

NBER WORKING PAPER SERIES

IDENTIFYING PROVIDER PREJUDICE IN HEALTHCARE

Amitabh Chandra  
Douglas O. Staiger

Working Paper 16382  
<http://www.nber.org/papers/w16382>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
September 2010

This research was funded by the National Institute of Aging (NIA) P01 AG19783-02. We thank Gary Becker, David Cutler, Kevin Murphy, Jonathan Skinner and many seminar participants 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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

© 2010 by Amitabh Chandra and Douglas O. Staiger. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Identifying Provider Prejudice in Healthcare  
Amitabh Chandra and Douglas O. Staiger  
NBER Working Paper No. 16382  
September 2010  
JEL No. I12,J15,J16,J71

### **ABSTRACT**

We use simple economic insights to develop a framework for distinguishing between prejudice and statistical discrimination using observational data. We focus our inquiry on the enormous literature in healthcare where treatment disparities by race and gender are not explained by access, preferences, or severity. But treatment disparities, by themselves, cannot distinguish between two competing views of provider behavior. Physicians may consciously or unconsciously withhold treatment from minority groups despite similar benefits (prejudice) or because race and gender are associated with lower benefit from treatment (statistical discrimination). We demonstrate that these two views can only be distinguished using data on patient outcomes: for patients with the same propensity to be treated, prejudice implies a higher return from treatment for treated minorities, while statistical discrimination implies that returns are equalized. Using data on heart attack treatments, we do not find empirical support for prejudice-based explanations. Despite receiving less treatment, women and blacks receive slightly lower benefits from treatment, perhaps due to higher stroke risk, delays in seeking care, and providers over-treating minorities due to equity and liability concerns.

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

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

## I. Introduction

A vast literature in social-science and medicine finds evidence of race and gender disparities in wages, hiring decisions, mortgage approvals, and medical treatments. But disparities cannot by themselves distinguish between two competing explanations for why they exist [Heckman (1998) and List (2004)]: are employers, lenders, or physicians, consciously or unconsciously withholding jobs, loans and medical treatment from minority groups despite similar benefit (prejudice), or is membership in a minority group associated with lower productivity or benefit (statistical discrimination)? In this paper, we use simple economic insights to develop a framework for using observational data to distinguish between prejudice and statistical discrimination.

We focus our inquiry on the enormous literature on racial and gender disparities in healthcare where treatment differences are not explained by access, preferences, or severity. While the explanation for disparities in this context may be different than those in labor and capital markets, given the scope for human wellbeing and longevity it is of tremendous importance. Disparities in medical treatment measured using regression adjustment, audit studies and Implicit Association Tests (IATs) are often ascribed to prejudicial providers [Healy (1991), Bogart et al. (1994), Schulman et al. (1999), van Ryn and Fu (2003), Fincher et al. (2004) and Green et. al. (2007)]. This view is shared by the Institute of Medicine's (IOM) *Unequal Treatment* report which concludes that provider bias was among the leading determinants of disparities (Smedley, Stith and Nelson, 2003).

The key to our model is the Beckerian intuition that under taste-based prejudice, providers may consciously or unconsciously use a higher benefit threshold before providing care to minority patients (for example, recommending a treatment to non-minority patients if it prolongs their life by at least three months, but only treating minority patients if it prolongs their life by at least five months). One consequence of this type of prejudice is that minority patients will have higher marginal returns from being treated. Another form of bias occurs when providers put too much or too little weight on race and gender or misinterpret other patient characteristics at the time of determining appropriateness (for example, incorrectly thinking that a black patient who actually receives 5 months of survival benefit only receives 3 months of benefit). Both of these forms of prejudice, taste-based and imperfect determination of appropriateness, result in minority patients receiving less care than is optimal, and both predict

higher benefits from treatment for minority patients. In contrast to this interpretation for why blacks and women (henceforth referred to as minorities) receive less care, membership in a minority group may predict lower benefit from treatment because of its statistical association with biology, pre-hospital delays, provider skill, and follow-up care. If providers are aware of the correlation between these factors and race and gender, the differential treatment of minorities may be due to statistical discrimination. While both explanations result in minority groups receiving less treatment, they imply vastly different policies.<sup>1</sup>

While it is intuitive to distinguish between the two possibilities by comparing outcomes for the marginal patient being treated for minority and majority groups, in the absence of observing exactly what the physician observes about the patient, it is impossible to identify the marginal patient and measure the benefit from treatment for this patient. Therefore, we address this complication by proposing a test for provider prejudice that relies on comparing the ‘treatment on the treated’ effect for men to women, and whites to blacks. In the absence of prejudice, we demonstrate that the treatment on the treated parameter is the same for both groups conditional on a patient’s appropriateness or propensity to receive treatment. This framework also allows us to ascertain whether physicians correctly determine patient appropriateness for the treatment. Our framework offers a test based on observational data that is complementary to that in List (2004) who used data from field experiments to make the distinction between prejudice and statistical discrimination.

Our approach to identifying provider prejudice provides several advantages over ostensibly simpler methods. Regression adjustment, audit-studies involving patient-actors, and Implicit Association Tests (IAT) are often used to assess the case for prejudice in lending or how the providers treat similar patients from different demographic backgrounds (Munnell et al. 1996, Schulman et al., 1999, Smedley, Stith and Nelson, 2003 and Green et al., 2007). But the fact that providers may offer fewer treatments to women and minorities is not by itself evidence of prejudice—if the benefit to these groups is in fact lower on average because of lower appropriateness stemming from biological differences in how the treatment is absorbed, or

---

<sup>1</sup> The literature often uses the terms prejudice, stereotyping, bias and discrimination interchangeably. We use the terms prejudice, animus, and bias to refer to provider behavior that results in allocative inefficiency. We avoid using the term stereotyping because both prejudice and statistical discrimination involve stereotyping—the key is whether physician behavior leads to greater efficiency or inefficiency in allocating treatment. We have chosen not to use the term discrimination to avoid confusion with the alternative hypothesis of statistical discrimination.

behavioral factors such as lower compliance or ambulance delays in being taken to the hospital (although it could be evidence of more macro prejudice against minorities). Yet because it is impossible to measure outcomes for simulated patients, it is not possible to distinguish between the two explanations, and consequently, their very different policy implications. Similarly, it may be tempting to see how physicians of one race (or gender) treat patients of their demographic versus others, as has been done in recent studies of racial bias in motor vehicle searches (Antonovics and Knight, 2005; Anwar and Fang, 2006). However, this approach can only identify whether there is differential bias across physicians of different race, and previous evidence has found no difference in the how white and black physicians treat heart attack patients (Chen et al., 2001). Moreover, since the care of minority patients is highly concentrated within a small set of physicians, studies (including audit studies) that focus on whether these physicians treat white and black patients differently exclude the substantial number of physicians who only treat patients of one race. In theory, such providers may be the most or the least prejudicial, and patients may sort into providers in way that makes this evidence very unrepresentative of the discrimination being faced by a typical patient (Heckman, 1998).

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 that support the statistical discrimination interpretation of treatment disparities. We do not find empirical support for alternative explanations based on prejudice or bias. If anything, despite receiving less treatment, women and blacks receive lower benefits, perhaps due to higher stroke risk, delays in seeking care or providers caring about equity or liability concerns. Differences in follow-up care or provider skill do not appear to explain our results. Section VII concludes.

## **II. Theory**

Our goal is to establish whether disparities in treatment rates represent prejudicial behavior by providers (implying underuse in blacks and women) or statistical discrimination against these groups. A simple model of patient treatment choice will guide 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 Vytlačil (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 race or gender could affect treatment choice: race or gender could be related to the expected benefit of treatment, or could alter the minimal threshold that must be met to receive care. We define taste-based prejudice to exist when the expected benefit from treatment for patients of a certain race or gender must exceed a higher threshold for them to receive treatment. When this is the case, the treatment of two patients with the same expected benefit from treatment will differ because of their race or gender. In this situation, there is underuse of treatment for the group facing a higher threshold. In the absence of any prejudice, treatment will differ by race or gender only if the expected benefits of treatment are different.

We illustrate this idea with the case of prejudicial behavior against women; the case for prejudice against blacks is analogous, with *female* being replaced by *black*. Let  $B$  represent the expected benefit from treatment for a given patient (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).<sup>2</sup> Suppose that the expected benefit from treatment depends on the patient's gender (ignoring race for the moment to simplify the presentation), observable patient characteristics ( $X$ ) such as age, medical history, and lab results (all of which are allowed to be interacted with gender and will be in the empirical work), and other factors that are known to the medical care provider when making the treatment decision but unobserved by the econometrician ( $\varepsilon$ ):

$$(1) \quad B = X\beta_1 + Female\beta_2 + \varepsilon$$

At this point, we are assuming that providers are not making errors in their estimation of patient benefit from treatment and that  $\beta_1$  and  $\beta_2$  are correctly estimated. Note that gender (or race) could be statistically related to the benefit of treatment in Equation (1) because of differences in biology, pre-existing medical conditions, follow-up care, presence of family

---

<sup>2</sup> 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 the differences in treatment costs by gender and race were estimated to be small and insignificant, so we focus on survival for simplicity.

member, or quality of the provider. Similarly, factors that are correlated with race and observed by providers but unobserved to the econometrician, will be absorbed into the coefficient on gender or race. All such differences by gender and race should be taken into account by the medical provider if the goal is to maximize the benefit to the patient. Thus, differences in treatment that arise from differences in expected benefit are not the result of prejudice in the current treatment decision, even though they may be the result of discrimination more broadly (for example, in ambulance response times or the fact that the treatment was primarily designed for males).

Each patient receives treatment if the expected benefit from treatment exceeds a minimal threshold ( $\tau$ ), where the threshold can vary by gender (or race):

$$(2) \quad \tau = \tau_1 + Female\tau_2$$

Women experience prejudice if  $\tau_2 > 0$  in Equation (2), i.e. if the expected benefits must be higher for women to receive treatment. Equations (1) and (2) follow directly from a Roy model of treatment allocation, where a patient receives treatment if the gain from the treatment exceeds a cost hurdle of  $\tau$ .

Note that we assume that providers use the right index in equation (1) to determine the benefit from care. But providers may be using the wrong index—because of clinical uncertainty, or inattention, they may place undue importance on gender or use the wrong weights on key comorbidities. In our model, these two forms of prejudice, taste-based and imperfectly determining appropriateness, are identical – adding an amount to the hurdle is equivalent to incorrectly subtracting the same amount from the benefit when determining whether to treat a patient.

Equation (1) and (2) imply 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:

$$(3) \quad \Pr(Treatment=1) = \Pr(B > \tau) = \Pr(-\varepsilon < I)$$

$$where I = X\beta_1 + Female(\beta_2 - \tau_2) - \tau_1$$

Equation (3) highlights why differences in treatment rates by gender (or race), holding all else equal, are not by themselves evidence of prejudice. Women may be less likely to receive

treatment because of lower expected benefits of treatment ( $\beta_2 < 0$ ) or because of prejudice ( $\tau_2 > 0$ ). Conversely, equation (3) demonstrates that it is premature to conclude that there is no prejudice against women based on estimating a coefficient of zero on female in a regression of treatment receipt. Such a result could mask the fact that there is prejudice against minority patients ( $\tau_2 > 0$ ) and that their benefit from treatment is higher ( $\beta_2 > 0$ ). Thus, contrary to much of the empirical literature, it is not possible to identify the effect of taste-based prejudice ( $\tau_2$ ) from the coefficient on *Female* in Equation (3) alone.

However, the effect of prejudice can be identified if information on the treatment effect among the treated population is available. The treatment-on-the-treated parameter is defined as:

$$(4) \quad E(B | Treatment = 1) = E(B | -\varepsilon < I) = X\beta_1 + Female\beta_2 + E(\varepsilon | -\varepsilon < I)$$

Noting that  $X\beta_1 + Female\beta_2 = I + \tau_1 + \tau_2 Female$ , we can rewrite Equation (4) as:

$$(5) \quad E(B | Treatment = 1) = \tau_2 Female + g(I)$$

where  $g(I) = \tau_1 + I + E(\varepsilon | -\varepsilon < I)$

Equation (5) states that in the absence of prejudice ( $\tau_2 = 0$ ), two patients receiving treatment who have the same propensity to get the treatment (same  $I$ ) will have the same expected benefit from the treatment. Since the propensity to get the treatment (or equivalently the index  $I$ ) can be estimated directly from Equation (3), we can therefore identify prejudice from an estimate of the treatment-on-the-treated parameter: If there is prejudice ( $\tau_2 > 0$ ) the treatment-on-the-treated effect is larger for women than it is for men with the same propensity to get treatment.

It is important to note that Equation (5) does not imply that the treatment-on-the-treated parameter is the same for all men and women in the absence of prejudice. In fact, Equation (5) implies that the treatment-on-the-treated effect will tend to be larger among men if men have a higher propensity to be treated (since  $g(I)$  is increasing in  $I$ ). The treatment effect is the same only for men and women with the same propensity to be treated, or equivalently for any population of men and women that have the same propensity distributions (e.g. are matched on propensity to be treated). Formally, one can see this by noting that the average benefit to men and women, across all values of  $I$  may be expressed as:

$$(6a) \quad B_m = \int E(\text{Benefit} | Male, I) dF(I|Male)$$

$$(6b) \quad B_f = \int E(\text{Benefit} | Female, I) dF(I|Female)$$



Equation (6) makes it clear that even if the treatment on the treated term is identical for men and women of the same  $I$ , integrating this benefit over different distributions of  $I$  will result in finding that  $B_f < B_m$  if women have a lower distribution of  $I$  (e.g., their characteristics make them less appropriate for the treatment). However, if the distributions of  $I$  are equalized for men and women (or whites and blacks) through reweighting, the comparison of average treatment benefits provides a test for prejudicial providers. This is the key empirical implication that we will test in our empirical work.

The above test 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 (5) depends only on the truncation point ( $I$ ). This would not be the case if the distribution of the unobservable factors determining treatment ( $\varepsilon$ ) differed by gender or by race. For example, if providers spend less time ascertaining a female patient's appropriateness for treatment, then the variance of the  $\varepsilon$ 's ( $\sigma_\varepsilon$ ) might be smaller for women. A smaller error variance for women would imply a lower truncated mean for any given truncation point – leading to lower expected benefits from treatment for women, despite facing the same threshold. This would look like prejudice against men (even though it is the opposite!) so it is important to rule out.

While we will maintain the single index assumption for our primary analyses, we will also test its validity empirically. One test for differences in the variance of unobservables follows naturally from assuming normality and estimating separate probit models by gender or race for receiving treatment. Because probit models identify  $\beta/\sigma$ , we would expect to observe a more muted relationship (smaller coefficients) between observable patient characteristics and the propensity to be treated for any group for which unobservable factors played a larger role in treatment. However, this is at best an indirect test since it relies on the assumption that the true  $\beta$  is the same across groups. For example, one might believe that inattention lowers the effect of both observables and unobservables on treatment for minorities, so that  $\beta/\sigma$  remains unchanged. Again, simply looking at the treatment propensity without estimating the benefits from treatment cannot clearly identify prejudice.

An alternative test for differences in the variance of the unobservables comes from estimating the effect of treatment on the treated patients (Equation 5). In our model, this treatment-on-treated effect is simply the mean of the truncated benefit distribution that lies above

the hurdle. Holding constant the proportion of patients lying above the hurdle (i.e., the propensity to be treated), the benefit for treated patients will tend to lie closer to the hurdle when there is less variance in the benefit distribution, resulting in a lower mean for the truncated distribution of benefits. In addition, smaller shifts in the benefit distribution will generate larger changes in the proportion above the hurdle when there is less variance in the benefit distribution, resulting in a flatter relationship between the propensity to be treated and the mean of the truncated benefit distribution. Therefore, if there is less variance in the unobservable factors used to determine treatment for minority groups, we would expect to see both smaller benefits from treatment and less of a relationship between benefits and treatment propensity among minority groups.

The graphical intuition for our model can be seen in Figure 1a for the case of no prejudice, and in Figure 1b for the case of prejudice against women. The expected benefit from treatment ( $B$ ) is given on the vertical axis, while the index  $I$  (which determines the propensity of being treated) is given on the horizontal axis. The thick curve in Figure 1a represents the treatment-on-the-treated effect for a patient with index  $I$ , that is it gives  $E(B | B > \tau)$ . Treatment-on-the-treated approaches the minimum threshold ( $\tau$ ) 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,  $X\beta_1 + Female\beta_2 = I + \tau$ . Figure 1b shows how treatment-on-the-treated differs for men and women when there is prejudice against women. The treatment effect among women is higher at every point, reflecting the fact that the benefit of treatment must exceed a higher minimum threshold for women ( $\tau_w > \tau_m$ ).

When there is no prejudice, as in Figure 1a, if men have a higher propensity to receive treatment ( $I_M > I_F$ ), then they will also have greater benefits from treatment ( $B_M > B_F$ ). But in the absence of prejudice, any two people with the same propensity will have identical treatment effects if treated. In contrast, when there is prejudice, as in Figure 1b, for any two people with the same propensity, the group who is discriminated against (females in this example) will always have greater treatment effects if treated. Thus, even if men have a higher propensity to receive treatment ( $I_M > I_F$ ), they may have smaller benefits from treatment ( $B_M < B_F$ ) if there is prejudice against women.

Our model shares the same notion of taste-based prejudice as Knowles, Persico and Todd (2001), and Balsa and McGuire (2003). In all three papers, the decision to take an action (search a motorist for contraband or treat a patient) is determined by whether the expected benefit exceeds a threshold, and prejudice exists when the threshold differs by race or gender. The model of racial bias in motor vehicle searches developed in Knowles et al. is conceptually similar to our model of prejudice, but differs in one important respect, which makes our empirical test of prejudice more complicated. In the Knowles et al. model, the returns to searching motorists in the absence of prejudice are equalized across all subgroups in equilibrium – otherwise, police would always search motorists with higher returns and motorists would react by carrying less contraband. The key difference in our setting is that patients are not choosing any action analogous to carrying contraband, so there will be some subgroups of patients (those with high propensity to be treated) who continue to have higher returns in equilibrium even in the absence of prejudice. By contrast, in Knowles et al. because of an equilibrium condition that relies on motorists responding to the threat of being searched, there is no distinction between the marginal and treatment on treated effects in the absence of prejudice. Thus, in the absence of prejudice, the Knowles et al. model implies that the returns to search are identical across subgroups unconditionally, while our model implies that the returns to treatment are identical across subgroups conditional only on the propensity to receive treatment ( $I$ ).

Our model differs from Balsa and McGuire (2003) in how we define statistical discrimination, or the alternative to prejudice. In Balsa and McGuire, statistical discrimination occurs when providers use different decision rules to determine a patient’s appropriateness for care; there is no outcomes based test. In our framework, finding evidence of these triage-rule differences is not informative about prejudice because it may be correct for physicians to use different triage rules, including using information on race and gender to determine appropriateness. This is analogous to how statistical discrimination is modeled by labor economists (Altonji and Pierret, 2001). In our framing, the test of prejudice is whether the benefit of treatment is larger for minority groups relative to others of the same appropriateness.

### **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, streptokinase 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, racial and gender disparities in treatments may directly translate into lost lives, and there is a rich tradition of studying racial and gender disparities in treatments and outcomes after heart attacks (Allison et al. (1996), Canto et al. (2000), Barnato et al. (2005), Vaccarino et al. (2005), Rathore et al. (2000), Skinner et al. (2005), Jha et al. (2007)).

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).

Third, because mortality post-AMI is high (survival rates at one year are less than 70 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 disparities 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 and receive some medical care, thereby allowing us to evaluate the case for provider prejudice comprehensively. “911” operators and emergency medical teams are especially trained to recognize the symptoms of heart attacks. This would be less true of chronic diseases that progress gradually. Nor do we believe that patient preferences matter as much for treating heart attacks—at least 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. While providers may specialize in the use of surgical or medical management of heart attacks, as in Chandra and Staiger (2007), the fact that patients are generally taken to the nearest hospital for treatment, renders the nature of treatment received as exogenous to the patient preferences. This feature of heart attack treatments would not be true of cancer therapies where two clinically identical patients may chose different therapies based on their idiosyncratic valuation of side-effects and treatment duration. Finally, one advantage of studying heart attacks in the Medicare population is that black/white differences in insurance (at least for in-hospital care) aren’t a source of confounding.

### ***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).

#### IV. Estimation

##### A. Basic Specifications

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 by obtaining the index from a probit model that regresses whether a patient received reperfusion within 12 hours of the heart attack on gender, race, and all the CCP risk-adjusters (X):

$$(7) \ Pr(Reperfusion_i) = F(X_i\Phi + (Female*X_i)\phi + (Black*X_i)\beta + (Female*Black*X_i)\theta + u_i)$$

In equation (7), the effect of race and gender on the probability of receiving reperfusion is allowed to vary across every comorbidity. Differences in the coefficients by race or gender may reflect ignorance, prejudice, or actual knowledge regarding the benefits of treatment (for example, the effect of diabetes or age may operate differently in blacks). Alternatively, larger coefficients on X for women and blacks could reflect lower variance in the unobservable factors determining treatment (since probit models estimate  $\theta/\sigma$ ). Finally, if providers maximize a different benefit in one group than another (for example, maximize survival in whites, but maximize survival minus costs in blacks) then we could observe different probit coefficients for blacks and whites. To account for these potential concerns in our empirical work, we allow for full gender-race interactions with X in equation 7. However, as we note in Section V, we find little support for these concerns in the empirical work: the predicted propensity to receive reperfusion from the fully interacted model was correlated 0.997 with the propensity from the model with no race or gender interactions.

Our test for taste-based prejudice is to determine, among patients with the same propensity to be treated, whether the survival benefit to reperfusion is greater for women relative to men, or blacks relative to whites. Survival is measured as a binary variable that measures survival to a certain date (e.g. survival to 7 days, or survival to 1 year). This suggests estimating models of the following type for women (equation 8a) and blacks (equation 8b), and focusing on the interaction term with reperfusion:

$$(8a) \ Survival_i = \alpha_0 + \alpha_1 Reperfusion + \alpha_2 (Reperfusion * Female) + X_i + e$$

$$(8b) \ Survival_i = \alpha_0 + \alpha_1 Reperfusion + \alpha_2 (Reperfusion * Black) + X_i + e$$

In equation 8a,  $\alpha_1$  is the survival gain from reperfusion for men,  $\alpha_2$  is the differential benefit for women ( $\alpha_1 = B_m$ ,  $\alpha_2 = B_f - B_m$ ), and X includes full gender-race interactions (including the main effect of race and gender) with the CCP risk-adjusters.<sup>3</sup> However, as we noted in the theory section,  $\alpha_2$  is estimated over different distributions of the propensity to receive reperfusion for men and women, and is therefore not a precise test of our model. In other words, evidence that  $\alpha_2 < 0$  is not sufficient evidence to conclude that there is no prejudice against women, if women are generally less appropriate for treatment and consequently, have lower treatment propensities. In the absence of prejudice, our model states that the treatment on treated (TT) effect is the same only for men and women with the *same propensity* to be treated, or equivalently for any population of men and women that have the same propensity distributions (e.g. are matched on propensity to be treated).

To test whether the treatment-on-treated effect is the same for men and women using the framework of equation (8) it is key that this parameter be estimated for men and women with similar propensities to treated women. Therefore, we reweight the above equations at the time of estimation using the method of Barsky et al. (2002) to make the propensity distribution of both treated and untreated males (or whites) look like the propensity distribution of treated females (or blacks), and also reweight untreated female (or black) propensities to look like the treated female (or black) distribution of propensities. In other words, the treatment effect for men and women (or whites and blacks) is estimated for a population with the propensity distribution observed among the treated in the female (or black) population. We compute the weights by calculating the value of the 100 percentiles of the propensity distribution among treated females. By construction 1 percent of treated women are in each of these percentiles. Suppose that  $m_p$  percent of one of the other three groups (treated or untreated men, or untreated women) are in the  $p$ th percentile of the treated female distribution of propensity. For such patients we will assign them a weight of  $1/m_p$  at the time of estimation. We provide evidence of the success of this strategy in the results section.

## **B. Allowing for Treatment Effect Heterogeneity**

---

<sup>3</sup> Estimating equation (8) with claims data as opposed to chart data would cause us to overestimate the effect of the treatment: not only do healthier patients receive treatment, patients also have to survive to the point of receiving treatment (see McClellan et al., 1994 for a discussion). In Section V, we demonstrate that our estimates are similar to those of randomized clinical trials.

An alternative to the reweighting strategy, which examines whether the treatment-on-treated effect is the same for men and women (or whites and black) with the same propensity to be treated, is to allow the treatment effect in equation 8 to depend directly each individual's propensity to be treated (i.e. to control for the interaction between the index and treatment directly). This requires us to estimate equations of the form:

$$(9a) \quad Survival_i = \alpha_0 + \alpha_1(Reperfusion * g(I)) + \alpha_2 (Reperfusion * Female) + XII + e$$

$$(9b) \quad Survival_i = \alpha_0 + \alpha_1(Reperfusion * g(I)) + \alpha_2 (Reperfusion * Black) + XII + e$$

In equation 9a,  $\alpha_1 * g(I)$  is the survival gain from reperfusion for men, and  $\alpha_2$  is the differential benefit for women. Thus, this specification allows the return to reperfusion to depend on  $g(I)$  for both men and women, and estimates whether women have different returns to reperfusion conditional on their propensity to be treated as captured by  $g(I)$ . 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. In this model, the coefficients of interest are those on *Reperfusion\*Female* or *Reperfusion\*Black* (the direct effect of reperfusion for each propensity percentile would be given by the *Reperfusion\*Percentile* indicators).<sup>4</sup> Moreover, we can recover the *Reperfusion\*Percentile* effects to examine whether the benefits of reperfusion are increasing in the index (as they should if the Roy-model framework underlying our theoretical model is a correct assumption.) Similarly, we explore the possibility that the  $g(I)$  function exerts different effects on survival by gender and race.

### C. Learning from Parametric Structure

Placing simple parametric structure on  $g(I)$  in equations 9a and 9b provides an alternative way to (a) evaluate whether the estimates of treatment of the treated vary by gender or race, (b) estimate the degree to which unobservable factors affect the benefit from treatment, and (c)

---

<sup>4</sup> Because we have controlled for  $g(I)$  in the regression, there is no further need to reweight the distributions of men or whites to look like that of treated women or blacks.



quantify any gender or racial differences in the importance of unobservable factors. We can estimate  $g(I)$  parametrically by assuming normality, which implies a tobit structure:<sup>5</sup>

$$(10) \quad g(I) = \tau_1 + \sigma \left[ I + \frac{\phi(I)}{\Phi(I)} \right]$$

We calculate the last term in brackets from the estimated propensity to receive reperfusion (equation 7), and we allow for the two unknown parameters ( $\tau_1$ ,  $\sigma$ ) by including reperfusion and reperfusion interacted with  $\left[ I + \frac{\phi(I)}{\Phi(I)} \right]$  in equations 9 (we also add  $\left[ I + \frac{\phi(I)}{\Phi(I)} \right]$  to the controls in X).

The direct effect of reperfusion identifies  $\tau_1$ , and the interaction of reperfusion with  $\left[ I + \frac{\phi(I)}{\Phi(I)} \right]$  identifies  $\sigma$ . In these models the coefficient on *Reperfusion\*Female* or *Reperfusion\*Black* continue to identify the presence of prejudice, but in a specification where the benefits to reperfusion are allowed to vary on the index. Moreover, because the second term picks up the effect of the index on the survival benefit from treatment, a positive coefficient is consistent with our model where providers use a single index to rank patients on the basis of their benefit from reperfusion, so that patients with a higher propensity to receive reperfusion also have a higher benefit from reperfusion.

Finally, we explore the possibility that the  $g(I)$  function exerts different effects on survival by gender and race. The parametric analog adds an interaction of *Reperfusion* \*  $\left[ I + \frac{\phi(I)}{\Phi(I)} \right]$  with gender/race. This three-way interaction provides a test of whether the variance in unobservable factors affecting the benefits to reperfusion ( $\sigma$  in equation 10) differs by gender or race. If unobservable factors are more important for one group, the effect of reperfusion on survival will depend more strongly on their propensity to receive care.

---

<sup>5</sup> Equation 10 is written in terms of the index (I) estimated from the probit model. Recall that  $\Pr(\text{Treatment}=1) = \Pr(-\varepsilon < I)$  and  $g(I) = \tau_1 + I + E(\varepsilon | -\varepsilon < I)$ . If we assume that  $\varepsilon \sim N(0, \sigma^2)$ , then the index (I) estimated from the probit corresponds to  $(I^*/\sigma)$ , where  $I^*$  is the index in our theoretical model. Inserting this into equation for  $g(I)$  yields its parametric version:

$$g(I) = \tau_1 + \sigma \left[ I + \frac{\phi(I)}{\Phi(I)} \right] = \tau_1 + \sigma \left[ \frac{I^*}{\sigma} + \frac{\phi(I^*/\sigma)}{\Phi(I^*/\sigma)} \right] = \tau_1 + I^* + \sigma \frac{\phi(I^*/\sigma)}{\Phi(I^*/\sigma)} = \tau_1 + I^* + E(\varepsilon | -\varepsilon < I^*)$$

Where the last term reflects the mean of a truncated normal.

## V. Results

In Table 1 we report some basic characteristics of our sample by sex and race. Women are older than men at the time of their first heart attack, and perhaps consequently, they're also substantially sicker, as measured by the presence of heart failure and hypertension. Blacks are younger than whites, but have significantly higher rates of heart failure, diabetes, and hypertension. We summarize how much sicker women and minorities are relative to men and whites by comparing predicted 30-day survival rates (where the prediction is made using all the CCP data, but not using race or gender). Predicted survival rates are similar for whites and blacks, and somewhat lower for women relative to men. Actual survival rates following a heart attack follow a similar pattern, with whites and blacks having similar survival while women have somewhat lower survival rates relative to men. Thus, these differences in underlying sickness translate into small if any difference in survival between the groups. However, there are large differences in actual reperfusion rates, as can be seen in the last row of Table 1. This average difference in the propensity to receive reperfusion between groups is the principal reason why estimation of equation 7 in the absence of weighting does not provide a powerful test for provider prejudice.

Table 2 reports probit estimates of the effect of sex and race on the receipt of reperfusion within 12 hours of arriving at the hospital—marginal effects are reported in brackets. Unadjusted, women are 4.5 percentage points less likely to receive reperfusion, while blacks are 6.1 percent points less likely to receive this procedure relative to whites. Adjusting for all the CCP risk-adjusters closes this gap for women but not for blacks. However, as noted in the theory section, the probit coefficients on race and sex capture both differences in the benefit from treatment, as well as potential differences in the treatment hurdle for these groups. As such, these negative coefficients do not provide us with a test of prejudice. Finally, in the last column of Table 2 we add a Female\*Black interaction. This interaction, which is insignificant, demonstrates that the racial disparity in treatment does not vary by gender.

In Table 3 we present our key results from the estimation of equation 8 for differences by gender (Panel A) and race (Panel B). The coefficient on Reperfusion reports the survival benefit of Reperfusion for men (or whites in Panel B), and the interaction effect with sex or race reports the differential benefit for women (and blacks in Panel B). Focusing on the unweighted results in Table 3, we see that Reperfusion improves the probability of surviving 30 days for men by 3.4

percentage points. But for women the effect is 1.4 percentage points less. If there were underuse in women, we would have expected the opposite – higher returns to reperfusion for women. The unweighted results by race are similar. The effect of reperfusion is 2.6 percentage points less for blacks. Thus, the unweighted results find significantly lower effects of reperfusion for both blacks and women, suggesting that if anything there is overuse among blacks and women relative to whites and men – that is, the lower returns for blacks and women suggest that they face a lower hurdle in the decision to treat.

However, as noted by our theory, this is not a sufficient test for prejudice because the treatment-on-the-treated effect will tend to be larger among men because men have a higher propensity to be treated (since  $g(I)$  is increasing in  $I$ ). In column 2, we use the method of Barsky et al. (2002) to reweight the distributions of propensities in untreated females, treated males, and untreated females to resemble that of treated females. Panel B performs the analogous weighting for blacks. As expected, the estimated treatment effect for males and whites (the direct effect of reperfusion) declines after reweighting, because we are now estimating the treatment effect in a (reweighted) population with a lower propensity to receive treatment. Correspondingly, the estimated differences in the treatment effect between men and women or whites and blacks (the interaction of reperfusion with female or black) is smaller, and we are no longer able to reject the null hypothesis that the effect of reperfusion for patients receiving reperfusion is the same by gender and race. These weighted results are the direct test of the theoretical model, and find no evidence of prejudice. Prejudice and underuse would imply higher returns to treatment for females and blacks, but our point estimates suggest the opposite. Even the upper bound of the 95% confidence interval on these estimates would suggest very little prejudice against females or blacks.

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. To ascertain the credibility of this assumption we can compare the estimates in Table 2 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 reduces 35-day mortality by 1.9 percentage points (an 18 percent reduction in mortality). In the

second column of Table 3 we estimate the effect of reperfusion (pooling across men and women) on 30-day survival as 1.9 percentage points (s.e.=0.2 percentage points). We focus on the second column because randomization balances the distribution of characteristics between the treatment and control arms of the trial.<sup>6</sup> We take this evidence as supporting the case for simpler regression models that invoke a ‘selection on observables’ assumption.

In Figure 2, we examine the success of our weighting strategy to equalize the distribution of propensities by gender (Panel A) and by race (Panel B). The unweighted and reweighted propensity distributions are plotted by gender and race for the untreated (left hand side) and for the treated (right hand side). For reference, we have also included the propensity distribution for treated females (or blacks), since this is the target distribution that is being matched with reweighting. In the unweighted distributions, it is apparent that whites and men have higher propensities to receive treatment, particularly among patients receiving treatment. Therefore, our model would predict that unweighted estimates of the treatment effect will tend to be larger for whites and men, even in the absence of any prejudice. In contrast, the re-weighted distributions (where greater weight is put on the observations that are similar to those of the treated minority group) look identical to those for treated women and blacks.

An alternative to rebalancing the distribution of propensities is to control for  $g(I)$  directly. In column (3) of Table 3 we include 100 indicator variables for the percentiles of  $I$  as a non-parametric way to control for appropriateness. We also included interactions of these indicators with Reperfusion to test whether the benefits of reperfusion increase with  $g(I)$  (this is why we cannot report the direct effect of reperfusion), as they should if providers are ranking patients correctly and then working down the distribution of appropriateness. The results of non-parametric specification column (3) is identical to those obtained from propensity score reweighting suggesting that concerns about imperfect reweighting aren’t a source of bias. Moreover, in Figure 3, we plot the estimated benefit from reperfusion at every percentile of  $I$ , using local linear regression to smooth the effect across percentiles. The benefit from reperfusion is increasing with  $I$ , as is required by our model and demonstrates that physicians are able to rank patients so that patients who benefit most from reperfusion are most likely to receive the

---

<sup>6</sup> Standard propensity score weighting to make the distribution of propensities for the untreated resemble that of the treated (that is, weighting the untreated by  $\Pr(\text{Reperfusion})/1-\Pr(\text{Reperfusion})$ ), estimates the effect of reperfusion as 2.1 percentage points which is very similar to the trial estimate of 1.9 points.

treatment. There is evidence of ‘harm’ for patients with very low propensities of being treated (suggesting that the hurdle is being set at a level where survival is reduced by the treatment), but there are very few patients getting treatment in this range of propensities.

In Table 4 we examine results that rely on parametric structure. In columns (1) and (2) we control for a (demeaned) parametric version of  $g(I)$  that was derived under the assumption of normality in equation (10). We demeaned the selection term in order to report estimates that are comparable to Table 3 – the estimated treatment effect for an average patient being treated (if the selection term were not demeaned, it would identify the value of the hurdle). The negative signs on the female/black\*reperfusion variables indicate that if anything, women and blacks have lower benefit from treatment, but that we cannot reject the null that the benefits are the same for minority and non-minority patients. In these models, the interaction of reperfusion\*selection term provides a parametric alternative to evaluating whether the benefits to reperfusion are increasing in physicians determination of appropriateness (which is captured by the selection term). The positive and significant coefficients on this interaction confirm that this is true and reinforce the message in Figure 3. None of these specifications alter the key results of Table 3— in fact, the estimated effect of reperfusion and estimated difference in this effect for women and blacks are almost identical to what was estimated with the reweighting strategy reported in Table 3.

The specifications reported in Table 4, column 2 allow  $g(I)$  to have different effects on the benefits to reperfusion for females and blacks through the triple interaction between  $g(I)$ , Reperfusion, and female (or black). There is no significant difference between whites and blacks:  $g(I)$  is associated with larger effects of reperfusion for both groups. In contrast, while the benefits to reperfusion are significantly increasing in  $g(I)$  for men, the effect for women is even larger. In the context of our model this implies that factors that are used by providers (but unobserved to us) play a larger role in explaining benefits for women. . We will examine this result in more detail in the section below.

## **VI. Understanding Differences in Benefits**

The above results may be summarized as finding no evidence of higher returns from treatment for women and blacks of similar appropriateness. If anything, we have consistent but statistically insignificant evidence that the benefits for these groups is lower than that for men

and blacks. Before concluding that this evidence represents evidence of no-prejudice, we have to rule out a number of other explanations. Some of these reasons may actually reflect prejudice and it is important to rule them out prior to concluding statistical discrimination.

### *A. Differences in Triage Rules*

First we consider the idea that physicians are incorrectly using different triage rules to rank patients (by sex or race) for treatment. If this were true, the reason for the slightly lower benefits for women and blacks could be that less appropriate patients from these populations were getting the treatment because physicians spend less time and effort in doing the triage for these groups. These pressures may be exacerbated by the demands of working in a high-pressure, time-sensitive environment where the potential for subconscious biases such as ‘implicit discrimination’ is high (Bertrand, Chugh and Mullainathan, 2005). In such an environment providers may subconsciously use a higher threshold for minority patients but also be biased in how patients are ranked as appropriate for care – leading to both fewer women and blacks receiving reperfusion and less benefit.<sup>7</sup> We can evaluate this concern directly by estimating separate probit models of receiving reperfusion by race and gender (e.g. allowing risk factors to have different impacts on propensity to treat by gender and race), and then examining whether the predicted values from this more flexible model is similar to a model in which race and gender enter only directly (as would be the case if they only effected the threshold). The predicted values are very similar, suggesting that physicians are not using different triage rules for women and blacks: in the full sample, the correlation between the fitted values is 0.996. For women it is 0.993 and for males it is 0.9958. Similarly, the correlations are 0.949 for blacks and 0.998 for whites.<sup>8</sup>

The similarity of rankings from the two models also helps us rule out other forms of triage-bias. If providers were maximizing different dimensions of benefit in different populations

---

<sup>7</sup> More malevolent versions of this story would involve providers offering treatment to the most appropriate white patients and randomly offering it to blacks, or, in the extreme, withholding it to the most appropriate black patients.

<sup>8</sup> In theory, it may be the case that on average, the common-effects model produces estimates of receiving reperfusion that are similar to those from the race-specific model, but there are significant departures in the tails of the appropriateness distribution. In Jha et al. (2007), we explore this possibility for a number of heart attack treatments, and note that the two models give very similar results across the entire distribution of appropriateness.

(e.g. maximizing survival in blacks but survival and quality of life in whites) then the coefficients from the separate probits would be different and result in weakly correlated predictions.

Finally, the similarity of predictions where the effects of comorbidities are allowed to vary by race and gender to those where they are not, supports the view that the variance of the unobservable characteristics (to clarify, unobserved to us, not to the physician) is similar across minority and majority groups (as we noted in Section IV, probit coefficients identify  $\beta/\sigma$  so models estimated separately by gender and race will produce larger coefficients for the group with the smaller variances, i.e. the group for whom providers spend less time ascertaining appropriateness). However, this would be a premature assessment if providers are more likely to underweight all information, both the parts that are observed to us and the parts that are not, at the time of triaging minority patients. In this case, *both* the betas and the sigmas would scale down and the predicted effects across models with and without race and gender specific effects would be similar. But, as discussed in Section IV, if the variance of unobservables were smaller in minority groups, then the slope of curve relating treatment benefits to the propensity of receiving treatment would also be smaller in minority groups. However, as was seen in column (2) of Table 4, while women and blacks benefit less overall from reperfusion, the benefit for blacks and women does not depend less on their propensity – and for women appears to depend more on the propensity to receive reperfusion.

### *B. Differences in Knowing Who Benefits From Treatment*

An alternative explanation for the fact that minorities and women receive lower returns is that providers are unaware of how to rank patients: in other words, even though the previous section demonstrates that providers use the same decision rule to rank patients, they should not. This form of bias would occur if medical textbooks and clinical trials are biased towards studying the etiology of disease in whites and men, and physicians assume that this knowledge applies equally to women and blacks. In this case, the return to reperfusion for women or blacks would be lower on average, and again would depend less on their propensity to receive care (and could even cross, e.g., with low propensity women benefiting more from treatment than low propensity men). However, as was just discussed in the previous section, the benefit for blacks and women does not depend less on their propensity – and for women appears to depend more

on the propensity to receive reperfusion. Figure 4 examines this possibility graphically by reporting the estimated benefit from reperfusion at every propensity to receive reperfusion, estimated as in Figure 3 but done separately by gender and race. The evidence in Figure 4 demonstrates that if anything the benefits to treatment are lower for female and blacks at all propensities (except the very highest for females). Moreover, while estimates for blacks and whites are roughly parallel, the relationship between propensity and the effect of reperfusion is steeper for females than for males – the opposite of what we would expect if physicians were using the wrong decision rule to treat women.

### *C. Differences in Follow-up Care*

Next we investigate the concern that reperfusion provides similar short-run benefits to blacks and women, but that a lack of follow-up care leads to worse outcomes after the patient is discharged from the hospital (usually within a week of the AMI). If this were the case, we would expect the effect of treatment on short-term mortality to be similar, while the effect of treatment on long-term mortality would diverge. We explore the time-path of survival in Table 5 and Figure 5, where we estimate separate models for survival at horizons of 3 days through 1 year. We find that for women, receiving reperfusion is actually associated with harm during the acute phase of treatment (the first 3 days). The clinical literature notes that strokes are a possible side-effect of reperfusion, and while we cannot directly verify that women are having strokes, the early mortality at 3 days is consistent with this explanation. After this initial phase, there is no further deterioration of benefits in women; lower benefits for women are largely the consequence of higher acute mortality associated with reperfusion. Thus, women appear to face greater acute risk from the procedure, but then experience similar survival benefits in the post-acute phase.

Blacks also experience somewhat greater mortality risk from the treatment in the acute phase, although the difference from whites at 3 days is not significant. In contrast to the result for women, however, blacks in our sample appear to receive lower (but not harmful) benefit in the post-acute phase—their estimated benefits of treatment remain small even at one year, and the gap in the treatment effect between whites and blacks continues to grow through 90 days. Thus, it appears that (like women) reperfusion may harm blacks more during the acute phase of treatment, but also (unlike women) blacks may benefit less from the treatment after the acute phase. One possible reason for the lack of post-acute benefit of reperfusion among blacks could



be a lack of follow-up care. But another possible explanation, which we discuss further in the conclusion, is that blacks are known to take longer to arrive at the hospital because of ambulance delays and other factors, and trial evidence suggests that the effectiveness of reperfusion declines sharply with time since the heart attack.

#### *D. Differences in Hospital Skill*

One potential explanation for the lower benefits for blacks is that blacks and whites go to different hospitals, and that hospitals which treat blacks are not good at the management of heart attacks. This explanation is motivated by Chandra and Skinner (2004) and Skinner et al. (2005) who present evidence that minority serving hospitals aren't particularly good at the management of heart-attacks: these hospitals exhibit lower 90-day survival for both black and white patients (this explanation cannot be a determinant of gender disparities in care, since men and women go to similar hospitals). To explore this theory, we modified equation 8 to include hospital fixed effects as well as hospital fixed effects interacted with whether a patient received reperfusion. With these fixed effects we allow hospitals to vary in quality and also in their reperfusion specific expertise. The estimates are reported in Table 6. Including these fixed-effects resulted in nearly identical estimates to the results without fixed-effects reported in Table 5, and demonstrates that the fact that women and minorities receive slightly lower returns is a within hospital phenomena.

#### *E. Differences in Cost of Treatment*

Finally, we explored the possibility that the benefit from treatment, and therefore the decision to treat, may also depend on the costs of treatment. Thinking about the costs as well as the health benefits of treatment is useful for two reasons. First, differences in the cost of treatment by race or gender may offset the survival differences, e.g. the larger impact of reperfusion on survival for men may be offset by higher costs of doing the procedure in men. Second, prejudice could appear in a more subtle form if medical care providers placed a smaller weight on costs in the decision to treat whites and men – implicitly placing a higher value on

their life. Empirically, we found that the story about costs did not affect the results – because the effect of reperfusion on cost was similar for men and women and for whites and blacks.<sup>9</sup>

## VII. Conclusions

We used a simple economic intuition to develop a framework that provides a way to distinguish between prejudice and statistical discrimination. Our framework relied on the Beckerian insight that prejudice can be ascertained by comparing outcomes for the marginal patient (a test that cannot be implemented without observing everything that providers observe), but recast this intuition by combining information on the propensity to receive treatment with estimates of the benefits from treatment for those patients who receive the treatment. This outcome-based test for prejudice is in contrast to much of the empirical literature on prejudice.

In contrast to what models of prejudice would predict, we find that providers appear to know of the lower benefits that accrue to women and blacks and consequently use gender and race to correctly determine patient appropriateness for care. Even the upper bound of our estimates does not support the case for prejudice against women and blacks. In fact, if anything, we find that women and blacks have slightly lower benefits relative to a model of pure statistical discrimination. There are a number of reasons that we might observe such a pattern. Medical care providers may be aware of the lower benefits of treatment in minorities, but choose to increase treatment rates for these groups (or lower treatment for other groups) because of perceived equity concerns. This behavior may be reinforced by malpractice concerns if providers worry that their decision not to treat a patient based on their race or gender could expose them to litigation. Alternatively, since few medical trials separately estimate the benefits of treatments by race or gender, medical care providers may be unaware of the extent to which the benefits of treatment are lower among women and minorities, or may be unwilling to make treatment decisions without a documented medical reason based solely on a statistical association

---

<sup>9</sup> To perform this test, we redefined benefit as the benefit net of costs from treatment:  $B = S - \lambda C$ , with  $S$  representing the survival benefit from treatment, and  $C$  the cost for a patient.  $\lambda$  is a measure of survival per dollar tradeoff (when  $\lambda=0$ , the medical care provider focuses solely on survival benefits). A minimum value for a life year commonly used in cost-effectiveness studies would be about \$20k per life year, which implies a value of  $\lambda=0.01$ . For reasonable values of  $\lambda$ , between 0 and .1, we were unable to reject the hypothesis that the returns to men and whites were similar to those for women and blacks. These results are available upon request.

between benefits and gender or race. In fact, the absence of such trials or medical knowledge may be a form of macro-prejudice against minorities (Healy, 1991), but this is a different explanation than bias in the clinical encounter.

Other mechanisms for the slightly lower benefits for women and blacks are less benign. For example, the returns to any treatment might be low unless other treatments are also being provided (for example, a provider may be less likely to provide reperfusion to black patients knowing that they are less likely to receive other complimentary therapies such as high-quality diabetes care or follow-up care after discharge from the hospital), leading to an equilibrium in which no treatments are ever provided to minorities. While this remains a possibility, in earlier work we noted that racial disparities in care within hospitals and within areas are not correlated (see Baicker et al., 2004). For example, the racial disparity in the receipt of high-quality care was correlated less than 0.10 with the disparity in colorectal cancer screening or cardiac catheterization. The lack of correlation in disparities would suggest that explanations grounded in self-fulfilling equilibria are not first-order. It is also possible that lower levels of provider effort cause lower benefits in minorities. But we did not find evidence that providers triage patients into care differently, as would be the case if they used all the information at the time of admission for men and whites, but not for women and blacks.

In future work it will be important to document the precise biological mechanism that leads to the slightly lower returns to treatment for female and black heart attack patients. For women, we interpret the risk of early mortality from reperfusion as evidence that they are at higher risk of having a stroke—and while the presence of stroke-risk is consistent with the trial results (see FTTCG, 1994) this would need to be established more rigorously. Such a mechanism suggests that alternative treatments must be developed for women that reduce the probability of early strokes. For blacks, lower long-term benefits from reperfusion may reflect delays in obtaining treatment. Sheifer et al. (2000) and Ting et al. (2008) demonstrate that it takes blacks longer to arrive at the hospital because of ambulance delays or because of delays in recognizing the symptoms of a heart attack. In our data 54 percent of whites arrived within 6 hours of the start of reported chest pain relative to 43 percent of blacks, and blacks were more likely to be coded as having no chest pain (22 percent versus 15 percent) or having unknown time since chest-pain at the time of arrival (9 percent versus 13 percent). Because providers do not know the exact status of the delay confronting minority patients they may treat black heart attacks

identically to white AMIs, but doing so could result in lower benefits for black patients if they are less appropriate for reperfusion by virtue of their delayed arrival. Efforts to reduce late arrivals for minority patients are very different from those that emphasize the prejudice channel. The policy prescription for prejudice is to reduce underuse by performing more treatment in minorities or better ranking of patients for treatment. For example, the IOM's report *Unequal Treatment* recommends increasing cultural competency for providers and hiring more minority physicians as ways to make progress on these channels. Under statistical discrimination, treatment disparities exist because of lower minority appropriateness for care. Thus, policies should be focused on improving their appropriateness, rather than increasing their treatment rates (which will only result in prejudice against the majority and could in theory even harm minority health).

While we reject the role of prejudice in explaining treatment disparities for heart attacks, we are cautious about generalizing these conclusions to treatment disparities across all medical conditions. It is certainly possible that prejudice may be more pertinent in an ambulatory setting where patient-doctor communication is more relevant for treatment decisions. Alternatively, prejudice may be less relevant in ambulatory settings where the physician has more time to make a decision, which helps to reduce the scope for implicit bias, which is most prevalent in a time-sensitive environment (Bertrand, Chugh and Mullainathan, 2005). Regardless of the specific context and explanation, our paper provides a general economic framework to evaluate the role of a variety of bias-based explanations for disparities in care, versus less pernicious alternatives. In theory, this framework can be used to assess the role of prejudice in other settings such as hiring, loan applications, and admissions decisions. In all of these settings, disparities in treatment rates alone cannot distinguish between prejudice and statistical discrimination, but can do so when combined with estimates of the benefits from treatment.

## References

- Altonji, Joseph G, and Charles R. Pierret. 2001. "Employer Learning and Statistical Discrimination." *Quarterly Journal of Economics* 116(1): 313-350.
- Allison, JJ, CI Kiefe, RM Centor, JB Box, RM Farmer. 1996. "Racial Differences in the Medical Treatment of Elderly Medicare Patients with Acute Myocardial Infarction." *Journal of General Internal Medicine* 11(12): 736-43.
- Anderson, R. D., and C. J. Pepine. 2007. "Gender Differences in the Treatment for Acute Myocardial Infarction: Bias or Biology?" *Circulation* 115(7): 823-826.
- Antonovics, Kate, and Brian Knight. 2009. "A New Look at Racial Profiling: Evidence from the Boston Police Department." *The Review of Economics and Statistics* 91(1): 163-177.
- Anwar, S, H Fang. 2006. "An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence." *The American Economic Review*. 96(1): 127-151.
- Ayanian, JZ, I Udvarhelyi, C Gatsonis, C Pashos, A Epstein. 1993. "Racial Differences in the Use of Revascularization Procedures After Coronary Angiography." *Journal of the American Medical Association* 269(20): 2642-6.
- Baicker, Katherine, Amitabh Chandra, Jonathan Skinner, and John E. Wennberg. 2004. "Who You Are and Where You Live: How Race and Geography Affect the Treatment of Medicare Beneficiaries." *Health Affairs*.
- Balsa, Ana I., and Thomas G. McGuire. 2003. "Prejudice, Clinical Uncertainty and Stereotyping as Sources of Health Disparities." *Journal of Health Economics* 22(1): 89-116.
- 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.
- Barsky, Robert B., John Bound, Kerwin Charles, and J.P. Lupton. 2002. "Accounting for the black-white wealth gap: A nonparametric approach." *Journal of the American Statistical Association* 97 (459): 663-673.
- Bertrand, Marianne, Dolly Chugh, and Sendhil Mullainathan. 2005. "Implicit Discrimination." *American Economic Review* 95(2): 94-98.
- Canto, JG, JJ Allison, CI Kiefe. 2000. "Relation of Race and Sex to the Use of Reperfusion Therapy in Medicare Beneficiaries with Acute Myocardial Infarction." *New England Journal of Medicine* 342:1094-1100
- 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*.
- Chen, Jersey, Saif S. Rathore, Martha J. Radford, Yun Wang, Harlan M. Krumholz. 2001.

- “Racial Differences in the Use of Cardiac Catheterization after Acute Myocardial Infarction.” *New England Journal of Medicine* 344: 1443-1449.
- 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.
- Fincher, C, JE Williams, V MacLean, JJ Allison, CI Kiefe, JG Canto. 2004. “Racial Disparities in Coronary Heart Disease: A Sociological View of the Medical Literature on Physician Bias.” *Ethnicity and Disease* 14(3): 360–71.
- Green, Alexander, Dana Carney, Daniel Pallin, Long Ngo, Kristal Paymond, Lisa Iezzoni, Mahzarin Banaji. 2007. “Implicit Bias among Physicians and its Predictions of Thrombolysis for Blacks and Whites Patients.” *Journal of General Internal Medicine*.
- Healy, Bernadine. 1991. “The Yentl Syndrome,” *New England Journal of Medicine* 325: 274-276.
- Heckman JH. 1998. “Detecting Discrimination.” *Journal of Economic Perspectives* 12(2):101-116.
- 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.
- Jha, AK, ES Fisher, Z Li, EJ Orav, and AM Epstein. 2005. “Racial Trends in the Use of Major Procedures Among the Elderly.” *New England Journal of Medicine* 353:683- 91.
- Knowles, John, Nicola Persico, and Petra Todd. 2001. "Racial Bias in Motor Vehicle Searches: Theory and Evidence." *Journal of Political Economy* 109(1): 203-232.
- List, John A. 2004. “The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field,” *Quarterly Journal of Economics* 119(1): 49-89.
- 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.
- Munnell, Alicia H. & Geoffrey M. B. Tootell & Lynn E. Browne & James McEneaney, 1996. "Mortgage Lending in Boston: Interpreting HMDA Data," *American Economic Review*, 86(1): 25-53.
- 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.
- Rathore, SS, AK Berger, KP Weinfurt, et al. 2000. “Race, Sex, Poverty, and the Medical Treatment of Acute Myocardial Infarction in the Elderly.” *Circulation* 102:642-8.
- Schulman, KA, JA Berlin, W Harless, et al. 1999. “The Effect of Race and Sex on Physicians’

- Recommendations for Cardiac Catheterization." *New England Journal of Medicine* 340(8): 618–26.
- Sheifer, S. E., S. S. Rathore, et al. 2000. "Time to Presentation With Acute Myocardial Infarction in the Elderly: Associations With Race, Sex, and Socioeconomic Characteristics." *Circulation* 102(14): 1651-1656.
- 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*.
- Smedley, B, A Stith, A Nelson, eds. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington DC: Institute of Medicine; 2003.
- Ting, H. H., E. H. Bradley, et al. 2008. "Factors Associated With Longer Time From Symptom Onset to Hospital Presentation for Patients With ST-Elevation Myocardial Infarction." *Arch Intern Med* 168(9): 959-968.
- Vaccarino, V, SS Rathore, NK Wenger. 2005. "Sex and Racial Differences in the Management of Acute Myocardial Infarction." *New England Journal of Medicine* 353:671-682
- Van Ryn, M, and SS Fu. 2003. "Paved With Good Intentions: Do Public Health and Human Service Providers Contribute to Racial/Ethnic Disparities in Health?" *American Journal of Public Health* 93(2): 248–55.

## 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  $\geq$  100 mm Hg, creatinine <1.5, 1.5-2.0 or >2.0 mg/dL) and included dummy variables for missing data.

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



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

Figure 1a: Illustration of Unprejudiced Provider Behavior

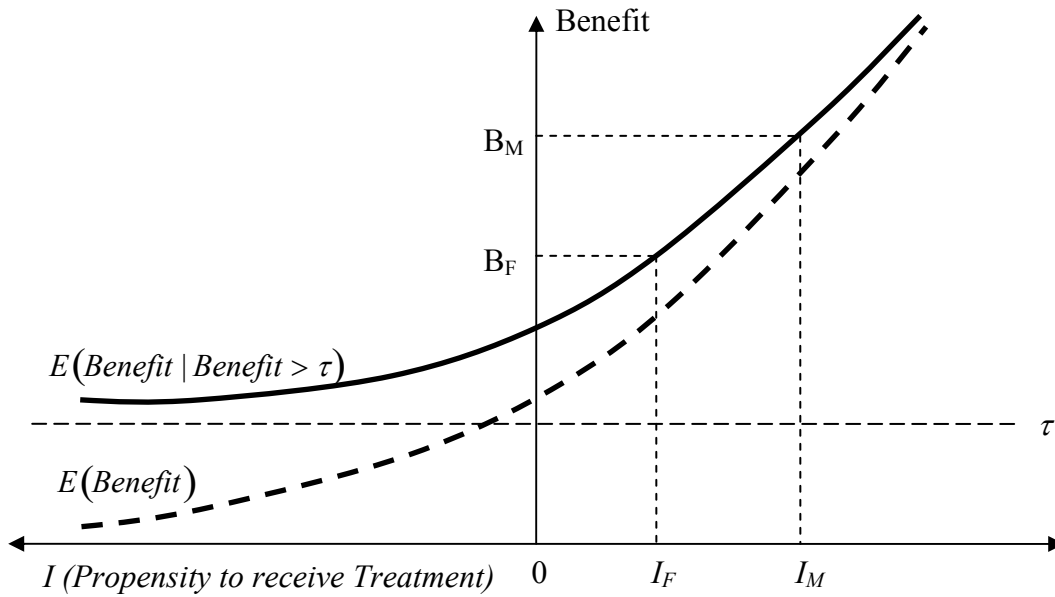


Figure illustrates the relationship between the expected benefit from treatment ( $B$ ) on the vertical axis, and the index  $I$  (which determines the propensity of being treated) on the horizontal axis. The thick curve represents the treatment-on-the-treated effect for a patient with index  $I$ . It approaches the minimum threshold ( $\tau$ ) for a patient with a low propensity of being treated.

Figure 1b: Illustration of Taste-Based Prejudice

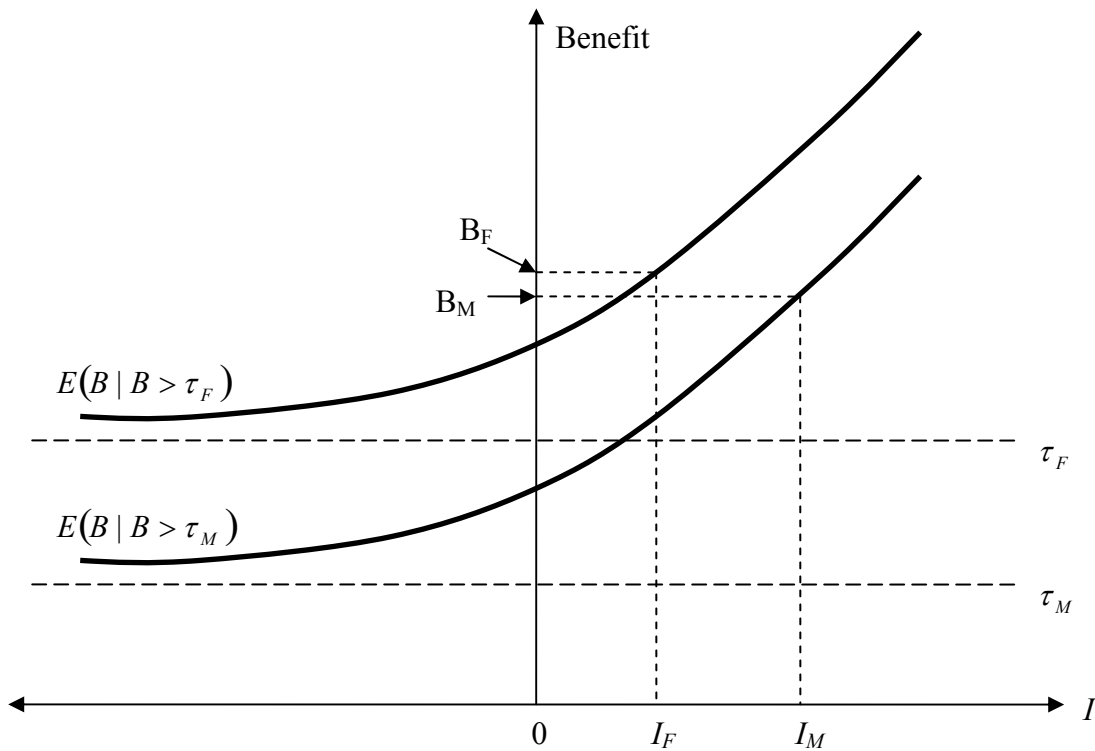
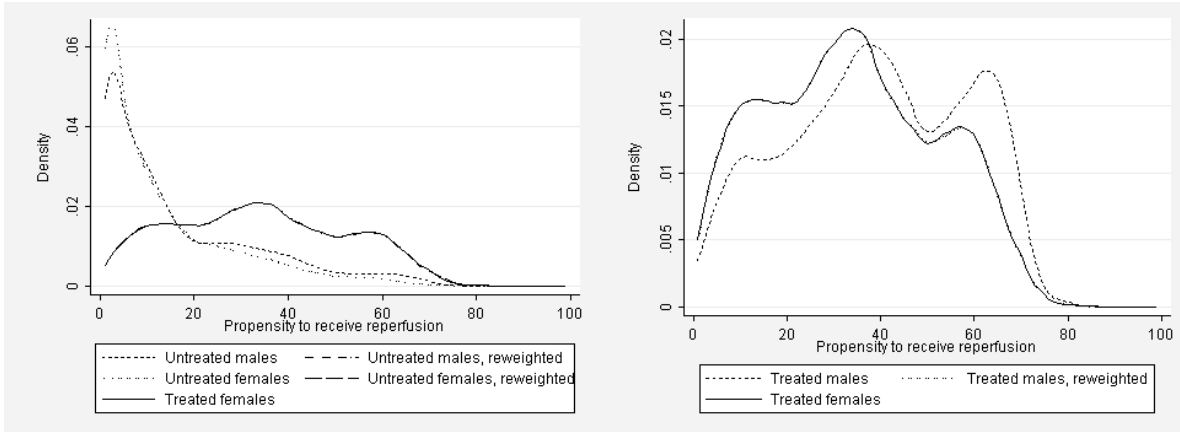


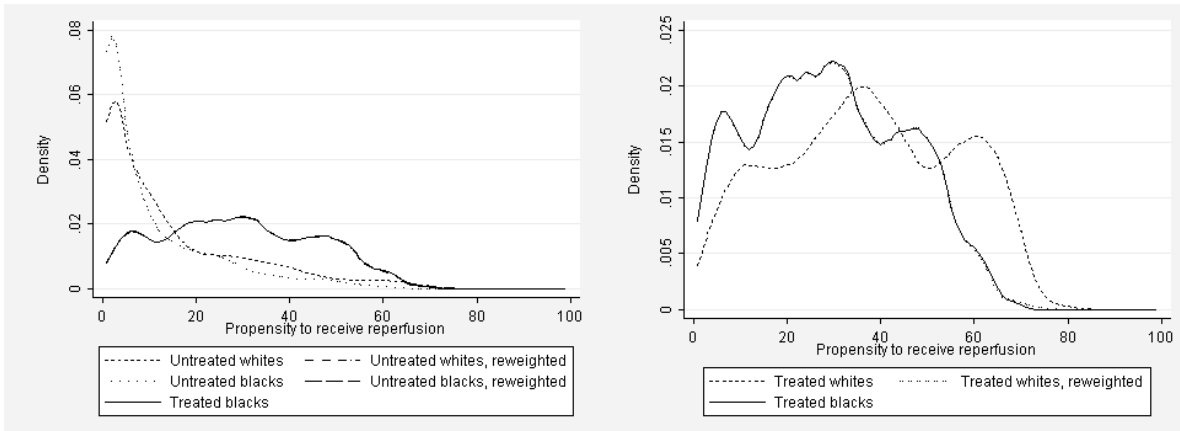
Figure illustrates how treatment-on-the-treated differs for men and women when there is prejudice against women. The treatment effect among women is higher at every point, reflecting the fact that the benefit of treatment must exceed a higher minimum threshold for women ( $\tau_w > \tau_m$ ). Because adding an amount to the hurdle because of prejudice is equivalent to incorrectly subtracting the same amount from the benefit when determining whether to treat a patient, this is also equivalent to providers incorrectly ascertaining treatment appropriateness for women.

Figure 2: Reweighting the Distributions of Propensities by Sex and Race

Panel A: Reweighting by Sex

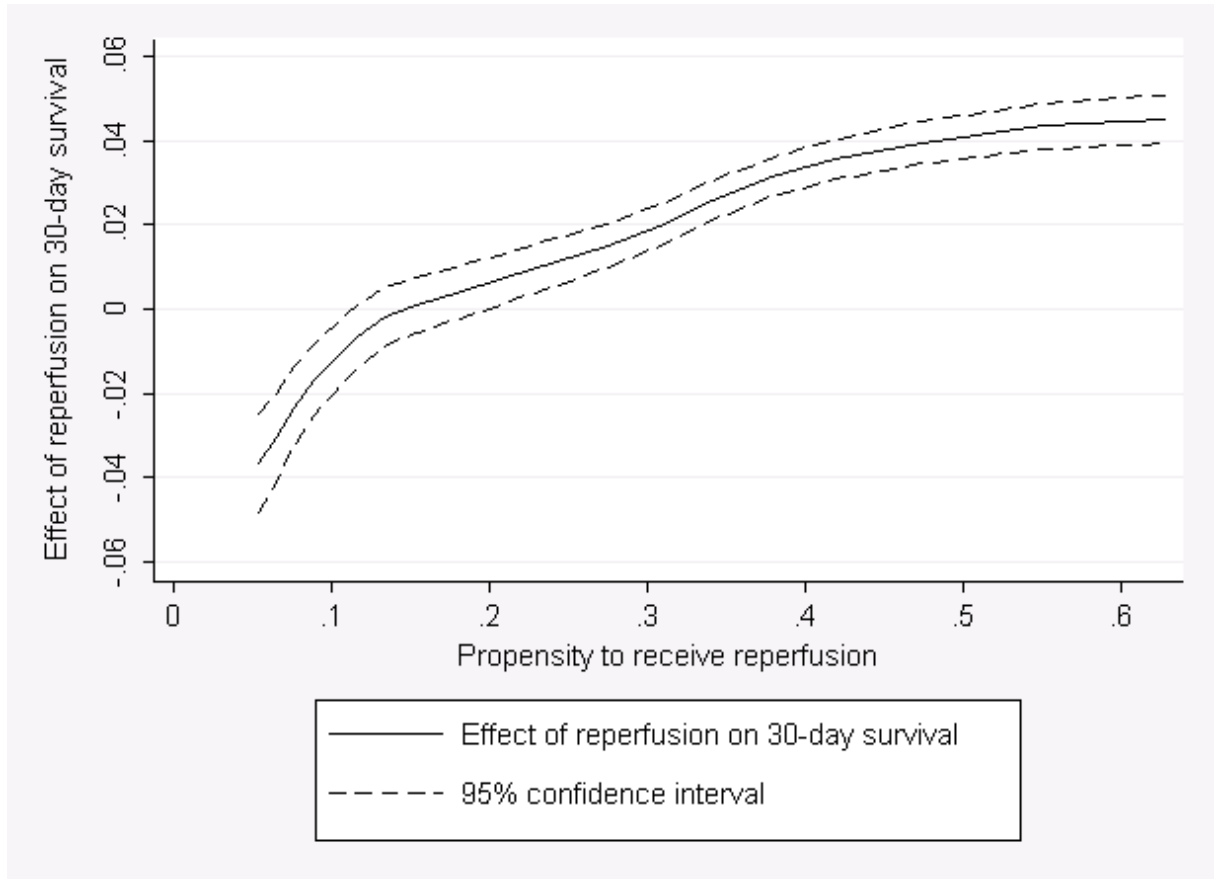


Panel B: Reweighting by Race



Note: the unweighted and reweighted kernel densities of the probability of receiving reperfusion are plotted by gender (panel A) and race (panel B) for the untreated (left hand side) and for the treated (right hand side). For reference, we have also included the propensity distribution for treated females (or blacks), since this is the target distribution that is being matched with reweighting. All reweighting is done using the method of Barsky et al. (2002).

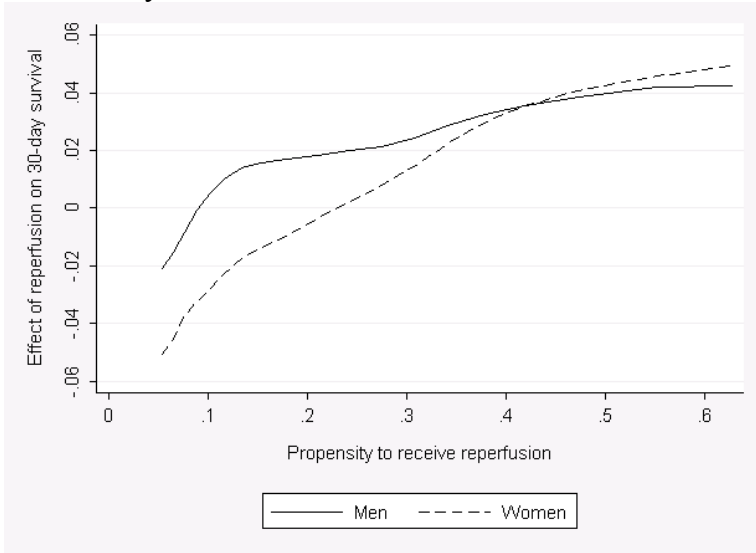
Figure 3: Effect of Reperfusion on 30 Day Survival, by propensity to receive reperfusion



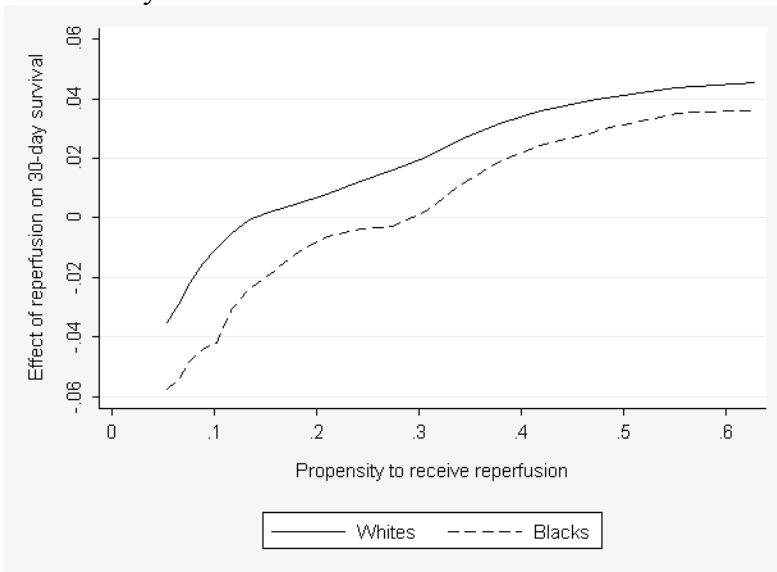
Note: Propensity to receive reperfusion is a patient's probability of receiving reperfusion within 12 hours of the heart-attack. It is calculated as fitted values from a probit model that includes sex, race and all covariates, and all interactions with sex and race. After ranking patients on this index, we calculated the effect of reperfusion on survival at each percentile of the index (see Table 3, column 3 and text for details).

Figure 4: Effect of Reperfusion on 30 Day Survival, by propensity to receive Reperfusion, and Sex and Race

Panel A: By Sex



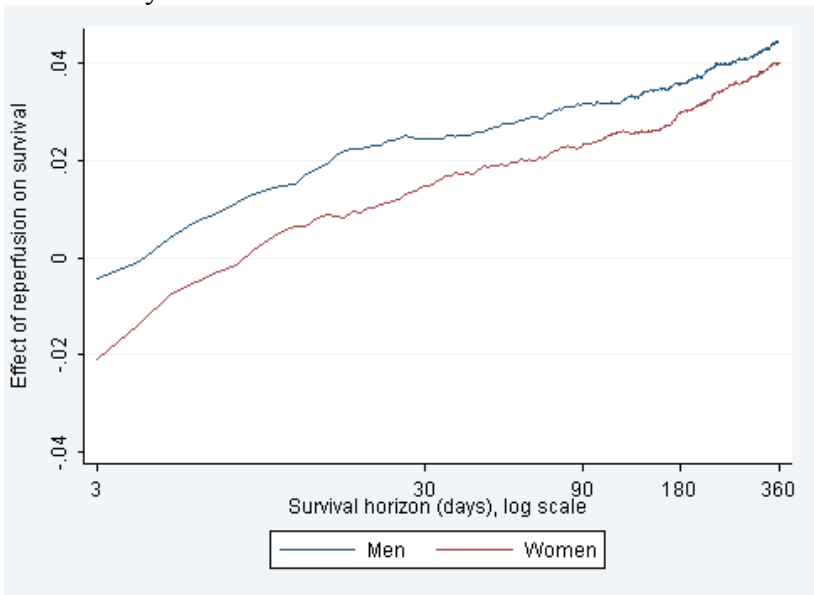
Panel B: By Race



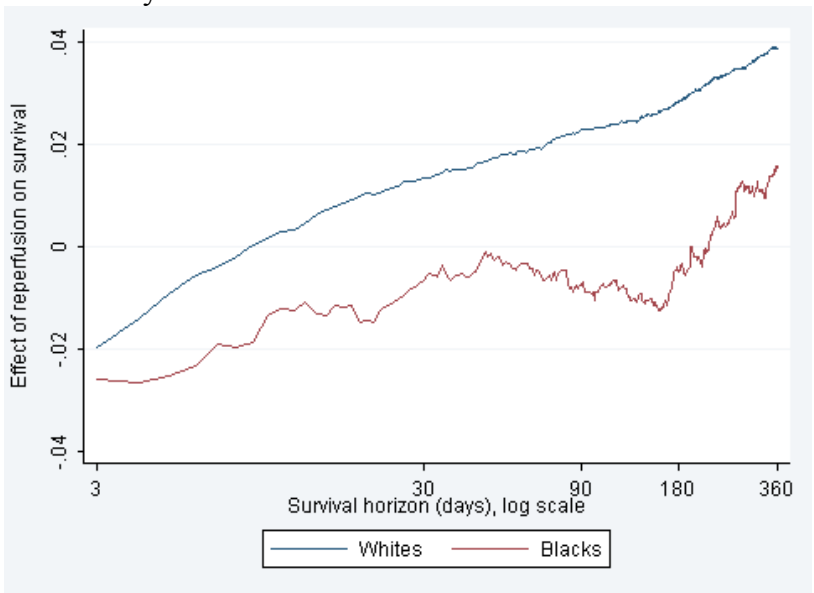
Note: Propensity to receive reperfusion is a patient's probability of receiving reperfusion within 12 hours of the heart-attack. It is calculated as fitted values from a probit model that includes sex, race and all covariates, and all interactions with sex and race. After ranking patients on this index, we calculated the effect of reperfusion on survival using local linear regression using a bandwidth of 10% of the sample and a linear kernel.

Figure 5: Effect of Reperfusion on Survival over Time

Panel A: By Sex



Panel B: By Race



Note: Figure reports the effect of reperfusion on the probability of survival at 3-365 days after the heart-attack. For each day of survival, we estimated a separate regression. The propensity of receive reperfusion is reweighted to resemble the distribution of treated minority patients (treated females in Panel A, and treated blacks in Panel B).

Table 1: Means by Sex and Race of Selected Variables

	Men	Women	Whites	Blacks
Sample Size	70,154	68,803	130,672	8,285
Age	75.25	78.08	76.72	75.60
Pr (CHF)	0.19	0.25	0.21	0.27
Pr (Diabetes)	0.28	0.33	0.30	0.42
Pr (Hypertension)	0.56	0.68	0.60	0.80
Predicted (30 day Survival)	0.82	0.80	0.81	0.81
Pr (3 day Survival)	0.92	0.90	0.91	0.92
Pr (30 day Survival)	0.83	0.80	0.81	0.83
Pr (360 day Survival)	0.70	0.65	0.68	0.67
Pr (Reperfusion <= 12 hours)	0.21	0.16	0.19	0.12

Predicted 30-day survival is a summary measure of patient sickness at the time of the heart-attack. The prediction is made using all the CCP data, but not using race or gender.



Table 2: Probit estimates of the Association between Gender and Race on the Probability of Receiving Reperfusion with 12 Hours

	(1) No Controls	(2) Full Controls	(3) Full Controls
Female	-0.170 (0.008) [-0.045]	-0.043 (0.009) [-0.008]	-0.044 (0.010) [-0.008]
Black	-0.255 (0.018) [-0.061]	-0.251 (0.021) [-0.042]	-0.260 (0.031) [-0.044]
Female*Black			0.017 (0.042) [0.003]
N	138,957	138,957	138,957

Note: Dependent variable is the whether patient received reperfusion within 12 hours. Table reports probit coefficients with marginal effects evaluated at the point of full sample means in square-brackets. Probit model includes all CCP risk-adjusters, but does not allow for further interactions between sex, race and these covariates.

Table 3: Gender and Race Differences in 30 day Survival from Reperfusion within 12 Hours, OLS Estimates using reweighting

	Unweighted	Reweighted	Non-Parametric
	(1)	(2)	
<u>A: Gender Difference</u>			
Reperfusion	0.034 (0.003)	0.024 (0.003)	
Reperfusion*Female	-0.014 (0.005)	-0.010 (0.006)	-0.009 (0.005)
N	138,957	138,957	138,957
<u>B: Race Difference</u>			
Reperfusion	0.029 (0.003)	0.013 (0.003)	
Reperfusion*Black	-0.026 (0.012)	-0.020 (0.012)	-0.019 (0.012)
N	138,957	138,957	138,957

Note: Results for each column and in each panel are from a linear regression in which the dependent variable is a dummy for 30-day survival, and that controls for sex, race and all covariates, and all interactions with sex and race. Regression in column (2) reweights everyone using the method of Barsky et al. (2002) to look like treated females (Panel A) or treated blacks (Panel B). Column 3 includes 100 percentiles of I interacted with the receipt of Reperfusion. Robust standard errors reported in parentheses.

Table 4: Gender and Race Differences in 30 day Survival from Reperfusion within 12 Hours, OLS Estimates using Parametric and Non-Parametric Controls for Propensity to Receive Reperfusion

	Parametric $g(I)$	Parametric $g(I)$
	(1)	(2)
<u>A: Gender Difference</u>		
Reperfusion	0.023 (0.003)	0.024 (0.003)
Reperfusion*Female	-0.009 (0.005)	-0.008 (0.005)
Reperfusion*Selection Term	0.147 (0.016)	0.097 (0.021)
Reperfusion*Selection Term*Female		0.124 (0.034)
N	138,957	138,957
<u>B: Race Difference</u>		
Reperfusion	0.013 (0.003)	0.013 (0.003)
Reperfusion*Black	-0.019 (0.012)	-0.016 (0.012)
Reperfusion*Selection Term	0.148 (0.016)	0.147 (0.017)
Reperfusion*Selection Term*Black		0.000 (0.091)
N	138,957	138,957

Note: Results for each column and in each panel are from a linear regression in which the dependent variable is a dummy for 30-day survival, and that controls for sex, race and all covariates, and all interactions with sex and race. The parametric selection term in columns (1) and (2) is  $(I + \phi(I) / \Phi(I))$ . Robust standard errors reported in parentheses.

Table 5: Gender and Race Differences in the Effect of Reperfusion within 12 hours on Survival at Different Time-points, OLS Estimates

Survival at:	3-day	7-day	30-day	90-day	360-day
<u>A: Gender Difference</u>					
Reperfusion	-0.004 (0.003)	0.009 (0.003)	0.024 (0.003)	0.032 (0.004)	0.045 (0.004)
Reperfusion*Female	-0.016 (0.004)	-0.012 (0.005)	-0.010 (0.006)	-0.009 (0.006)	-0.005 (0.006)
N	138,957	138,957	138,957	138,957	138,957
<u>B: Race Differences</u>					
Reperfusion	-0.020 (0.002)	-0.004 (0.003)	0.013 (0.003)	0.023 (0.003)	0.039 (0.004)
Reperfusion * Black	-0.006 (0.009)	-0.015 (0.010)	-0.020 (0.012)	-0.030 (0.013)	-0.023 (0.015)
N	138,957	138,957	138,957	138,957	138,957

Note: Results for each column and in each panel are from a linear regression in which the dependent variable is a dummy for survival as indicated, and that controls for sex, race and all covariates, and all interactions with sex and race. Regression reweight everyone using the method of Barsky et al. (2002) to look like treated females (Panel A) or treated blacks (Panel B). Robust standard errors reported in parentheses.

Table 6: Gender and Race Differences in the Effect of Reperfusion within 12 hours on Survival at Different Time-points, OLS Estimates with Hospital and Hospital Expertise Fixed-Effects

Survival at:	3-days	7-days	30-days	90-days	360-days
<u>A: Gender Gap in Benefit</u>					
Reperfusion*Female	-0.017 (0.004)	-0.013 (0.005)	-0.010 (0.006)	-0.009 (0.006)	-0.005 (0.006)
Hospital FE	Yes	Yes	Yes	Yes	Yes
Hospital FE * Reperfusion	Yes	Yes	Yes	Yes	Yes
N	138,957	138,957	138,957	138,957	138,957
<u>B: Race Gap in Benefit</u>					
Reperfusion * Black	-0.010 (0.009)	-0.019 (0.010)	-0.020 (0.012)	-0.027 (0.012)	-0.039 (0.014)
Hospital FE	Yes	Yes	Yes	Yes	Yes
Hospital FE * Reperfusion	Yes	Yes	Yes	Yes	Yes
N	138,957	138,957	138,957	138,957	138,957

Note: Results for each column and in each panel are from a linear regression in which the dependent variable is a dummy for survival as indicated. Controls are included for sex, race and all covariates, and all interactions with sex and race, along with hospital and hospital\*reperfusion fixed effects. Regression reweight everyone using the method of Barsky et al. (2002) to look like treated females (Panel A) or treated blacks (Panel B). Robust standard errors reported in parentheses.