

NBER WORKING PAPER SERIES

TESTING FOR THE ROLE OF PREJUDICE IN EMERGENCY DEPARTMENTS
USING BOUNCEBACK RATES

Shamena Anwar
Hanming Fang

Working Paper 16888
<http://www.nber.org/papers/w16888>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2011

We would like to thank Dennis Epple, Nicola Persico, Daniel Polsky, Andy Postlewaite and Jason Wong for helpful comments. All remaining errors are ours. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

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 They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2011 by Shamena Anwar and Hanming Fang. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

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

Shamena Anwar and Hanming Fang

NBER Working Paper No. 16888

March 2011

JEL No. I11,J7

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 the patients who were discharged after receiving diagnostic tests during their initial ED visits. A bounceback is defined as a return to the ED within 72 hours of being initially discharged. Based on a plausible 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 the racial prejudice of the physicians. 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. Our finding suggests that, at least in the emergency department setting, taste based discrimination does not play an important role in the racial disparities in health care.

Shamena Anwar
Carnegie Mellon University
5000 Forbes Ave
Heinz College, Hamburg Hall Room 2116D
Pittsburgh, PA 15213
shamena@andrew.cmu.edu

Hanming Fang
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104
and NBER
hanming.fang@econ.upenn.edu

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 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,

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 et al., 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).

$\Pr(\text{disease})$, 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 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). 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

⁴[Balsa, McGuire and Meridith \(2005\)](#) referred to the two forms of statistical discrimination respectively as a *prevalence hypothesis* and a *miscommunication hypothesis*.

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.

⁶For example, [Goldin and Rouse \(2000\)](#) and [Bertrand and Mullainathan \(2004\)](#), among others, studied disparities in employment; [Ross and Yinger \(2002\)](#) 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 \(2010\)](#) in jury decisions; [Ayres and Waldfoegel \(1994\)](#) and [Bushway and Gelbach \(2010\)](#) in bail setting; [Alesina and La Ferrara \(2009\)](#) in prosecution and capital sentencing; [Mechoulan and Sahuguet \(2011\)](#) in parole releases; and [Price and Wolfers \(2007\)](#) in sports refereeing.

by doctors. For example, [Strumpf \(2010\)](#) studied the impact of concordance on quality of care received by patients of different races. She 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, even though [Chandra and Staiger \(2008\)](#) used the outcome test, their test is valid only under the assumption that the distributions of the unobserved component of the treatment effect is 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 racial 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.

Our approach of inferring about the role of racial prejudice from the bounceback rates of discharged patients differs from most of the existing literature in health disparities, which attempts to infer racial prejudice from racial disparities in care prescribed by the physicians to patients. For example, [Schulman et al. \(1999\)](#) 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 ob-

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.

served among black patients reflect an effort by the physicians to provide more appropriate care to these patients. Barnato et al. (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.

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 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 “infra-marginality 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 infra-marginality problem will not be an issue for our inference about racial prejudice. In Section 3, we formalize the infra-marginality 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. Our finding suggests that, at least in the emergency department setting, taste based discrimination does not play an important role in the racial disparities in health care.

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 having diagnostic tests done 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 infra-marginality problem. We also discuss some recent attempts to deal with the infra-marginality problem and explain why our use of bounceback rates conditional on diagnostic tests resolves the infra-marginality problem. In Section 4 we describe the data sets used in our empirical application. In Section 5 we present descriptive statistics of our sample, the basic test of our model, and our main results regarding the role of racial prejudice in the ED. In Section 6 we conclude.

2 A Model of Emergency Department Physicians' Behavior

We now present a plausible behavioral model of ED physician behavior and describe how the model can allow us to design an outcome-based test of prejudice that is not subject to the infra-marginality problem. Consider a patient with race r and other characteristics c who comes to the Emergency Department. 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. Let $F_r(c)$ be the cumulative distribution function of c among race- r patients.

Suppose that such a patient 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 assessment that a patient with race r and characteristics c has a major problem.

Determination of the Discharging Thresholds. We assume that 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 $\pi_h(r, c) > \pi_l(r, c)$. 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 .⁹ However, the ED doctor must decide on the threshold $\pi_l(r, c)$ below which they will discharge the patient from the ED.

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 $R''(\pi_l) < 0$;¹⁰ (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

⁹ Assuming 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, we will only be using the outcomes of a subset of the discharged patients to infer about racial prejudice of the physicians. If the ED physicians also control the thresholds for admission, then it is possible that the outcomes of the patients who are admitted to the hospitals may also be informative about physicians' prejudice. It is not clear, however, what will be the appropriate outcome measure for those admitted to the hospitals. Moreover, one also needs to have data of ED patients linked to their treatment and outcomes in the hospitals.

¹⁰ A 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.

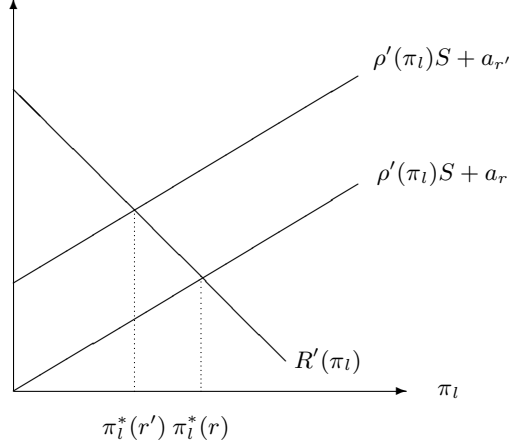


Figure 1: Graphical Illustration of the First-Order Condition: $a_r < a_{r'}$.

suffer a penalty $S > 0$, and we assume that $\rho'(\pi_l) > 0$ and $\rho''(\pi_l) > 0$;¹¹ (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 $\pi_l^*(r) = \pi_l^*(r')$.

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:

¹¹ S can reflect the cost of a lawsuit, damage compensation, as well as lost future revenues and increased malpractice insurance premiums.

- 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 costs of the diagnostic tests are born by the patients.

Under Assumption 2, the ED doctors will simply choose diagnostic tests that can allow them to make decisions upon receiving the test outcomes, without having to worry about the costs of diagnostic tests. 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 positive, will just hit π_h^* ; similarly, the doctors' posterior assessment that a race- r patient has a major problem given a negative result on the tests will just hit $\pi_l^*(r)$ where $\pi_l^*(r)$ is the optimal threshold as defined in (2) for race- r patients. Given Assumption 1, the doctors will, for race- r patient with characteristics c , choose the test- (n_f, p_f) that satisfies:

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

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

Equation (3) says that the doctor chooses test- (n_f, p_f) such that the doctor's posterior for a race- r patient with characteristics c having a major problem after observing a *positive* test outcome should just hit the upper threshold π_h^* for being admitted to the hospital. Using tests any more precise than this is a waste of resources, as there is no benefit to the ED of getting a patient to exceed the admission threshold. Similarly, Equation (4) says that the doctor would choose test- (n_f, p_f) such that the doctor's posterior for a race- r patient with characteristics c having a major problem after observing a negative test outcome should just hit the lower threshold $\pi_l^*(r)$ for being discharged home. Again, it is not necessary for doctors to use more precise testing, since the threshold $\pi_l^*(r)$ was chosen optimally by definition.

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

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

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

2.1 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 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 he/she encounters a major problem following discharge in the previous ED visit.*

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:

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

$$= \frac{\int_{\{c:\pi(r,c) \in (\pi_h^*, \pi_l^*(r))\}} \pi_l^*(r) \left\{ n_f^*(r, c) \pi(r, c) + [1 - p_f^*(r, c)] [1 - \pi(r, c)] \right\} dF_r(c)}{\int_{\{c:\pi(r,c) \in (\pi_h^*, \pi_l^*(r))\}} \left\{ n_f^*(r, c) \pi(r, c) + [1 - p_f^*(r, c)] [1 - \pi(r, c)] \right\} dF_r(c)} \quad (8)$$

$$= \pi_l^*(r). \quad (9)$$

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 5 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. In this way, our mathematical model of physician behavior allows us to develop a simple outcome test that gets around the omitted variable problem previous studies have run into.

2.2 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, seemingly related, 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 give by

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

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 $D(r|\text{Diagnostic Tests})$], or unconditional [denoted by $D(r)$], are not informative of the physicians' racial prejudice. To see this, note that the conditional and uncondi-

tional discharge rates described above are respectively:

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

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

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 $a_r = a_{r'}$. We say that the doctors engage in statistical discrimination if, conditional on any diagnostic tests being prescribed, $(n_f^*(r,c), p_f^*(r,c)) \neq (n_f^*(r',c), p_f^*(r',c))$ if $r \neq r'$.

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 $\pi_l^*(r) = \pi_l^*(r')$. 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 the types or the numbers of diagnostic tests done, which should reveal something about the test- (n_f, p_f) used, are not informative about the role of physicians' racial prejudice. The reason is simple. From expressions (5) and (6), we know that the doctors' choices of $n_f^*(r,c)$ and $p_f^*(r,c)$ depend 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, the comparison of diagnostic tests 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 the types and amount of diagnostic tests done are informative about the physicians' racial prejudice.

2.3 Discussion of the Model

So far, 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. It is useful to point out that 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) to be indexed by any *vector* of observable patient characteristics. For example, the group may be indexed by race, gender, age and insurance status, instead of just race. The logic of our proposed test for prejudice based on comparisons of conditional bounceback rates remains valid.

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 .

2.4 Testable Implications of Our Model

It is also important to recognize that our model has two testable implications. First, our model predicts that those 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.

Second, our model 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. While patients will have different tests done depending on what is necessary to get them to the discharge threshold $\pi_i^*(r)$, all patients of the same race should bounce back at exactly the same rate. The implementation of this test is not easy, though, because we are typically unable to observe the n_f and p_f directly. However, in the empirical section, we proxy for n_f and p_f with the type of test done, and show evidence in support of this prediction.

3 The Empirical Test

In this section, we describe in more details the advantages of outcome-based tests, as well as the well-known infra-marginality problem associated with the outcome test. We then explain how our model of the ED physicians' behavior allows us to avoid the infra-marginality 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.¹²

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 the ED physicians, are prejudiced against a group of patients, then that group of patients are likely to be prematurely released relative to 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 “*infra-marginality 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).

The Infra-marginality Problem and Our Proposed Solution. Figure 2 illustrates the infra-marginality 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 infra-marginality 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 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 $\pi_l^*(r') > \pi_l^*(r)$,

¹²The “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).

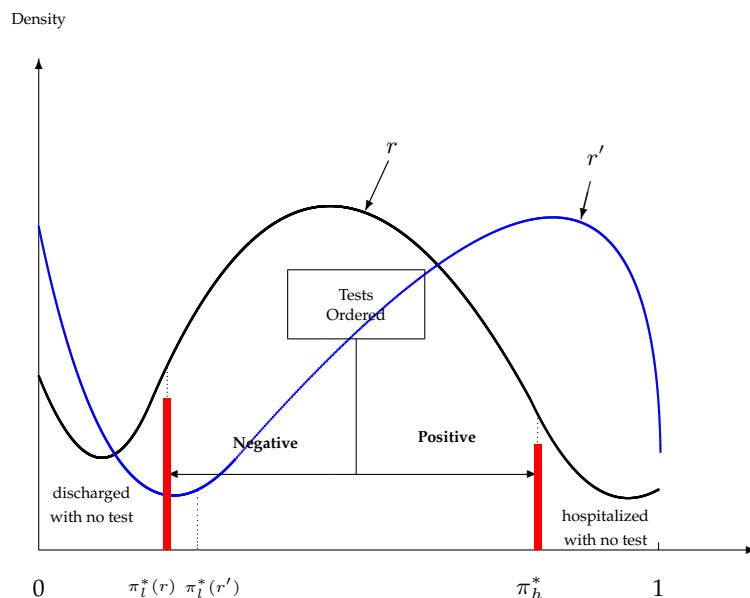


Figure 2: The Inframarginality Problem and The Proposed Solution

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 infra-marginality 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).

We should mention that the idea that continuous control variables by decision-makers may alleviate the infra-marginality 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.¹³ 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 is *no* infra-marginality 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.¹⁴

¹³The 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 \(2005, p. 14\)](#).

¹⁴One 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. However, [Mechoulan and Sahuguet \(2011\)](#) argue that police discrimination should be invariant to how parolees were granted parole. In the data, though, they found that there is a *larger* discrepancy in violations between black and white parolees who have been released through discretionary parole relative to those released through mandatory parole in states where both types of parole release coexist. These two findings are somewhat

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 infra-marginality 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 infra-marginality 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 2007.¹⁵ The HCUP databases collect patient-level hospital data from the majority of U.S. states 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 2007.¹⁶ 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, 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

contradictory and puzzling at first glance. The first finding suggests that judges are being lenient toward black prisoners in terms of the threshold probability used in releasing them, despite the fact that black prisoners typically serve a larger fraction of their sentences before obtaining parole release. One would then have expected that when parole boards are not allowed to strategically and differentially time the release of black and white prisoners, the gap in the parole violation rates would become larger, *not smaller*. One likely reason is that the crimes for which the mandatory parole rule is used differ from those for which the discretionary parole rule is used.

¹⁵See <http://www.ahrq.gov/data/hcup> and <http://www.state.nj.us/health> for more information about these data sets.

¹⁶See <http://www.oshpd.ca.gov> for more information on the data available from the OSHPD.

patient's visits to the same ED; for California, this indicator is unique across hospitals, which allow 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.¹⁷ If a patient 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.

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.^{18,19} We also restrict

¹⁷The 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 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.

¹⁸For 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.

¹⁹Note 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 bounceback 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.

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 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.²⁰ 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 patients were admitted to the hospital, or they were discharged by the ED doctor to a different facility, or they left against medical advice, or they 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 than any of the diagnoses from their first visit. 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.²¹

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 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).²²

²⁰For a list of all possible patient diagnoses please visit www.ahrq.gov/data/hcup.

²¹In Table 9 this is referred to as the *baseline* bounceback definition. Table 9 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.

²²Suppose, for example, a white and minority patient come in for a problem and are both correctly discharged home. Suppose

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.^{23, 24} 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 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.

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

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 overestimating the successful visits for minority patients.

²³Note 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.

²⁴Our 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.

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 A1 how we arrived at our analysis sample from the raw data sets we obtained from New Jersey and California.²⁵

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. For this purpose 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 and hospital restrictions. We drop all visits to hospitals that did not always record the diagnostic procedures.²⁶

From January 2006 through July 2007, there were a total of over 3.5 million visits to the relevant emergency departments in New Jersey. Of all these patients, 28.44% of them were discharged without any diagnostic tests, 49.1% were discharged after some diagnostic tests were performed, and about 20% of them were admitted to the hospital either immediately or after diagnostic tests. A small fraction (1.03%) were discharged elsewhere and some (1.36%) left the emergency department against medical advice. Finally, 0.13% died in the ED. However, the dispositions differ substantially when we stratify the sample by racial groups. We restricted our attention to white, black, and Hispanic patients. Due to their relative number, we also included Asian patients for California. For example, 23.24% of the white patients, but only 16.96% of Blacks and 13.93% of Hispanics were admitted to hospitals; 24.25% of whites, but 33.5% of Blacks and 34.73% of Hispanics, were discharged without diagnostic tests. The fractions of death are respectively 0.16% for whites, 0.12% for Blacks and 0.06% for Hispanics. The fractions of patients discharged with diagnostic tests were quite similar for whites and Hispanics at 49.97% and 49.11% respectively, and for Blacks at a somewhat lower rate of 33.5%. The overall disposition pattern is quite similar in California, though the fraction discharged without diagnostic tests at 33.41% in California is higher than, and the fraction discharged with diagnostic tests at 42.77% is lower than, their respective counterparts in New Jersey. However, the total fractions of patients discharged are remarkably similar in the two states. Racial disparities in California are also quite large. Asians seem to be much more likely to be admitted to the hospital than other patients, and much less likely to be discharged without diagnostic

²⁵The file we use for our California analysis actually has five more observations than is listed in Table A1. It is impossible to exactly replicate the file we used because the code involves grouping the patient visits. If a patient continues to return to the ED on the same day, we don't know which visit was first, and so the ordering of the visits can change, which leads to the discrepancy in observations. Of course, none of the results should be sensitive to five out of more than 2.6 million discharges.

²⁶See the subsection "Diagnostic Tests" in Section 4.1 above for the hospitals that are dropped.

tests. The death rate in ED is also much higher for Asians and whites than Blacks and Hispanics.

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 56.5% in New Jersey and 53.8% 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 Medicare and 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 older patients than New Jersey does.

Table 3 shows the breakdown of insurance status and age by race. About 60% of white patients in New Jersey have private insurance, while only 45% of black patients do. Overall, one can see that white and Asian patients are more likely to have private insurance or Medicare. Black and Hispanic patients are more likely to have Medicaid or no insurance. In terms of age, black and Hispanic patients tend to be younger than white and Asian patients.

Table 4 provides the descriptive statistics on the diagnostic tests received by ED patients in New Jersey and California by demographic and insurance status. In our analysis sample, 47.3% of ED patients in NJ and 42.6% of ED patients in CA received at least one diagnostic test before being discharged home, with the unconditional mean number of tests being 2.43 in NJ and 2.25 in CA respectively. The mean number of diagnostic tests conditional on having at least one test are also quite similar in the two states, with 5.13 in NJ and 5.27 in CA. In both states, female patients are more likely to receive diagnostic tests than male patients: 52.4% of the female patients, and 41.4% of the male patients in NJ received at least one diagnostic tests, and in CA, the respective fractions are 46.9% for females and 37.3% for males. The fraction of patients receiving diagnostic tests also differ by insurance status and age. Medicare patients are most likely to receive diagnostic tests, which is partly explained by their age as older patients are also most likely to receive diagnostic tests. The patterns are very similar in NJ and CA.

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 .05% of the visits in New Jersey and .1% of the visits in California result in a bounceback. 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.²⁷ 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

²⁷As 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%.

Dispositions	New Jersey (Jan. 2006 - July 2007)				California (Jan. 2006 - Sept. 2007)				
	All	White	Black	Hispanic	All	White	Black	Hispanic	Asian
Died in ED	0.13	0.16	0.12	0.06	0.18	0.22	0.16	0.10	0.24
Admitted	19.93	23.24	16.96	13.93	18.69	20.31	16.40	14.90	29.70
Discharged Elsewhere	1.03	1.24	0.88	0.62	3.07	3.15	3.25	3.03	2.02
Left Against Medical Advice	1.36	1.15	1.74	.54	1.89	1.83	2.55	1.82	1.56
Discharged with Diagnostic Tests	49.10	49.97	46.80	49.11	42.77	42.42	42.73	43.77	40.41
Discharged without Diagnostic Tests	28.44	24.25	33.50	34.73	33.41	32.06	34.91	36.39	26.07
No. of Obs.	3,521,366	2,022,556	763,040	735,770	4,127,704	2,248,260	428,151	1,261,229	190,064
[percentage]	[100%]	[57.44%]	[21.67%]	[20.89%]	[100%]	[54.47%]	[10.37%]	[30.56%]	[4.60%]

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.

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

(3). The null that the percentages in each row are equal is rejected against the alternative that they differ at p -value almost zero, for all rows.

Demographic Variables		New Jersey	California
Race	White	0.565	0.538
	Black	0.211	0.101
	Hispanic	0.224	0.319
	Asian		0.041
Gender	Female	0.532	0.556
Insurance Status	Private	0.530	0.320
	Medicare	0.117	0.174
	Medicaid	0.092	0.238
	None	0.226	0.177
	Other	0.036	0.091
Age	0-40	0.634	0.546
	41-64	0.265	0.307
	65+	0.101	0.147
Sample Size		2,413,874	2,611,233

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

Category	Variable	New Jersey			California			
		White	Black	Hispanic	White	Black	Hispanic	Asian
Insurance Status	Private	0.597	0.451	0.437	0.379	0.186	0.249	0.417
	Medicare	0.152	0.085	0.056	0.227	0.125	0.098	0.202
	Medicaid	0.052	0.160	0.127	0.170	0.352	0.325	0.177
	None	0.160	0.276	0.345	0.143	0.235	0.224	0.112
	Other	0.039	0.029	0.035	0.081	0.102	0.105	0.093
Age	0-40	0.579	0.678	0.731	0.465	0.581	0.678	0.484
	41-64	0.282	0.266	0.219	0.340	0.346	0.240	0.295
	65+	0.139	0.056	0.050	0.195	0.073	0.082	0.221

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

	New Jersey				California			
	Fraction with Tests	Mean	Mean if > 0		Fraction with Tests	Mean	Mean if > 0	
Full Sample	47.3%	2.43	5.13		42.6%	2.25	5.27	
Race								
White	48.6%	2.60	5.36		43.1%	2.36	5.47	
Black	45.4%	2.18	4.80		42.3%	21.4	5.06	
Hispanic	45.6%	2.21	4.85		41.3%	2.04	4.94	
Asian					47.8%	2.60	5.45	
Gender								
Male	41.4%	1.96	4.74		37.3%	1.91	5.12	
Female	52.4%	2.83	5.41		46.9%	2.52	5.36	
Insurance Status								
Private	48.0%	2.37	4.95		45.8%	2.38	5.19	
Medicare	57.0%	3.85	6.75		56.0%	3.68	6.57	
Medicaid	40.8%	1.81	4.43		36.6%	1.66	4.53	
None	45.2%	2.26	5.00		35.9%	1.72	4.79	
Other	34.5%	1.23	3.55		34.7%	1.59	4.59	
Age								
0-40	43.9%	1.94	4.42		37.7%	1.62	4.30	
41-64	51.2%	2.97	5.79		44.1%	2.54	5.76	
65+	58.0%	4.03	6.96		57.8%	3.95	6.83	
Sample Size	2,413,874	2,413,874	1,140,732		2,611,233	2,611,233	1,113,415	

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	0.05	0.03	0.07	0.000		0.10	0.07	0.14	0.000	
Race										
White	0.06	0.04	0.08	0.000		0.11	0.09	0.15	0.000	
Black	0.04	0.02	0.06	0.000		0.09	0.06	0.12	0.000	
Hispanic	0.03	0.02	0.05	0.000		0.08	0.05	0.12	0.000	
Asian						0.15	0.12	0.19	0.002	
<i>p</i> -value	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	
Gender										
Male	0.05	0.03	0.08	0.000		0.11	0.08	0.17	0.000	
Female	0.05	0.03	0.06	0.000		0.09	0.07	0.12	0.000	
<i>p</i> -value	0.737	0.07	0.000	0.000		0.000	0.705	0.000	0.000	
Insurance Status										
Private	0.04	0.02	0.07	0.000		0.09	0.05	0.13	0.000	
Medicare	0.12	0.14	0.10	.999		0.25	0.29	0.22	1.000	
Medicaid	0.03	0.01	0.06	0.000		0.06	0.03	0.11	0.000	
None	0.03	0.02	0.06	0.000		0.06	0.04	0.09	0.000	
Other	0.01	0.01	0.02	0.009		0.05	0.03	0.10	0.000	
<i>p</i> -value	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	
Age										
0-40	0.02	0.01	0.05	0.000		0.05	0.02	0.08	0.000	
41-64	0.06	0.03	0.09	0.000		0.10	0.08	0.15	0.000	
65+	0.15	0.19	0.13	0.9995		0.24	0.33	0.26	1.000	
<i>p</i> -value	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	
Sample Size	2,413,874	1,273,142	1,140,732		2,611,233	1,497,818	1,113,415			

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

NOTE: (1). All numbers, except for the *p*-values and sample sizes, are percentages.

(2). The row *p*-values is 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 null of equal bounceback rates for different rows within the same column against the alternative that they are not equal.

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.

One can see from these descriptive statistics that whites are actually more likely to bounce back than blacks and Hispanics, and Medicare patients are significantly more likely to bounce back than patients with other types of insurance. One likely reason for this is because white patients and Medicare patients are older on average than the other patients in their respective grouping. As one can see from the last panel of Table 5, the bounceback rate increases significantly as patients get older. While this might reflect age discrimination, it more likely reflects the fact that doctors are less likely to get sued if they miss a serious problem in an older patient. The consequences are more severe when they miss a serious problem in a relatively young patient. If the probability of getting sued decreases as patient age increases, doctors will rationally allow for a higher threshold in dismissing older patients. The regression analysis will allow us to determine if the racial differences in the bounceback rate result from discrimination, or are simply due to the different age distributions among the different races.

It is also important to note that, with the exception of Medicare patients and patients 65 and older, 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 Main Result

Our main results are reported in Panel B of Tables 6 and 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 the bounceback rate depends on the race of the patient. We also need to control for all other variables the bounceback rate could depend on, such as age, gender and insurance status.²⁸

Panel B of Table 6 shows that among the patients who were discharged home with diagnostic tests in New Jersey, the bounceback rates do not differ by race of the patients. If anything, Black and Hispanic patients have somewhat lower bounceback rates than the omitted racial group, the White patients. Female patients have a significantly lower bounceback rates than male patients. Interestingly, we find that patients with Medicare and other insurance, once we control for age, have lower bounceback rates. As ex-

²⁸The bounceback rate can depend on age, gender and insurance status either because doctors discriminate against people based on these variables, or because it is acceptable for the bounceback rates to differ within these variables.

Variables	Panel A: All Discharges			Panel B: Discharges with Diagnostic Tests		
	OLS (1)	Logit (2)	Probit (3)	OLS (4)	Logit (5)	Probit (6)
Diagnostic Dummy	2.96e-4*** (0.3e-4)	2.18e-4*** (0.2e-4)	2.47e-4*** (0.2e-4)			
Black	-0.67e-4 (0.4e-4)	-0.28e-4 (0.3e-4)	-0.32e-4 (0.3e-4)	-0.62e-4 (0.7e-4)	-0.43e-4 (0.6e-4)	-0.45e-4 (0.6e-4)
Hispanic	-0.58e-4 (0.4e-4)	-0.43e-4 (0.3e-4)	-0.44e-4 (0.3e-4)	-0.99e-4 (0.7e-4)	-0.86e-4 (0.6e-4)	-0.84e-4 (0.6e-4)
Female	-1.14e-4*** (0.3e-4)	-0.66e-4*** (0.2e-4)	-0.72e-4*** (0.2e-4)	-2.20e-4*** (0.5e-4)	-1.77e-4*** (0.4e-4)	-1.85e-4*** (0.4e-4)
Medicare	0.65e-4 (0.7e-4)	-0.96e-4*** (0.2e-4)	-0.95e-4*** (0.2e-4)	-4.34e-4*** (1.1e-4)	-2.88e-4*** (0.4e-4)	-3.09e-4*** (0.5e-4)
Medicaid	0.19e-4 (0.5e-4)	-0.05e-4 (0.4e-4)	-0.05e-4 (0.4e-4)	0.16e-4 (1.0e-4)	0.1e-4 (0.9e-4)	0.08e-4 (0.9e-4)
Other Insurance	-3.85e-4*** (0.4e-4)	-2.38e-4*** (0.3e-4)	-2.43e-4*** (0.3e-4)	-5.39e-4*** (0.9e-4)	-3.98e-4*** (0.7e-4)	-4.09e-4*** (0.7e-4)
No Insurance	-1.44e-4*** (0.3e-4)	-0.73e-4** (0.2e-4)	-0.78e-4** (0.3e-4)	-1.47e-4* (0.6e-4)	-1.02e-4 (0.5e-4)	-1.08e-4* (0.5e-4)
Age	0.16e-4*** (0.0e-4)	0.11e-4*** (0.0e-4)	0.11e-4*** (0.0e-4)	0.18e-4*** (0.0e-4)	0.14e-4*** (0.0e-4)	0.15e-4*** (0.0e-4)
Constant	-0.76e-4* (0.3e-4)			2.97e-4*** (0.6e-4)		
Hospital Fixed Effect?	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.000461			0.00032		
No. of Obs.	2,413,874	2,355,727	2,355,727	1,140,732	1,113,424	1,113,424

Table 6: The Relationship between Bounceback Rates and Covariates in New Jersey.

NOTES: (1). The standard errors, reported in parenthesis, are clustered at the hospital level; (2). For the Logit and Probit specifications, the coefficients reported are the marginal effects; (3). *, **, *** respectively represent statistical significance at 10%, 5% and 1%.

Variables	Panel A: All Discharges			Panel B: Discharges with Diagnostic Tests		
	OLS	Logit	Probit	OLS	Logit	Probit
	(1)	(2)	(3)	(4)	(5)	(6)
Diagnostic Dummy	4.13e-4*** (0.4e-4)	2.88e-4*** (0.3e-4)	3.40e-4*** (0.4e-4)			
Black	0.05e-4 (0.7e-4)	0.4e-4 (0.6e-4)	0.39e-4 (0.6e-4)	0.46e-4 (1.3e-4)	0.7e-4 (1.2e-4)	0.69e-4 (1.2e-4)
Hispanic	0.67e-4 (0.5e-4)	0.34e-4 (0.4e-4)	0.42e-4 (0.4e-4)	0.74e-4 (0.9e-4)	0.55e-4 (0.8e-4)	0.67e-4 (0.8e-4)
Asian	3.55e-4** (1.3e-4)	2.19e-4** (0.8e-4)	2.43e-4** (0.9e-4)	4.23e-4* (2.1e-4)	3.18e-4* (1.6e-4)	3.43e-4* (1.7e-4)
Female	-3.14e-4*** (0.4e-4)	-2.10e-4*** (0.3e-4)	-2.28e-4*** (0.3e-4)	-5.45e-4*** (0.8e-4)	-4.41e-4*** (0.6e-4)	-4.67e-4*** (0.7e-4)
Medicare	6.90e-4*** (0.9e-4)	0.5e-4 (0.5e-4)	0.85e-4 (0.5e-4)	-1.47e-4 (1.4e-4)	-2.71e-4*** (0.8e-4)	-2.76e-4** (0.9e-4)
Medicaid	0.53e-4 (0.5e-4)	-0.31e-4 (0.5e-4)	-0.31e-4 (0.5e-4)	0.55e-4 (1.0e-4)	0.14e-4 (0.9e-4)	0.08e-4 (1.0e-4)
Other Insurance	-1.91e-4** (0.6e-4)	-1.87e-4*** (0.5e-4)	-1.97e-4*** (0.5e-4)	-1.97e-4 (1.3e-4)	-1.77e-4 (1.2e-4)	-1.79e-4 (1.2e-4)
No Insurance	-1.53e-4** (0.5e-4)	-1.48e-4** (0.5e-4)	-1.56e-4** (0.5e-4)	-2.23e-4* (1.1e-4)	-2.05e-4* (1.0e-4)	-2.19e-4* (1.0e-4)
Age	0.28e-4*** (0.0e-4)	0.21e-4*** (0.0e-4)	0.22e-4*** (0.0e-4)	0.33e-4*** (0.0e-4)	0.28e-4*** (0.0e-4)	0.29e-4*** (0.0e-4)
Constant	-2.11e-4*** (0.6e-4)			2.96e-4** (1.1e-4)		
Hospital Fixed Effect?	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.000863			0.00069		
No. of Obs.	2,611,233	2,611,233	2,611,233	1,113,415	1,110,808	1,110,808

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

NOTES: (1). The standard errors, reported in parenthesis, are clustered at the hospital level; (2). For the Logit and Probit specifications, the coefficients reported are the marginal effects; (3). *, **, *** respectively represent statistical significance at 10%, 5% and 1%.

pected, we find that older patients are more likely to bounceback than younger patients.²⁹ These findings are quantitatively and qualitatively similar across the OLS, Logit and Probit specifications.

The results for California, reported in Panel B of Table 7 are mostly similar to those for New Jersey. In all specifications, we find that the bounceback rates for Black and Hispanic patients who were discharged from their initial ED visit with diagnostic tests are somewhat higher than that of the White patients, but the difference is not statistically different from zero. The only exception is Asian patients, whose bounceback rate is higher than that for the White patients, and the difference is statistically significant at 10% level. We also find that, similar to New Jersey, Female patients have a significantly lower bounceback rates than male patients, and patients with Medicare and other insurance have lower bounceback rates than those with private insurance; finally, we find that older patients are more likely to bounceback than younger patients.

Using the conceptual framework we outlined in Section 3, we can conclude that in our data set, there is no evidence that the ED physicians exhibit prejudice in their treatment decisions against black and Hispanic patients relative to white patients, though there is some evidence of prejudice against Asian patients in California as they exhibit a higher bounceback rate.

5.2 Testing the Model's Implications

Our test result is credible only if the behavioral model of physician behavior in Section 2, on which our test is based, is plausible. Fortunately, as discussed in Section 2 our model has some testable implications. Firstly, the model predicts that 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 Panel A of Tables 6 and 7, 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 all specifications, 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 0.05% and 0.10% in NJ and CA.

The second implication of our model is that, among those discharged with diagnostic tests, the accuracy of the diagnostic tests used (as measured by n_f and p_f) should *not* affect the bounceback rate conditional on race (and/or any observable characteristics that the physicians may base their prejudice on). Unfortunately, while we know the diagnostic test codes for each patient, we do not know the accuracy of these tests as measured by the false positive and false negative rates. To proxy for this, we stratified the various diagnostic tests into six groups: lab tests (which primarily include blood and urine

²⁹Thus the higher bounceback rates of Medicare patients reported in Table 5 is completely driven by the older ages of the Medicare patients.

Variables	New Jersey		California	
	(1)	(2)	(3)	(4)
Accurate	-0.57e-4 (0.5e-4)	0.01e-4 (0.5e-4)	1.45e-4* (0.7e-4)	0.62e-4 (0.7e-4)
Black	-0.63e-4 (0.7e-4)	-0.15e-4 (0.7e-4)	0.47e-4 (1.3e-4)	1.63e-4 (1.2e-4)
Hispanic	-0.99e-4 (0.7e-4)	-0.62e-4 (0.7e-4)	0.75e-4 (0.9e-4)	1.41e-4 (0.9e-4)
Asian			4.23e-4** (2.1e-4)	5.17e-4** (2.2e-4)
Female	-2.24e-4*** (0.5e-4)	-1.81e-4*** (0.5e-4)	-5.38e-4*** (0.8e-4)	-5.32e-4*** (0.8e-4)
Medicare	-4.36e-4*** (1.1e-4)	-4.63e-4*** (1.1e-4)	-1.43e-4 (1.4e-4)	-3.52e-4** (1.7e-4)
Medicaid	0.12e-4 (1.0e-4)	0.04e-4 (0.9e-4)	.58e-4 (1.0e-4)	0.21e-4 (1.0e-4)
Other Insurance	-5.34e-4*** (0.9e-4)	-5.14e-4*** (0.9e-4)	-1.99e-4 (1.3e-4)	-2.33e-4* (1.3e-4)
No Insurance	-1.51e-4** (0.6e-4)	-1.06e-4* (0.6e-4)	-2.21e-4** (1.1e-4)	-2.02e-4* (1.1e-4)
Age	0.18e-4*** (0.0e-4)	0.13e-4*** (0.0e-4)	0.33e-4*** (0.0e-4)	0.24e-4*** (0.0e-4)
Constant	3.4e-4*** (0.7e-4)	3.55e-4*** (0.8e-4)	2.09e-4 (1.3e-4)	5.09e-4*** 1.3e-4
Hospital Fixed Effect?	Yes	Yes	Yes	Yes
Sample	All	Age < 65	All	Age < 65
R^2	0.0003	0.0002	0.0007	0.0005
No. of Obs.	1,140,732	998,664	1,113,415	890,838

Table 8: The Relationship between Bounceback Rates and the Accuracy of the Diagnostic Tests.

NOTES: (1). All specifications are OLS with hospital fixed effects; (2) The standard errors, reported in parenthesis, are clustered at the hospital level and are heteroskedasticity-robust; (3). *, **, *** respectively represent statistical significance at 10%, 5% and 1%.

tests), electrocardiograms (EKG's), ultrasounds, x-rays, CT scans, and other tests. The costs of these tests can be measured by both how long they take to perform and get the results back (time-intensiveness), and how invasive they are to patients. Overall, lab tests and EKG's are considered to be the least costly, as opposed to ultrasounds which are more time-intensive, x-rays which are more invasive, and CT scans which are both time-intensive and invasive. Thus, if lab tests and EKGs are helpful in screening for a major problem the doctor is concerned about, doctors should always want to do these tests first. The only reason ED doctors would choose to do other tests (either alone or in combination with these tests) is if they were more accurate. Thus, we create the indicator variable "*accurate*" which equals zero if the patient had either an EKG only, lab tests only, or both an EKG and lab tests; "*accurate*" equals one if the patient had any of the more costly tests (either by themselves or in combination with lab tests and EKG's). Table 8 presents the results from regressing the bounceback indicator on the traditional covariates, as well as the *accurate* variable for the sample of patients that received diagnostic tests. If our model is correct the coefficient on *accurate* should be insignificant, which implies that the accuracy of the tests used does not affect the bounceback rate. This is indeed what we find for New Jersey in Column (1). We repeat this analysis for California in Column (3), but do find the accuracy of the tests is significant at the 10% level. However, when we restrict to the sub-sample of patients who are younger than 65, shown in Columns (2) and (4), we find that the accuracy of the tests has no effect on bouncebacks for both New Jersey and California. Recall that earlier in Table 5 we found evidence that the first test of our model does not work on patients over 65, as the bounceback rate is higher for those patients without diagnostic tests. Taken together, these results provide support for our model for patients less than 65.³⁰

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. The model checks performed above implied that our model seems to work best for patients younger than 65. We thus check if the same general results we found in Panel B of Tables 6 and 7 for the full sample holds up for patients younger than 65. We present the OLS regression results in Columns (1) and (2) of Table 9. For New Jersey, none of the coefficients change appreciably. For California, the only changes are that the coefficients on Medicaid and other insurance now become significant, although the signs remain the same. The coefficients for Blacks and Hispanics continue to be insignificant for both states, implying that our conclusion of no racial prejudice against these groups

³⁰We also tested the second implication of our model by using the number of diagnostic tests a patient received as a proxy for the accuracy of the tests done, making the plausible assumption that the more tests done, the higher the overall accuracy should be. However, we found that the coefficient estimates on the number of diagnostic tests are positive and statistically significant, which goes against the model's predictions. One potential reason for this positive correlation is that a high number of tests can indicate the doctor is very uncertain about which of the various major problems the patient can potentially have, and thus tests them for many different things. Our model works best in the situation where a doctor has narrowed down the potential major problems to only a few and tests for those. Using the number of diagnostic tests can stratify patients according to doctor's uncertainty about what the major problems are, which might be why the test fails. In contrast, the regressions shown in Table 8 that use the type of test as a proxy for accuracy do not necessarily stratify patients according to doctor's uncertainty.

Variables	New Jersey		California	
	(1)	(2)	(3)	(4)
Black	-0.15e-4 (0.7e-4)	1.62e-4 (1.2e-4)	-0.09e-4 (1.0e-4)	0.67e-4 (1.0e-4)
Hispanic	-0.62e-4 (0.7e-4)	1.40e-4 (0.9e-4)	0.60e-4 (0.8e-4)	0.58e-4 (0.8e-4)
Asian		5.17e-4** (2.2e-4)	4.68e-4** (1.9e-4)	4.80e-4*** (1.9e-4)
Female	-1.81e-4*** (0.5e-4)	-5.37e-4*** (0.8e-4)	-2.79e-4*** (0.6e-4)	-2.89e-4*** (0.6e-4)
Medicare	-4.64e-4*** (1.1e-4)	-3.54e-4** (1.7e-4)	-2.29e-4** (1.1e-4)	-5.44e-4*** (1.3e-4)
Medicaid	0.04e-4 (0.9e-4)	0.20e-4 (1.0e-4)	-0.52e-4 (0.8e-4)	-1.00e-4 (0.8e-4)
Other Insurance	-5.14e-4*** (0.9e-4)	-2.32e-4* (1.3e-4)	-1.19e-4 (1.2e-4)	-1.32e-4 (1.1e-4)
No Insurance	-1.06e-4* (0.6e-4)	-2.03e-4* (1.1e-4)	-2.21e-4** (0.9e-4)	-2.22e-4** (0.9e-4)
Age	0.13e-4*** (0.0e-4)	0.25e-4*** (0.0e-4)	0.22e-4*** (0.0e-4)	0.15e-4*** (0.0e-4)
Constant	3.56e-4*** (0.7e-4)	5.47e-4*** (1.2e-4)	2.82e-4*** (1.0e-4)	5.13e-4*** (1.0e-4)
Hospital Fixed Effect?	Yes	Yes	Yes	Yes
Bounceback Definition	Baseline	Baseline	Return to Same Hospital	Return to Same Hospital
Sample	Age < 65	Age < 65	All	Age < 65
R^2	0.0002	0.0005	0.0004	0.0003
No. of Obs.	998,664	890,838	1,113,415	890,838

Table 9: Robustness of the Results to Restrictions to Patients Less than 65 Only and to Using an Alternative Bounceback Definition for California.

NOTES: (1). 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 Columns (3) and (4), we also code a bounceback to occur in CA if the discharged patient returns to the same ED within three days; (2). All specifications are OLS with hospital fixed effects; (3) The standard errors, reported in parenthesis, are clustered at the hospital level and are heteroskedasticity robust; (4). *, **, *** respectively represent statistical significance at 10%, 5% and 1%.

continues to hold for this subsample. Another 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 examine 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 full sample in Column (3) of Table 9, and for the age-restricted sample in Column (4). 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

In this section, we report in Tables 10 and 11 the results from other descriptive 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, Column 1 in these two tables test for racial difference 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 following: the discharge rate of patients (Column 2), the proportion of discharged patients receiving diagnostic tests (Column 3), the number of diagnostic tests discharged patients received given they received at least one (Column 4), and the accuracy of the diagnostic tests done (Column 5). 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. The results of these OLS regressions are shown in Tables 10 and 11 for New Jersey and California, respectively, where we proxy for the accuracy of the diagnostic tests done by the indicator variable "*accurate*" dummy described earlier.³¹ Table 10 shows while there are no racial differences in the bounceback rates among all patients, there are significant racial differences among the other descriptive indicators. The race results do not, however, all go in the same direction. Specifically, black patients are more likely to be discharged, but are also more likely to have had at least one diagnostic test done. Conditional on having at least one test done, blacks get less tests done, and also get less accurate ones. Meanwhile, Hispanics are less likely to be discharged and are more likely to have gotten at least one diagnostic test. They also get more diagnostic tests done, although the tests done are less accurate. Table 11 shows a similar pattern for California, in that there are no racial differences among bounceback rates for all discharged patients (except for Asians), while there are significant racial

³¹The sample used in the specifications in Column 2 are all hospital discharges and admissions, where we have excluded those patients that either died in the ED or left against medical advice.

Variables	Dependent Variable				
	Unconditional	Discharge	Diagnostic Test	Number of	Accurate
	Bounceback	Dummy	Dummy	Diagnostic Tests	Dummy
(1)	(2)	(3)	(4)	(5)	
Black	-0.63e-4 (0.4e-4)	8.07e-4* (4.9e-4)	0.011*** (9.1e-4)	-0.182*** (0.01)	-0.023*** (0.001)
Hispanic	-0.5e-4 (0.4e-4)	-0.008*** (5.0e-4)	0.026*** (9.5e-4)	0.023*** (0.01)	-0.010*** (0.001)
Female	-0.86e-4*** (0.3e-4)	0.029*** (3.4e-4)	0.095*** (6.4e-4)	0.602*** (0.007)	-0.060*** (8.6e-4)
Medicare	0.5e-4*** (0.9e-4)	0.580*** (4.8e-4)	-0.049*** (0.001)	-0.133*** (0.015)	-0.067*** (0.002)
Medicaid	0.09e-4 (0.5e-4)	0.217*** (4.7e-4)	-0.034*** (0.001)	0.143*** (0.013)	-0.069*** (0.002)
Other Insurance	-4.28e-4*** (0.4e-4)	0.313*** (5.1e-4)	-0.144*** (0.002)	-1.62*** (0.02)	0.075*** (0.002)
No Insurance	-1.50e-4*** (0.3e-4)	0.307*** (3.5e-4)	-0.019*** (8.4e-4)	0.074*** (0.01)	-0.067*** (0.001)
Age	0.17e-4*** (0.0e-4)	-0.008*** (0.0e-4)	0.003*** (0.2e-4)	0.052*** (2.1e-4)	0.002*** (0.2e-4)
Constant	0.17e-4 (0.3e-4)	0.974*** (4.3e-4)	0.315*** (8.0e-4)	2.91*** (0.009)	0.668*** (0.001)
Sample	All Discharges	All Discharges and Hospital Admissions	All Discharges	All Discharges with at least one Diag. Test	All Discharges with at least one Diag. Test
R ²	0.000417	0.3882	0.0449	0.1243	0.0495
No. of Obs.	2,413,874	3,468,865	2,413,874	1,140,732	1,140,732

Table 10: Results from Inappropriate Tests for Racial Prejudice using New Jersey Data.

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

Variables	Dependent Variable				
	Unconditional		Diagnostic Test		Number of
	Bounceback	Discharge	Dummy	Dummy	Diagnostic Tests
(1)	(2)	(3)	(4)	(5)	
Black	0.14e-4 (0.7e-4)	0.012*** (4.1e-4)	0.023*** (0.001)	0.033** (0.014)	-0.010*** (0.002)
Hispanic	0.81e-4 (0.5e-4)	0.007*** (3.0e-4)	0.035*** (7.7e-4)	0.050*** (0.01e-4)	-0.005*** (0.001)
Asian	3.76e-4** (1.3e-4)	-0.024*** (5.8e-4)	0.048*** (0.002)	0.088*** (0.02)	0.9e-4 (0.002)
Female	-2.80e-4*** (0.4e-4)	-0.002*** (2.2e-4)	0.082*** (5.9e-4)	0.224*** (0.008)	-0.050*** (8.9e-4)
Medicare	6.98e-4*** (0.9e-4)	0.111*** (2.7e-4)	0.019*** (0.001)	-0.158*** (0.01)	-0.032*** (0.001)
Medicaid	0.27e-4 (0.5e-4)	-0.038*** (1.7e-4)	-0.064*** (8.4e-4)	-0.162*** (0.01)	-0.022*** (0.001)
Other Insurance	-2.39e-4*** (0.6e-4)	-0.669*** (4.6e-4)	-0.116*** (0.001)	-0.605*** (0.02)	0.012*** (0.002)
No Insurance	-1.89e-4* (0.5e-4)	-0.019*** (1.9e-4)	-0.088*** (9.2e-4)	-0.208*** (0.01)	-0.014*** (0.002)
Age	0.29e-4*** (0.0e-4)	-0.004*** (0.0e-4)	0.003*** (0.2e-4)	0.055*** (2.2e-4)	0.003*** (0.3e-4)
Constant	-0.9e-4 (0.6e-4)	1.14*** (3.2e-4)	0.294*** (9.4e-4)	2.88*** (0.01)	0.600*** (.001)
Sample	All	All Discharges and	All	All Discharges	All Discharges
R^2	Discharges	Hospital Admissions	Discharges	with at least one Diag. Test	with at least one Diag. Test
No. of Obs.	0.000826	0.7006	0.1112	0.1504	0.0546
	2,611,233	4,042,262	2,611,233	1,113,415	1,113,415

Table 11: Results from Inappropriate Tests for Racial Prejudice using California Data.

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

differences among the other indicators that go in different directions.

The results from these tests show the consequences from running incorrect tests. Overall, the results show that patient race has a significant effect on the discharge rate, whether diagnostic tests are done, and the amount and accuracy of the tests 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

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 the patients who are 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 the racial prejudice of the physicians. Applying the test to large data sets from California and New Jersey, we do not find evidence of racial prejudice. Our finding suggests that, at least in the emergency department setting, taste based discrimination does not play an important role in the racial disparities in health care.

This paper contributes to the literature on the outcome-based test 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 infra-marginality problem for the use of the outcome test can be avoided. In our setting, we show that the bounceback rates are the same for same-race discharged patients if they received diagnostic tests in the initial ED visits, and thus applying the outcome test to this subsample is not subject to the infra-marginality problem.

Importantly, we also provide evidence broadly consistent with the 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, two testable implications of our model are empirically examined and supported. First, 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. Second, we show that a measure of the accuracy of the diagnostic tests does not predict the bounceback rates once conditioning on race and other characteristics of the patients. Finally, we show that the conclusions from our conditional bounceback rate test differ from other commonly used, but inappropriate (according to our model), tests.

References

- Alesina, Alberto and Eliana La Ferrara**, "A Test of Racial Bias in Capital Sentencing," mimeo, Bocconi University and Harvard University, 2009.
- Antonovics, Kate L. and Brian G. Knight**, "A New Look at Racial Profiling: Evidence from the Boston Police Department," *Review of Economics and Statistics*, 2009, 91, 163–177.
- Anwar, Shamena and Hanming Fang**, "An Alternative Test of Racial Profiling in Motor Vehicle Searches: Theory and Evidence," *American Economic Review*, 2006, 96, 127–151.
- , **Patrick J. Bayer, and Randi Hjalmarsson**, "Jury Discrimination in Criminal Trials," mimeo, ERID Working Paper No. 55, Duke University, 2010.
- Arrow, Kenneth**, "Uncertainty And The Welfare Economics Of Medical Care," *American Economic Review*, 1963, 53, 941–973.
- Ayres, Ian**, "Outcome Tests Of Racial Disparities In Police Practices," *Justice Research and Policy*, vol. 4, Special Issue, Fall 2002, 2002, 4, Special Issue.
- , "Three Tests for Measuring Unjustified Disparate Impacts in Organ Transplantation: the Problem of "Included Variable" Bias," *Perspectives in Biology and Medicine*, 2005, 48, S68–S87.
- , "Testing for Discrimination and the Problem of "Included Variable Bias"," mimeo, Yale Law School, 2010.
- **and Joel Waldfogel**, "A Market Test for Race Discrimination in Bail Setting," *Stanford Law Review*, 1994, 46.
- Balsa, Ana I., Thomas G. McGuire, and Lisa S. Meridith**, "Testing for Statistical Discrimination in Health Care," *Health Service Research*, 2005, 40, 227–252.
- Barnato, Amber E, F. Lee Lucas, Douglas Staiger, David E. Wennberg, and Amitabh Chandra**, "Hospital-Level Racial Disparities In Acute Myocardial Infarction Treatment And Outcomes," *Medical Care*, 2005, 43 (4), 308–319.
- Becker, Gary S.**, *The Economics of Discrimination*, University of Chicago Press: Chicago, 1957.
- , "The Evidence against Blacks Does't Prove Bias," *Business Week*, 1993.
- , "Nobel Lecture: The Economic Way of Looking at Behavior," *Journal of Political Economy*, 1993, 101 (3), 385–409.

- Bertrand, Marianne and Sendhil Mullainathan**, "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, 2004, 94 (4), 991–1013.
- Bushway, Shawn D. and Jonah B. Gelbach**, "Testing for Racial Discrimination in Bail Setting Using Nonparametric Estimation of a Parametric Model," mimeo, Yale Law School, 2010.
- Center for Disease Control and Prevention**, "CDC Health Disparities and Inequalities Report - United States, 2011," 2011, <http://www.cdc.gov/mmwr/pdf/other/su6001.pdf>.
- Chandra, Amitabh and Douglas O. Staiger**, "Identifying Provider Prejudice in Health Care," Working Paper, Harvard University and Dartmouth College, 2008.
- Coate, Stephen and Glenn C. Loury**, "Will Affirmative-Action Policies Eliminate Negative Stereotypes?," *American Economic Review*, 1993, 83, 1220–1240.
- Eisenberg, J.M.**, *Doctor's Decisions and the Cost of Medical Care*, Health Administration Press, Ann Arbor, Michigan, 1986.
- Goldin, Claudia and Cecilia Rouse**, "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians," *American Economic Review*, 2000, 90 (4), 715–741.
- Gordon, James A., Lawrence C An, Rodney A Hayward, and Brent C Williams**, "Initial Emergency Department Diagnosis and Return Visits: Risk Versus Perception," *Annals of Emergency Medicine*, 1998, 32 (5), 569–573.
- Grogger, Jeffrey and Greg Ridgeway**, "Testing for Racial Profiling in Traffic Stops From Behind a Veil of Darkness," *Journal of the American Statistical Association*, 2006, 101, 878–887.
- Institute of Medicine**, *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*, The National Academies Press. Washington, DC., 2002.
- Knowles, John, Nicola Persico, and Petra Todd**, "Racial Bias in Motor Vehicle Searches: Theory and Evidence," *Journal of Political Economy*, 2001, 109, 203–228.
- Mechoulan, Stephane and Nicolas Sahuguet**, "Assessing Racial Discrimination in Parole Release," mimeo, HEC Montreal, 2011.
- O'Dwyer, F and GG Bodiwala**, "Unscheduled Return Visits by Patients to the Accident and Emergency Department," *Archives of Emergency Medicine*, 1991, 8, 196–200.
- Persico, Nicola**, "Racial Profiling? Detecting Bias Using Statistical Evidence," *Annual Review of Economics*, 2010.

- Phelps, Charles**, "Information Diffusion and Best Practice Adoption," in J. Newhouse and A. Culyer eds., *Handbook of Health Economics*, Vol. 1, North Holland, Amsterdam, 2000.
- Pierce, John M, Arthur L Kellerman, and Catherine Oster**, "Bounces: An Analysis Of Short-term Return Visits To A Public Hospital Emergency Department," *Annals of Emergency Medicine*, 1990, 19 (7), 752–757.
- Polsky, Daniel, Ashish K. Jha, Judith Lave, Mark V. Pauly, Liyi Cen, Heather Klusaritz, Zhen Chen, and Kevin G. Volpp**, "Short- and Long-Term Mortality after an Acute Illness for Elderly Whites and Blacks," *Health Service Research*, 2008, 43, 1388–1402.
- Price, Joseph and Justin Wolfers**, "Racial Discrimination Among NBA Referees," *NBER Working Paper* 13206, 2007.
- Ross, Stephen L. and John Yinger**, *The Color of Credit: Mortgage Discrimination, Research Methodology, and Fair-Lending Enforcement*, MIT Press, 2002.
- Schulman, Kevin A., Jesse A. Berlin, William Harless, Jon F. Kerner and Shyrl Sistrunk, Bernard J. Gersh, Ross Dub, Christopher K. Taleghani, Jennifer E. Burke, Sankey Williams, John M. Eisenberg, William Ayers, and Jos J. Escarce**, "The Effect of Race and Sex on Physicians' Recommendations for Cardiac Catheterization," *New England Journal of Medicine*, 1999, 340, 618–626.
- Simeonova, Emilia**, "Doctors, Patients, and the Racial Mortality Gap: What Are the Causes?," Working Paper, Columbia University, 2007.
- Strumpf, Erin**, "Racial/Ethnic Disparities in Outpatient Primary Care: The Role of Physician-Patient Concordance," Working Paper, McGill University, 2010.
- Weinstock, Michael B. and Ryan Longstreth**, *Bouncebacks! Emergency Department Cases: ED Returns*, Anadem Publishing: Columbus, OH, 2007.
- Wilkins, P.S. and M.W. Beckett**, "Audit of Unexpected Return Visits to an Accident and Emergency Department," *Archives of Emergency Medicine*, 1992, 9, 352–256.
- Williams, Richard Allen**, *Eliminating Healthcare Disparities in America: Beyond the IOM Report*, Humana Press, New Jersey, 2007.

Appendix: Accounting for the Sample Selection

Table A1 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 2,413,874 discharges. For California, we did not impose any initial sample restrictions; Panel B shows that imposing the necessary sample restrictions leaves us with 2,611,228 observations.³²

Action	Observations Lost	Observations 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
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

Table A1: From the Raw Data to the Analysis Sample: Details of the Sample Selections in New Jersey and California.

³²As we explained in Section 4.1, the analysis sample for California actually has five more observations than is listed in Table A1 (2,611,233 instead of 2,611,228).