Testing for Statistical Discrimination in Health Care[#]

Ana I. Balsa
Center for Multicultural Mental Health Research
Cambridge Health Alliance

Thomas G. McGuire
Department of Health Care Policy
Harvard Medical School

Lisa S. Meredith
RAND Health

February 24, 2004

[#] This research was supported by grant PO1 MH59876 from the National Institute of Mental Health. We are grateful to John Ayanian, Carlos Blanco, Randy Ellis, Kevin Lang, Sharon Lise-Normand, Marc Rysman, Jon Skinner, Milt Weinstein, Alan Zaslavsky, HSR editors and refereesfor helpful comments. This paper was presented at the International Conference on Health Policy Research (American Statistical Association), the Boston University-Harvard-MIT Health Economics Seminar, the Latino Research Program Project Working Conference and the International Health Economics Association Conference. We thank seminar and conference participants for their useful comments. E-mails: balsaa@bellsouth.net; mcguire@hcp.med.harvard.edu; seidel@rand.org.

Testing for Statistical Discrimination in Health Care

Objective. To examine the extent to which doctors' rational reactions to clinical uncertainty ("statistical discrimination") can explain racial differences in the diagnosis of depression, hypertension and diabetes.

Data Sources. Main data are from the Medical Outcomes Study (MOS), a 1986 study conducted by RAND Corporation in three US cities. The study compares the processes and outcomes of care for patients in different health care systems. Complementary data from NHANESIII and NCS are also used.

Study design. Within each system of care (staff HMO, multi-specialty groups and solo practices), the MOS selected 523 health care clinicians. A representative cross-section (21,480) of patients was then chosen from a pool of adults who visited any of these providers during a 9-day period.

Data Collection. We analyze a sub-sample of the MOS data consisting of patients of white family physicians or internists (11,664 patients). We obtain variables reflecting patients' health conditions and severity, demographics, SES and insurance from the patients' screener interview (administered by MOS staff prior to the patient's encounter with the clinician). We use the reports made by the clinician after the visit to construct indicators of doctors' diagnoses. We obtain prevalence rates from NHANESIII and NCS.

Findings. We find evidence consistent with statistical discrimination for doctors' diagnoses of hypertension, diabetes and depression for Whites and Minorities. In particular, we find that different priors held by the physician about the prevalence of the disease across racial groups account for racial differences in the diagnosis of hypertension and diabetes. In the case of depression, we find evidence that race affects decisions through differences in communication patterns between doctors and white and minority patients.

Conclusions. To contend effectively with inequities in health care it is necessary to understand the mechanisms behind the problem. Discrimination stemming from prejudice is of a very different character than discrimination stemming from the application of rules of conditional probability as a response to clinical uncertainty. If miscommunication is the culprit, then efforts should be aimed at reducing disparities in the ways doctors understand and communicate with patients.

Key Words. Health care disparities, Race/ethnicity, statistical discrimination, clinical decision-making, clinical uncertainty.

INTRODUCTION

In the US context, a "disparity" refers to the unfair treatment of patients on the basis of race or ethnicity. In its recent report, "Unequal Treatment," the Institute of Medicine (2002) defines a racial disparity as a difference in treatment provided to members of different racial (or ethnic) groups not justified by the underlying health conditions or preferences about treatment of the patient. Disparities by this definition have many sources. Disparities can stem from an array of social factors, such as the patients' socioeconomic status, insurance coverage or geography, with the key one probably being that Minorities are much more likely than Whites to be uninsured or to be in plans with restrictive payment policies (Monheit and Vistnes, 2000; Phillips, Mayer and Aday, 2000). Social factors do not, however, fully account for all of the unjustified differences between Whites and Minorities. Disparities also emerge from the face-to-face decisions of doctors when insurance and other social factors can be ruled out: such differences have been referred to as discrimination.

An increasing body of literature has documented and condemned disparities originating at the clinical encounter (Bach et al., 1999; Shapiro et al, 1999; Schulman et al, 1999; see Geiger, 2001, or Mayberry, Mili and Ofili, 2000 for a review). Also, conceptual research has identified potential sources of health disparities within the clinical encounter (Einbinder and Schulman, 2000; Bloche, 2001; Balsa and McGuire, 2001, 2003; van Ryn, 2001). However, we are aware of no paper trying to measure empirically the magnitude and importance of these different mechanisms. In this paper, we intend to understand and shed light on some of the processes that may be activating the observed discriminatory patterns in the provision of health care to patients of different racial and ethnic groups.

The Institute of Medicine (2002) identifies three mechanisms that may be behind discrimination at the medical encounter: simple prejudice against members of a minority group, stereotypes a doctor holds about the health-related behavior of Minorities (such as, "Blacks don't comply with treatment

recommendations"), or the rational application of probabilistic decision rules when uncertainty surrounds the doctor's estimate of a patient's health status.[1] In the same way the term is applied in labor economics and related fields, we refer to this latter form of discrimination as "statistical." The basic idea of statistical discrimination in the health context is that uncertainty about the patient's severity of illness can induce the doctor to behave differently with otherwise identical members of different race/ethnic groups. If the underlying prevalence of the illness is associated with race, the doctor might take race into account in deciding about the diagnosis and treatment of a particular patient. Or, race/ethnicity might be associated with poor doctor-patient communication (Balsa and McGuire, 2001), interfering with the doctor's ability to discern and respond appropriately to the patient's health status (see Cooper and Roter, 2001, for a recent review of the literature on communication and disparities). Our objective in this paper is to test whether statistical discrimination can explain a race effect in clinical decisions and the extent to which this mechanism can account for the observed racial differentials in health care data. Our tests of statistical discrimination revolve around whether a race effect can be interpreted as working through the route of information in its influence on the physician's decision to diagnose a certain condition.

Methodologically, our paper integrates the normative literature on clinical decision making (Weinstein, Fineberg et al., 1980) with economic literature on statistical discrimination. Decision-theoretic approaches to diagnosis are common in the medical literature (Mushlin et al (1990) or Fendrick et al (1995)), although the connection to discrimination and disparities appears not to have been made previously. In economics, attempts to distinguish statistical discrimination from prejudice (or "taste discrimination" as it is sometimes called) have been made in the area of wage differentials (Altonji and Pierret, 2001) and differences in vehicle detention rates (Knowles, Persico and Todd, 2001). Our paper is in the same spirit as these investigations.

We use data from the Medical Outcomes Study (MOS), a study conducted by RAND Corporation

[1] Disparities from within the clinical encounter could stem from discrimination by doctors (the subject of this paper) or from higher patient refusal rates among members of ethnic/racial minorities. There is some literature on patient refusal, although Geiger (2001), and IOM (2002) discount this as a major factor.

that focused on the processes of care in four diseases: depression, hypertension, diabetes and heart disease. We chose the MOS to conduct our tests because it has information on both what the patient reports about his or her own health, and independently on what the doctor reports about the patient's health. Another strength for this study is that the MOS contains information about diseases for which the physician must rely more or less on patient reports. Depression is assessed through an interview whereas hypertension, on the other extreme, is assessed primarily by a pressure cuff. We expect communication to be more of a problem for Minorities for depression than for other conditions, and the MOS allows us to test this comparison.

Disparities and discrimination are complex, with multiple sources and operating through multiple mechanisms. In this paper we study the diagnostic phase of a clinical encounter, chosen by us to be the domain of decision-making where information-related forces on the clinician are most likely to play out. Most of the literature on disparities is concerned, by contrast, with decisions about treatment. The role of statistical discrimination in relation to other possible causes of discrimination may well differ in other parts of the decision making sequence.

STATISTICAL DISCRIMINATION AS A MECHANISM BEHIND DISPARITIES

Uncertainty, Bayes' Rule and differential predictions by race

The basic idea behind statistical discrimination is that doctors, unencumbered by prejudice or stereotypic beliefs, and in the presence of uncertainty about patients' underlying condition, may use race in making a diagnosis of a patient. As doctors are instructed, "The starting point of any diagnostic process is the patient presenting with a constellation of symptoms and signs." (*User's Guide to the Medical Literature* (UG) web site, American Medical Association). The immediate symptoms must be put in context of the clinical information the doctor brings to the encounter. A doctor hears a symptom report from the patient and weighs this along with her prior belief about the likelihood the patient has the condition given what else the doctor knows. This process is formalized by Bayes' rule, whereby the doctor's posterior assessment of the probability that the patient has the disease, given an observed symptom equals:

$$\Pr(disease \mid symptom) = \frac{\Pr(symptom \mid disease) * \Pr(disease)}{\Pr(symptom)} \quad (1)$$

After observing a symptom a doctor can be regarded as revising a prior estimate and coming to a decision about diagnosis and treatment, even if she does not engage in an explicit conditional probabilistic analysis. The doctor may have a threshold probability above which he/she decides that diagnosis is not warranted.

Within this model, race can matter in two ways. One form of statistical discrimination results when physicians use race as an indicator of the underlying likelihood a patient has a condition. For instance, the doctor may believe that the prevalence of the disease (*Pr(disease)* in the formula above) differs by racial/ethnic group. A second form arises when a physician reads diagnostic signals with more noise from members of minority groups (different *Pr(symptom|disease)* for patients of different ethnic/racial groups in (1)). If doctors fail to understand certain patients, because of differing language, culture or communication patterns, they may be more likely to fail to detect symptoms when they are present or vice-versa. We refer to these forms of statistical discrimination as a *prevalence* hypothesis and a *miscommunication* hypothesis about the role of race.[2] The rest of this section illustrates the concept of statistical discrimination with an example.

An example of a Symptom-Disease Relationship that differs by racial groups

Table 1 shows the relationship between a symptom report and the presence of major depression for African American and White respondents according to data from NHANESIII, a national survey designed by the National Center for Health Statistics to estimate the prevalence of selected diseases and risk factors.[3] We use this information to illustrate how clinical uncertainty may lead to clinical discrimination. The symptom question is "Have you ever had two weeks or more during which you felt sad, blue, depressed, or

[2] Both hypotheses draw heavily on the labor economics literature on statistical discrimination. A first strand asserts that differences in wages across ethnic groups are related to the existence of group differences in the quality of employers' information (Aigner and Cain, 1977, Lundberg and Startz, 1983). The other strand attributes differences in outcomes to decision-makers' use of priors (about any characteristic of the individual) that differ across racial groups (Arrow, 1973; Coate and Loury, 1993). Both lines of work share the underlying concept that uncertainty can generate disparities.

[3] See the data section for a description of NHANESIII. Within the medical examination part, a Diagnostic Interview Schedule (DIS) was administered to individuals aged 15-39 to detect the presence of mental illnesses.

when you lost all interest and pleasure in things that you usually cared about or enjoyed?"[4] The 2x2 matrix for each group shows the percent of respondents that fell into the four possible combinations of reporting the symptom yes/no and having a DSMIII diagnosis of depression or not. The matrices are obviously different, and some of these differences are summarized in the lower part of Table 1. The prevalence of major depression is greater for Whites, 10.4% compared to 6.2% for African Americans. The properties of the symptom-test differ for both groups as well. The test is more sensitive for African Americans, meaning the probability of reporting the symptom given the disease is present is greater for African Americans (6.0/6.2 = 96.6% versus 9.7/10.4= 93.2% for Whites), whereas the test is more specific for Whites, meaning the probability of not reporting the symptom given the disease is absent is higher for Whites (59.8/89.6 = 66.8% versus 57.6/93.8 = 61.4%).

Table 1: Symptom - Disease Relationship in Depression (NHANESIII)
Symptom: Felt sad, blue or disinterested for 2 weeks or more in the past
Disease: Current Major Depression

	African Americans			*Whites*			*Comparison* (t values)	
In %	**Disease**			**Disease**			**Disease**	
	Present	Absent	Total	Present	Absent	Total	Present	Absent
Symptom Report								
Positive	5.96	36.19	42.15	9.73	29.74	39.47	-5.11	4.94
Negative	0.21	57.64	57.85	0.71	59.82	60.53	-2.70	-1.59
Total	6.17	93.83	100.00	10.44	89.56	100.00		
Prevalence			6.17			10.44		-5.65
Sensitivity			96.60			93.20		1.50
Specificity			61.43			66.79		-3.85
Positive predictive value			14.14			24.65		-6.24

From the standpoint of the doctor, when hearing a patient endorse the symptom, "yes there has been a two week period….," the positive predictive value (given by equation (1) when symptom is positive) is the key number. When the doctor hears the report from an African American, the right inference is to say that there is a 14.1% chance (5.96/42.15) this patient has major depression, whereas for a white patient who says

[4] This symptom is one of many symptoms used to construct the DIS diagnoses.

the same thing, the right inference is that there is a 24.6% chance (9.73/39.47), nearly double the African American number.

Decisions about collecting more information or about treatment depend on the likelihood a doctor thinks the patient has a condition. The doctor could well recommend treatment for a patient for whom the chances of having the condition are one in four, but might not make the recommendation for a patient with a 14% chance, roughly the magnitude of the racial differences here on the basis of the answer to one symptom question. This would be discrimination. Assuming all else about these patients is the same, and even with exactly the same test result, the doctor recommends treatment for the white and no treatment for the African American patient.

What explains the better predictive power of these symptoms for Whites than for African Americans? There are two factors at work in the example in Table 1: the higher prevalence rates for Whites, and the more reliable signal contained in the symptom. Whites start off being more likely than African Americans to be depressed in a ratio of 10.4/6.2 even prior to the doctor hearing the symptom. Both percentages go up after a positive report, to 24.6/14.1. Whites go up a bit more because more information is conveyed for this group. We can illustrate these two effects by simple simulations. If African Americans are given the same sensitivity and specificity as Whites, but keeping their underlying prevalence lower, their positive predictive value would go to 15.6%, erasing about one-tenth of the difference between them and Whites. If they were to keep their own sensitivity/specificity but have the higher prevalence of Whites, the positive predictive value would move to 22.6%, erasing about 80% of the difference. Both factors contribute substantially to the worse predictive value of the test for African Americans.

METHODS

Our tests of statistical discrimination are based on the comparison of a *traditional disparities regression*, a regression in which race effects are only accounted for through the coefficient of a race dummy, with a *statistical discrimination regression*, in which new variables that account for information-related processes are introduced. A traditional disparities regression would consist, for instance, of a probit

or logistic regression explaining the likelihood that a doctor diagnoses certain condition, given indicators of the patient's disease (S_i), patient's race (R_i), patient's characteristics (X_i) and doctor's characteristics (Y_j). Formally, the coefficient of interest would be that of R_i in the following expression:

$$\Pr(diagnosis) = \Phi\left(S_i, R_i, X_i, Y_j\right) \quad (2)$$

While the usual claim is that a negative and significant sign of the race coefficient is evidence for discrimination, the traditional disparities regression offers no information for why diagnosis is lower for some racial groups. A negative sign on the race variable admits many possible interpretations. Doctors might be prejudiced against minority groups, might hold stereotypes about minority patients or might have different information about members of different racial groups.

A statistical discrimination regression accounts in addition for the doctor's process of inference when information about the patient's health status is imperfect. Assume the doctor's posterior assessment of the probability that the patient has the disease (the equivalent to equation (1)), is given by expression $p(\theta_i, S_i)$, where θ_i is the prevalence of the disease among individuals similar to patient *i*, and S_i is a symptom experienced by patient *i* (which could be observed with more or less accuracy by the doctor). The statistical discrimination expression would be of the form:

$$\Pr(diagnosis) = \Phi\left(p(\theta_i, S_i), R_i, X_i, Y_j\right) \quad (3)$$

Race effects in the diagnosis decision could work through the race dummy R_i, but also through different prevalence rates (θ_i's) across races (the *prevalence* hypothesis) or through different weights given by the provider to the patient's signal, S_i, when doctor-patient communication patterns differ systematically across patients of diverse racial/ethnic groups (the *miscommunication* hypothesis).

To make (3) operational we construct and include as a regressor a prior $\hat{\theta}_i$ that is common for all doctors across race, age and gender categories. While the assumption that physicians share a single prior distribution on the prevalence of each disease is strong, it helps us to test a relevant empirical question: the degree to which "rational" priors could explain a clinician's decision-making process. It is true that factors

like the physician's experience (such as age or volume of practice) and geographic location are likely to influence physicians' beliefs. Our regressions may control for some of the systematic variation in the priors by considering physician characteristics or physician fixed effects. In addition to the prior, we include as regressors interactions of race and signal and interactions of prior and signal. The basic estimating equation is of the form[5]:

$$\Pr(diagnosis) = \Phi\left(\lambda_1 S_i + \lambda_2 S_i\hat{\theta}_i + \lambda_3\hat{\theta}_i + \lambda_4 R_i + \lambda_5 R_i S_i + X_i\beta + Y_j\eta\right) \quad (4)$$

with $\lambda_1 ... \lambda_5, \beta, \eta$ parameters.

To test for the *prevalence effect* of statistical discrimination we evaluate whether the prior captures some of the effects of race, age and gender in our equations for diagnosis of the two illnesses. We test the model in (4) versus the traditional disparities regression in (2) - a competing model that assumes no significant effect of the prior. The restrictions we test are whether λ_2 and λ_3 are jointly equal to zero. If we find (4) to be a superior model, we then test the extent to which the prior terms account for the effects of race, gender and sex as found in a traditional disparities regression. A finding that the terms in the prior are significant and positive and that the main effects of race, gender and sex diminish once we account for the prior is consistent with the hypothesis of statistical discrimination in its *prevalence* form. We also use a Bayesian Information Criterion to compare the fit of the data to the alternative models. [6]

The *miscommunication* hypothesis implies that the way a doctor perceives a signal of disease differs across racial/ethnic groups. Poor communication between doctor and patient is like noise between the patient's report and the signal actually observed by the doctor. If a race effect is due to miscommunication between the white doctor and minority patients, then the doctor's diagnostic decisions should be less related

[5] The estimating equations can be derived from a structural model of decision-making under imperfect information. Such model can be made available by authors upon request.

[6] The limitation of usual log-likelihood based measures is that the simpler model (null) has to be a subset of the more complicated model (alternative) – the models have to be nested. Since our models are not nested, we use an alternative model selection criterion, the Bayesian Information Criterion (BIC)= -2log($\hat{L}$) +log (N) K+…
where $\hat{L}$ is the maximum likelihood, K the number of parameters to be estimated in the model, and N the sample size. High-order terms are ignored here. A model is better than another model if it has a smaller BIC value.

to the patient's signal in the case of non-white patients. To account for that, we interact the patient's signal of the disease S_i with a race (minority) dummy, R_i. We expect miscommunication to be reflected in a lower weight on the signal for Minorities ($\lambda_5 < 0$).

For each of the chosen conditions in the study, we run a pair of probit regressions based on the structure in (4). The first specification uses physician characteristics as regressors in a regular probit model. The second one uses a fixed effect for each physician. Clustering and weighting of the data are discussed in the results section, once the data have been described.

DATA

Data Sources: The MOS, NHANESIII and NCS

The main data we use are from the screener sample (cross-section) of patients in the Medical Outcome Study (MOS). The MOS was an observational study of adult patients who received medical care under three different systems of health care (staff model health maintenance organizations (HMOs), large multi-specialty groups, and solo practices) at three US sites (Los Angeles, Boston and Chicago). The MOS was designed to compare the processes and outcomes of care within these systems for patients with hypertension, diabetes, depression and heart disease. Within each system of care, a representative sample of physicians (general internists, family physicians, cardiologists, endocrinologists, psychiatrists), psychologists and other mental health providers was selected (523 in total). A representative cross-section (21,481) of English-speaking patients was chosen from a pool of adults (ages 18 and older) who visited one of the MOS participating clinicians during a 9-day period in the autumn of 1986. (See Rogers et al. (1992) for more details about the study design and sample.)

The MOS has several advantages for us in comparison to other health care data sets. Most importantly, data were collected by researchers on the clinical characteristics of patients on the same day that these patients were seen and evaluated by a physician. Patients completed pre-visit screening questionnaires and this information can be used to assess the conditions described above. Having measures

of the patient's health status and morbidities allows us to control for these factors in a study of what doctors report. Furthermore, the MOS contains extensive information about the physician, not just the patient, and includes an average of 40 observations for each physician in the study, allowing us to use fixed effect estimation methods. Finally, the MOS contains information about diseases for which the physician has different tools for assessment and for which the importance of clinical uncertainty is likely to differ.

In this paper we use only a sub-sample of the MOS data consisting of patients of white family physicians or internists. We consider in our analysis two groups of patients: Whites and Minorities (including Blacks, Hispanics, Asians, Native American and other Minorities). Although black patients were adequately represented in the MOS, the number of patients in other minority groups was too small to study separately. Also, there were too few black physicians or physicians from other racial/ethnic groups to allow us to reliably study other racial/ethnic combinations.[7] Our analysis compares three of the conditions surveyed in the MOS: depression, hypertension, diabetes. We have our reservations about the diabetes analysis because the designation covers two conditions that are difficult to distinguish: juvenile and adult-onset diabetes. This explains the lower prominence we give to the diabetes results in the discussion. We do not consider an analysis of heart disease because few patients had this condition. The final sample we use consists of 11,664 patients (9,441 white and 2,223 minority patients). The screener questions about hypertension were distributed only to around half of the sample, explaining the different numbers of observations used to study hypertension versus depression and diabetes.

We use complementary data from the National Health and Examination Survey III (NHANESIII) and from the National Comorbidity Survey (NCS) to compute external prevalence rates for hypertension, diabetes, and depression respectively. The NHANESIII is a national survey designed by the National Center for Health Statistics to estimate the national prevalence of selected diseases and risk factors. The survey was implemented in two phases (1988-1991 and 1991-1994) in households selected from 81 counties (approximately 33,000 individuals) and consisted of both an interview and a medical examination. We use

[7] In the results section we consider black physicians and distinguish black from other minority patients.

the first phase of the NHANESIII to perform our computations because it is most proximate to the date of the MOS data. The NCS, fielded from the Fall of 1990 to the Spring of 1992, constitutes the first nationally representative mental health survey in the U.S. that uses a fully structured research diagnostic interview to assess the prevalence and correlates of psychiatric disorders. It is based on a stratified, multi-stage area probability sample of persons aged 15 to 54 years in the non-institutionalized civilian population in the 48 coterminous states. A total of 8098 respondents participated in the baseline survey.

Variables

We describe below the variables used in the analysis. Most of the variables are constructed on the basis of the MOS data, so we make no explicit reference to the data source when that is the case. We explicitly indicate when NHANESIII or NCS are the sources of variables.

Signals of severity. We assess whether a patient emits a positive signal for a certain condition by analyzing patients' responses to a set of questions asked in the screener form distributed to them before the visit. We assume a patient signals positive for hypertension when the patient answers "yes" to the question "Have you ever been told you have hypertension?" For depression, we employ a screener based on the patient's report about his/her mental health history and recent symptoms. [8] (For a description of this screener, see Wells et al. (1996)). Patients with a screener value greater than a certain threshold are considered to have a high probability of current major depression. We considered a patient to be diabetic when he/she acknowledged having been told that he/she had diabetes or if he/she reported taking insulin injections. For each condition, we construct a variable we denote as the "signal", which takes the value of 1 if the patient has rated positive and zero otherwise. Notice we treat the patient's response to screener questions as a signal. Whether the doctor ends up receiving this information or not will depend on both the patient's and the doctor's willingness and ability to communicate. Also, we do not assume the patient's response represents "truth".

[8] The questions asked were: Have you had 2 weeks or more in the past year during which you felt sad, blue or depressed? Have you felt depressed or sad much of the time in the past year? During the past week, how often have you felt depressed/ had crying spells / felt sad / enjoyed life / slept restlessly / felt that people disliked you?

Doctors' diagnoses. We assume there is a positive identification of hypertension or diabetes if the doctor either indicates the condition as the main reason for the visit or answers positively to the question: "Does this patient have hypertension/diabetes?" We assume the doctor diagnoses depression whenever she either indicates depression as the main problem addressed in the visit, counsels the patient for depression, refers the patient to a mental health specialist or acknowledges that the patient was depressed for two weeks or more in the past year. This is based on a definition of recognition of depression used for the MOS in Borowsky et al (2000).

Patients' demographics and general clinical information. We control in our analysis for age, gender, race, marital status, employment status, education (highest grade complete), income, location and patient's insurance status (whether insured or not and whether the plan is an HMO or fee for service). In addition to the disease-specific information, we also control for some other clinical information reported by the patient before the visit: the patient's self report of health status, the patient's body mass index and if the patient is a smoker.

Clinicians' information. We first account for the clinician's age, his/her practice type (staff, solo or group) and specialty (family practice or internist). We also take advantage of the multiple observations available for each doctor and check for robustness by using doctor fixed effects.

Doctors' priors. The "prior" in equation (3) is θ_i, the doctor's assessment of the probability that patient *i* has the illness before the doctor reads any patient signal. As mentioned before, we assume all physicians share the same "prior" estimate of a patient in an age-gender-race cell. 16 different priors are computed (4 age categories, 2 gender categories and 2 race categories).

We first construct the priors using epidemiological (external) information. For hypertension, we use both the Blood Pressure Section of the Physician's exam and the Blood Pressure section of the Interview Questionnaire in NHANESIII, phase 1 (1988-1991) and define a person as hypertensive when systolic blood pressure average (average of 3 readings) is above 140 and / or diastolic average is above 90 and / or the person I s taking medication for hypertension. After excluding pregnant women, we compute the prevalence

rates as weighted averages of the hypertension indicator by race, age and sex categories (the weights are nationally representative). In the case of diabetes, we construct the priors on the basis of three questions of the Interview Questionnaire in NHANESIII, phase 1 (1988-1991): "Were you ever told you had diabetes" "Are you now taking diabetes pills?", "Are you now taking insulin"? For depression, the rates are defined as the number of people with major depression disorder within the total sample. The diagnostic interview used to generate the diagnosis of major depression in the NCS was a modified version of the Composite International Diagnostic Interview. Since the NCS only surveyed people aged below 59, when using this data we exclude from the analysis patients aged 60 or more.

An alternative way to construct priors is by using the information reported by the patient in the screener questionnaire of the MOS. [9] Within this second approach, we define priors as averages, for each condition, of the patients' signals within cells defined by age, gender and race categories. Formally, $\hat{\theta}_i^k = \frac{\sum_i S_i^k}{n_k}$, where S_i^k is the signal of individual i in cell k, n_k is the number of patients in cell k, and cells are defined by age, race and sex categories. In computing these priors, we use not only observations from patients visiting white general physicians or internists (the core of our analysis), but also observations from patients visiting minority GP's or internists (13,000 observations for depression and 6,764 for hypertension). The average number of observations per cell exceeds 400 for hypertension, and the smallest cell contains 93 observations, allaying concerns about small number problems.

There are advantages and disadvantages of using either priors from epidemiological data or priors based on the MOS screener. On the one hand, epidemiological data are completely independent of doctor's diagnoses and of patient's perceptions. Unlike the MOS priors, they are free of potential biases generated by doctors' attitudes or patients' knowledge of their diseases. Moreover, epidemiological data are likely to be at

[9] Another way would be using physicians' diagnosis. One could argue that a prior is a belief on the part of the physician, and the average rate of diagnosis in an age-gender-race cell might better capture the physician's prior belief about the prevalence of some disease for that set of people. Using diagnosis presents two major problems. First, we would be using as a regressor the mean of the variable we are seeking to explain, raising identification problems. Second, we want the prior to represent the underlying distribution of disease in the population without incorporating other beliefs (such as stereotypes or prejudice) that may be reflected in doctors' diagnoses rates.

least as strong an influence on the prior than any information obtained from the patient during the screening interview. On the other hand, the external prevalence rates we use in this paper represent the US population as a whole, and not the population that visits a primary physician. These prevalence rates wouldn't capture, for instance, doctors' awareness of sick Minorities being less likely to visit a physician than Whites. Because neither of these measures satisfies us completely, we perform our main analysis using priors from epidemiological data and then check the robustness of the results using MOS priors.

A description of the data by minority status

Table 2 summarizes the prevalence rates constructed on the basis of external data, the MOS indicators of disease as screened before the visit with the doctor (the signals), doctors' diagnoses for each of the three conditions in the study, relevant demographic and clinical characteristics of patients, and clinicians' characteristics. Depression prevalence rates, constructed on the basis of NCS, average 6.9% for Whites and 7.6% for Minorities (a significant difference). The NHANES data, on the other hand, indicate higher rates of diabetes in Minorities than in Whites (6.4% vs. 5.1%) but similar rates of hypertension (23%) for both groups.[10] All measures take as a basis an adult population (age 18 or more). The rates of disease derived from the MOS signals maintain in general the sign of the white-minority differences shown for the external prevalence rates. But because patients visiting a clinician are more likely to be sick than the average person in the population, rates of disease portrayed by the MOS screener are higher than those constructed on the basis of NHANESIII or NCS data. For hypertension and diabetes the differences in diagnostic rates across groups run in the same direction as those derived from the patients' signals. But for depression, the difference runs in the opposite way: doctors are less likely to diagnose depression in minority patients than in white patients. An interesting observation is that diagnosis rates are usually lower than the rates of disease derived from the MOS patient's screener.

Minorities in the MOS are younger, more likely to be female, less likely to be married, more likely to be unemployed and have a lower income than Whites. They are more likely to have no insurance, more

[10] Prevalence rates for hypertension are not significantly different across racial groups. Higher rates of hypertension at higher ages for Minorities are compensated by a younger age structure of this population.

likely to be in an HMO and more likely to come from LA or Chicago rather than Boston. Minorities are more likely to assess their health as poor and to be overweight. Also, they see staff physicians and group-practice physicians at higher rates than Whites.

RESULTS

Table 3 and Table 4 show the results for hypertension and depression respectively. The first column in each table depicts a traditional disparities regression in which the physician's diagnosis is estimated as a function of patients' socio-demographics (including race, age, gender) and other factors, physicians' characteristics and a control for the presence of the disease obtained from the patient's report in the screening survey before the visit. We evaluate this model against that of a more comprehensive one that allows for information-related effects, as specified in equation (4). Columns 2 to 6 show the results for the statistical discrimination model. [11]

Hypertension

In a traditional disparities regression (column 1 in Table 3), detection of hypertension is significantly affected by gender (p=0.003) and age (p=0.012; p=0.000 and p=0.000 for each age category): men and elderly patients are more likely to be diagnosed with hypertension. Minority status is positive, but non-significant (p=0.167). The second column in the table shows the results of the full statistical discrimination specification, given by equation (4). The terms in the prior turn out to be very significant and of the right sign (jointly, they are significant at 0.1%, with a chi square (2) of 13.32). We also notice that the main effects of the gender and age variables become insignificant once the prior is accounted for. Using the Bayesian Information Criterion, the comparison of these two models leads to a rejection of the traditional

[11] Observations were weighted to account for the study design effects of sampling probabilities for patients and clinicians. The MOS weights were derived from three components. The first one adjusted for the different quota for clinicians in each specialty in the solo and small group system of care (e.g., sampling probabilities varied across specialty type and site). In the large multi-specialty groups and HMO systems, every clinician was selected if they were eligible (e.g., weight=1). The second component adjusted for differential probability of being sampled within the screening period (the weight is inversely proportional to the length of the screening period). The third component adjusted for the differential probability of patients visiting the office within a two-week period (again the weight is inversely proportional to the visit probability). These three weights were multiplied together for a total weight. We recognized clustering of observations at the clinician level and computed robust standard errors.

disparities regression for hypertension. In other words, the statistical discrimination full specification fits the data better than the traditional disparities regression. To ensure that multicollinearity is not behind these results, we next check each variable singly. Columns (4), (5) and (6) contain the results of adding race, age and gender to the statistical discrimination expression in turn. When comparing these columns to column (1), we observe that the effects of gender and age virtually disappear or are reduced in magnitude. The race effect also goes to zero.[12] These results are evidence that age, gender and race effects can be understood as working through *prevalence* in the doctor's diagnosing of hypertension.

On the other hand, the coefficient on the race times signal interaction is negative but insignificant for hypertension. The null result on such interaction is not surprising given the lack of race effect to begin with in the traditional disparities regression. Miscommunication is unlikely to be an issue in the diagnosis of hypertension.

Depression

Similar to the results in Borowsky et al. (2000), also using MOS data to study only depression, the traditional disparities regression in column 1 of Table 4 shows that Minorities are less likely to have depression diagnosed than Whites. Women are more likely to be diagnosed with depression and, as in hypertension, middle-aged and older patients have a higher probability of being diagnosed than younger patients. While not shown in Table 4, depressed patients are more likely to be diagnosed when they have reported they are in bad health, but are less likely to be recognized in the first visit with the doctor. Also, being single or unemployed and seeing a general physician instead of an internist increases the chances of a depression diagnosis.

The statistical discrimination model contains mixed results. With respect to the prevalence hypothesis, the coefficients on the prior variables are jointly significant but of the wrong sign (column 2, Table 4). Using a Bayesian Information Criterion, the traditional disparities regression cannot be rejected against the full specification of the statistical discrimination model. The effects of age and gender remain

[12] When the model is run with doctor fixed effects, the traditional disparities regression is still rejected against the "Bayesian" model. Also, main effects show a similar pattern to that described in the regular probit.

significant even when the prior is in the model (p=0.042 and p=0.033 for two age categories and p=0.001 for gender), and the magnitudes of the coefficients on race, gender and age do not diminish.[13] The analysis suggests that either the doctor is not acting in a Bayesian fashion, or the priors she held are not captured in our estimate of θ_i.

We do find, however, evidence of miscommunication in the case of depression. The coefficient on the race times signal interaction is negative and marginally significant at p=0.084 in the full specification of the statistical discrimination regression. Because the *prevalence* hypothesis did not fit the data in the case of this condition, we run a new specification of the depression regression excluding the terms in the prior (Table 4, column 7). A significant coefficient of the minority-signal interaction in such specification (p= 0.049) suggests that doctors dealing with cases of depression rely less on the minority patients' reports.[14] Miscommunication between white doctors and minority patients appears to be one of the forces behind the disparities observed in the diagnosis of depression. The fact that we find miscommunication to be relevant for depression but not for hypertension, is reasonable given that the detection of depression relies more on communication than the detection of hypertension.

Communication between doctors and patients requires that a signal be emitted, be heard and be internalized. What we call *miscommunication* in this paper conflates these. Another way to say what we have found is that an available signal is not responded to by a physician. We cannot tell with the data available whether the patient did express the signal at all or whether the doctor understood it but ignored. But we can address a related concern: if the doctor's misunderstanding stemmed from real communication problems with the patient or from less effort or time put in the encounter. It is possible that prejudice shows itself in a lower level of effort or time on behalf of certain patients, leading to miscommunication. As a

[13] The prior for depression does not vary much across age-gender-race cells. Econometrically, if the variance in the prior is too small, it would be tantamount to having another constant term in the regression. This could result in imprecise estimates of the prior and its interactions. To dismiss this problem, we run new regressions assuming no constant. Results are still imprecise and far from what theory would predict.

[14] We also estimated the miscommunication model for different ranges of the prior, seeking to find a stronger effect of miscommunication as the prior approached 0.5. We find, in effect, more evidence of miscommunication when the prior is higher than when it is nearer to 0. We also run a regression with a continuous signal of depression. The interaction between this continuous variable and race remains negative and significant.

response to this issue, we introduce into the analysis a variable that may be indicative of the physician's effort: the amount of time that the physician spent with the patient (as reported by the patient). We run a regression of "duration of visit" as the dependent variable on all patient and physician characteristics, including race and controls for different comorbidities. The coefficient on race is significant and positive (p=0.032) when controlling for the main comorbidities in the study, diabetes, depression, hypertension and myocardial infarction, and remains positive, though insignificant (p=0.162) when adding other comorbidities, as cancer, arthritis, paralysis, amputations and others. This implies that doctors are not spending less time with minority patients than with white ones. We take this as evidence that the observed miscommunication is unlikely to be a manifestation of prejudice, at least if prejudice is measured by a shorter duration of the medical visit.

Diabetes

We also tested the statistical discrimination hypotheses for diabetes. Results for this condition are very similar to those reported for hypertension. (Results not shown). There appears to be some prevalence effect in the case of age, but no evidence for a miscommunication effect. We believe that this analysis may be murky due to the underlying heterogeneity in health status (adult onset versus juvenile diabetes) that is captured poorly by the data available in the MOS.[15]

Robustness

We performed several robustness checks in the analysis of prevalence that support the previous results. First, we run the same equations described above with new priors constructed on the basis of the MOS signals. The conclusions obtained in the main framework remained valid, both for hypertension and for depression. Second, we tried different specifications for the priors (both for the MOS as for the NHANES and NCS priors), like considering different priors by location. While with the MOS priors the previous results still held, results were mixed with the external priors. Third, to account for the fact that we were running a second-stage regression with a prior estimated in a first stage, we bootstrapped the standard

[15] Results are available from authors.

errors of the joint estimation. The patterns described above were confirmed with the bootstrapped errors. Fourth, we rerun the prevalence equations with doctor fixed effects, confirming the results. Finally, we refined the estimation of the posteriors by adding a quadratic term in the prior, and prevalence continued to successfully account for gender, age and race effects in the case of hypertension, but not in the case of depression.

Black patients; non-white physicians

Results are unchanged when we include only black patients in the minority category. We also checked the data for evidence of "racial" determinants of physicians' diagnosis decisions in the case of non-white physicians, running separate models for this group of 40 physicians and 735 and 1399 patients for hypertension and depression respectively. There is no evidence of a race effect in either of the two traditional disparities regressions we run for the conditions in the study. This may be due to imprecise estimates from a smaller sample, a different pattern of behavior for non-white physicians, or systematically different patients in unobserved factors.[16] Unfortunately, interpreting this finding does not add much to the identification of miscommunication versus prejudice discrimination. While minority physicians may be less likely to feel prejudice against minority patients, they may also be more likely to communicate better with them.[17] For depression, age and gender appear to have a significant input on doctors' diagnoses decisions, while only age is significant in the case of hypertension. Once we account for the statistical discrimination structure, we find some evidence of *prevalence* for age and gender in hypertension. We find no evidence of heightened *miscommunication* between non-white doctors and their patients for depression. This result is not surprising since we found no significant race effects in the traditional disparities regression for minority physicians.

[16] We comment on the "selection" of patients to minority physicians in the Conclusions.

[17] We also run a regression that considers jointly all physicians (white and minority). The likelihood of diagnosing depression depends now on doctor's race, on interactions of patient's race and doctor's race, on interactions of patient's race, signal and doctor's race, and on the other characteristics used previously. None of the variables associated with the race of the physician is significant in explaining diagnosis decisions.

CONCLUSIONS

Race/ethnicity effects are being intensively investigated in health care data, where findings of lower rates of diagnosis and treatment for Minorities document a fundamental inequity in health care. Our paper proposes and tests a specific mechanism for why race matters. We analyze two implications of the statistical discrimination idea. The first one claims that racial effects on doctors' diagnostic decisions operate through the priors doctors hold about the patient's condition, even before observing any particular signal from the patient. The other variant states that "race effects" appear because of communication problems between white doctors and minority patients. We find strong evidence that race works through a "prior," Bayesian fashion, for hypertension and diabetes. Furthermore, in the case of depression, we find evidence that race affects decisions through communication. While we do not explore it in this manuscript, communication may play a role in diagnosis also through other variables, like age and gender.

Our findings of statistical discrimination should not be taken as implying that Bayesian behavior is responsible for most of health care disparities. We only study the diagnostic decision, the precursor to decisions about resource use. If disparities emerge in the diagnostic process, empirical research conditional upon diagnosis may understate the full magnitude of disparities. We chose to begin our analysis of the role of information in discrimination with the stage of treatment where information matters most. At the same time we recognize that disparities are mainly about treatment differences. A more complete analysis would follow patients as they and their clinicians make decisions about treatment. In these next stages of the process, stereotypes or prejudices may play more of a role.

There are some limitations to this study we have to consider. First, the model we estimate assumes that prejudice enters as a cost and affects the physician's utility in a linear way, and in particular does not interact with severity of the patient. If prejudice is different for patients of different underlying severities, it could offer another interpretation of our Race*Signal coefficient. Second, our measure of diagnosis may not always capture the fact that the doctor has recognized the condition in the patient. For instance, in the case of depression, physicians may have recognized a mental health problem as a secondary, rather than a primary reason for a visit. Third, we assume that these screeners are good approximations of the patients'

underlying health conditions and that they are equally valid for patients from all groups. This assumption may be incorrect for several reasons. The same questions may have different significance for members of different cultural groups and fail to capture the same objective measures. Additionally, if patients' knowledge of their condition depend on earlier contacts with providers, and prejudice affected this previous exposure, then the signal (constructed on the basis of patients' reports) would already reflect a bias. Furthermore, patients may not be aware of their own condition and the lack of correlation between patient's signals and doctor's diagnoses may signal "good diagnosticians" rather than problems of communication.[18] Finally, there might be some self-selection biasing the results. Since some patients can choose the doctor they want to be treated by, possibly on the basis of physician communication styles, the results we obtain are likely to be milder than they would be were patients randomly assigned to doctors. Also, our sample excludes non-English speaking patients. Even with such a bias working against us, we find evidence for miscommunication in the case of depression.

To contend effectively with inequities in health care and outcomes, it certainly is necessary to know the source of the problem. Discrimination stemming from prejudice is of a very different character than discrimination stemming from application of rules of conditional probability. When doctors indulge prejudices, they are not acting in the best interests of their patients; when they apply rules of conditional probability they are doing the best they can, given the information available.[19] Sorting out the role of these two classes of explanation is critical for empirical research and policy making. But statistical reasons for discrimination may also be unfair. A finding in the case of depression that miscommunication is the explanation for a race effect does not ameliorate the disparity in services or outcomes that results, in spite of not stemming from doctors' ill will. Minorities still fare worse than Whites in the clinical encounter, and attempts to bridge this gap through policies that improve doctor-patient communication should be encouraged. Better communication between doctors and patients will improve doctors understanding of

[18] If Minorities are less likely to have contact with the medical system, they are less likely to be aware of their health problems. In this case, the lower correlation between signal and diagnosis for this group of patients would indicate lack of trust in the medical system, but not necessarily miscommunication in the medical encounter.
[19] The information available may not always reflect the "truth". Stereotypes may lie behind priors held by doctors.

patients' condition, reduce reliance on population averages in clinical decisions, and thereby improve the match between treatment and the health needs of the patient.

References

Aigner, D., Cain, G., 1977. Statistical Theories of Discrimination in Labor Markets. Industrial and Labor Relations Review 30(2), 175-187.

Arrow, K., 1973. The Theory of Discrimination, in Ashenfelter O.A., Rees, A. (Eds), Discrimination in Labor Markets. Princeton, N.J: Princeton University Press, 3-33.

Altonji, J., Pierret, C., 2001. Employer Learning and Statistical Discrimination. Quarterly Journal of Economics 116(1): 313-350.

Ayanian, J., Cleary, P., Weissman, J., and Epstein, A., 1999. The Effect of Patient Preferences on Racial Differences in Access to Renal Transplantation. New England Journal of Medicine 341(22):1661-1669.

Bach, P., Cramer, L., Warren, J., Begg, C., 1999. Racial Differences in the Treatment of Early-Stage Lung Cancer. The New England Journal of Medicine 341(16): 1198-1205.

Balsa, A.I., McGuire, T.G., 2001. Statistical Discrimination in Health Care. Journal of Health Economics, 20: 881-907.

Balsa, A.I., McGuire, T.G., 2003. Prejudice, Clinical Uncertainty and Stereotyping as Sources of Health Disparities, Journal of Health Economics (22): 89-116.

Benson, V., Marano M.A., 1998. Current Estimates from the National Health Interview Survey, 1995, *Vital and Health Statistics*, 10(1999): 1-428.

Bloche, M.G., 2001. Race and Discretion in American Medicine, Yale Journal of Health Policy, Law and Ethics I: 95-131.

Borowsky, S., Rubenstein, L.V., Meredith, L.S., Camp P., Jackson-Triche, M., Wells K.B., 2000. Who is at risk for non-detection of mental health problems in primary care? Journal of Global Information Management 15: 381-388.

Coate, S., Loury, G., 1993. Will Affirmative-Action Policies Eliminate Negative Stereotypes? American Economic Review, 83(5), 1220-1240.

Cooper, L.A., Roter, D.L., 2001. Patient-Provider Communication: The Effect of Race and Ethnicity on Process and Outcomes of Health Care. Unpublished, Johns Hopkins University.

Einbinder, L.C., Schulman, K.A., 2000. The Effect of Race on the Referral Process for Invasive Cardiac Procedures. Medical Care Research and Review. 57(1), 162-180.

Fendrick AM, Chernew ME, Hirth RA, Bloom BS, 1995. Alternative management strategies for patients with suspected peptic ulcer disease. Annals of Internal Medicine, 123: 260-268

Geiger, H. J., 2001. Racial and Ethnic Disparities in Diagnosis and Treatment: A Review of the Evidence and a Consideration of Causes. Unpublished.

Harrison, M. and N.K. Thurston, 2002. Racial Matching Among African-American and Hispanic Physicians and Patients: Causes and Consequences. Journal of Human Resources, Spring.

Institute of Medicine, 2002. Title and exact publication data TBD, National Academy of Sciences, Washington, DC.

Knowles, J., Persico N., Todd, P., 2001. Racial Bias in Motor Vehicle Searches: Theory and Evidence, Journal of Political Economy 109(1): 203-229.

Lundberg, S., Startz, R., 1983. Private Discrimination and Social Intervention in Competitive Labor Markets. American Economic Review 73, 340-47.

Mayberry, R., Mili, F., Ofili, E., 2000. Racial and Ethnic Differences in Access to Medical Care. Medical Care Research and Review 57(1), 108-145.

Mushlin AI, Mooney C, Holloway RG, Detsky AS, Mattson DH, Phelps CE, 1997. The cost-effectiveness of magnetic resonance imaging in patients with equivocal neurological symptoms. International Journal of Technology Assessment in Health Care 1997; 13:21-34

Monheit, A.C., Vistnes, J., 2000. Health Insurance Coverage and Access to Care. Medical Care Research and Review 57(1): 11-35.

Phillips, K.A., Mayer M.L., Aday L.A., 2000. Barriers to Care Among Racial/Ethnic Groups Under Managed Care. Health Affairs 19(4), 65-75.

Rogers W., McGlynn E., Berry S., et al., 1992. Creation of Data bases on the Health of Individuals. In: Stewart A.L., Ware J.E. Jr., eds. Measuring Functioning and Well-Being: The Medical Outcomes Study Approach, Durham, NC: Duke University Press, 27-47.

Schulman, K.A., Berlin, J.A., Harless, W., et al, 1999. The Effects of Race and Sex on Physicians' Recommendations for Cardiac Catheterization. New England Journal of Medicine 340(8), 618-626.

Shapiro, M., Morton, S., McCaffrey, D., Senterfitt, J.W., Fleishman, J.A., Perlman J.F., Athey, L.A., Keesey, J.W., Goldman, D.P., Berry, S.H., Bozzette, S.A., additional authors from the HCSUS Consortium, 1999. Variations in the Care of HIV-Infected Adults in the United States: Results From the HIV Cost and Services Utilization Study. Journal of the American Medical Association 281, 2305-2315.

Tarlov A., Ware J.E., Greenfield S., Nelson, E.C. et al, 1989. The Medical Outcomes Study: an application of methods for monitoring the results of medical care. Journal of the American Medical Association, 262: 925-930.

van Ryn, M., 2001. The Implications of Social Cognition Research and Theory for the Provider Contribution to Race/Ethnicity Disparities in Health Care. Unpublished, University of Minnesota.

Weinstein, M., Fineberg, H., Elstein, A., Frazier, H., Neuhauser, D., Neutra, R., McNeil, B., 1980. Clinical Decision Analysis, W.B. Saunders Company, Philadelphia.

Wells, K.S., Sturm, R., Sherbourne, C.D., Meredith, L.S., 1996. Caring for Depression. Harvard University Press, Cambridge, Massachussets.

Table 2: Description of the data by minority status

	Whites		Minorities		t test #
	Mean	Std.dev.	Mean	Std.dev.	
1. PRIORS (NHANESIII, NCS)					
Depression	0.069	0.017	0.076	0.023	**
Hypertension	0.227	0.214	0.225	0.203	
Diabetes	0.051	0.042	0.064	0.053	**
2. SIGNALS OF DISEASE (MOS)					
Depression	0.163	0.369	0.238	0.426	**
Hypertension	0.308	0.462	0.342	0.475	
Diabetes	0.094	0.292	0.149	0.356	**
3. DIAGNOSES (MOS)					
Depression	0.136	0.342	0.102	0.302	**
Hypertension	0.234	0.423	0.247	0.431	*
Diabetes	0.055	0.228	0.083	0.275	**
4. PATIENTS' DEMOGRAPHICS and GENERAL CLINICAL INFORMATION (MOS)					
Age	44.058	16.888	40.685	14.515	**
Female	0.587	0.492	0.655	0.476	**
Married	0.575	0.494	0.489	0.500	**
Unemployed	0.024	0.153	0.048	0.213	**
Education	14.154	2.731	13.556	2.858	**
Income ($ 1986)	38540	25546	29948	21539	**
Los Angeles	0.233	0.423	0.349	0.477	**
Boston	0.484	0.500	0.260	0.439	**
Chicago	0.283	0.451	0.391	0.488	**
No insurance	0.036	0.186	0.037	0.188	**
HM0	0.526	0.499	0.666	0.472	**
Patient's first visit	0.234	0.423	0.235	0.424	**
Health fair/poor	0.123	0.329	0.207	0.405	**
Current smoker	0.221	0.415	0.225	0.418	
Body mass index	25.301	5.515	26.251	5.992	**
5. CLINICIANS' INFORMATION (MOS)					
Clinician's Age	38.390	6.100	37.894	7.018	**
Staff Practice	0.176	0.381	0.288	0.453	**
Solo Practice	0.450	0.498	0.210	0.407	**
Group Practice	0.374	0.484	0.502	0.500	**
Family practice	0.688	0.463	0.689	0.463	
Internist	0.312	0.463	0.311	0.463	
Observations	6505		1359		

- t-tests testing for significance of mean differences across groups
* Significant at 5% level; ** significant at 1% level

Table 3: Hypertension

Dep. variable: Diagnosis of hypertension	**(1) Traditional disparities**	**(2) Statistical discrimin.** Full specification	**(3) Statistical discrimin.** Excluding X* effects from (2)	**(4) Statistical discrimin.** Adding race effects to (3)	**(5) Statistical discrimin.** Adding gender to (3)	**(6) Statistical discrimin.** Adding age to (3)
Signal	2.305	2.216	2.214	2.252	2.212	2.177
	(24.88)**	(13.87)**	(13.39)**	(14.37)**	(13.38)**	(12.93)**
Prior		4.691	1.986	2.018	1.962	4.728
		(3.11)**	(6.40)**	(6.73)**	(6.12)**	(4.21)**
Prior X signal		0.381	0.327	0.302	0.332	0.416
		(0.96)	(0.84)	(0.79)	(0.84)	(1.04)
Race X signal		-0.166		-0.169		
		(0.62)		(0.63)		
Race (minority)	0.175	0.045		0.191		
	(1.38)	(0.18)		(0.86)		
Female	-0.219	-0.055			-0.133	
	(3.01)**	(0.67)			(1.78)	
Age: 31-41	0.512	0.169				0.159
	(2.52)*	(0.78)				(0.77)
Age: 42-60	0.959	-0.482				-0.507
	(5.45)**	(1.05)				(1.42)
Age: 60+	1.377	-1.334				-1.361
	(7.33)**	(1.65)				(2.28)*
OTHER CONTROLS						
Health status	yes	yes	yes	yes	yes	yes
Patient's SES	yes	yes	yes	yes	yes	yes
Insurance, Type of visit	yes	yes	yes	yes	yes	yes
Geography	yes	yes	yes	yes	yes	yes
Clinician charact.	yes	yes	yes	yes	yes	yes
Observations	4066	4066	4066	4066	4066	4066
Baysian inform. Criterion	-2301	-2305	-2326			

Table 4: Depression

Dep. variable: Diagnosis of depression	**(1) Traditional disparities**	**(2) Statistical discrimin.** Bayesian model	**(3) Statistical discrimin.** Excluding X* effects from (2)	**(4) Statistical discrimin.** Adding race effects to (3)	**(5) Statistical discrimin.** Adding gender to (3)	**(6) Statistical discrimin.** Adding age to (3)	**(7) Miscommunication only**
Signal	0.917	1.290	1.397	1.373	1.301	1.343	0.974
	(14.17)**	(5.61)**	(6.86)**	(6.56)**	(5.70)**	(6.21)**	(14.13)**
Prior		-9.847	2.049	2.171	-19.273	8.707	
		(1.62)	(1.05)	(1.06)	(5.11)**	(4.11)**	
Prior X signal		-3.723	-6.126	-4.873	-4.729	-5.475	
		(1.13)	(2.16)*	(1.60)	(1.50)	(1.81)	
Race X signal		-0.291		-0.285			-0.303
		(1.73)		(1.69)			(1.97)*
Race (minority)	-0.177	-0.044		-0.081			-0.081
	(2.09)*	(0.38)		(0.72)			(0.79)
Female	0.287	0.631			0.914		0.288
	(5.11)**	(3.39)**			(7.64)**		(5.11)**
Age: 31-41	0.413	0.229				0.525	0.414
	(5.59)**	(2.04)*				(7.15)**	(5.62)**
Age: 42-60	0.500	0.272				0.642	0.500
	(6.19)**	(2.14)*				(7.29)**	(6.17)**
Age: 60+	0.376						0.375
	(4.43)**						(4.42)**
OTHER CONTROLS							
Health status	yes	yes	yes	yes	yes	yes	yes
Patient's SES	yes	yes	yes	yes	yes	yes	yes
Insurance, Type of visit	yes	yes	yes	yes	yes	yes	yes
Geography	yes	yes	yes	yes	yes	yes	yes
Clinician charact.	yes	yes	yes	yes	yes	yes	yes
Observations	7864	6251	6251	6251	6251	6251	7864
Bayesian inform. criterion	-403	-394	-321				-400
Robust z statistics in parentheses * significant at 5%; ** significant at 1%							

[1] Note: the number of observations for columns (2)-(6) differs from those on columns (1) and (7) because NCS data, used to compute the priors didn't sample persons aged 60 or more.
X*=(race, gender, age)