tests are conditional on estimates of the patient's propensity ($\left(\hat{I}_{ih} = X_{ih}\hat{\beta} + \hat{\theta}_h\right)$) and the hospital's effect in the propensity equation ($\hat{\theta}_h$). In our large sample, the coefficients on patient risk-adjusters ($\hat{\beta}$) are consistently estimated and can be thought of as known (asymptotically). However, the hospital effect ($\hat{\theta}_h$) is a posterior estimate and based on small samples of patients for many hospitals – half of the hospital's in our sample have fewer than 20 AMI patients in the sample. Therefore, $\hat{I}_{ih}$ will only imperfectly control for a patient's propensity, and after conditioning on $\hat{I}_{ih}$ there may be remaining variation in treatment effects across hospitals due to expertise (which was not fully reflected in the estimated propensity).

An alternative approach that avoids this problem is to estimate the treatment propensity equation and the survival equation jointly, treating the hospital intercept in the propensity equation and the hospital-specific effect of reperfusion in the survival equation as correlated random coefficients (along

with the hospital intercept in the survival equation). We implement this approach by taking a linear approximation to Equation 4 ($g(I_{ih}) = \lambda_0 + \lambda_1(I_{ih})$) and estimating the following equations jointly:

$$\begin{aligned} &\Pr(reperfusion_{ih} = 1) = F(X_{ih}\beta + \theta_h) \\ &\Pr(Survival_{ih} = 1) = F(reperfusion_{ih}\lambda_0 + (reperfusion_{ih}X_{ih}\beta)\lambda_1 + reperfusion_{ih}\mu_h + X_{ih}\gamma + \varphi_h) \end{aligned} \tag{8}$$

The three hospital-level parameters ($\theta_h, \mu_h, \varphi_h$) are treated as random coefficients that are distributed as a joint normal with variance and covariance to be estimated. As discussed in the theory section, hospital-level expertise (α_h) and the minimum threshold (τ_h) are linear combinations of the hospital intercept in the propensity equation (θ_h) and the hospital-level coefficient on reperfusion (μ_h), where $\alpha_h = \mu_h + (1-\lambda_1)\theta_h$ and $\tau_h = \mu_h - \lambda_1\theta_h$. Thus, estimates of the joint distribution of expertise and the minimum threshold can be derived from estimates of the joint distribution of the random effects (θ_h, μ_h) combined with an estimate of λ_1.

Equation 8 is a hierarchical logistic model (patients nested within hospitals) with random coefficients at the hospital level. To simplify estimation, we estimated the coefficients on the patient-level risk adjusters (β,γ) in a first stage, using simple logit models of each equation estimated separately (omitting the hospital effects). We then interacted the first-stage estimate of $X_{ih}\hat{\beta}$ with reperfusion, and included the estimated indexes ($X_{ih}\hat{\beta}, X_{ih}\hat{\gamma}$) directly in Equation 8 with a single coefficient (expected to equal 1) rather than including all of the X variables individually. The remaining parameters determining the effect of reperfusion (λ_0, λ_1) and the variance and covariance of the hospital-level random coefficients were estimated by maximum likelihood using xtmelogit in Stata 11.1. All of the reported standard errors are conditional on the first-stage estimates ($X_{ih}\hat{\beta}, X_{ih}\hat{\gamma}$), 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 in the first-stage logit models.

V. Results

In Table 1 we report some basic characteristics of our sample overall, and by whether the patient received reperfusion within 12 hours of admission to the hospital. In our sample, 19% of patients received reperfusion within 12 hours of admission for a heart attack. Overall, 81% of patients were still alive 30 days after admission, but survival was higher for patients receiving reperfusion (86%) than for patients who did not receive reperfusion (80%). However, much of the difference in survival between these two groups was due to differences in underlying health and pre-existing conditions, rather than the result of

reperfusion. Patients receiving reperfusion were younger, and much less likely to have pre-existing conditions such as congestive heart failure, hypertension, diabetes, and dementia.

In the appendix table, we report estimates from a random effect logit model predicting reperfusion as a function of the full list of patient risk-adjusters (also listed in the appendix) and a random hospital-level intercept (Equation 5). These results were used to form posterior estimates of the hospital random effects $\left(\hat{\theta}_h\right)$ and an estimate of the propensity index for each patient $\left(\hat{I}_{ih} = X_{ih}\hat{\beta} + \hat{\theta}_h\right)$. The coefficients on the patient-level variables are consistent with the medical literature, with reperfusion being less likely among patients with pre-existing conditions and who are older (based on full age-sex-race interactions not reported in the table), and also depending on the location and severity of the heart attack. The estimated standard deviation of the hospital effect is 0.44 (Std. Err. = 0.01), which implies that a one standard deviation in the hospital effect increases the logodds of receiving reperfusion by 0.44, which would increase an average patients probability of receiving reperfusion from 19% to 26%. Thus, there is sizable variation across hospitals in the rate at which they provide reperfusion to observationally similar patients. The model is able to predict much of the hospital-level variation, with the posterior prediction of each hospital's effect on reperfusion having a standard deviation of 0.30 in our data.

In the appendix table we also report estimates from a simple logit model estimating the impact of reperfusion on 30-day survival, controlling for the patient risk-adjusters. Our estimation strategy relies on using the richness of the CCP data to invoke a `selection on observables' assumption to estimate treatment on the treated for reperfusion therapy. We compared the estimates from this logit model to those obtained from clinical trials to evaluate the plausibility this assumption. A summary of nine trials was published in the journal *Lancet* by the Fibrinolytic Therapy Trialists' Collaborative Group (FTTCG, 1994). This was the same time-period as the CCP data and each trial evaluated fibrinolytic 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 reperfusion reduced the logodds of mortality by 0.20. The logit model controlling for the CCP risk-adjusters estimates a nearly identical effect, with reperfusion increasing the logodds of survival (equivalently reducing the logodds of mortality) by 0.206 (S.E. = 0.023). We take this evidence as supporting the case that these logit models provide unbiased estimates of the treatment effect.

Estimates of Treatment Effects Conditional on the Estimated Propensity

In Table 2 we present results from estimating the treatment effect of reperfusion conditional on the estimated propensity using Equation 7:

$$(7) \qquad \Pr\left(Survival_{ih} = 1\right) = F\left(reperfusion_{ih}\lambda_0 + \left(reperfusion_{ih}\hat{I}_{ih}\right)\lambda_1 + \left(reperfusion_{ih}\hat{\theta}_h\right)\lambda_2 + X_{ih}\gamma + \varphi_h\right)$$

Recall from the earlier discussion that the key parameters of interest are those associated with reperfusion ($\lambda_0, \lambda_1, \lambda_2$). Table 2 reports estimates of these parameters, which together determine the treatment effect: $\delta_{ih} = \lambda_0 + \lambda_1 \hat{I}_{ih} + \lambda_2 \hat{\theta}_h$. Coefficients on the patient risk-adjusters are not reported, but similar to those reported in the appendix table. The top panel reports coefficients from various specifications using simple logit models (clustering standard errors at the hospital level), while the bottom panel reports coefficients from the same specifications adding hospital fixed effects (conditional logit models).

As a baseline, the first column of Table 2 reports estimates that do not condition on the patient's propensity, but allow the effect of reperfusion to interact with the estimated hospital effect from the propensity equation ($\hat{\theta}_h$). Since this effect is mean zero, the coefficient on reperfusion gives the treatment effect for an average hospital, and is 0.229 and highly significant in the model without hospital fixed effects. The coefficient on the interaction with the hospital effect from the propensity equation is negative, but not statistically significant. The results with hospital fixed effects are very similar. Since these specifications do not condition on the propensity, the coefficient on the interaction with the hospital effect (which is part of the propensity) is biased in the positive direction, and not a strong test of whether hospitals differ in their minimum treatment threshold.

The second column of Table 2 allows the effect of reperfusion to interact with the patient's propensity to receive reperfusion (the index, $\hat{I}_{ih}$). To help with interpretation, we have normed the 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 treated at an average hospital. The coefficient on the interaction of reperfusion with the propensity index is positive and highly significant in specifications with or without hospital fixed effects, implying that the treatment effect of reperfusion on survival is increasing in the patient's propensity index ($\lambda_1 > 0$) as predicted by our model. The coefficient on this interaction implies that an increase in the propensity index of one (about one standard deviation of the propensity index in the treated population) is associated with roughly a doubling of the treatment effect. Thus, it appears that hospitals are choosing patients for treatment based on the benefit of the treatment, and the heterogeneity in the treatment effect is large relative to the average treatment effect.

As expected, the coefficient on the interaction of reperfusion with the hospital effect from the propensity equation becomes more negative and statistically significant in the specifications that condition on the propensity index. The coefficient is similar in column 3, where we non-parametrically control for the interaction of reperfusion with a set of 100 dummies for each propensity percentile. In other words, conditional on a patient's propensity, the treatment effect is smaller in aggressive hospitals

with a high propensity to treat ($\lambda_2 < 0$). As discussed earlier, this finding suggests that hospitals vary in their minimum threshold, and more aggressive hospitals with lower minimum thresholds for treatment treat more patients (conditional on *X*) and have lower benefits to treatment (conditional on *I*). The estimated coefficients suggest that a one standard deviation increase in the hospital effect from the propensity equation (about 0.3) lowers the return to reperfusion by about .06-.09.

As discussed above, the treatment effect for patients with a low propensity index is an (upper bound) estimate of each hospital's minimum threshold, and can be used to identify overuse ($\tau_h < 0$). Based on the estimates from Table 2, we can use $\delta_{ih} = \lambda_0 + \lambda_1 \hat{I}_{ih} + \lambda_2 \hat{\theta}_h$ to calculate the expected treatment effect for patients with a low propensity index admitted to a given hospital. For example, consider a "low propensity" patient who is one standard deviation below average in their propensity index (about -1). Using estimates from the parametric model without hospital fixed effects, a low propensity patient at an average hospital would have a treatment effect near zero (0.345-1*0.299=.006). But a low propensity patient at a hospital that was one standard deviation above average (about 0.3) in its hospital effect from the propensity equation would have a negative treatment effect (0.345-0.299*1-0.258*0.3=-0.077). In other words, for many patients in our sample, our estimates imply overuse (a negative treatment effect).

The parametric model estimated in Table 2 imposes linearity, and may not provide accurate predictions of treatment effects for patients with low propensity. Therefore, we explore our results non-parametrically in Figures 2-4 (see the appendix for details of how this was done). In the left-hand panel of figure 2 we plot 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 of Figure 2 is the analogous plot but 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 random effects in the propensity equation. Among all patients (the left-hand plot), the estimated survival benefit from reperfusion is positive for all hospitals, although it is small and not significant in hospitals with the highest treatment rates (those 2 standard deviations above average, with $\hat{\theta}_h$=0.6). In contrast, among the lowest propensity patients (the right-hand plot), only hospital's with the lowest treatment rates are estimated to have survival benefits from reperfusion that are near to zero. The estimated survival benefit from reperfusion is negative and significant in hospitals with the highest treatment rates, suggesting that there is overuse in these hospitals and, as a result, we were able to identify substantial subsets of patients who were harmed by reperfusion treatment.

Figure 3 plots the estimated survival benefit from reperfusion against a patient's treatment propensity index for hospital's in the lowest (left-hand side) and highest (right-hand side) terciles of the estimated hospital effect from the propensity equation ($\hat{\theta}_h$). Both plots show a strong upward slope, with higher benefit from treatment for patients with a higher propensity to receive reperfusion. But at every propensity, the benefits of reperfusion are lower in the top-tercile hospitals. 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 estimate 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.

Joint Estimates of Treatment Effect and Propensity Equations

In Table 3 we present results from estimating the treatment propensity equation and the survival equation jointly using Equation 8:

$$\begin{aligned} &\Pr(reperfusion_{ih}=1) = F(X_{ih}\beta + \theta_h) \\ &\Pr(Survival_{ih}=1) = F(reperfusion_{ih}\lambda_0 + (reperfusion_{ih}X_{ih}\beta)\lambda_1 + reperfusion_{ih}\mu_h + X_{ih}\gamma + \varphi_h) \end{aligned} \tag{8}$$

Recall from the earlier discussion that the key parameters of interest are the standard deviations and correlation of the random hospital intercept in the propensity equation and the hospital-specific effect of reperfusion, and the coefficient (λ_1) on the interaction between reperfusion and $X_{ih}\hat{\beta}$. Table 3 reports estimates of these parameters (along with the reperfusion parameters in the survival equation), which (as described earlier) were then used to construct estimates reported at the bottom of the table of the variance across hospitals in both expertise and the minimum treatment threshold, along with their correlation. We focus our discussion on these transformed estimates at the bottom of the table, since they are directly interpretable in the context of our model (Equations 1-4 in Section II).

The estimates in Table 3 suggest that there is considerable variation across hospitals in both expertise (Std. Dev. = 0.451) and the minimum threshold for treatment (Std. Dev. = 0.327). The standard deviation in expertise captures the difference in the benefit from reperfusion across hospitals for a randomly chosen patient. Because hospitals with low benefits from reperfusion will select fewer patients for reperfusion (and only the most appropriate), the treatment-on-treated effect will vary less across hospitals. In our model, one can multiply expertise by λ_1 to get the variation across hospitals in the treatment-on-treated effect. Thus, variation in expertise across hospitals accounts for a standard deviation across hospitals in the observed treatment effect of 0.125 (0.451*0.276). This calculation suggests that a large portion of the variation across hospitals in the observed treatment effect (with Std. Dev. = 0.307) is the result of variation in the treatment threshold rather than expertise. Interestingly, the standard deviation

in the hospital intercept in the reperfusion equation (0.442) is less than one might expect because a hospital's minimum treatment threshold is positively correlated with hospital expertise – hospitals with low expertise also tend to overuse reperfusion, making their overall reperfusion rate less low than it would be otherwise.

VI. Conclusions

The reasons for variation in treatment rates across hospitals serving similar patient may be due to greater benefits of treatment in some hospitals (expertise), withholding of beneficial treatment in some hospitals (underuse), or providing harmful treatment in other hospitals (overuse). Our empirical results distinguish between these explanations, using a simple behavioral model in which hospitals choose to treat patients if the benefit from treatment exceeds a hospital-specific threshold. We distinguished between expertise, underuse, and overuse based on differences across hospitals in both their reperfusion rates and the effect of reperfusion on patient survival in a sample of heart attack patients. Our results suggest that expertise varies considerably across hospitals, but a substantial amount of variation in treatment and treatment effectiveness in our data was due to overuse.

References

Barnato, AE, FL Lucas, D Staiger, A Chandra. 2005. "Hospital-Level Racial Disparities in Acute Myocardial Infarction Treatment and Outcomes." *Medical Care* 43: 308-19.

Chandra, Amitabh and Jonathan S. Skinner, "Geography and Racial Health Disparities" in Anderson, Norman B., Rodolfo A. Bulatao, and Barney Cohen (Eds). 2004. *Critical Perspectives: on Racial and Ethnic Differences in Health in Late Life*. Washington D.C.: The National Academies Press.

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

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

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

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

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

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

Skinner, Jonathan, Amitabh Chandra, Douglas Staiger, Julie Lee, and Mark McClellan. 2005. "Mortality After Acute Myocardial Infarction in Hospitals that Disproportionately Treat African Americans." *Circulation*.

Appendix

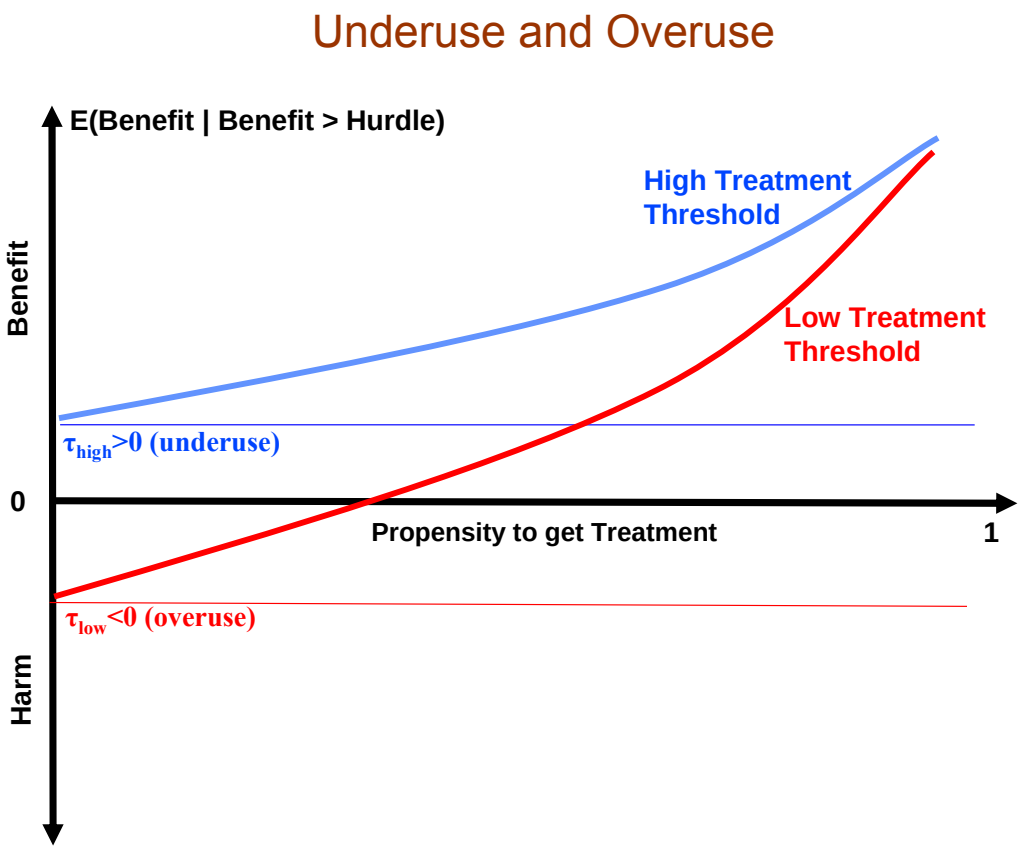
Construction of CCP Estimation Sample:

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

Our choice of variables was based on those selected by Fisher et al. (2003a,b) and Barnato et al. (2005). With the exception of two variables that are both measured by blood-tests, albumin and bilirubin (where the rates of missing data were 24 percent), we do not have a lot of missing data (rates were less than 3 percent). Included in our model are the following risk-adjusters (APPENDIX TABLES TO BE INCLUDED IN LATER DRAFT):

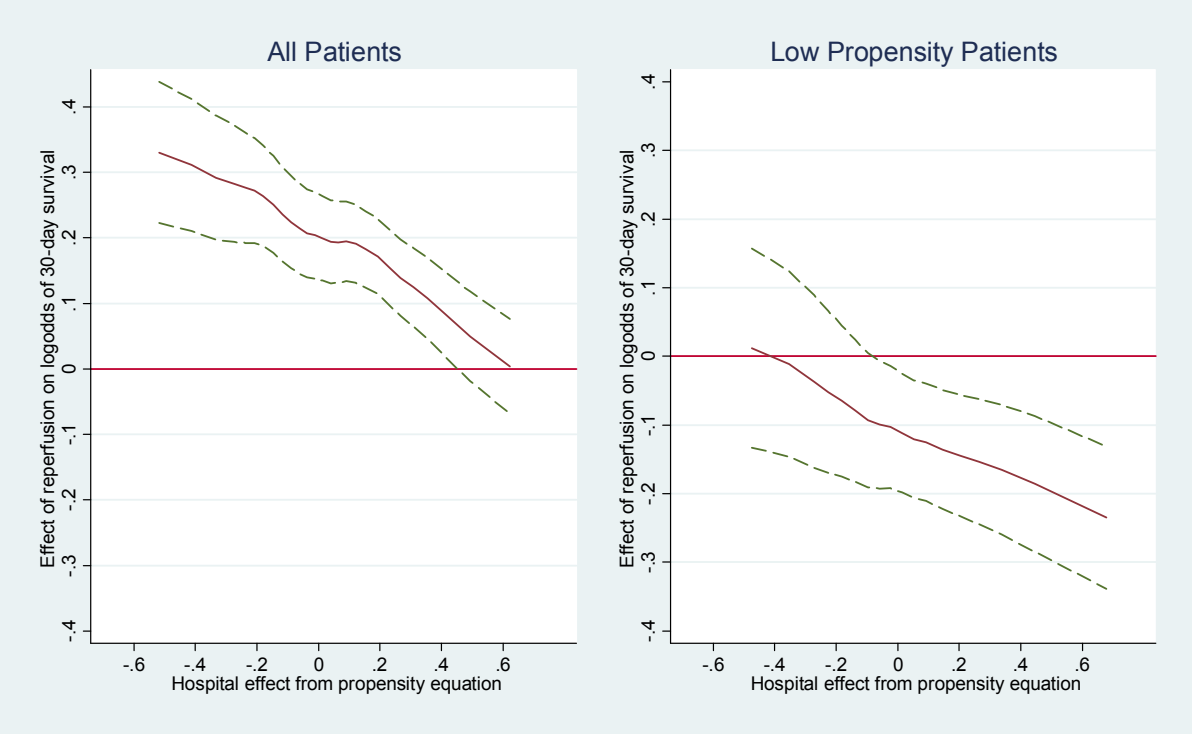
Age, Race, Sex (full interactions)
previous revascularization (1=y)
hx old mi (1=y)
hx chf (1=y)
history of dementia
hx diabetes (1=y)
hx hypertension (1=y)
hx leukemia (1=y)
hx ef <= 40 (1=y)
hx metastatic ca (1=y)
hx non-metastatic ca (1=y)
hx pvd (1=y)
hx copd (1=y)
hx angina (ref=no)
hx angina missing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

Figure 1: How the Expected Benefit from Treatment Varies with the Propensity to Get Treatment and the Treatment Threshold.



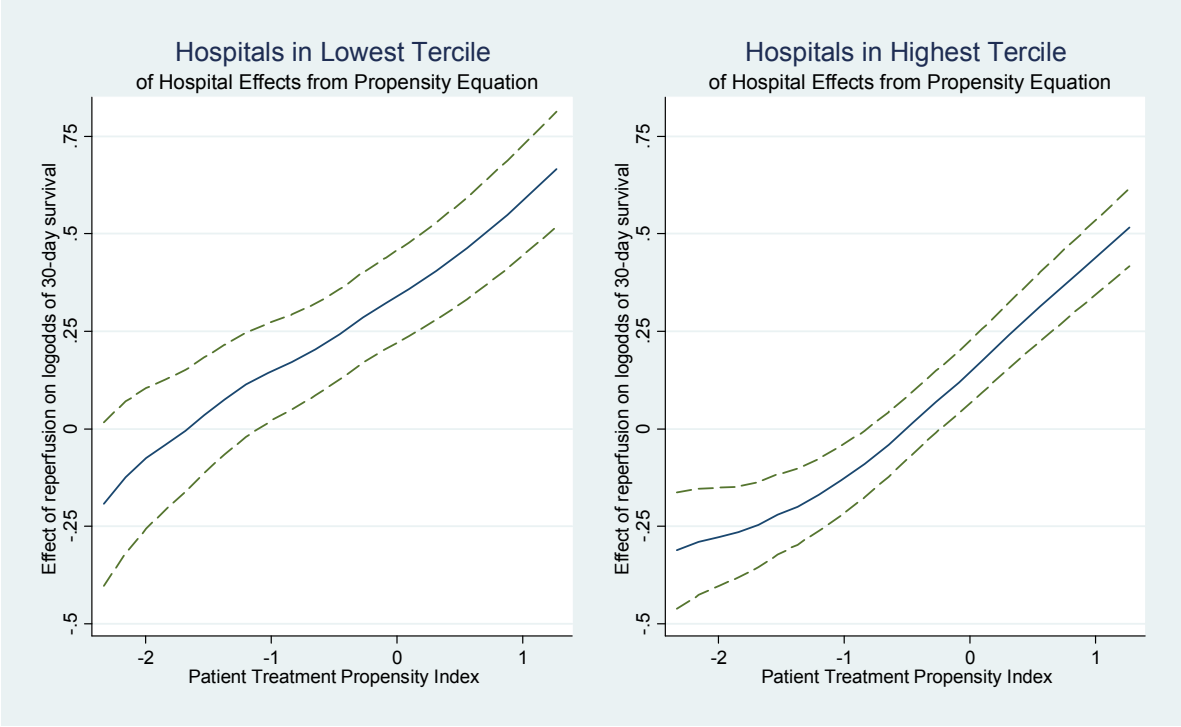
The figure illustrates the relationship between the expected benefit from treatment, $E(B|B> \tau)$, on the vertical axis, and the propensity index I (which determines the propensity of being treated) on the horizontal axis. The thick curves represent the treatment-on-the-treated effect for a patient 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 2: Survival Benefit from Reperfusion According to Hospital Effect on Treatment Propensity, All patients (Panel A) and Low-propensity patients (Panel B).



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.

Figure 3: Survival Benefit from Reperfusion According to Patient's Treatment Propensity, Low-Treatment-Rate (Panel A) and High-Treatment-Rate (Panel B) Hospitals.



The figures plot the estimated survival benefit (and 95% confidence intervals) from reperfusion against a patient's treatment propensity index for hospital's in the lowest (left-hand side) and highest (right-hand side) terciles of the estimated hospital effect from the propensity equation

Table 1: Means of Selected Variables, Overall 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 w/in 12 hours	19%	100%	0%
Age	77	73	77
Previous diagnoses:			
Congestive Heart Failure	22%	7%	25%
Hypertension	62%	56%	63%
Diabetes	30%	23%	32%
Dementia	6%	2%	7%
Number of observations	138,957	25,876	113,081

Table 2. Logit Estimates of the Effect of Reperfusion on 30-day Survival, Conditional on the Estimated Propensity and the Hospital Effect from the Propensity Equation.

	Not Conditional on Propensity	Conditional on Propensity (Parametric)	Conditional on Propensity (Non-Parametric)
Without Hospital Fixed Effects			
Reperfusion <12 hours	0.229	0.345	non-parametric
	(0.027)	(0.027)	
Reperfusion <12 hours * propensity index		0.299	non-parametric
		(0.018)	
Reperfusion <12 hours * Hospital effect from propensity equation	-0.094	-0.258	-0.310
	(0.078)	(0.080)	(0.080)
With Hospital Fixed Effects			
Reperfusion <12 hours	0.214	0.328	non-parametric
	(0.026)	(0.027)	
Reperfusion <12 hours * propensity index		0.291	non-parametric
		(0.018)	
Reperfusion <12 hours * Hospital effect on treatment propensity	-0.052	-0.211	-0.254
	(0.074)	(0.076)	(0.077)

Note: Dependent variable is the whether survived to 30 days. The top panel reports coefficients from various specifications using simple logit models (clustering standard errors at the hospital level), while the bottom panel reports coefficients from the same specifications with hospital fixed effects (conditional logit models). Standard errors in parentheses. Models include all CCP risk-adjusters. Column 3 includes 100 percentiles of I interacted with the receipt of Reperfusion.

Table 3. Joint Estimates of The Reperfusion Equation Determining Treatment Propensity and the Survival Equation.

	Estimate	(Standard Error)
Reperfusion Equation:		
Std. Dev. Of Hospital-Level intercept	0.442	(0.013)
30-day Survival Equation:		
Reperfusion <12 hours	0.265	(0.026)
Reperfusion <12 hours * ($X\beta$)	0.276	(0.018)
Hospital-level intercept		
Standard Deviation	0.199	(0.017)
Correlation with reperfusion intercept	-0.100	(0.073)
Hospital-level coefficient on reperfusion		
Standard deviation	0.307	(0.056)
Correlation with reperfusion intercept	0.035	(0.112)
Correlation with survival intercept	-0.381	(0.151)
Transformed Estimates:		
Hospital minimum treatment threshold		
Standard deviation	0.327	(0.055)
Correlation with survival intercept	-0.321	(0.150)
Hospital expertise		
Standard deviation	0.451	(0.045)
Correlation with minimum threshold	0.390	(0.144)
Correlation with survival intercept	-0.330	(0.115)

Note: See text for discussion of estimation method and equations.