

12-2012

Testing for the Role of Prejudice in Emergency Departments Using Bounceback Rates


Shamena Anwar

Carnegie Mellon University, shamena@andrew.cmu.edu

Hanming Fang

University of Pennsylvania

Follow this and additional works at: <http://repository.cmu.edu/heinzworks>

 Part of the [Databases and Information Systems Commons](#), and the [Public Policy Commons](#)

Published In

The B.E. Journal of Economic Analysis & Policy, 12, 3, Article 4.

This Article is brought to you for free and open access by the Heinz College at Research Showcase @ CMU. It has been accepted for inclusion in Heinz College Research by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

The B.E. Journal of Economic Analysis & Policy

Advances

Volume 12, Issue 3

2012

Article 4

FORENSIC ECONOMICS

Testing for the Role of Prejudice in Emergency Departments Using Bounceback Rates

Shamena Anwar*

Hanming Fang[†]

*Carnegie Mellon University, shamena@andrew.cmu.edu

[†]University of Pennsylvania, hfang@econ.upenn.edu

Recommended Citation

Shamena Anwar and Hanming Fang (2012) "Testing for the Role of Prejudice in Emergency Departments Using Bounceback Rates," *The B.E. Journal of Economic Analysis & Policy*: Vol. 12: Iss. 3 (Advances), Article 4.

DOI: 10.1515/1935-1682.3247

Copyright ©2012 De Gruyter. All rights reserved.

Testing for the Role of Prejudice in Emergency Departments Using Bounceback Rates*

Shamena Anwar and Hanming Fang

Abstract

We propose and empirically implement a test for the presence of racial prejudice among emergency department (ED) physicians based on the bounceback rates of patients discharged after receiving diagnostic tests during their initial ED visit. A bounceback is defined as a return to the ED within 72 hours of being initially discharged. Applying the test to administrative data of ED visits from California and New Jersey, we do not find evidence of prejudice against black and Hispanic patients, but we find evidence of prejudice against Asians in California. We also find evidence of prejudice against male patients.

KEYWORDS: racial prejudice, statistical discrimination, outcome test, bounceback rates

*We would like to thank Eric Zitzewitz (the Editor), two anonymous referees, Dennis Epple, Nicola Persico, Daniel Polsky, Andy Postlewaite, Nicolas Sahuguet, Mel Stephens and seminar participants at Drexel, Federal Reserve Board of Governors, HEC-Montreal and Washington University at St. Louis (Olin) for helpful comments and suggestions. We are particularly grateful to Dr. Jason Wong who generously provided us with many useful insights about the workings in Emergency Departments. Fang gratefully acknowledges financial support from NSF Grant SES-1122902. All remaining errors are ours.

1 Introduction

The presence and pervasiveness of racial disparities in health care and health outcomes have been abundantly documented.¹ It is conceptually useful to broadly group the various potential channels for racial disparities in health outcomes into three categories. First, patients of different races may contract various illnesses at different rates. Such differences may result from different exposures to environmental hazards, different life style choices, and different genetic dispositions toward illnesses. This category of mechanisms will lead to racial disparities in health *prior to* the interactions between patients and the health care system. Second, patients of different races may have differential *access* to health care facilities and physicians. The differential health care access can result from different rates of health insurance, different proximity of health care facilities, and different qualities of available health care facilities. Third, patients of different races may receive differential *quality* of care even if they have access to the same health care facility and physicians.² Two major pathways for the racial disparity in the quality of health care delivered by health care providers are *statistical discrimination*, and *racial prejudice*. This paper contributes to the literature on understanding the roles of statistical discrimination and racial prejudice in explaining the racial disparities in health outcomes, in the context of emergency care.

Statistical discrimination (or stereotyping) by health care providers may cause racial disparities in health care because almost all of the physicians' decisions are made under uncertainty (Arrow, 1963; Eisenberg, 1986; Phelps, 2000). Physicians typically cannot perfectly observe the disease and its severity and do not precisely know the effectiveness of a treatment on a particular patient. They have to make treatment decisions based on information collected during their encounter with the patient and possibly other noisy signals from diagnostic tests.³ A benevolent physician who aims solely to maximize the net payoff of the patient may rationally choose to use the average of the patient group (i.e. stereotype) in forming his/her prior. Specifically, the doctor's posterior assessment of the probability that the patient has a particular disease given an observed symptom is, according to the

¹Institute of Medicine (2002) provides the most comprehensive review of the literature, and Center for Disease Control and Prevention (2011) presents the most updated information regarding disparities and inequalities in health access and health outcomes in the U.S.

²Health outcome disparities could also result from different post-treatment behavior by patients of different races (see, e.g., Simeonova, 2007 and Polsky, Jha, Lave, Pauly, Cen, Klusaritz, Chen, and Volpp, 2008). It can be argued, however, that physicians should have anticipated such racial differences in post-treatment behavior in a more integrated care delivery system.

³As such, a physician's problem is similar to that of an employer who needs to decide whether to hire a job applicant (e.g., Coate and Loury, 1993), or a highway trooper who needs to decide whether to search a motor vehicle (e.g., Anwar and Fang, 2006).

Bayesian rule,

$$\Pr(\text{disease}|\text{symptom}) = \frac{\Pr(\text{symptom}|\text{disease}) \times \Pr(\text{disease})}{\Pr(\text{symptom})}.$$

Thus, statistical discrimination can appear in two instances. First, doctors may believe the *prevalence* of a disease differs by racial/ethnic group, and thus the ex ante probability of a patient having a disease, $\Pr(\text{disease})$, differs by race. Second, a physician may believe that the accuracy (or the signal/noise ratio) of a given diagnostic test differs by race, i.e., $\Pr(\text{symptom}|\text{disease})$ may depend on race.⁴ Notice that to the extent that $\Pr(\text{symptom}|\text{disease})$ and $\Pr(\text{disease})$ depend on race, doctors may make diagnosis decisions differently for minority patients even if they exhibit symptoms identical to those of white patients. If doctors' beliefs regarding prevalence of a disease and the accuracy of diagnostic tests are accurate, such disparate treatment will then reflect a desire for effective medicine, and not an intent to discriminate.

In contrast, physicians that harbor racial prejudice against minority patients will care less about the wellbeing of minority patients (relative to whites). This will lead to worse health outcomes for minorities.⁵

In order to effectively reduce racial inequities in health care and health outcomes, it is vitally important to know the causes for the racial disparities. Obviously, disparities due to the racial differences in the propensity to contract illnesses will call for different policy responses than disparities due to racial differences in access to health care; likewise, disparities that result from racial prejudice would call for a very different policy intervention than disparities due to statistical discrimination. For disparities caused by physicians' prejudice, policymakers would like to identify those physicians with prejudice and replace them with physicians without racial animus. On the other hand, if racial disparities in health care are caused by statistical discrimination, policymakers may want to provide accurate information regarding $\Pr(\text{symptom}|\text{disease})$ and $\Pr(\text{disease})$ within patients of different races to physicians.

Thus understanding whether racial disparities result from racial prejudice or from statistical discrimination is at least as important in the health care setting

⁴Balsa, McGuire, and Meridith (2005) referred to the two forms of statistical discrimination respectively as a *prevalence* hypothesis and a *miscommunication* hypothesis. See Phelps (1972) for the earliest theoretical exposition of statistical discrimination, and Fang and Moro (2010) for a recent survey of the theoretical literature.

⁵It can also reflect *unfounded* stereotypes that a doctor may hold about the health-related behavior of minorities. We will not attempt to distinguish unfounded stereotypes from racial animus in our test.

as in other settings that have attracted more academic attention.⁶ However, most of the existing literature in health economics has focused on documenting racial disparities in health care (both in diagnosis and treatment) and health outcomes, as well as documenting how much of the racial disparities could be explained by socio-economic and health insurance status. The racial disparities are still significant after controlling for these variables (see Institute of Medicine, 2002; Williams, 2007 and references cited therein). Another approach commonly taken is to infer prejudice from racial disparities in care prescribed by the physicians to patients. For example, Schulman, Berlin, Harless, and Shyrl Sistrunk, Gersh, Dub, Taleghani, Burke, Williams, Eisenberg, Ayers, and Escarce (1999) assessed physicians' recommendations for management of chest pain after they viewed vignettes of "patients" who complained of symptoms of coronary artery disease. "Patients" varied only in race, sex, age, level of coronary risk and the results of an exercise stress test. The authors found that physicians were less likely to recommend cardiac catheterization procedures for women and African Americans than for whites and men. However, it is possible that the lower catheterization utilization rates observed among black patients reflect an effort by the physicians to provide more appropriate care to these patients. Barnato, Lucas, Staiger, Wennberg, and Chandra (2005) examined the within-hospital racial disparities in the treatment of acute myocardial infarction (AMI) among Medicare beneficiaries, and found that within-hospital analyses narrowed or erased black-white disparities for medical treatments received during the acute hospitalization, but widened black-white disparities for follow-up surgical treatments, and augmented the survival advantage among blacks.

There are surprisingly few studies that attempt to examine whether the racial disparities reflect some degree of racial prejudice or is merely statistical discrimination. One paper in this vein is Balsa, McGuire, and Meridith (2005) who tested whether doctors' diagnosis is affected by the prevalence of the disease (hypertension, diabetes and depression) in the racial group, which they interpret as the priors of the doctors. They found evidence consistent with statistical discrimination. Some have tried to test whether racial and ethnic concordance between physicians and patients can affect health care disparities by reducing the racial differences in $Pr(\text{symptom}|\text{disease})$ held by doctors. For example, Strumpf (2011) studied the impact of concordance on quality of care received by patients of different races. She

⁶For example, Goldin and Rouse (2000) and Bertrand and Mullainathan (2004), among others, studied disparities in employment; Ross (1997) and Ross and Yinger (1999, 2002) in mortgage lending; Knowles, Persico, and Todd (2001), Anwar and Fang (2006), Grogger and Ridgeway (2006) and Antonovics and Knight (2009) in motor vehicle stops and searches; Anwar, Bayer, and Hjalmarsson (2012) in jury decisions; Ayres and Waldfogel (1994) and Bushway and Gelbach (2010) in bail setting; Alesina and La Ferrara (2009) in prosecution and capital sentencing; Mechoulam and Sahuguet (2011) in parole releases; and Price and Wolfers (2010) in sports refereeing.

found that concordance is not generally an important predictor of outcomes. The most related study in the health literature is probably Chandra and Staiger (2008). They attempt to identify provider prejudice in the setting of heart attack treatments based on a model where they show that if providers are prejudiced against minority patients, then one would expect to find that minority patients should have higher returns from being treated, whereas under statistical discrimination the expected return from treatment, conditional on the treatment being received, should be equalized across patients of different races. They did not find evidence of prejudicial behavior against women or minorities by providers. However, their test is valid only under the assumption that the distributions of the unobserved component of the treatment effect are identical across the racial and gender groups (see page 6 of Chandra and Staiger, 2008).

In this paper, we propose and implement an “outcome test” for the role of prejudice vs. statistical discrimination in the Emergency Department (ED) setting. The outcome test, first proposed by Becker (1957, 1993a,b), attempts to infer about the role of prejudice using patients’ outcomes. In our setting, we measure patients’ outcomes by whether or not they “bounce back” subsequent to being discharged from their ED visit. A “bounceback” is defined in the medical literature as a return to the Emergency Department after being discharged home from the initial ED visit within 72 hours. According to Weinstock and Longstreth (2007), each year there are approximately 115 million visits to Emergency Departments in the United States. Approximately 3% of these patients will “bounce back” (about 3.3 million occurrences per year) and 0.6% will bounce back and require admission (660,000 occurrences per year). Of the patients who return, 18-30% return due to a *possible medical error* made during the initial visit (600,000 to 1 million occurrences per year).^{7, 8} Given the vital role of emergency departments in the U.S. health care system, it is important to examine whether there is evidence of disparities in the quality of care received by patients of different races; and more importantly, whether racial prejudice plays an important role in the racial disparities in emergency departments.

By examining bounceback rates in the ED we can determine whether the different diagnoses and care that patients of different races receive lead to different health outcomes: if they do, then the differential treatment of patients of different races is likely due to racial prejudice; otherwise, the differences in treatment are

⁷See Gordon, An, Hayward, and Williams (1998), Pierce, Kellerman, and Oster (1990), Wilkins and Beckett (1992) and O’Dwyer and Bodiwala (1991) for the original articles for the above statistics.

⁸As we describe below, our definition of bounceback is similar to the restriction that the return to the ED is due to a possible medical error made during the initial visit. Thus, our bounceback rates of 0.05% in New Jersey and 0.10% in California are within the bounds of those reported in the literature.

likely driven by statistical discrimination of the physicians trying to provide more appropriate care to patients of different races. Formally, we present in Section 2 a model which justifies the use of the comparison of the bounceback rates as a test for racial prejudice by the doctors. The basic idea is that if doctors are prejudiced against minority patients, then they are more willing to release them from the ED. This will lead to more bouncebacks for minority patients. Since our test belongs to the class of “outcome tests”, it has to deal with the well-known “inframarginality problem” in its application. We argue, based on a plausible model of ED physician behavior, that conditional on the patients *receiving diagnostic tests during their initial ED visit*, the bounceback rates for blacks and whites should be equal if physicians are not racially prejudiced. In other words, restricting ourselves to the sample of discharged patients who received diagnostic tests during their ED visits, the inframarginality problem will not be, or at least will be less of, an issue for our inference about racial prejudice. In Section 3, we formalize the inframarginality problem associated with the outcome-based test idea and explain our proposed solution in detail.

In Section 5, we apply our proposed test for prejudice to administrative data of ED visits from California and New Jersey. We do not find evidence of prejudice against black and Hispanic patients, but we find evidence of prejudice against Asians in California. We also find evidence of prejudice against male patients. We also show that if researchers were to use other descriptive tests, such as whether discharge probabilities differ by race and gender they would have concluded that there is racial prejudice against black and Hispanic patients. These tests, we argue, suffer from the inframarginality problem and thus they are inappropriate for the purpose of detecting prejudice.

The remainder of the paper is structured as follows. In Section 2 we present a plausible model of Emergency Department physicians’ behavior and argue that bounceback rates of patients who are discharged after receiving diagnostic tests can be used as the basis for an outcome test to detect racial prejudice. In Section 3 we describe the outcome test for racial prejudice and highlight the main difficulty in its empirical implementation—the inframarginality problem. We also discuss some recent attempts to deal with the inframarginality problem and explain why our use of bounceback rates conditional on diagnostic tests resolves the inframarginality problem. In Section 4 we describe the data sets used in our empirical application. In Section 5 we present descriptive statistics of our sample, a basic test of our model, and our main results regarding the role of racial prejudice in the ED. In Section 6 we conclude. In the Appendix, we provide additional information about our data and sample selection.

2 A Model of Emergency Department Physicians' Behavior

2.1 Overview of Model

In this section we present a brief overview of our model and intuitively explain how it leads to our empirical test of racial prejudice. We formally derive these results and expound on the key assumptions in Sections 2.2 and 2.3.

Consider a patient with race r and other characteristics c who comes to the Emergency Department.⁹ The characteristics included in c could encompass variables that researchers may have about patients such as gender, age, insurance status, etc., as well as other variables that may not be collected in a typical dataset such as the patients' past medical history (including comorbidities) and the patient's current complaint that led them to ED. Each patient that comes to the ED can either have a minor problem (N) whereby they can be treated in the ED and discharged home, or a major problem (J) for which they will need to be admitted to the hospital. Let $\pi(r, c) > 0$ be the doctors' initial probability assessment that a patient with race r and characteristics c has a major problem.¹⁰

We assume the ED physician will admit the patient to the hospital if their assessment that the patient has a major problem exceeds a threshold $\pi_h(r, c) \in (0, 1)$, and will release the patient if the assessment is lower than $\pi_l(r, c) \in (0, 1)$, where $\pi_h(r, c) > \pi_l(r, c)$. If the assessment is between $\pi_h(r, c)$ and $\pi_l(r, c)$, the ED physician will administer diagnostic tests in order to make an admission/discharge decision. We assume doctors have a continuum of diagnostic tests available to them with respect to their false positive and false negative rates. Because more specific tests are costly, doctors will want to perform the minimum testing necessary in order to make a decision. Doctors will thus choose the test with the false positive and false negative rate such that if the patient turns up positive, their probability of having a major problem will be exactly $\pi_h(r, c)$ and they can be admitted; likewise, if the patient turns up negative on the test, their probability of having a major problem

⁹We single out race r from other characteristics because here we are illustrating the basic ideas of our model assuming that race is the observable patient characteristic on which doctor's prejudice is based. Clearly, if we are interested in studying gender prejudice, then we should single out the patients' gender. See Section 2.3 below.

¹⁰Note that c will also include things such as the symptoms and pain level the patient reports, which can be subjective. We assume doctors optimally adjust for this when forming $\pi(r, c)$. For example, if males tend to under-report pain as compared to females, we assume doctors will assign a higher π for a male patient who reports the same pain level as a female patient.

will be exactly $\pi_l(r, c)$ and they can be discharged.¹¹

The above result implies that the outcomes of race- r patients who are discharged after having at least one diagnostic test will identify the discharge threshold $\pi_l(r, c)$ used on them, because *every patient* that is discharged after having diagnostic tests will have a probability of having a major problem that is exactly equal to $\pi_l(r, c)$. If every patient that is discharged with a major problem returns to the ED (i.e. bounces back), we can identify each racial groups discharge threshold $\pi_l(r, c)$ simply by determining the average bounceback rate among the group. We show that if ED physicians are prejudiced against race- r patients they use a higher discharge threshold for them (i.e. they are willing to release race- r patients with a higher probability of having a major problem). Thus, our test for racial prejudice is just a simple comparison of these bounceback rates.

The validity of our test relies on two key assumptions: (1) the diagnostic tests available to ED physicians are indeed continuous; and (2) patients' bounceback rate to the ED can exactly identify the probability with which they were discharged with a major problem. We discuss the reasonableness of the first assumption in Section 2.3 and show in Section 3 what the implications are if the available diagnostic tests are more discrete. In Section 4.1 we discuss the sample restrictions that are necessary in order satisfy the second assumption.

2.2 Formal Model

In this section we formally model how ED doctors determine their discharge thresholds, and decide which diagnostic tests to use on patients. We then show how our model of physician behavior implies a relatively simple test for racial prejudice.

2.2.1 Determination of the Discharging Thresholds

For simplicity, we assume that $\pi_h(r, c)$ is set by the physician in charge of admitting patients to the hospital, so that ED doctors take this as given. Thus, we will set $\pi_h(r, c) = \pi_h^*$ for all r, c .¹² However, the ED doctor must decide on the threshold $\pi_l(r, c)$ below which they will discharge the patient from the ED.

¹¹Another way to think of this "continuum of tests" is that doctors are performing a series of infinitesimal tests to get patients to either threshold.

¹²Assuming that the threshold for admission to the hospitals $\pi_h(r, c)$ is not controlled by the attending ED physician is without loss of generality. As will be clear from our analysis below, our test for prejudice involves identifying $\pi_l(r, c)$, which will not be affected by whether or not the ED physician affects the upper threshold, nor whether this upper threshold reflects prejudice.

The ED doctor chooses discharging standard $\pi_l(r, c)$ to maximize his expected utility, which is given by:

$$\max_{\{\pi_l \geq 0\}} U_r(\pi_l) = R(\pi_l) - \rho(\pi_l)S - \pi_l a_r, \quad (1)$$

where (i). the first component $R(\pi_l)$ represents the total revenue (i.e., the benefit) to the doctors of using the discharge standard π_l , and we assume $R'(\pi_l) > 0$ and $R''(\pi_l) < 0$;¹³ (ii). the second term, $-\rho(\pi_l)S$, represents the loss in payoff if the doctor is successfully sued by the patient in the event that a major problem occurs following the discharge, where $\rho(\pi_l)$ is the probability that the patient who experiences a major problem following the discharge would file *and* win a lawsuit, in which case the ED doctor will suffer a penalty $S > 0$, and we assume that $\rho'(\pi_l) > 0$ and $\rho''(\pi_l) > 0$;¹⁴ (iii). the last component, $-\pi_l a_r$ measures the expected amount of *affinity* ED doctors have towards race- r patients if they discharge a race- r patient for whom a major problem can arise with probability π_l . In a sense, this measures how much doctors personally care about the outcomes of their patients aside from worries about the probability the patient will sue them.

Definition 1. We say that the doctors are racially prejudiced if $a_r \neq a_{r'}$ for $r \neq r'$. We say that the ED doctor is racially prejudiced against race- r patients if $a_r < a_{r'}$, i.e. if the ED doctor feels less affinity for the race- r patient's sufferings.

From problem (1), it is clear that the ED doctors will choose the threshold π_l toward race- r patients to satisfy the first order condition:

$$R'(\pi_l) = \rho'(\pi_l)S + a_r. \quad (2)$$

Figure 1 shows the determination of π_l for race- r and race- r' patients for which $a_r < a_{r'}$. From (2), the result below immediately follows:

Proposition 1. If the ED doctor is racially prejudiced against race- r patients relative to race- r' patients according to Definition 1, i.e., if $a_r < a_{r'}$, then the doctor will set $\pi_l^*(r) > \pi_l^*(r')$; if the doctor is not racially prejudiced, i.e., if $a_r = a_{r'}$, then $\pi_l^*(r) = \pi_l^*(r')$.

¹³A rationale for these assumptions is as follows. The higher the threshold to discharge patients π_l is, the less time ED doctors have to spend with each patient, and the more patients they can see in a given time period. Since ED doctors have profit incentives to see as many patients as possible, their total revenue will increase as π_l increases. However, each subsequent increase in π_l should increase total revenue by less.

¹⁴ S can reflect the cost of a lawsuit, damage compensation, as well as lost future revenues and increased malpractice insurance premiums.

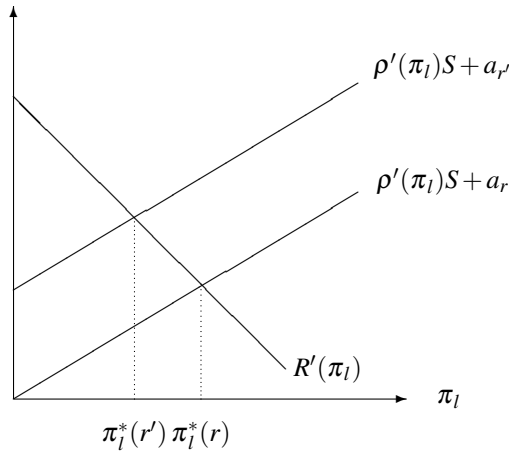


Figure 1: Graphical Illustration of the First-Order Condition: $a_r < a_{r'}$.

2.2.2 Determination of the Diagnostic Tests

Now that we have obtained the lower bound the doctor will use to discharge patients, we can describe the optimal behavior of the ED doctor towards a patient they initially assess with probability $\pi(r, c)$ of having a major problem:

- if $\pi(r, c) \geq \pi_h^*$, the ED doctor will immediately admit the patient to the hospital *without* any additional diagnostic tests;
- if $\pi(r, c) \leq \pi_l^*(r)$, they will immediately discharge the patient *without* any additional diagnostic tests;
- however, if $\pi(r, c) \in (\pi_h^*, \pi_l^*(r))$, the ED doctor will have to *perform diagnostic tests* before they can decide whether to admit or discharge the patient. We describe the decisions about what diagnostic tests to perform below.

Definition 2. *Diagnostic tests are indexed by two numbers (n_f, p_f) where $n_f = \Pr(\text{negative}|J) > 0$ is the false negative probability and $p_f = \Pr(\text{positive}|N) > 0$ is the false positive probability.*

We make two plausible assumptions about the diagnostic tests:

Assumption 1. *ED doctors have a continuous battery of diagnostic tests available to them, so that they can choose any diagnostic test $(n_f, p_f) \in (0, 1)^2$.*

Assumption 2. *The monetary costs of the diagnostic tests are born by the patients.*

Under Assumption 2, the ED doctors do not worry about monetary costs when they choose what tests to run. It is important to emphasize, however, Assumption 2 does *not* imply that the ED doctors will choose the most precise tests for the patients. The reason is that these tests are still costly to the ED physician in terms of *time*, with more accurate tests taking longer to run and to analyze the results.¹⁵ Assumption 2 does imply that doctors will want to do the minimum testing necessary in order to make a decision; specifically, the doctors will choose (n_f, p_f) such that the doctors' posterior assessment that a race- r patient has a major problem, given that the test- (n_f, p_f) turns up negative, will just hit π_l^* , where $\pi_l^*(r)$ is the *optimal* threshold as defined in (2) for race- r patients. It is not necessary for doctors to use more precise testing, since the threshold $\pi_l^*(r)$ was chosen optimally for their objective function (1) by definition. Similarly, the doctors' posterior assessment that a race- r patient has a major problem given a positive result on the tests will just hit $\pi_h^*(r)$. Thus, given Assumption 1, the doctors will, for race- r patient with characteristics c , choose the test- (n_f, p_f) that satisfies:

$$\pi_h^* = \Pr(J|\text{positive}) \equiv \frac{(1 - n_f)\pi(r, c)}{(1 - n_f)\pi(r, c) + p_f[1 - \pi(r, c)]}; \quad (3)$$

$$\pi_l^*(r) = \Pr(J|\text{negative}) \equiv \frac{n_f\pi(r, c)}{n_f\pi(r, c) + (1 - p_f)[1 - \pi(r, c)]}. \quad (4)$$

Solving the above two equations for n_f and p_f , we have:

$$n_f^*(r, c) = \frac{\pi_l^*(r) [\pi_h^* - \pi(r, c)]}{[\pi_h^* - \pi_l^*(r)] \pi(r, c)}, \quad (5)$$

$$p_f^*(r, c) = \frac{[\pi(r, c) - \pi_l^*(r)] (1 - \pi_h^*)}{[1 - \pi(r, c)] [\pi_h^* - \pi_l^*(r)]}. \quad (6)$$

2.2.3 Main Implication

Our empirical test is based on the implication of Equation (4). It says the following: after a doctor observes a race- r patient with characteristics c , they first determine the initial probability of a major problem $\pi(r, c)$. If $\pi(r, c) \in (\pi_h^*, \pi_l^*(r))$, so that the doctor needs diagnostic tests to determine the course of actions (discharge home, or admit to the hospital), they will choose the optimal diagnostic test $(n_f^*(r, c), p_f^*(r, c))$ according to the formulas given by (5) and (6). Under test- $(n_f^*(r, c), p_f^*(r, c))$, Equation (4) guarantees that every race- r patient discharged

¹⁵This time cost is captured in our assumption that the term $R(\pi_l)$ in expression (1) is decreasing in π_l .

home after undergoing diagnostic tests has a probability of a major disease that is equal to their discharge threshold $\pi_l^*(r)$, independent of other characteristics c .

Assumption 3. *A patient will return to the ED, i.e., bounce back, if and only if he/she encounters a major problem following discharge in the previous ED visit.*

Assumption 3 requires that all patients with a missed major problem return to the ED, and that all return visits to the ED occur because a major problem was missed on the first visit. In order to satisfy this assumption we will have to restrict our definition of a bounceback to one where a patient returns to the ED within three days and is immediately admitted to the hospital and diagnosed with a major problem that is different than what they were diagnosed with on their first visit. In Section 4 we discuss in detail why this restriction should satisfy Assumption 3.

Since every patient of race- r that is discharged home after undergoing diagnostic tests has probability $\pi_l^*(r)$ of having a major problem, Assumption 3 ensures that we can estimate $\pi_l^*(r)$ by computing the proportion of bounceback patients among discharged race- r patients who underwent diagnostic tests prior to their discharges. Denote the bounceback rate for discharged race- r patients, *conditional on* them obtaining additional diagnostic tests while in the ED, as $B(r|\text{Diagnostic Tests})$, which we can express as:

$$\begin{aligned}
 & B(r|\text{Diagnostic Tests}) \\
 = & \frac{\int_{\{c:\pi(r,c)\in(\pi_l^*(r),\pi_h^*)\}} n_f^*(r,c) \pi(r,c) dF_r(c)}{\int_{\{c:\pi(r,c)\in(\pi_l^*(r),\pi_h^*)\}} \left\{ n_f^*(r,c) \pi(r,c) + [1 - p_f^*(r,c)] [1 - \pi(r,c)] \right\} dF_r(c)} \quad (7) \\
 = & \frac{\int_{\{c:\pi(r,c)\in(\pi_l^*(r),\pi_h^*)\}} \pi_l^*(r) \left\{ n_f^*(r,c) \pi(r,c) + [1 - p_f^*(r,c)] [1 - \pi(r,c)] \right\} dF_r(c)}{\int_{\{c:\pi(r,c)\in(\pi_l^*(r),\pi_h^*)\}} \left\{ n_f^*(r,c) \pi(r,c) + [1 - p_f^*(r,c)] [1 - \pi(r,c)] \right\} dF_r(c)} \quad (8) \\
 = & \pi_l^*(r) \quad (9)
 \end{aligned}$$

where $F_r(c)$ is the cumulative distribution function of c among race- r patients. To understand the above expression, note that in line (7) the numerator is the total measure of race- r patients who actually have major problems but are discharged home because the diagnostic tests yield a false negative outcome. The denominator is the total measure of race- r patients who are discharged after getting a negative test result. Line (8) follows from the definition of $n_f^*(r,c)$ as defined in (4).

Together with Proposition 1, we immediately have the following result:

Proposition 2. *Under Assumptions 1-3, ED doctors are racially prejudiced against race- r patients relative to race- r' patients if and only if $B(r|\text{Diagnostic Tests}) > B(r'|\text{Diagnostic Tests})$.*

Proposition 2 provides the basis of our empirical test that we explain in detail in Section 3 below. Note that to implement this test, we only require information on the race of the patient, whether diagnostic tests were done, and whether they returned within three days after being discharged home. All of this data is readily available. Our test does not rely on knowing the information contained in c , which is vital, since no data set contains information that detailed.

2.3 Discussion of the Model

We have established that comparisons of the conditional bounceback rates as defined in (7) are informative about the physicians' racial prejudice: physicians are prejudiced against race- r patients if and only if their bounceback rate is higher conditional on having received diagnostic tests in the initial ED visit. Here we discuss some important points regarding the model.

- The most important assumption of our model is Assumption 1, which states that doctors have access to a continuous array of diagnostic tests which differ in their false positive and false negative rates. This strong assumption is what we rely on to ensure that the probability of having a major problem among those who were discharged with some diagnostic tests is independent of potentially unobserved (by econometricians) characteristics c . We believe from our discussions with emergency department doctors that this is a plausible assumption in practice. For example, if a patient comes in with chest pains, they will be screened for a heart attack using either an EKG, several different blood tests, chest x-ray, CT-scan, and/or a cardiac stress test. Depending on the patient's initial risk of having a heart attack, various different combinations of these tests will be ordered. This effectively gives doctors a wide array of test sets to choose from. Although this is not exactly continuous, the sheer number of potential test sets make it a reasonably close approximation.¹⁶ However, even if this assumption is not strictly satisfied, the heterogeneity in the posterior probability of having a major problem among the discharged patients who received tests will be much less than that among discharged patients in general. Because our empirical analysis only focuses on the former group, the inframarginality problem will be alleviated, even if not completely

¹⁶It is important to note that our model's main implication (given in (4)) would not be met if doctors were just performing perfunctory testing. However, this does not seem to be the case, as in the data we observe that the types of tests done vary greatly from patient to patient. This implies patients are getting tests done that are specific to their problem, as opposed to just getting a general set of tests done.

eliminated. We will explain this notion in more detail in Section 3, when the inframarginality problem is discussed in depth.

- Up to now we have couched our discussion strictly in terms of race-based prejudice. However, it is obvious that we can allow the affinity parameter a_r in the doctor's problem (1), instead of being indexed just by race r , to be indexed by any *vector* of observable patient characteristics, e.g. combinations of race and gender. The logic of our proposed test for prejudice based on comparisons of conditional bounceback rates remains valid when applied to test for prejudice in more finely defined groups.
- In problem (1), it is also possible that other components of the utility function can depend on the patient's insurance status and age. Specifically, the doctor might believe the revenue function $R(\pi_l)$ depends on the patient's insurance status. Age is likely to affect the expected loss from being successfully sued $\rho(\pi_l)S$, because malpractice payouts typically depend on the patient's expected future earnings. This means doctors face a lower expected loss for older patients. The notion that older patients sue less has been documented in a study by Burstin, Johnson, Lipsitz, and Brennan (1993). Because of this, bounceback rates can be different across different age and insurance groups for reasons other than prejudice. Thus, to effectively test for racial and gender prejudice, we will need to control for age and insurance status in a regression framework.
- We are implicitly assuming that the probability a patient files a malpractice suit does not depend on their race and gender. Empirical support for this assumption can also be found in Burstin, Johnson, Lipsitz, and Brennan (1993). The study identifies all of the hospital records in New York in 1984 where there was evidence of malpractice. Within this subsample they find that race and gender have no predictive power over who subsequently filed a malpractice lawsuit.

2.4 Testable Implications of Our Model

It is also important to recognize that our model has a key testable implication—patients discharged without any diagnostic tests should have *lower* bounceback rates than those who were discharged with diagnostic tests. This implication follows from the threshold behavioral rule of the physicians in our model, as the only discharged patients that don't get diagnostic tests done are ones with bounceback rates that are below the lower threshold. We will provide evidence in support of this prediction in our empirical results below.

Our model also predicts that, conditional on race (and/or any observable characteristics that physicians may base their prejudice on), the accuracy of the diagnostic tests, as measured by n_f and p_f , should *not* affect the bounceback rate. Unfortunately, there is no good way to implement this test, because we are unable to observe the n_f and p_f of patients' tests directly. While we have information on the types of tests done (lab tests, EKGs, x-rays, etc.), the relative accuracy of each of these tests depends on the specific problem the doctor is screening for, which is not available in our data.¹⁷

2.5 Other Implications

Having shown in Proposition 2 that comparisons of the bounceback rates conditional on receiving diagnostic tests, $B(r|\text{Diagnostic Tests})$, across patients of different races can be informative of the ED doctors' racial prejudice, we now show that three other alternative tests that researchers might be tempted to do are not informative about doctors' prejudice.

First, the comparison of the bounceback rates across patients of different races *without* restricting to the sub-sample of discharged patients who received diagnostic tests in the initial visits is *not* informative of physicians' racial prejudice. To see this, note that the unconditional bounceback rate of discharged race- r patients, denoted by $B(r)$, is given by

$$B(r) = \frac{\int_{\{c:\pi(r,c) \in (\pi_l^*(r), \pi_h^*(r))\}} n_f^*(r,c) \pi(r,c) dF_r(c) + \int_{\{c:\pi(r,c) \leq \pi_l^*(r)\}} \pi(r,c) dF_r(c)}{\int_{\{c:\pi(r,c) \in (\pi_l^*(r), \pi_h^*(r))\}} \left\{ \begin{array}{l} n_f^*(r,c) \pi(r,c) \\ + [1 - p_f^*(r,c)] [1 - \pi(r,c)] \end{array} \right\} dF_r(c) + \int_{\{c:\pi(r,c) \leq \pi_l^*(r)\}} dF_r(c)} \quad (10)$$

Note that the difference between the expression for the unconditional bounceback rate $B(r)$ above and that for $B(r|\text{Diagnostic Tests})$ in (7) is the extra term $\int_{\{c:\pi(r,c) \leq \pi_l^*(r)\}} \pi(r,c) dF_r(c)$ in the numerator and $\int_{\{c:\pi(r,c) \leq \pi_l^*(r)\}} dF_r(c)$ in the denominator. These, as we will discuss in Section 3 below, represent the *infra-marginally* discharged patients for whom the doctors' initial assessment $\pi(r,c)$ is sufficiently low not to warrant a diagnostic test. The addition of these infra-marginal patients results in $B(r)$ depending on c . Since the distributions $F_r(c)$ are likely to vary by race, unconditional bounceback rates can differ either because doctors use different discharge thresholds or because patients of different races have different underlying disease prevalence. A comparison of $B(r)$ across r will thus not be informative of the relationship between a_r and $a_{r'}$.

Second, comparisons of discharge rates (or, equivalently, hospital admission rates) whether conditional on diagnostic tests [denoted by $D(r|\text{Diag. Tests})$],

¹⁷For example, while an EKG can be helpful in determining whether a patient is having a heart attack, it is not helpful in screening for appendicitis.

or unconditional [denoted by $D(r)$], are not informative of the physicians' racial prejudice. To see this, note that the conditional and unconditional discharge rates described above are respectively:

$$D(r|\text{Diag. Tests}) = \int_{\{c:\pi(r,c)\in(\pi_l^*(r),\pi_h^*(r))\}} \left\{ \frac{n_f^*(r,c)\pi(r,c)}{1-p_f^*(r,c)} + [1-p_f^*(r,c)][1-\pi(r,c)] \right\} dF_r(c); \quad (11)$$

$$D(r) = \int_{\{c:\pi(r,c)\in(\pi_l^*(r),\pi_h^*(r))\}} \left\{ \frac{n_f^*(r,c)\pi(r,c)}{1-p_f^*(r,c)} + [1-p_f^*(r,c)][1-\pi(r,c)] \right\} dF_r(c) + \int_{\{c:\pi(r,c)\leq\pi_l^*(r)\}} dF_r(c). \quad (12)$$

Note that cross-race differences in either of the discharge rates calculated above mix together the three sources for racial differences: the first channel is that the groups may have ex ante differences in the probability of major problems, as represented by the potential difference between $F_r(c)$ and $F_{r'}(c)$, and the fact that the initial assessment $\pi(r,c)$ depends on c and r ; the second channel is racial prejudice, which leads to differences in $\pi_l^*(r)$ which appear in the region of integration; and the third channel is potential statistical discrimination, which we define below:

Definition 3. *Suppose that $a_r = a_{r'}$. If two patients with $r \neq r'$ underwent diagnostic tests such that $(n_f^*(r,c), p_f^*(r,c)) \neq (n_f^*(r',c), p_f^*(r',c))$, then we say that doctors engage in statistical discrimination.*

To understand why Definition 3 captures the notion of statistical discrimination, note that if ED doctors do not have racial prejudice, then they will choose $\pi_l^*(r) = \pi_l^*(r')$. In this case, according to Equations (3) and (4), $(n_f^*(r,c), p_f^*(r,c)) \neq (n_f^*(r',c), p_f^*(r',c))$ can occur only if $\pi(r,c) \neq \pi(r',c)$, i.e., the ED doctor forms different assessments for race- r and race- r' patients with identical characteristics c , which is exactly the commonly used definition of statistical discrimination. Because both the conditional and unconditional discharge rates defined above mix all three channels for racial differences, they are unable to be directly informative about the role of racial prejudice.

Finally, we should emphasize that comparisons of whether diagnostic tests are done are not informative about the role of physicians' racial prejudice. The reason is simple. The doctors' decision to do diagnostic testing depends not only on $\pi_l^*(r)$ [which is reflective of racial prejudice as we show in Proposition 1], but also on $\pi(r,c)$, which reflects both underlying differences in c and statistical discrimination. Thus, comparing whether diagnostic tests are done suffers exactly the same problem as the comparison of discharge rates in inferring about racial prejudice.

The following proposition summarizes the above discussions:

Proposition 3. *Without further assumptions on the distributions of initial assessment $\pi(r, c)$ across patients of different races, neither the cross-race comparisons of the unconditional bounceback rates (10), nor the discharge rates [whether conditional (11) or unconditional (12)], nor whether diagnostic tests are done, are informative about the physicians' racial prejudice.*

In Section 5.4 we run these incorrect tests and show that the results we obtain from these are quite different than the results we obtain from our proposed test of racial prejudice.

3 The Empirical Test

In this section, we describe in more details the advantages of outcome-based tests, as well as the well-known inframarginality problem associated with the outcome test. We then explain how our model of ED physician behavior allows us to avoid the inframarginality problem when we focus on the sub-sample of discharged patients who received diagnostic tests in their initial ED visit.

Outcome Test for Prejudice. There is a large literature in economics that attempts to distinguish the contributions of statistical discrimination and racial prejudice to racial disparities in a variety of settings, including employment, health care, mortgage and other lending situations, motor vehicle stops and searches as well as all phases of law enforcement such as jury selection, prosecution and sentencing. The standard approach of using regression analysis to infer bias would regress, as the left side variable, an indicator of the actions taken by the treater, on a list of variables, including race and/or gender, that are thought to be possibly related to the treater's decision. It is well recognized, however, the regression approach suffers from both the "omitted" and "included" variable biases.¹⁸

More recently, a growing literature has advocated the use of an "outcome test", first proposed by Becker (1957, 1993a,b). The idea of the outcome test is quite intuitive. If decision-makers, say ED physicians, are prejudiced against a group of patients, then that group of patients are likely to be prematurely released relative to

¹⁸The "omitted variable" bias arises if there are variables that are legitimately related to the decision making, but not included in the regression. If there is correlation between race/gender with the omitted variable, the race/gender coefficient may be picking up the effect of the omitted variable. The "included variable" bias arises if variables correlated with race that should not have legitimately mattered are included as regressors (see Ayres, 2010 for a discussion).

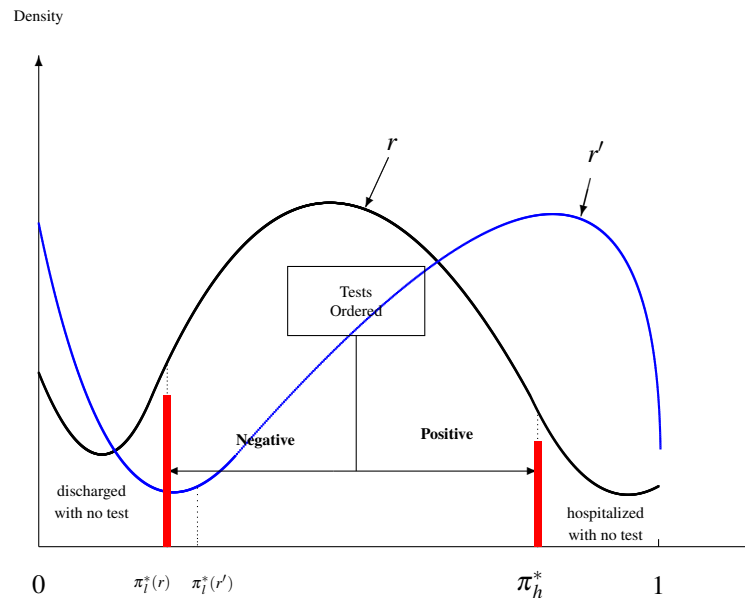


Figure 2: The Inframarginality Problem and The Proposed Solution

other groups of patients, resulting in a higher bounceback rate for the prejudiced-against group. Thus, the comparisons of the outcomes of different groups of patients, i.e. the bounceback rates, would be informative of the racial prejudice of the physicians. The application of the outcome test, however, is plagued by the “inframarginality *problem*,” which refers to the difference between the comparisons of the average and marginal outcomes across racial or gender groups (see Knowles, Persico, and Todd, 2001, Anwar and Fang, 2006 and Persico, 2010 for descriptions of this problem).¹⁹

The Inframarginality Problem and Our Proposed Solution. Figure 2 illustrates the inframarginality problem in our setting if we were just to compare the bounceback rates of all discharged patients across patient races. It also explains how our model of physician behavior allows us to avoid the inframarginality problem if we focus on the sub-sample of discharged patients who received diagnostic tests in their initial ED visits. The dark curve in Figure 2 depicts the distributions of the initially assessed probability by physicians that race- r patients have major

¹⁹Charles and Guryan (2008) describes an alternative description of the inframarginality problem in the context of employers with heterogenous levels of prejudice. In their paper, they propose to approximate the “marginal” level of prejudice in a State by the p -th percentile of the prejudice distribution in that state where p is the percentage of the state workforce that is black.

problems, i.e., $\pi(r, c)$. As we describe in Section 2, the ED physicians will observe the $\pi(r, c)$ for a particular race- r patient and will then decide upon the course of action according to where $\pi(r, c)$ lies relative to the two thresholds $\pi_l^*(r)$ and π_h^* : if $\pi(r, c) \leq \pi_l^*(r)$, the patient will be discharged without any additional tests; if $\pi(r, c) \geq \pi_h^*$, the patient will be admitted to the hospital without any additional tests; however, if $\pi(r, c) \in (\pi_l^*(r), \pi_h^*)$, then diagnostic tests $(n_f^*(r, c), p_f^*(r, c))$ will be ordered for the patient and the physicians will discharge the patient if and only if the outcomes from the diagnostic tests are negative.

Notice, as we highlighted in expression (10) for the unconditional bounceback rates, the comparisons of the average bounceback rates for race- r and race- r' patients may not reveal the ranking of $\pi_l^*(r)$ and $\pi_l^*(r')$. In Figure 2, the discharge thresholds for race- r and race- r' patients are such that $\pi_l^*(r') > \pi_l^*(r)$, i.e., the physicians are prejudiced against race- r' patients. However, because the distribution of $\pi(r', c)$ has a higher lower tail than that of $\pi(r, c)$, the average bounceback rate for race- r' patients is lower than that for race- r patients. This is exactly the inframarginality problem.

However, if we restrict ourselves to the comparisons of the bounceback rates to patients discharged after receiving diagnostic tests, their posterior assessments are all concentrated at $\pi_l^*(r)$ and $\pi_l^*(r')$ respectively for race- r and race- r' patients. This is ensured by the physicians' optimal choices of the diagnostic tests as described by (5) and (6).

When Assumption 1 – the assumption that the ED doctors have access to a continuous battery of diagnostic tests – does not strictly hold, Figure 2 also shows that the doctors will choose from the available tests so that their posterior upon receiving a negative test results about the patient having a major problem is *as close as possible* to the discharge threshold $\pi_l^*(r)$. In this sense, there is much less heterogeneity among the discharged patients with diagnostic tests within a racial group. Therefore, our test *alleviates* the inframarginality problem in the application of the outcome test even when Assumption 1 is not strictly satisfied.

We should mention that the idea that continuous control variables by decision-makers may alleviate the inframarginality problem in the outcome test is independently developed in Mechoulan and Sahuguet (2011), where they use the outcome test idea to test for the role of racial prejudice by parole boards.²⁰ They argue that, to the extent a parole board can choose the time of release for a parolee to minimize the number of parole violations, it implies that all released parolees should have the same probability of a parole violation. Thus from a researcher's perspective, there

²⁰The potential that continuous control variables available to the treators may alleviate the inframarginality problem has also been discussed in Ayres and Waldfogel (1994), Ayres (2002) and Ayres (2005, p. 14).

is no inframarginality problem. They find that in almost every state with a discretionary parole board, African American parolees are more likely to violate parole than White parolees by about ten percentage points, suggesting that parole boards are more lenient in their releasing decisions when they face African American prisoners.²¹

Difference from KPT's Justification for the Outcome Test. It is also useful to distinguish our justification for the use of the outcome test from the justification provided in the seminal paper by Knowles, Persico, and Todd (2001) in the context of racial profiling in motor vehicle searches. Knowles, Persico, and Todd (2001) develop a simple but elegant theoretical model about motorist and police behavior and show that in *equilibrium* the inframarginality problem may not arise. In their model, motorists differ in their characteristics, including race and possibly other factors that are observable to troopers but may or may not be available to researchers. Troopers decide whether or not to search motorists while motorists decide whether or not to carry contraband. In this “matching pennies”-like model they show that if troopers are not racially prejudiced, all motorists, if they are searched at all, must in equilibrium carry contraband with equal probability regardless of their race and other characteristics. Thus in their model there is *no difference* between the marginal and the average search success rates.

In contrast, the key for us and for Mechoulan and Sahuguet (2011) to address the inframarginality problem is that the decision makers, in our case the ED physicians and in Mechoulan and Sahuguet (2011)'s case the parole board, have continuous controls that can affect the relevant outcomes (bounceback rates in our case and the parole violation rates in Mechoulan and Sahuguet, 2011).

4 Data

The data sets we use to implement our proposed test for racial prejudice in Emergency Departments using bounceback rates come from New Jersey and California. The New Jersey data was obtained by combining data from the Healthcare Cost and Utilization Project (HCUP) and the New Jersey Department of Health and Senior Services, and covers the period from January 2006 through July 2007.²² The HCUP databases collect patient-level hospital data from the majority of U.S. states

²¹One objection to their study is that parole violations are not objectively measured; instead they are determined by police officers, who may be discriminatory against black parolees. See Anwar and Fang (2012) for an alternative test of prejudice in parole releases.

²²See <http://www.ahrq.gov/data/hcup> and <http://www.state.nj.us/health> for more information about these data sets.

and organize the data in a unified framework. It represents the largest collection of longitudinal hospital data in the U.S. The California data was obtained from the Office of Statewide Health Planning and Development (OSHPD) and covers the period from January 2006 through September 2007.²³ In both data sets, we have information on all Emergency Department (ED) visits that occurred during their respective coverage period. For both states, we observe a patient's admission and discharge date (for both outpatient ED visits and ED visits that led to hospital admissions), the procedures done, the diagnoses and the final disposition of the patient (i.e., whether they were admitted to the hospital or discharged home). In both data sets there is a patient indicator which allows patients' visits to be tracked over time. However, for New Jersey, this indicator is not unique across hospitals, and thus we can only track a particular patient's visits to the same ED; for California, this indicator is unique across hospitals, which allows us to follow all of a patient's ED visits even if the return visits are to an ED in a different hospital.

The sizes of our samples are very large, with about 3.86 million and 11.7 million ED discharge observations in NJ and CA respectively. Such large samples are necessary to examine bounceback rates because bouncebacks occur with quite small probability (due to their severe consequences). However, in order to use this admission data to identify missed major problems in a way that is robust to potential behavioral differences between white and minority patients, we must restrict our analysis to some subsamples. We explain our sample selection criterion below and describe the construction of some of the key variables.

4.1 Sample Selection

In order to test for discrimination we need to identify the exact proportion of patients given diagnostic tests that are mistakenly discharged home with a major problem. The data we have only includes information on patients' ED visits. In order to use this data to identify the patients where a major problem was missed, we first identify the proportion of patients discharged from the ED that bounce back. In the ED literature (see, for example, Weinstock and Longstreth, 2007), bounceback patients are ones that return to the ED within three days of being discharged. If on the second visit, the patient is admitted to the hospital with a major problem that is different than what they were diagnosed with on their first visit, then this is a strong indication that a major problem was missed on the patient's first visit.²⁴ If a patient

²³See <http://www.oshpd.ca.gov> for more information on the data available from the OSHPD.

²⁴The return window of three days is somewhat arbitrary, as some definitions of a bounceback allow the patient to return within seven days. The key is that the return window needs to be short enough that one can assume the problem was present on the first visit, but not caught. As the

does not bounce back, it implies they do not have a major problem, and the doctor was correct in their decision to discharge them home.

In order for us to use bouncebacks to identify the patients where a major problem was missed, it must be the case that the two subsequent visits to the ED were for the same underlying problem. For most patients, this is a plausible assumption, as it would be rare for an individual to have two separate issues requiring ED treatment within three days. However, this assumption may not be plausible for older patients, as they tend to be more sickly and conceivably could return to the ED within three days for two unrelated issues. Because of this, we exclude all patients from our main analysis that are older than 65, as well as those that are on Medicare.²⁵

Major Problems. Patients may return to the ED after being discharged for various reasons with or without major problems, and importantly, the return rates may differ by race. In order to use the proportion of bounceback patients to identify the exact proportion of patients that had a *missed major problem*, two requirements must be satisfied: (1) everyone that has a missed major problem must result in a bounceback; and (2) any patient that bounces back does so because the doctor missed a major problem on their first visit. In order to satisfy these two requirements we restrict our definition of “major” problem to only include extremely serious problems which would require a patient to return to the ED.^{26,27} We also restrict this definition to only include underlying problems that cannot be affected by a patient’s behavior. For example, suppose a patient is diagnosed and discharged with a simple infection and told to take antibiotics. If they do not follow these instructions properly, the infection can turn into sepsis, and the patient will need to return and be admitted. This bounceback, however, is not because the doctor misdiagnosed the patient on the first visit, and thus should not be counted. After consulting

problems we are trying to detect are quite serious, it is likely that patients with missed diagnoses will return to the ED sooner rather than later, which is why we use three days.

²⁵Excluding patients age 65 or older effectively excludes most Medicare patients. However we also exclude the Medicare patients under 65 because these are patients that can have either permanent disabilities or congenital physical disabilities, and are thus also likely to be sickly.

²⁶For less serious problems, a patient with a missed diagnosis may choose to go to their general practitioner, who might correctly diagnose them. Because they never return to the ED, we have no way of knowing that their case was missed. In contrast, when we only examine serious problems like heart attacks, the patient will be forced to return to the ED no matter who they see.

²⁷Note that a patient might choose to return to a different ED. Because we can track patients across hospitals in California, as long as the patient returns to a California hospital, we will observe their bounceback. In New Jersey, however, we can only track patients’ visits to the same hospital, and thus if they bounce back to a different hospital we will not observe it. We will do some robustness checks with the California data to see whether this is likely to affect the results.

with an ED physician about the diagnoses that jointly satisfy both requirements, we settled on the following major problems: meningitis, encephalitis, heart attack, cardiac dysrhythmia, stroke, aneurysm, embolism, pulmonary collapse, appendicitis, intestinal obstruction, peritonitis, gastrointestinal hemorrhage, and intracranial injury.²⁸ We thus define a bounceback as a patient that returns to the ED within three days and is subsequently admitted to the hospital with, or dies from, one of these major problems.

Discharge and Bounceback. Our test requires us to identify the proportion of patients receiving diagnostic tests discharged home by an ED doctor that bounce back. This means that any patient visit whereby either the patient was admitted to the hospital, discharged by the ED doctor to a different facility, left against medical advice, or died in the ED, is not an eligible visit to be a bounceback. The only visits that are eligible to be bouncebacks occur when the ED doctor discharges the patient home. The bounceback variable is coded as one if they return to the ED within three days, are admitted to the hospital, have a principal diagnosis that is one of the major problems listed above, and their principal diagnosis is *different* from any of the diagnoses from their first visit.²⁹ For all other eligible visits, the bounceback variable is coded as zero. Importantly, since the NJ and CA data sets differ in the ability to track patients across different hospitals, a bounceback occurs in NJ if the patient returns to the *same* ED as the initial visit within three days of being discharged, but in CA a bounceback occurs if the patient returns to *any* ED because we can track patients across hospitals there.³⁰

We then arrange the visits for each patient into visit sets, where a visit set consists of all of the patient's ED visits that are within three days of each other. If a patient only has one ED visit in a three day period, there will be only one visit in the

²⁸For a list of all possible patient diagnoses please visit www.ahrq.gov/data/hcup.

²⁹There are several reasons that we require in our definition of a bounceback that the principal diagnosis in their return visits differs from any of the diagnoses from their first visit. First, doctors would often times ask patients with some major problems to return home but watch out for any worsening of the symptoms. For example, a patient diagnosed with a gastrointestinal hemorrhage may be told by the doctor to come back if it gets worse; and the patient may naturally get admitted in the second visit. We believe that we should not consider this as a misdiagnosis. Second, for many other major problems we consider, doctors would never discharge a patient home, and thus we would never observe a return visit with the same diagnosis. For example, doctors would never send home a patient with a heart attack, thus we would not realistically observe in the data that a patient is discharged home with a heart attack diagnosis in the initial visit and then comes back with a heart attack diagnosis again.

³⁰In Table 8 this is referred to as the *baseline* bounceback definition. Table 8 shows that the results do not change if we instead also define a bounceback to occur in CA if the patients return to the same ED.

visit set. Because patient visits within a visit set are likely to be related to the same underlying problem, we only include one of the visits. We assume that different visit sets for the same patient correspond to a different underlying problem. If there is no bounceback in the visit set, we only include the first eligible visit. If there is a bounceback in the visit set, we only include the ED visit that directly led to the bounceback. All other visits are dropped. Only including one visit in the visit set allows us to determine what proportion of underlying problems discharged are successfully handled, as opposed to what proportion of patient visits are successfully handled. This allows our test to be robust to any differences between minorities and whites in terms of the frequency of their visits for a given underlying problem (i.e., the number of visits in a visit set).³¹

Diagnostic Tests. To implement our empirical test, we also need to identify patients that received diagnostic tests before being discharged, as our theoretical model only predicts that bounceback rates are the same among patients of the same race who were discharged after receiving diagnostic tests.^{32, 33} The diagnostic tests patients are likely to receive to screen for these major problems include lab tests, CT scans, chest x-rays, and/or EKG's. Identifying patients that receive any of these procedures is somewhat problematic because different hospitals have different definitions of what a procedure is. The hospitals in our data are only required to record procedures that are surgical in nature or carry a procedural or anesthetic

³¹Suppose, for example, a white and minority patient come in for a problem and are both correctly discharged home. Suppose the white patient chooses to follow up with their general practitioner, but the minority patient returns back to the ED to follow up, and thus ends up with more visits in the visit set. If we counted all visits in the visit set, then the minority patient would be credited with two 'successful' visits, while the white patient would only be credited with one. This would result in us over-estimating the successful visits for minority patients.

³²Note that running our empirical test only on patients that receive diagnostic tests helps deal with the fact that minorities might use the ED differently than a white patient does. If minorities are less likely to have a general practitioner, they might go to the ED to receive treatment for more minor problems than white patients will. One might worry this will reduce the proportion of missed serious problems for minorities since these visits have extremely low risk of there being a major problem, and thus the proportion of successful problems treated will increase. However, by requiring that diagnostic tests be done we can effectively eliminate these types of low-risk visits as they will typically not be serious enough to merit diagnostic tests.

³³Our model predicts that all patients of the same race discharged after having diagnostic tests done will have the same probability of having a major problem. These major problems are mutually exclusive, and it is assumed that doctors are only screening for one of these problems (the specific one tested for depends on the patient's initial complaint). We assume doctors set the same discharge threshold across all of these major problems. As these are all extremely serious problems, this is a rational assumption. These assumptions ensure that everyone discharged has the same probability of bouncing back, and does not require us to separate out the analysis by visit reason.

risk. Because the diagnostic tests listed above are not invasive, some hospitals in New Jersey and California do not record these procedures at all. These hospitals are somewhat easy to identify, however, because none of their patients are recorded as having these procedures, which is unrealistic and implies they just do not count these diagnostic tests as procedures. Thus any hospital that records no lab tests, no CT scans, no chest x-rays or no EKG's was dropped. We also dropped hospitals where less than 10% of the patients discharged from the ED underwent any kind of diagnostic test. This included about 68% of the CA hospitals and about 25% (21 out of 83 hospitals) of the New Jersey hospitals. If a hospital is dropped, all of the corresponding eligible visits for that hospital are also dropped.^{34, 35}

One remaining issue with the above diagnostic test restrictions is that not all patients receiving diagnostic tests are actually screened for a major problem. For example, a patient that comes in with a broken leg will typically be x-rayed to aid in fixing the fracture. However, the doctor is using the x-ray test for treatment purposes, not to screen for any of the major problems. Our test requires that we identify patients that have had diagnostic tests for the purpose of screening for a major problem (since these are the patients among which the bounceback rate will be the same). To that end, we recode patients that are discharged with a diagnosis which implies they likely would not have been screened for a major problem as having zero diagnostic tests done. We consulted with an ED physician to determine the diagnoses that fit this criteria, which primarily include skin and tissue infections, bone fractures, and open wounds.

With our definition of a bounceback, we should be able to accurately identify the proportion of patients where one of the above major problems was missed. This means we can only identify whether doctors engage in discrimination when they diagnose these particular diseases. Because of the nature of our data, we cannot determine whether doctors discriminate in their diagnosis of other diseases. The strength of our test for prejudice, however, is that it is robust to underlying differences between minority and white patients, such as their propensity to use the ED. In the appendix, we describe in Table A how we arrived at our analysis sample from the raw data sets we obtained from New Jersey and California.

³⁴In Appendix B, we show that the dropped hospitals and the kept hospitals in CA and NJ do not seem to exhibit any systematic differences.

³⁵We only drop original visits that are from non-eligible hospitals. Thus, if a patient's first visit is to an eligible hospital and they bounce back to a non-eligible hospital, their original visit will be in the data set and coded as a bounceback.

4.2 Descriptive Statistics

In this section, we provide some descriptive statistics of our data set. Table 1 reports the disposition of emergency department patients in New Jersey and California. The sample used in Table 1 includes not just the patients that were discharged home (as would be in our sample in the main analysis below), but *all* ED patients including those that were admitted to the hospital, those that died in the ED, those who left the ED against medical advice, and those who were discharged elsewhere. These data sets were formed by combining ED discharges with ED visits that led to hospital admissions, and then making race, hospital, age and insurance restrictions. We restricted our attention to white, black, and Hispanic patients in New Jersey. Due to their relative number, we also included Asian patients for California. We drop all patients 65 and older, as well as those on Medicare. We also drop all visits to hospitals that did not always record the diagnostic procedures.³⁶

The overall disposition pattern is quite similar in New Jersey and California, although California has a slightly higher percentage of patients that are discharged without diagnostic tests and a lower percentage of patients discharged with diagnostic tests than New Jersey. The disposition results are then stratified by patient race. In New Jersey, we find white patients are more likely to be discharged with diagnostic tests, and less likely to be discharged without diagnostic tests. There are no substantial racial differences in California between whites, blacks, and Hispanics. However, Asians are more likely to be admitted to the hospital than other patients, and less likely to be discharged without diagnostic tests.

Table 2 reports the race, gender, age, and insurance status for the emergency department visits we included in our analysis (those that ended up in being discharged home), for both California and New Jersey. With the sample restrictions discussed above, we end up with over two million visits for both California and New Jersey. For both states, whites make up the majority (with 53.9% in New Jersey and 50.2% in California), although black and Hispanic patients make up a sizable proportion of the visits in both states. There are substantial differences between the insurance makeup of the patients in New Jersey and California. Patients in New Jersey are much more likely to have private insurance, while patients in California are more likely to be on Medicaid. In terms of age, the majority of patients in both California and New Jersey are young (age 40 and under). However, California has a higher prevalence of patients between ages 41 and 64 than New Jersey does.

Table 3 shows the breakdown of insurance status and age by race. Overall, one can see that white and Asian patients are more likely to have private insurance.

³⁶See the subsection “Diagnostic Tests” in Section 4.1 above for the hospitals that are dropped.

Dispositions	New Jersey (Jan. 2006 - July 2007)				California (Jan. 2006 - Sept. 2007)				
	All	White	Black	Hispanic	All	White	Black	Hispanic	Asian
Died in ED	0.06	0.07	0.07	0.04	0.08	0.09	0.08	0.05	0.10
Admitted	11.90	12.75	11.89	10.01	11.37	11.60	11.61	10.41	16.19
Discharged Elsewhere	0.77	0.88	0.73	0.57	2.68	2.45	3.13	2.93	1.99
Left Against Medical Advice	1.46	1.25	1.83	1.56	2.10	2.11	2.72	1.92	1.87
Discharged with Diag. Tests	52.58	55.40	48.51	50.31	44.16	44.04	43.64	44.37	45.44
Discharged w/o Diag. Tests	33.23	29.66	36.96	37.51	39.61	39.71	38.82	40.32	34.40
No. of Obs. [percentage]	2,753,194 [100%]	1,455,944 [52.88%]	645,880 [23.46%]	651,370 [23.66%]	3,037,274 [100%]	1,515,882 [49.91%]	346,658 [11.41%]	1,055,227 [34.74%]	119,507 [3.93%]

Table 1: Disposition of Emergency Department Patients in New Jersey and California.

NOTE: (1). The sample includes all ED visits to hospitals that always record diagnostic procedures, unless race information of the patient is missing or does not belong to the groups considered in the paper. Patients 65 or older and Medicare patients were also excluded.

(2). All numbers in the main table are the percentages.

Demographic Variables		New Jersey	California
Race	White	0.539	0.502
	Black	0.221	0.109
	Hispanic	0.240	0.351
	Asian		0.039
Gender	Female	0.524	0.549
Insurance Status:	Private	0.597	0.382
	Medicaid	0.105	0.291
	None	0.257	0.216
	Other	0.041	0.111
Age	0-40	0.723	0.663
	41-64	0.278	0.337
Sample Size		2,088,414	2,106,705

Table 2: Summary Statistics of Demographic Variables in New Jersey and California in the Analysis Samples.

Variable	New Jersey			California			
	White	Black	Hispanic	White	Black	Hispanic	Asian
<u>By Insurance Status:</u>							
Private	0.700	0.491	0.464	0.483	0.208	0.275	0.523
Medicaid	0.062	0.176	0.134	0.224	0.406	0.360	0.214
None	0.191	0.302	0.365	0.188	0.269	0.248	0.143
Other	0.047	0.031	0.037	0.104	0.117	0.117	0.120
<u>By Age:</u>							
0-40	0.691	0.737	0.781	0.603	0.656	0.754	0.631
41-64	0.310	0.263	0.219	0.397	0.344	0.246	0.369

Table 3: Insurance Status and Ages by Race in New Jersey and California Analysis Samples.

Black and Hispanic patients are more likely to have Medicaid or no insurance.³⁷ In terms of age, black and Hispanic patients tend to be younger than white and Asian patients.

Table 4 provides descriptive statistics on the amount of diagnostic tests received by ED patients in New Jersey and California by demographic and insurance status.³⁸ In our analysis sample, 45.8% of ED patients in New Jersey and 39.6% of ED patients in California received at least one diagnostic test before being discharged home, with the unconditional mean number of tests being 2.21 and 1.91, respectively. The mean number of diagnostic tests conditional on having at least one test done are essentially the same. The fraction of patients receiving diagnostic tests differs by demographics, as female patients are more likely to receive diagnostic tests than male patients.

Table 5 shows the bounceback rate for all eligible emergency department visits, as well as only the visit sets where diagnostic tests were done. Overall, only .03% of the visits in New Jersey and .07% of the visits in California result in a bounceback.³⁹ Part of the reason for the higher bounceback rates in California is likely due to the fact that we are using a broader definition of bounceback for CA patients (return to any ED) than that for NJ patients (return to the same ED). These bounceback rates were quite consistent with that described in the medical literature.⁴⁰ As a bounceback is a mistake that can have extremely serious consequences, we would expect the rate to be quite low. The remainder of the table breaks down the bounceback rate by race, gender, age and insurance status. The column *p*-value under each grouping comes from a Chi-Square test of whether the bounceback rate depends on the categories in that grouping; the row *p*-value tests whether the bounceback rates for discharges with and without diagnostic tests are equal against the one-sided alternative that the bounceback rate is higher for discharges with diagnostic tests.

³⁷The category "Other Insurance" includes CHAMPUS(Civilian Health and Medical Program of the Uniformed Services), Veterans Affairs Plan, Worker's Compensation, Department of Vocational Rehabilitation, other federal and non-federal programs, no charge by hospitals and others (which includes payments by governments of other countries and payments by charities).

³⁸Appendix Table C provides the summary statistics on the percentages of ED patients in our analysis receiving the nine types of diagnostic tests recorded in the data set.

³⁹In Table 5, the row *p*-value refers to the *p*-value for the null of equal bounceback rates with and without diagnostic tests against the alternative that the bounceback rate with diagnostic tests is higher than without for the sample listed in the row heading. The column *p*-value refers to the *p*-value for the null of equal bounceback rates for different rows within the same column against the alternative that they are not equal.

⁴⁰As we discussed in the introduction, Weinstock and Longstreth (2007) estimated the "serious" bounceback rates (for which hospital admission was required) due to possible medical error in the initial visit to be between 0.1% to 0.18%.

	New Jersey			California		
	Fraction with Tests (%)	Mean	Mean if > 0	Fraction with Tests (%)	Mean	Mean if > 0
Full Sample:	45.8	2.21	4.83	39.6	1.91	4.83
<u>By Race:</u>						
White	46.9	2.33	4.97	39.2	1.94	4.94
Black	44.2	2.04	4.61	40.2	1.93	4.80
Hispanic	44.7	2.10	4.69	39.4	1.83	4.65
Asian				44.5	2.23	5.02
<u>By Gender:</u>						
Male	39.5	1.73	4.39	33.7	1.56	4.64
Female	51.5	2.64	5.14	44.4	2.20	4.95
<u>By Insurance Status:</u>						
Private	47.8	2.34	4.89	45.6	2.34	5.13
Medicaid	40.6	1.79	4.40	36.4	1.64	4.49
None	45.1	2.24	4.98	35.8	1.70	4.76
Other	34.4	1.22	3.53	34.4	1.57	4.55
<u>By Age:</u>						
0-40	43.8	1.93	4.41	37.6	1.61	4.29
41-64	51.0	2.94	5.76	43.5	2.50	5.75
Sample Size	2,088,414	2,088,414	956,111	2,106,705	2,106,705	833,766

Table 4: Descriptive Statistics On the Number of Diagnostic Tests in New Jersey and California Analysis Samples, with and without Diagnostic Tests.

	New Jersey				California			
	All Visits	No Diag Tests	Diag Tests	<i>p</i> -value	All Visits	No Diag Tests	Diag Tests	<i>p</i> -value
Full Sample:	0.03	0.01	0.06	0.000	0.07	0.04	0.11	0.000
<u>By Race:</u>								
White	0.04	0.01	0.07	0.000	0.07	0.04	0.11	0.000
Black	0.03	0.02	0.06	0.000	0.06	0.04	0.10	0.000
Hispanic	0.03	0.01	0.05	0.000	0.06	0.03	0.10	0.000
Asian					0.09	0.04	0.16	0.000
<i>p</i> -value	0.004	0.134	0.034		0.001	0.092	0.012	
<u>By Gender:</u>								
Male	0.04	0.01	0.07	0.000	0.07	0.04	0.14	0.000
Female	0.03	0.01	0.05	0.000	0.06	0.04	0.09	0.000
<i>p</i> -value	0.201	0.322	0.001		0.000	0.220	0.000	
<u>By Insurance Status:</u>								
Private	0.04	0.01	0.06	0.000	0.08	0.05	0.12	0.000
Medicaid	0.03	0.01	0.06	0.000	0.06	0.03	0.10	0.000
None	0.03	0.01	0.05	0.000	0.06	0.04	0.09	0.000
Other	0.01	0.005	0.02	0.021	0.05	0.03	0.09	0.000
<i>p</i> -value	0.000	0.400	0.012		0.000	0.012	0.007	
<u>By Age:</u>								
0-40	0.02	0.008	0.05	0.000	0.05	0.02	0.08	0.000
41-64	0.06	0.03	0.09	0.000	0.10	0.07	0.15	0.000
<i>p</i> -value	0.000	0.000	0.000		0.000	0.000	0.000	
Sample Size	2,088,414	1,132,303	956,111		2,106,705	1,272,939	833,766	

Table 5: Bounceback Rates in New Jersey and California with and without Diagnostic Tests.

NOTE: All numbers, except for the *p*-values and sample sizes, are percentages.

One can see from these descriptive statistics that whites are actually slightly more likely to bounce back than blacks and Hispanics. In order to test for racial prejudice, however, we need to simultaneously control for other demographic and insurance variables in a regression framework. This will be done in Section 5.

It is also important to note that, for every subgroup listed in Table 5, the bounceback rate for those discharged with diagnostic tests is always higher than for those discharged without diagnostic tests. This is consistent with our model's implication that only patients for whom physicians' initial probability assessment that they have a major problem is sufficiently low are discharged without diagnostic tests. In a subsection below, we will present more formal tests to confirm this basic implication of the physicians' behavioral model.

5 Empirical Results

5.1 Testing the Model's Implications

Our empirical test results are only credible if the model of physician behavior in Section 2, on which our test is based, is plausible. We thus first present the results from our model's key testable implication: the bounceback rate for patients discharged after having diagnostic tests should be higher than for patients discharged without diagnostic tests, since the latter were discharged with a bounceback rate that is below the lower threshold. In Columns (1) and (2) of Tables 6, we explicitly check this by regressing the dummy of whether a discharged patient bounces back on a set of covariates, including the dummy variable of whether the patient is discharged after receiving diagnostic tests. In both states we find that the diagnostic dummy is positive and statistically significant at the 1% level. That is, controlling for the other covariates, individuals who are discharged with diagnostic tests are indeed more likely than those discharged without diagnostic tests to return to the ED within 72 hours. The magnitude of the diagnostic dummy is also quite large because the baseline average bounceback rates of all patients are respectively 0.03% and 0.07% in NJ and CA.

5.2 Main Result

Our main results are reported in Table 7. The sample used in these regressions is the set of patients who were discharged from the initial ED visit *with* at least one diagnostic test. In order to test for racial prejudice we need to determine whether

Variables	New Jersey	California
	(1)	(2)
Diagnostic Dummy	4.32*** (0.3)	7.13*** (0.4)
Black	-0.20 (0.4)	0.14 (0.826)
Hispanic	-0.53 (0.4)	0.42 (0.5)
Asian		2.64** (1.2)
Female	-1.2*** (0.3)	-2.73*** (0.4)
Medicaid	0.12 (0.5)	0.25 (0.5)
Other Insurance	-3.07*** (0.4)	-1.71*** (0.6)
No Insurance	-0.97*** (0.3)	-1.32** (0.5)
Age	0.09*** (0.0)	0.18*** (0.0)
Constant	-0.09 (0.3)	-0.43 (0.6)
Hospital Fixed Effect?	Yes	Yes
R^2	0.0003	0.0004
No. of Obs.	2,088,414	2,106,705

Table 6: The Relationship between Bounceback Rates and the Presence of Diagnostic Tests.

NOTES: (1). All reported coefficients and standard errors are the actual estimates multiplied by 10^4 ; (2). The omitted insurance category is “Private Insurance”; (3). All specifications are OLS with hospital fixed effects; (4). The standard errors, reported in parenthesis, are clustered at the hospital level and are heteroskedasticity-robust; (5). *, **, *** respectively represent statistical significance at 10%, 5% and 1%.

Anwar and Fang: Prejudice and Bouncebacks

Variables	Panel A: New Jersey			Panel B: California		
	OLS	Logit	Probit	OLS	Logit	Probit
	(1)	(2)	(3)	(4)	(5)	(6)
Black	-0.17 (0.7)	-0.09 (0.6)	-0.12 (0.6)	0.55 (1.3)	0.52 (1.2)	0.48 (1.2)
Hispanic	-0.78 (0.7)	-0.68 (0.6)	-0.65 (0.6)	1.00 (0.9)	0.84 (0.8)	0.89 (0.8)
Asian				5.39** (2.2)	4.75** (2.0)	4.80** (2.0)
Female	-1.99*** (0.5)	-1.66*** (0.5)	-1.73*** (0.5)	-5.60*** (0.8)	-4.76*** (0.7)	-4.90*** (0.7)
Medicaid	0.09 (1.0)	0.1 (0.8)	0.08 (0.8)	0.46 (1.0)	0.37 (0.8)	0.28 (0.9)
Other Insurance	-5.17*** (0.9)	-3.81*** (0.6)	-3.89*** (0.6)	-2.33* (1.3)	-1.76* (1.0)	-1.76* (1.0)
No Insurance	-1.08* (0.6)	-0.76 (0.5)	-0.82 (0.5)	-1.89* (1.1)	-1.48* (0.9)	-1.62* (0.9)
Age	0.14*** (0.0)	0.11*** (0.0)	0.12*** (0.0)	0.25*** (0.0)	0.21*** (0.0)	0.22*** (0.0)
Constant	3.64*** (0.7)			5.78*** (1.2)		
R^2	0.0002			0.0006		
No. of Obs.	956,111	932,194	932,194	833,766	828,287	828,287

Table 7: The Relationship between Bounceback Rates and Covariates in New Jersey and California.

NOTES: (1). All reported coefficients and standard errors are the actual estimates multiplied by 10^4 ; (2). Hospital fixed effect is included in all specifications; (3). The omitted insurance category is "Private Insurance"; (4). The standard errors, reported in parenthesis, are clustered at the hospital level; (5). For the Logit and Probit specifications, the coefficients reported are the marginal effects; (6). *, **, *** respectively represent statistical significance at 10%, 5% and 1%.

the bounceback rate depends on the race of the patient, while simultaneously controlling for all other variables the bounceback rate could depend on, such as age, gender and insurance status (see Section 2.3 for justification).

The results are quite similar across both states. In both New Jersey and California, the bounceback rates for blacks and Hispanics are not significantly different than the rates for whites, implying there is no racial prejudice against those groups. However, one of the downsides of using an outcome test with an event as rare as a bounceback is that we have low statistical power in detecting racial prejudice. Consequently, the confidence intervals for our race estimates are consistent with both a practically significant amount of racial prejudice both in favor of and against blacks and Hispanics. It bears pointing out, though, that while the statistical power of our test is low, more than half of our coefficients are statistically significant. Thus it is not the case that a bounceback is so rare that no coefficient will ever obtain statistical significance. Specifically, in California, we find that Asians are significantly more likely to bounce back, implying there is racial prejudice against them.

We also find significant evidence of prejudice against males. There are about 1.99 fewer bouncebacks among every 10,000 female patients discharged with diagnostic tests than among male patients; in California, this number is 5.60. Finally, as expected, the age coefficient is positive. As discussed in Section 2.3, because there is a lower expected loss from older patients suing, it is rational that doctors allow higher bounceback rates for them.⁴¹

Using the conceptual framework we outlined in Section 3, we can conclude that in our data set there is evidence of prejudice against Asian patients in California and against male patients in both states. It is important to point out that our model cannot distinguish between racial prejudice and unfounded stereotypes that a doctor may hold about the health-related behavior of certain demographic groups. For example, it is possible that Asian patients and males under-report their symptoms and the pain they are experiencing, which is reflected in c . Knowing this, doctors should adjust their prior perceptions accordingly. However if they do not take this into account enough when forming their prior perceptions, it could result in higher bounceback rates for males and Asians. Another potential reason why Asian patients have higher bounceback rates could be that they are less able to communicate their symptoms in English.⁴² This could result in physicians either not correctly ascertaining the patient's prior, or screening for the wrong issue. Future work could

⁴¹In the regressions reported in Table 7, as well as subsequent tables, the R^2 is low. This is to be expected: if the R^2 is high, it means that the ED physicians could have used the included regressors to predict who is likely to bounce back with a high degree of certainty, which then would have implied that the patient would not have been discharged.

⁴²This may not be an issue for Hispanics because there are likely to be Spanish speakers on-site. Due to the plethora of Asian languages, it is more unlikely that Asian patients have access to on-site

potentially examine whether the white-Asian bounceback differential is eliminated if we look among hospitals that have on-site translators for Asian patients

Finally, as pointed out in Section 2.3—in order to effectively test for racial and gender prejudice, we must control for all other factors that can legitimately affect bounceback rates and are correlated with race and gender. While we do control for insurance, our controls may be too coarse. For example, there are many different types of private insurance that will carry different deductibles. Another potentially important control is the exact problem the ED physician is screening for. We have been implicitly assuming that the physician allows the same bounceback rate for each of these problems, but that may not be the case if the seriousness of missing the various problem differs greatly.⁴³ Although we attempt to deal with both of these issues in our robustness checks in the next section, future work could improve our study by using better controls for both insurance and the problem the patient is being screened for.

5.3 Robustness of Results

In this subsection, we provide evidence that our basic finding above is robust to some different sample and econometric specifications. Throughout the analysis we have excluded patients older than 65 and/or on Medicare because they are unlikely to satisfy the assumption that two subsequent visits to the ED are for the same underlying problem. In Columns (1) and (3) of Table 8 we re-run our main specification for New Jersey and California, respectively, including only Medicare patients (and not making any age restrictions).^{44, 45} Although we caution that the make-up of these patients might not fit our model, one nice thing about this specification is that Medicare is one of the only insurance groups whereby all patients get comparable coverage. Thus the issue of not including fine enough insurance

translators. However, all hospitals have access to translators for every language through telephone services.

⁴³Because we have restricted our definition of bounceback to only include missing extremely serious problems, it is not unlikely that doctors will use the same discharge thresholds for each of these problems. If instead our definition of bounceback included a wide variety of missed problems, controls for the problem being screened for would be much more important.

⁴⁴In Table 8, the baseline definition of bounceback is to return to the same ED for NJ and to return to any ED for CA, within three days after being discharged from the initial ED visit; in Column (5), we code a bounceback to occur in CA if the discharged patient returns to the same ED within three days. In Columns (1) and (3) we include patients over 65, while Columns (2), (4), and (5) exclude both Medicare patients, as well as those over 65.

⁴⁵There are fewer Medicare patients than might be expected because our sample includes only ED patients that are *discharged* from the hospital after undergoing diagnostic tests. Medicare patients make up a much higher fraction of ED patients that are admitted to the hospital.

Variables	New Jersey		California		
	(1)	(2)	(3)	(4)	(5)
Black	-3.08 (1.9)	0.04 (0.9)	2.34 (3.9)	0.52 (1.3)	0.20 (1.1)
Hispanic	-4.62 (3.2)	-1.02 (0.9)	-1.78 (2.7)	0.24 (0.9)	0.40 (0.8)
Asian			1.73 (5.4)	4.31* (2.3)	5.01** (2.0)
Female	-2.40 (1.6)	-3.22*** (0.7)	-4.87** (1.9)	-6.59*** (0.9)	-3.08*** (0.7)
Medicaid		0.97 (1.2)		0.96 (1.0)	-0.83 (0.8)
Other Insurance		-2.47** (1.2)		-1.13 (1.3)	-1.37 (1.1)
No Insurance		0.46 (0.8)		-1.20 (1.1)	-2.17** (0.9)
Age	0.21*** (0.0)	0.14*** (0.0)	0.48*** (0.0)	0.19*** (0.0e-4)	0.15*** (0.0)
Constant	-2.66*** (0.6)	3.82*** (0.9)	-9.10** (3.8)	8.19*** (1.3)	5.23*** (1.0)
Fixed Effects?	Hospital	Hosp. & Visit Reason	Hospital	Hosp. & Main Diag.	Hospital
Bounceback Definition	Baseline	Baseline	Baseline	Baseline	Return to Same Hosp.
Sample	Medicare Patients	Patients w/ a Recorded Visit Reason	Medicare Patients	Patients w/ a Recorded Main Diag.	Baseline
R^2	0.0014	0.0141	0.0012	0.0294	0.0004
No. of Obs.	160,282	607,804	254,831	818,504	833,766

Table 8: Robustness of the Results.

NOTES: (1). All reported coefficients and standard errors are the actual estimates multiplied by 10^4 ; (2). The omitted insurance category is "Private Insurance"; (3). All specifications are OLS; (4) The standard errors, reported in parenthesis, are clustered at the hospital level and are heteroskedasticity robust; (5). *, **, *** respectively represent statistical significance at 10%, 5% and 1%.

controls that was discussed above is alleviated here. The results imply that our previous conclusion of no racial prejudice against blacks and Hispanics continues to hold here. However, we now find no evidence of racial prejudice against males in New Jersey and against Asians in California. Both of these coefficient estimates have the same sign as before, but are now statistically insignificant. It is unclear whether this change in significance is due to the much smaller sample size used here, or because male and Asian Medicare patients differ from their non-Medicare counterparts.

As discussed in the previous section, we would ideally like to control for the problem the ED physician is screening for in case the severity (and hence the allowable bounceback rate) differs across problems. Although this information does not exist in our data set, we can proxy for this by including fixed effects for patients' visit reason in New Jersey, and patients' main diagnosis in California. (We do not observe patients' visit reason in California, or patients' main diagnosis in New Jersey.) This implicitly assumes that patients that come in with the same visit reason in New Jersey, or are discharged with the same main diagnosis in California, were screened for the same problem. Note that the visit reason is available for less than two-thirds of the patients in New Jersey, and thus we lose a lot of observations. However, our key gender and race conclusions in both states remain unchanged.

The last robustness check we perform is to determine if the results are sensitive to the fact that in the California data, we can link patients across hospitals, while in New Jersey we cannot. So far, we have treated any Californian patient returning to *any hospital* within 72 hours of being discharged as a bounceback, while for New Jersey patients, we only treat patients returning to the *same hospital* within 72 hours of being discharged as a bounceback. It will be useful to examine whether the data limitation in New Jersey might make a difference in our inference about racial prejudice of the ED physicians. To help determine this, we examine whether the results in California change when we use the New Jersey definition of a bounceback (i.e., any bounceback whereby the patient returned to a different hospital is coded as a successful visit). Results from OLS regressions are shown for the baseline sample in Column (5) of Table 8. Once again, the coefficients do not change much as compared to Column (4) of Table 7. In particular, the coefficients on Black and Hispanic continue to be insignificant.

5.4 Results from “Inappropriate” Tests

The main contribution of our paper is to propose a test for the presence of racial prejudice among emergency department (ED) physicians based on the bounceback rates of patients discharged *after receiving diagnostic tests during their initial ED*

Variables	Dependent Variable					
	New Jersey			California		
	Uncond. Bounceback (1)	Discharge Dummy (2)	Diag. Test Dummy (3)	Uncond. Bounceback (4)	Discharge Dummy (5)	Diag. Test Dummy (6)
Black	-0.18 (0.4e)	0.020*** (5.5e-4)	0.005*** (9.7e-4)	0.27 (0.6)	0.016*** (0.001)	0.019*** (0.001)
Hispanic	-0.42 (0.4)	0.021*** (5.5e-4)	0.025*** (0.001)	0.64 (0.5)	0.005*** (4.6e-4)	0.031*** (8.3e-4)
Asian				2.95** (1.2)	0.030*** (0.001)	0.042*** (0.002)
Female	-0.70*** (0.3)	0.022*** (3.9e-4)	0.108*** (6.8e-4)	-2.05*** (0.4)	0.036*** (3.7e-4)	0.097*** (6.4e-4)
Medicaid	0.01 (0.5)	-0.028*** (6.9e-4)	-0.029*** (0.001)	-0.20 (0.5)	-0.035*** (5.1e-4)	-0.064*** (8.6e-4)
Other Insurance	-3.69*** (0.4)	0.041*** (0.001)	-0.146*** (0.002)	-2.56*** (0.6)	0.012*** (6.8e-4)	-0.118*** (0.001)
No Insurance	-1.05*** (0.3)	0.042*** (4.8e-4)	-0.020*** (8.5e-4)	-1.95*** (0.5)	0.085*** (4.9e-4)	-0.088*** (9.3e-4)
Age	0.12*** (0.0)	-0.004*** (0.0e-4)	0.004*** (2.0e-5)	0.20*** (0.0)	-0.004*** (0.1e-4)	0.003*** (2.0e-5)
Constant	1.18*** (0.3)	0.968*** (4.6e-4)	0.296*** (8.4e-4)	1.63*** (0.6)	0.985*** (5.6e-4)	0.289*** (0.001)
Sample	All Discharges	All Discharges & Hosp. Adm.	All Discharges	All Discharges	All Discharges & Hosp. Adm.	All Discharges
R ²	0.0002	0.0631	0.0469	0.0003	0.0800	0.1194
No. of Obs.	2,088,414	2,711,318	2,088,414	2,106,705	2,971,000	2,106,705

Table 9: Results from Inappropriate Tests for Racial Prejudice using New Jersey and California Data.

NOTES: (1). The reported coefficients and standard errors in Column (1) and (4) are the actual estimates multiplied by 10⁴; (2). All specifications include hospital fixed effects; (3). All specifications use OLS; (4). The standard errors, reported in parenthesis, are clustered at the hospital level and are heteroskedasticity-robust; (5). *, **, *** respectively represent statistical significance at 10%, 5% and 1%.

visit. In this section, we report in Table 9 results from other descriptive, and possibly more “standard”, tests researchers might be tempted to do when testing for racial prejudice in the ED, as we described in the text preceding Proposition 3. Specifically, Columns 1 and 4 test for racial differences in the unconditional bounceback rates among *all* discharged patients, as opposed to the subsample who were discharged home with diagnostic tests. Other potential descriptive tests include testing for racial differences in the discharge rate of patients (Columns 2 and 5) as well as the proportion of discharged patients receiving diagnostic tests (Columns 3 and 6). All of these tests were shown to be “inappropriate” tests for racial prejudice in Proposition 3. There might be racial differences in these variables either because doctors are racially prejudiced, or because there are underlying differences in the patient’s condition that are correlated with race.⁴⁶ The results show that while there are no racial differences in the bounceback rates among all patients (except for Asians), there are significant racial differences among the other descriptive indicators.⁴⁷ The race results do not, however, all go in the same direction. Specifically, black, Hispanic and Asian patients are less likely to be discharged, but given that they are discharged, are more likely to have had at least one diagnostic test done.

The results from these tests show the consequences from running incorrect tests. Overall, the results show that patient race has a significant effect on both the discharge rate and whether diagnostic tests are done. This would lead researchers using these descriptive tests to conclude racial prejudice was occurring, while our correct test implies there is none.

6 Conclusion and Discussion

In this paper we propose and empirically implement a test for the presence of racial prejudice among emergency department physicians based on the bounceback rates of patients discharged from their initial ED visit after getting diagnostic tests. A bounceback is defined as a return to the ED within 72 hours of being initially discharged. Based on a plausible theoretical model of physician behavior, we show that differential bounceback rates across patients of different racial groups who are discharged *after receiving diagnostic tests* from their ED visits are informative of

⁴⁶The sample used in the specifications in Columns 2 and 5 are all hospital discharges and admissions, where we have excluded those patients that either died in the ED or left against medical advice.

⁴⁷Note that these results which include infra-marginal patients happen to give us results similar to our previous results which did not suffer from the infra-marginality problem. While Figure 2 shows these results are likely to be different, this does not necessarily have to be the case, depending on the underlying distributions.

the racial prejudice of physicians. Applying the test to large data sets from California and New Jersey, we do not find evidence of prejudice against black and Hispanic patients, although our confidence intervals of these estimates are consistent with an economically significant amount of prejudicial behavior, both in favor of and against black and Hispanics. We do find evidence of prejudice against Asians in California, as well as against male patients in both states.

This paper contributes to the literature on outcome-based tests for racial prejudice by providing an explicit model in which the availability of continuous control variables by the decision maker – in our case the ED physicians – may generate *subsamples* in which the inframarginality problem can be avoided. In our setting, we show that bounceback rates are the same for same-race discharged patients if they received diagnostic tests in their initial ED visit, and thus applying the outcome test to this subsample is not subject to the inframarginality problem. We argue that even when the continuous control assumption is not strictly satisfied, our conditional bounceback rates test is likely to be less subject to the inframarginality problem. It is also worth emphasizing that, while we proposed our conditional bounceback rate test for prejudice in the context of ED visits, the idea of using continuous controls by the decision-makers to alleviate the inframarginality problem in the application of outcome tests can be more generally applied in other contexts.

Importantly, we also provide evidence consistent with the key testable implications of our model, lending credibility to our model and thus our empirical findings that there is no evidence of racial prejudice in the ED. Specifically, the data shows that the bounceback rates are higher for those who were discharged without diagnostic tests than those who were discharged with diagnostic tests. We also show that the conclusions from our conditional bounceback rate test differ from other commonly used, but somewhat “inappropriate” (according to our model), tests.

Finally, we should mention that, as in any empirical analysis using observational data, our test is valid only under a set of maintained assumptions, some explicit and others implicit. For example, we have implicitly assumed that the ED physicians are monolithic in their prejudice. This may not be true in practice, but we are restricted by the lack of information about the characteristics of the attending physicians in our data set. Also, we implicitly assumed that patients of different races are not sorting into Emergency Departments based on their beliefs about the prejudice in different EDs. This may also be violated in practice, but our data set does not contain patients’ home addresses. These are limitations of this study, and are fruitful directions for future research.

Appendix

In this appendix, we describe in more detail about our sample selection in Section A, the comparison of the dropped and kept hospitals in Section B, and the diagnostic tests in Section C.

A Accounting for the Sample Selection

Table A shows how the primary data samples used for both New Jersey and California were formed from the original files of all ED discharges. For New Jersey we first dropped all patients that were not either white, black or Hispanic. Panel A shows how imposing the sample restrictions discussed in Section 4 leaves us with 2,088,414 discharges. For California, we did not impose any initial sample restrictions; Panel B shows that imposing the necessary sample restrictions leaves us with 2,106,705 observations.

B Comparisons Between the Kept and Dropped Hospitals in California and New Jersey

In Section 4.1, we explained that in order to implement our conditional bounceback rate test for prejudice, we need to be able to identify patients who received diagnostic tests during their initial ED visits. As a result, we have to drop a significant number of hospitals, particularly in California, from our analysis. Table B compares the demographic characteristics of the patients, as well as their unconditional bounceback rates for the dropped and kept hospitals. There do not seem to be systematic differences between the demographic characteristics of the patients in the dropped and kept hospitals.

C Types of Diagnostic Tests Done

Table C provides the summary statistics on the percentages of ED patients in our analysis receiving the nine types of diagnostic tests recorded in the data set. The types of tests received in New Jersey and California are quite similar, with the exception of "Laboratory tests-other" and "X-ray-other", which were recorded much less frequently in California.

Action	Obs. Lost	Obs. Left
Panel A: New Jersey		
(1) Begin (<i>all ED discharges</i>)		3,858,717
(2) Drop if medical record number (MRN) is missing	481	3,858,236
(3) Drop MRNs with different race/gender	256,798	3,601,438
(4) Drop MRNs that have one or more missing admit date	151,114	3,586,324
(5) Drop duplicate observations	2073	3,584,251
(6) Drop if left AMA or discharged somewhere besides home	114,403	3,469,848
(7) Drop multiple visits in the visit set	111,362	3,358,486
(8) Drop if visit in the last three days	18,172	3,340,314
(9) Drop if died in ED	5792	3,334,522
(10) Drop if hospitals do not record certain diagnostic tests	920,648	2,413,874
(11) Drop patients older than 65	245,115	2,168,759
(12) Drop patients younger than 65 on Medicare	80,345	2,088,414
Panel B: California		
(1) Begin (<i>all ED discharges</i>)		11,659,094
(2) Drop MRNs that are different people	1,737,678	9,921,416
(3) Drop visit set with a missing return diagnosis	6,681	9,914,416
(4) Drop if left AMA or discharged somewhere besides home	539,120	9,375,615
(5) Drop multiple visits in the visit set	554,640	8,820,975
(6) Drop if visit in the last three days	42,559	8,778,416
(7) Drop if hospitals do not record certain diagnostic tests	6,043,772	2,734,644
(8) Drop if patient race is not white, black, Hispanic or Asian	116,722	2,617,922
(9) Drop if patient gender is missing	66	2,617,856
(10) Drop if died in ED	6628	2,611,228
(11) Drop patients older than 65	385,075	2,226,153
(12) Drop patients younger than 65 on Medicare	119,448	2,106,705

Table A: From the Raw Data to the Analysis Sample: Details of the Sample Selections in New Jersey and California.

Anwar and Fang: Prejudice and Bouncebacks

Variables		CA		NJ	
		Dropped Hospitals	Kept Hospitals	Dropped Hospitals	Kept Hospitals
Race	White	0.530	0.538	0.590	0.565
	Black	0.130	0.101	0.270	0.211
	Hispanic	0.30	0.319	0.140	0.224
	Asian	0.04	0.041		
Gender	Female	0.555	0.556	0.544	0.532
Insurance Status	Private	0.390	0.320	0.550	0.530
	Medicare	0.180	0.174	0.120	0.117
	Medicaid	0.210	0.238	0.090	0.092
	None	0.150	0.177	0.190	0.226
	Other	0.070	0.091	0.050	0.036
Age	0-40	0.538	0.546	0.649	0.634
	41-64	0.305	0.307	0.244	0.265
	65+	0.157	0.147	0.108	0.101
Bounceback Rates	Full Sample	0.11%	0.10%	0.04%	0.05%
	White	0.13%	0.11%	0.05%	0.06%
	Black	0.09%	0.09%	0.03%	0.04%
	Hispanic	0.09%	0.08%	0.03%	0.03%
	Asian	0.14%	0.15%		
Sample Size		5,410,296	2,611,233	920,648	2,413,874

Table B: Comparisons of Dropped and Kept Hospitals in California and New Jersey.

Notes: (1). The dropped hospitals are those that do not record diagnostic tests; (2) "Bouncebacks" are serious bouncebacks as defined in Section 4.1. Note that for CA, bouncebacks include returns to any hospital within 72 hours of initial discharge, while in NJ, bouncebacks include returns to the same hospital within 72 hours of initial discharge.

Type of Diagnostic Tests	New Jersey	California
Laboratory tests–chemistry and hematology	0.296	0.318
Laboratory tests–bacteriology and microbiology	0.192	0.168
Laboratory tests–other	0.202	0.065
CT scan–head	0.047	0.047
CT scan–other	0.064	0.054
EKG	0.097	0.092
Chest X-ray	0.117	0.120
X-ray-other	0.254	0.161
Ultrasound	0.032	0.051
Sample Size	2,088,414	2,106,705

Table C: Percentage of ED Patients Receiving Various Diagnostic Tests in New Jersey and California in the Analysis Samples.

References

- Alesina, A. and E. La Ferrara (2009): "A test of racial bias in capital sentencing," mimeo, Bocconi University and Harvard University.
- Antonovics, K. L. and B. G. Knight (2009): "A new look at racial profiling: Evidence from the boston police department," *Review of Economics and Statistics*, 91, 163–177.
- Anwar, S., P. J. Bayer, and R. Hjalmarrsson (2012): "The impact of jury race in criminal trials," *Quarterly Journal of Economics*, 127 (2), 1017–1055.
- Anwar, S. and H. Fang (2006): "An alternative test of racial profiling in motor vehicle searches: Theory and evidence," *American Economic Review*, 96, 127–151.
- Anwar, S. and H. Fang (2012): "Testing for racial prejudice in the parole board release process: Theory and evidence," NBER Working Paper 18239.
- Arrow, K. (1963): "Uncertainty and the welfare economics of medical care," *American Economic Review*, 53, 941–973.
- Ayres, I. (2002): "Outcome tests of racial disparities in police practices," *Justice Research and Policy*, vol. 4, Special Issue, Fall 2002, 4, Special Issue.
- Ayres, I. (2005): "Three tests for measuring unjustified disparate impacts in organ transplantation: the problem of "included variable" bias," *Perspectives in Biology and Medicine*, 48, S68–S87.
- Ayres, I. (2010): "Testing for discrimination and the problem of "included variable bias"," mimeo, Yale Law School.
- Ayres, I. and J. Waldfogel (1994): "A market test for race discrimination in bail setting," *Stanford Law Review*, 46.
- Balsa, A. I., T. G. McGuire, and L. S. Meridith (2005): "Testing for statistical discrimination in health care," *Health Service Research*, 40, 227–252.
- Barnato, A. E., F. L. Lucas, D. Staiger, D. E. Wennberg, and A. Chandra (2005): "Hospital-level racial disparities in acute myocardial infarction treatment and outcomes," *Medical Care*, 43 (4), 308–319.
- Becker, G. S. (1957): *The Economics of Discrimination*, University of Chicago Press: Chicago.
- Becker, G. S. (1993a): "The evidence against blacks does't prove bias," *Business Week*.
- Becker, G. S. (1993b): "Nobel lecture: The economic way of looking at behavior," *Journal of Political Economy*, 101 (3), 385–409.
- Bertrand, M. and S. Mullainathan (2004): "Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination," *American Economic Review*, 94 (4), 991–1013.
- Burstin, H., W. Johnson, S. Lipsitz, and T. Brennan (1993): "Do the poor sue more?:"

- A case-control study of malpractice claims and socioeconomic status,” *Journal of the American Medical Association*, 270 (14), 1697–1701.
- Bushway, S. D. and J. B. Gelbach (2010): “Testing for racial discrimination in bail setting using nonparametric estimation of a parametric model,” mimeo, Yale Law School.
- Center for Disease Control and Prevention (2011): “Cdc health disparities and inequalities report - united states, 2011,” <http://www.cdc.gov/mmwr/pdf/other/su6001.pdf>.
- Chandra, A. and D. O. Staiger (2008): “Identifying provider prejudice in health care,” Working Paper, Harvard University and Dartmouth College.
- Charles, K. and J. Guryan (2008): “Prejudice and wages: An empirical assessment of becker’s *The Economics of Discrimination*,” *Journal of Political Economy*, 116 (5), 773–809.
- Coate, S. and G. C. Loury (1993): “Will affirmative-action policies eliminate negative stereotypes?” *American Economic Review*, 83, 1220–1240.
- Eisenberg, J. (1986): *Doctor’s Decisions and the Cost of Medical Care*, Health Administration Press, Ann Arbor, Michigan.
- Fang, H. and A. Moro (2010): “Theories of statistical discrimination and affirmative action: A survey,” Chapter 5 in *Handbook of Social Economics*, Vol. IA, edited by J. Benhabib, A. Bisin and M. Jackson, North-Holland, 133–200.
- Goldin, C. and C. Rouse (2000): “Orchestrating impartiality: The impact of “blind” auditions on female musicians,” *American Economic Review*, 90 (4), 715–741.
- Gordon, J. A., L. C. An, R. A. Hayward, and B. C. Williams (1998): “Initial emergency department diagnosis and return visits: Risk versus perception,” *Annals of Emergency Medicine*, 32 (5), 569–573.
- Grogger, J. and G. Ridgeway (2006): “Testing for racial profiling in traffic stops from behind a veil of darkness,” *Journal of the American Statistical Association*, 101, 878–887.
- Institute of Medicine (2002): *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*, The National Academies Press. Washington, DC.
- Knowles, J., N. Persico, and P. Todd (2001): “Racial bias in motor vehicle searches: Theory and evidence,” *Journal of Political Economy*, 109, 203–228.
- Mechoulan, S. and N. Sahuguet (2011): “Assessing racial discrimination in parole release,” mimeo, HEC Montreal.
- O’Dwyer, F. and G. Bodiwala (1991): “Unscheduled return visits by patients to the accident and emergency department,” *Archives of Emergency Medicine*, 8, 196–200.
- Persico, N. (2010): “Racial profiling? detecting bias using statistical evidence,” *Annual Review of Economics*.

- Phelps, C. (2000): "Information diffusion and best practice adoption," in J. Newhouse and A. Culyer eds., *Handbook of Health Economics*, Vol. 1, North Holland, Amsterdam.
- Phelps, E. (1972): "The statistical theory of racism and sexism," *American Economic Review*, 62, 659–661.
- Pierce, J. M., A. L. Kellerman, and C. Oster (1990): "bounces: An analysis of short-term return visits to a public hospital emergency department," *Annals of Emergency Medicine*, 19 (7), 752–757.
- Polisky, D., A. K. Jha, J. Lave, M. V. Pauly, L. Cen, H. Klusaritz, Z. Chen, and K. G. Volpp (2008): "Short- and long-term mortality after an acute illness for elderly whites and blacks," *Health Service Research*, 43, 1388–1402.
- Price, J. and J. Wolfers (2010): "Racial discrimination among nba referees," *Quarterly Journal of Economics*, 125(4), 1859–1887.
- Ross, S. L. (1997): "Mortgage lending discrimination and racial differences in loan default: A simulation approach," *Journal of Housing Research*, 8 (2), 277–197.
- Ross, S. L. and J. Yinger (1999): "The default approach to studying mortgage discrimination: A rebuttal," in Margery A. Turner and Felicity Skidmore ed., *Mortgage Lending Discrimination: A Review of Existing Evidence*, Urban Institute, Washington DC.
- Ross, S. L. and J. Yinger (2002): *The Color of Credit: Mortgage Discrimination, Research Methodology, and Fair-Lending Enforcement*, MIT Press.
- Schulman, K. A., J. A. Berlin, W. Harless, J. F. K. and Shyrl Sistrunk, B. J. Gersh, R. Dub, C. K. Taleghani, J. E. Burke, S. Williams, J. M. Eisenberg, W. Ayers, and J. J. Escarce (1999): "The effect of race and sex on physicians' recommendations for cardiac catheterization," *New England Journal of Medicine*, 340, 618–626.
- Simeonova, E. (2007): "Doctors, patients, and the racial mortality gap: What are the causes?" Working Paper, Columbia University.
- Strumpf, E. (2011): "Racial/ethnic disparities in outpatient primary care: The role of physician-patient concordance," *In Press, Medical Care*.
- Weinstock, M. B. and R. Longstreth (2007): *Bouncebacks! Emergency Department Cases: ED Returns*, Anadem Publishing: Columbus, OH.
- Wilkins, P. and M. Beckett (1992): "Audit of unexpected return visits to an accident and emergency department," *Archives of Emergency Medicine*, 9, 352–256.
- Williams, R. A. (2007): *Eliminating Healthcare Disparities in America: Beyond the IOM Report*, Humana Press, New Jersey.