NBER WORKING PAPER SERIES

BEWARE OF UNAWARENESS:
RACIAL/ETHNIC DISPARITIES IN AWARENESS OF CHRONIC DISEASES

Pinka Chatterji
Heesoo Joo
Kajal Lahiri

Working Paper 16578
http://www.nber.org/papers/w16578

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

This research was supported by the National Center on Minority Health and Health Disparities, National Institutes of Health (grant number P20MD003373). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Center on Minority Health and Health Disparities, the National Institutes of Health, or the National Bureau of Economic Research. The HRS 2006 Biomarker data is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and was conducted by the University of Michigan.

Beware of Unawareness: Racial/Ethnic Disparities in Awareness of Chronic Diseases
Pinka Chatterji, Heesoo Joo, and Kajal Lahiri
NBER Working Paper No. 16578
December 2010
JEL No. I1

ABSTRACT

This paper studies racial/ethnic disparities in awareness of chronic diseases using biomarker data from the 2006 HRS. We estimate a 3-step sequential probit model which accounts for selection into: (1) participating in biomarker collection; (2) having illness (hypertension or diabetes); (3) being aware of illness. Contrary to studies reporting that African-Americans are more aware of having hypertension than non-Latino whites, we do not find this conclusion holds after self-selection and severity are considered. Likewise, African-Americans and Latinos are less aware of having diabetes compared to non-Latino whites. Disparities in unawareness are exacerbated when we limit the sample to untreated respondents.

Pinka Chatterji
State University of New York at Albany
Economics Department
1400 Washington Avenue
Albany, NY 12222
and NBER
pchatterji@albany.edu

Heesoo Joo
State University of New York at Albany
Economics Department
1400 Washington Avenue
Albany NY 12222
hj528176@albany.edu

Kajal Lahiri
State University of New York at Albany
Economics Department
1400 Washington Avenue
Albany, NY 12222
klahiri@albany.edu

1.0 Introduction

There are well-documented racial/ethnic disparities in the incidence of chronic diseases in the US. Numerous studies show that African-Americans and Latinos are more likely than non-Latino whites to develop and have adverse consequences related to chronic health conditions such as heart disease, cancer, diabetes, arthritis, obesity, and hypertension (see NHDR, 2003 for a review). Racial/ethnic disparities in chronic diseases can result from a variety of mechanisms, including differences across groups in access to effective medical care, insurance status, SES, geography, and patient/provider interactions (Adler & Stewart, 2010; Gray et al., 2009; Bahr, 2007; Caillier, 2006; LêCook et al., 2010; Alegria et al., 2002; Balsa et al., 2005; Balsa & McGuire, 2003; Chandra & Skinner, 2003). Recent work suggests that factors related to health knowledge and information also contribute to health disparities (Cutler & Lleras-Muney, 2010; Aizer & Stroud, 2010; Goldman & Lakdawalla, 2005; Meara, 2001).[1] If patient-level factors such as therapy compliance or behavioral response to health information vary across groups, these differences ultimately may lead to disparities in effective treatment and health outcomes even when groups have equal access to medical care.

Early awareness of having a chronic health condition is another aspect of health knowledge that influences an individual's ability to manage the progression of a disease, and may contribute to health disparities. Without early preventative intervention, the course of a chronic disease is a continuum from the disease-free state to asymptomatic biological change,

[1] Lleras-Muney & Cutler (2010), for example, report that knowledge and cognitive ability explain a portion of the education gradient in health behaviors. Aizer & Stroud (2010) find that more educated mothers are more likely than less educated mothers to quit smoking in response to new information about the dangers of cigarettes, explaining a portion of the education gradient in newborn health. Goldman & Lakdawalla (2005) find that education disparities in HIV outcomes can be attributed in part to differences across groups in adherence to a complicated treatment regimen. Meara (2001) reports that differential response to knowledge by SES can explain about one third of the education gradient in maternal smoking.

clinical illness, impairment, disability and ultimately death. Due to the long latency period, early recognition and treatment of chronic illnesses is critical for preventing a rapid progression of the condition to disability. In the present study, we examine whether there exist racial and ethnic disparities in the awareness of two highly prevalent, costly chronic diseases that are typically asymptomatic in their early stages -- diabetes and hypertension. We use new, state-of-the-art data from the Health and Retirement Study (HRS), which includes biomarkers for diabetes and measurement of hypertension. We build on prior work in this area by using a three-step sequential probit model, which accounts for the possibility that there are factors (both observed and unobserved) that drive an individual's choice to participate in the medical examination portion of the HRS survey, the likelihood the individual has chronic health conditions, and the probability that the individual is aware of having a chronic health condition if he does indeed have one. As we discuss below, these factors are likely to be correlated with race/ethnicity, and thus may obscure the identification of racial/ethnic disparities in awareness of chronic conditions if we do not adjust for them.

In contrast to recent research, which indicates that African-Americans are more aware of having hypertension than non-Latino whites, our findings suggest that while relatively more African-Americans receive treatment for this illness, African-Americans still are disproportionately unaware of having the condition compared to individuals from other racial/ethnic groups. With respect to diabetes, both African Americans and Latinos are more likely than non-Latino whites to have the disease and be treated for it, but the remaining untreated minorities are less likely to be aware of their condition compared to non-Latino whites. We find no income gradient in awareness of hypertension or diabetes, but more post-secondary education is associated with higher likelihood of awareness among those with hypertension.

2.0 Background

Studies of disease awareness require data that include both self-reports of illness and objective measurement of the condition. That is, classifying an individual as being unaware of having a disease requires that the individual reports not having the disease, and also that an almost simultaneous medical examination shows that the individual does indeed have the disease. In the US context, the National Health and Nutrition Examination Survey (NHANES) is one of the few large-scale surveys which provide both self-reported and objective measurements of chronic illness.[2] As a result, most previous studies of disease awareness in the US are based on the NHANES.

Several recent studies examine racial/ethnic differences in awareness of chronic disease, with some results related to hypertension showing greater awareness among minorities compared to non-Latino whites. Hertz et al. (2005), for example, use the 1999-2002 NHANES and find that 82 percent of African-Americans over 60 years of age are aware that they have hypertension compared to 72 percent of non-Latino whites in the same age group (Hertz et al., 2005). Ong et al. (2007) report that awareness of hypertension is higher among African-Americans than non-Latino whites in the 2001-2002 NHANES (African-Americans 73.2% vs. non-Latino whites 59.9%). Similarly, Howard et al. (2006), using the REasons for Geographic And Racial Differences in Stroke (REGARDS) database, find that African-Americans are more likely to be aware of their high blood pressure than non-Latino whites (odds ratio: 1.31).

Other findings, however, suggest that minorities are less aware of having chronic illness compared to non-Latino whites, or that there is no difference between minorities and non-Latino

[2] This information is available in the NHANES III and in later versions of these data.

whites in awareness. Based on data from the 2003-2006 NHANES and the 2003-2007 Behavioral Rick Factor Surveillance System (BFRSS), Danaei et al.(2009) report that Latinos are more likely than non-Latino whites to have undiagnosed diabetes (odds ratio: 2.03), but there are no statistically significant differences between African-Americans and non-Latino whites. Pierce et al. (2009) report no racial/ethnic differences in awareness of diabetes using 2004-2005 data from the English Longitudinal Study of Aging (ELSA).

Several empirical challenges arise in estimating racial/ethnic disparities in disease awareness. First, individual-level factors such as disease severity and health status may confound an observed association between race/ethnicity and awareness in a cross-sectional sample. Hertz et al. (2005) and Ong et al. (2007), for example, find that African-Americans who objectively meet criteria for hypertension are more likely to be aware of their condition than non-Latino whites who meet objective criteria. The models, however, do not adjust for health status and severity of disease. As a consequence, the results may reflect the fact that African-Americans at a particular age simply are more likely than non-Latino whites to have advanced, symptomatic disease, and, as a result, are more likely to be under treatment and be aware of the disease. If that is the case, greater awareness of disease observed in the sample is not necessarily indicative of better access to screening and treatment services.

Moreover, awareness status is known only for those who objectively meet criteria for disease. That is, one cannot observe whether a healthy person would be aware of his disease if he had the disease. This censoring issue potentially confounds estimates of disparities in prior work. For example, Howard et al. (2006) adjust for observable aspects of severity using a logit model, and still find that African-Americans are more likely to be aware of having hypertension compared to non-Latino whites. However, this single-equation approach does not account for the possibility that unmeasured aspects of health status and health behaviors may be correlated with race/ethnicity, and also related to both the likelihood of having the disease and the likelihood of being aware of the disease if it exists. This issue may confound estimates of racial/ethnic disparities in unawareness.

Most recently, Johnston et al. (2009), using the Health Survey for England (HSE), examine the income/health gradient using self-reported and objective measures of hypertension. They estimate a censored bivariate probit model to account for the possibility that measured and unmeasured factors may affect both an individual's propensity to meet objective criteria for hypertension and an individual's likelihood of misreporting his hypertension status (what we term "being unaware" in our study). The findings show that income is negatively related to misreporting hypertension status, but they do not find evidence of racial/ethnic disparities (Johnston et al., 2009).

In the present study, we estimate racial/ethnic disparities in health awareness among older individuals in the US and build on prior work in a number of ways. First, we take advantage of the rich data available in the HRS to adjust for a range of individual-level factors that may be correlated with both race/ethnicity and awareness, including disease severity. Second, like Johnson et al. (2009), we account for the possibility that factors exist which affect both the likelihood of having chronic illness and the likelihood of being aware of the illness if it exists. By jointly estimating equations modeling the probability of having the illness and the probability of being aware of the illness if it exists, we can account for measured and unmeasured variables that may confound an observed association between race/ethnicity and awareness. We apply this method using data from the US, while Johnson et al. (2009) uses data from England and focuses on the income/health gradient.

Our third contribution is that we take into account a second form of censoring that may be particularly important in the estimation of racial/ethnic disparities - - censoring that results from respondents refusing to participate in the collection of health examination and biomarker data. We only observe objective and self-reported measures of chronic illness (and thus unawareness status) for HRS respondents who: (1) agree to participate in the collection of health examination and biomarker data; and (2) provide a self-report about chronic illness. We allow for the possibility that individuals select into survey participation along measured and unmeasured factors that also affect the existence of disease and awareness of disease if it exists.

As discussed below, participation in the HRS biomarker data collection effort involves a medical examination, which includes blood pressure measurement and a blood draw. Many factors may affect an individual's decision to participate in the medical examination, and some of these factors are also likely to affect disease prevalence and awareness. For example, an individual who has strong mistrust of the health care system may be reluctant to participate in the medical examination, and also may be more likely to have chronic illness and more likely to be unaware of it if it exists. On the other hand, an individual who is knowledgeable about health may be both more likely to participate in the medical examination, less likely to have illness, and more likely to be aware of it if it exists.[3]

There are numerous reasons to believe that this type of censoring may be related to race/ethnicity and SES. In the case of African-Americans, there is a well-documented history of mistrust of medicine and medical research, stemming in part from the Tuskegee Syphilis Study and other violations of medical ethics in the US that were targeted at minority patients (LaVeist

[3] The HRS provides both the respondent and the respondent's physician with the results of the diabetes and hypertension examinations. Thus, more health conscious individuals may have an incentive to participate. However, it's also possible that respondents with poor health behaviors may be more likely than others to participate if these individuals believe that their health behaviors put them at elevated risk for disease.

et al., 2000; Achter et al., 2005; and Halbert et al., 2009). For all minority groups, culture, language, immigration status, education, and health knowledge are just some of factors that may affect participation in the HRS biomarker data collection as well as disease and awareness of disease. For example, in a study of prostate cancer patients, Halbert et al. (2009) report that African-Americans and low SES individuals report greater mistrust of the healthcare system compared to non-Latino whites and higher SES individuals. Achter et al. (2005) find that African- Americans are more likely to believe that clinical trials are not safe compared to non-Latino whites. LaVeist et al.(2000), in a study of cardiac patients, find that lower satisfaction of medical care among African-Americans compared to non-Latino whites partly comes from mistrust of medical care. Thus, adjusting for this type of censoring is particularly relevant to the estimation of racial/ethnic disparities.

3.0 Data, definitions, and sample statistics

3.1 The Health and Retirement Study

Data for this study come from the 2006 Health and Retirement Study (HRS), a biannual, nationally representative, longitudinal household survey initiated in 1992. The HRS started with a sample of households in which the household heads were 51 to 61 years old (born between 1931 and 1941). The sample also included spouses of household heads that were of any age. Since the first wave, the HRS has added samples which include different age cohorts. As of 2006, the HRS contained 18,469 individuals from AHEAD (born before 1924), CODA (born in 1924-1930), HRS (born in 1931-1941), War Babies (born in 1942-1947), Early Baby Boomers (born in 1948-1953), and Late Baby Boomers (1954-1959). The HRS collects information through face-to-face interviews, phone interviews, and proxy interviews.

The 2006 HRS core interview included an enhanced face-to-face interview with a medical examination by a trained interviewer. The medical examination involved measurement of height, weight, mobility, strength, blood pressure, and lung capacity, as well as collection of biomarker data through saliva and blood samples. The sampling process used for the enhanced interview is as follows. First, the HRS randomly assigned half of the households in the 2006 Core sample to an enhanced face-to-face interview in 2006. Those households who were not assigned to an enhanced face-to-face interview in 2006 wave were assigned to an enhanced face-to-face interview in the 2008 wave. Among those respondents who were assigned to an enhanced face-to-face interview in 2006 (N=9,570), the HRS excluded respondents who were living in a nursing home, who chose a proxy interview, or who chose a phone interview (12.4 percent of the 2006 enhanced face-to-face interview sample). Based on this process, 8,379 respondents in 2006 (45.4% of the entire HRS 2006 sample) were eligible for an enhanced interview which involved a medical examination.

Our analysis sample includes 8,051 respondents. We drop 328 respondents who were eligible for an enhanced interview but have missing data for any socio-demographic and economic variables we use in the analysis. The body mass index (BMI) variable, which is based on respondents' self-reported height and weight, had a large number of missing values. We replace missing BMI values with the sample mean and use a missing data indicator to flag respondents with missing BMI information.

The first column of Table 1 shows the un-weighted mean characteristics for the analysis sample. The mean age is 67 years old (min: 30, max: 104).[4] The sample is 76 percent non-

[4] Those who are younger than the late baby boomer cohort (under 47 years old in 2006) are the spouses of household heads who are older than 47 years old. These individuals comprise only 1.6 percent of the sample.

Latino white, 14 percent African-American, 8 percent Latino, and 2 percent American Indian, Alaskan Native, Asian and Pacific Islander (Table 1). In the sample, 22 percent are high school dropouts, 34 percent have graduated from high school, 22 percent have completed some college and 22 percent have completed four years of college or more. The average household income is $69,922 (Table 1). Based on body mass index (BMI) calculated from self-reported height and weight, 34 percent of the sample is overweight and 28 percent is obese (11 percent have missing BMI information).

Among HRS respondents who were eligible for the enhanced interview, some respondents refused to participate in the medical examination and/or blood draw, or could not provide usable information from the medical examination. There was a separate consent process for the saliva and blood draws that was conducted just before the samples were collected. Respondents only participated if they affirmed both that they understood the directions and that they felt safe participating.[5] "Agreed to participate" in our sample includes eligible respondents who agreed to participate in the medical examination and provided usable data from the medical examination. "Refused to participate" includes those respondents who refused, as well as those who accepted the medical examination but provided unusable measurements. For hypertension, 537 respondents who smoked, exercised, or consumed alcohol or food within the 30 minutes prior to completing the blood pressure measure could not provide usable measurements of blood pressure.

If the respondents in the refusal group have very similar observable characteristics as respondents in the acceptor group, it would be less likely that a sample selection problem exists that would affect our estimates of racial/ethnic disparities. In Table 1, we show sample

[5] More information is available at: http://hrsonline.isr.umich.edu/sitedocs/userg/HRS2006BiomarkerDescription.pdf.

characteristics by refusal/acceptance status for both hypertension and diabetes. In the case of hypertension, which involves a physical measurement of blood pressure status (described in more detail below), the refusal group disproportionately includes African-Americans and respondents with low income, less education, and poor health. The same general pattern is true for diabetes, which involves a blood draw. This result differs from Johnston et al. (2009), who report no selection problem with regard to agreeing to participate in measurement of chronic illness. However, in our case, Table 1 strongly suggests the existence of a self-selection problem in the 2006 HRS that must be addressed in the estimation of racial/ethnic disparities.

3.2 Measurement of chronic illness: hypertension and diabetes

The HRS asks respondents to provide a self-report of hypertension and diabetes. If a respondent is new in the 2006 wave, the respondent is asked "*Has a doctor ever told you that you have high blood pressure or hypertension [diabetes]?*" If a respondent participated in a prior HRS wave and reported hypertension (diabetes) in the last interview, the interviewer asks him/her whether s/he wishes to dispute the prior report of illness. If not, the respondent is asked whether s/he currently takes medication for hypertension (diabetes).

If a respondent participated in a prior HRS interview but did not report hypertension (diabetes), the respondent is asked "*Since we last talked to you, has a doctor told you that you have high blood pressure or hypertension [diabetes]?*" If the respondent reports high blood pressure or hypertension (diabetes) in the current period, s/he is asked a follow-up question about whether s/he takes any medication for the illness.[6] In Table 2, we see that 56 percent of our

[6] This question structure has the advantage of reducing the likelihood of false-positive reporting due to 'no longer having the illness' since the data is updated every two years. Even though hypertension or diabetes is very unlikely to be completely cured, there is the possibility of complete recovery due to proper management. Update of the status every two years certainly reduces the likelihood of this kind of error.

sample self-reports hypertension and 50 percent reports currently taking medication for hypertension. Table 2 also shows that 20 percent of our sample reports having diabetes and 17 percent reports taking medication for diabetes.

The HRS enhanced interview included measurement of each consenting participant's systolic blood pressure (SBP) and diastolic blood pressure (DBP), which we consider to be objective measurement of hypertension. The blood pressure reading was taken three times during the interview. We use the mean of the 2nd and the 3rd readings of systolic blood pressure and diastolic blood pressure (Johnston et al., 2009). Based on standard definitions (see Hertz et al., 2005; Morenoff et al., 2007; Angell et al., 2008; and Johnston et al., 2009), we consider a person to be hypertensive if s/he has over 140 mmHg systolic blood pressure (SBP) or over 90 mmHg diastolic blood pressure (DBP). The HRS enhanced interview also included a blood draw and blood analysis. As an objective measure of diabetes, we use the A1C level, a measure of the average glucose level in the respondent's blood over the past 2-3 months.[7] Following guidelines from the American Diabetic Association, we consider an A1C of higher than 6 percent to be an indicator of diabetes (Buell et al., 2007; and Ginde et al., 2008)[8].

As seen in Table 2, about 32 percent of our sample meets objective criteria for hypertension and about 24 percent meets objective criteria for diabetes. Note that these individuals have uncontrolled or undiagnosed disease – that is, these individuals include both who are aware of their disease but have not been able to effectively treat it, as well as people who are unaware that they have the illness. These rates, however, would not include individuals

[7] One merit of the A1C test is that test results are insensitive to the timing of measurement and show broad shot of an individual's diabetes status. A normal A1C for people without diabetes is 4-6 percent. However, A1C level is not usually used for the diagnosis of diabetes but rather used for checking blood sugar level among diagnosed diabetes patients.

[8] Buell et al.(2010) propose that A1C≤6% are normal and recommend glucose measurement for A1C level ≥5.8%. Ginde et al.(2008) present that A1C≥6.1% are high risk group with undiagnosed diabetes.

who have the disease but who are being treated effectively with medication and/or lifestyle changes.

Co-morbidity between chronic illnesses is common, particularly among racial/ethnic minorities. Among non-Latino whites, 18 percent have both hypertension and diabetes, compared to co-morbidity rates of 35 percent among African-Americans, and 32 percent among Latinos. The high co-morbidity among African-Americans and Latinos may reflect that African-Americans and Latinos are in worse health overall compared to non-Latino whites. In Figures 1 and 2, we show the prevalence of hypertension and diabetes over different age groups in our sample (based on the criteria that the patient has an objective measurement that indicates disease, and/or self-reports medication use) by race/ethnicity. We see in Figure 1 that the largest hypertension prevalence gap between African-Americans and non-Latino whites occurs in the under-50 age group, which suggests that African-Americans have earlier onset of illness compared to other groups. Figure 2 shows that in the case of diabetes, the prevalence gap between African-Americans and non-Latino whites increases until the 65-70 year old age group and decreases after that point. There is a similar pattern in the prevalence gap between Latinos and non-Latino whites (Figure 2).

3.3 Definition of awareness

Our primary definition of awareness is if conditional on having the disease as determined by medical examination, respondents self-report that they do indeed have the illness. Following a recent report of the Seventh Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC7, 2003) and previous studies (Hertz et al., 2005; Primatesta & Poulter, 2006; Morenoff et al., 2007; and Angell et al., 2008), our criteria for “having the disease” is if the respondent meets objective measurement criteria (e.g., has uncontrolled illness) and/or the individual self-reports taking medication for the disease. We include individuals who are taking medication as being part of the disease group because many respondents in our sample have disease that is controlled by medication. For example, in our sample, among those who report taking medication for hypertension, 63 percent have normal blood pressure measurement and, among those who report taking medication or using insulin pump for diabetes, 25 percent have normal blood sugar measurement.

Table 2 shows descriptive statistics for disease prevalence and disease awareness. Among those with hypertension, 83 percent of respondents are aware of having hypertension, and, among those with diabetes, 64 percent of respondents are aware of having diabetes (Table 2).[9] We see in Table 2 that awareness of hypertension is 83 percent for non-Latino whites, 88 percent for African-Americans and 83 percent for Latinos. These findings show that awareness is higher among African-Americans compared to non-Latino whites, which is consistent with previous studies (Hertz et al., 2005; and Ong et al., 2007). For diabetes, awareness is 63 percent for non-Latino whites, 64 percent for African-Americans and 67 percent for Latinos. This result, which shows small differences in awareness between racial and ethnic groups for diabetes, is also consistent with previous studies from NHANES data (Danaei et al., 2009 and Pierce et al., 2009).

[9] In Table 3, we compare the weighted versions of these rates to published rates based on NHANES respondents over 60 years old. The prevalence of hypertension appears to be somewhat higher in the NHANES than the HRS, but awareness of hypertension very close across the two samples; 83 percent of our sample is aware of having hypertension vs. 81 percent in the NHANES. In the case of diabetes, the prevalence of illness is similar in the NHANES, but awareness of diabetes appears to be somewhat higher in the NHANES compared to the HRS. These differences may simply reflect the differences in the demographic characteristics of the two samples, and/or the somewhat different blood analyses used to diagnose diabetes in the two surveys.

Our main definition of awareness follows the epidemiological literature in this area, which considers an individual to be aware if s/he meets the “gold standard” criteria for disease (based on objective measurement and usage of medication) and conditional on this fact, self-reports having the disease. It is best understood in the context of the schematic diagram depicting the flow of HRS respondents through the medical examination and self-reports (Figure 3). The absolute counts and the percentages are also reported at each stage yielding the unawareness rates of 17% for hypertension and 38% for diabetes under Group C. This definition focuses on false-negative responses and does not consider respondents who give false-positives to be unaware, even though this form of misreporting may be less relevant from the standpoint of public health. In Table 4, we present a cross-classification table between self repots and medical examination reports. In the context of this concordance table, the awareness is defined as {a/a+c}, which is the *sensitivity* of self-reports. In our sample, about 7 percent gives false-positives for hypertension and about 3 percent give false positives for diabetes. These cases can arise if, following a doctor’s report, the individual is controlling the disease by diet, exercise and other healthy lifestyle changes such that the actual medical examination does not identify the disease. In addition some errors in measurement can never be ruled out, particularly in case of hypertension.

3.4 Concordance between “self-reports” and “medical examinations”

Two types of mismatches between self reports and medically-tested objective measurements are possible – an individual may fail to a report a disease that exists, or an individual may incorrectly report a disease that does not exist (see Table 4). If there are a reasonably small number of mismatched cases, the self-reported information has strong power to predict actual chronic health conditions, and there is little room for improvement in the awareness of chronic conditions. In this section, we examine whether this is the case, and we also test whether dependency between self-reports and objective measurements differs by race/ethnicity.

First, we proceed with the bellwether test for complete independence of self-report and objective measurement using Chi-squared tests. If there is complete independence, self-reports of chronic illness do not give us any information about whether individuals actually have disease. The test statistics are shown in Table 5, and are based on the concordance tables such as Table 4. Not surprisingly, we reject the null hypothesis of complete independence for all racial/ethnic groups. More meaningfully, next we use an index of concordance $\hat{I}$ proposed by Harding and Pagan (2002) to test complete dependency between self-reports and objective measurements:

$$\hat{I} = \frac{1}{N}\{\sum_{i=1}^{N} y_{pi}y_{si} + \sum_{i=1}^{N}(1 - y_{pi})(1 - y_{si})\} \qquad (1)$$

where y_{pi} takes value 1 if individual i is truly in the disease group (i.e., hypertension or diabetes) and 0 otherwise, and y_{si} is 1 if the same individual i self-reports having the disease and 0 otherwise. We use a t-test with null hypothesis $H_0\colon \hat{I} = 1$ which means that y_{pi} and y_{si} are completely dependent. Table 5 presents estimated indices $\hat{I}$, and associated t-test statistics by race/ethnicity.

For both hypertension and diabetes, indices of concordance by each racial/ethnic group are significantly different from zero (hypertension: 0.86-0.88, and diabetes: 0.83-0.89), which means that self-reports and objective measurements are not perfectly correlated. We also test whether the dependency differs by race/ethnicity. If the index $\hat{I}$ for non-Latino whites is significantly greater than the index $\hat{I}$ for African-Americans (Latinos), the result supports the idea that non-Latino whites are more aware of chronic conditions compared to African-

Americans (Latinos). Let us denote $\hat{I}_w$, $\hat{I}_A$, and $\hat{I}_L$, as the indices for non-Latino whites, African-Americans and Latinos, respectively. In Table 5 we find that the null hypotheses $H_0: \hat{I}_w = \hat{I}_A$ and $H_0: \hat{I}_w = \hat{I}_L$ are not rejected for hypertension. For diabetes, however, the null hypotheses for both African-Americans and Latinos are rejected against the alternatives $H_1: \hat{I}_w > \hat{I}_A$ and $H_1: \hat{I}_w > \hat{I}_L$. This finding suggests that non-Latino whites are more aware of their diabetic conditions than African-Americans or Latinos.

Even though the index of concordance is appropriate for testing complete dependency, it does not provide information regarding whether the incomplete dependency comes from false-negative or false-positive reporting, since both type of misreporting decrease concordance index. A regression method proposed by Benitez-Silva et al. (2004) is as an alternate way to test for complete dependency. This method has an advantage that it can distinguish the effects of false-positives and false-negatives. This is because the procedure regresses $y_p - y_s$, which has a different sign for false-positive and false-negative reporting, on a constant term and racial/ethnic dummies. The results in Table 5 show the constant term is significantly positive, which shows that there are more people in the false-negative reporting group than in the false-positive reporting group for both hypertension and diabetes. The significantly positive coefficients on the racial/ethnic dummy variables in the OLS model for diabetes support the idea that non-Latino whites are more aware of diabetes than African Americans and Latinos. For hypertension, the race-ethnic dummies are not statistically significance at even 10% significance level. These results are consistent with the those related to the index $\hat{I}$.

In sum, the analysis in this section suggests that the mismatch between self-reports and objective measures of chronic illness in HRS is not ignorable because of the statistically significant partial dependency between self-reports and objective measurements. In the case of hypertension, the dependency is not significantly different by race/ethnicity, but for diabetes, the match is significantly higher among non-Latino whites compared to African-Americans and Latinos. Furthermore, non-Latino whites are more aware of diabetes than African-Americans and Latinos. In the sections that follow, we focus on racial/ethnic disparities in false-negative reporting (e.g., unawareness) of chronic illness that will also correct for possible sample selection problems.

4.0 Model and Methods

As introduced in Figure 3, our definition of awareness implies a 3-step sequential model in which HRS respondents potentially may be screened out at two steps. Figure 3 maps out this model with sample sizes at different levels separately for hypertension and diabetes. In the first step, HRS respondents either agree to participate in the medical examination or they refuse; if they refuse, they are screened out at this stage since we do not have information on whether or not they have the disease (Figure 3, Step 1: Accept medical exam). In the second step, HRS respondents who participate either have the disease based on objective criteria or they do not (Figure 3, Step 2: Have disease). At this stage, respondents who do not have the disease are screened out since we cannot observe what their awareness status would have been if they had the disease. Finally, in the third step, among those who have the disease, we can ascertain awareness of disease status based on the respondent's self-report of illness (Figure 3, Step 3: Aware of disease). Respondents who self-report that they have the disease are aware, and those who self-report not having the disease are unaware of illness.

Essentially, we observe four mutually exclusive outcomes, labeled *A*, *B*, *C* and *D* in Figure 3. The sets of sample individuals belonging to these outcomes are denoted by *a*, *b*, *c*, and *d* respectively. The outcome D includes respondents who do not agree to participate $(h_1 = 0)$, and the outcome B includes respondents who agreed to participate and do not have the disease $(h_1 = 1, and\ h_2 = 0)$. The outcome C includes respondents who agreed to participate, have the disease, and are not aware of disease $(h_1 = 1, h_2 = 1, and\ h_3 = 0)$, and finally, outcome A includes respondents who agreed to participate, have the disease, and are aware of the disease $(h_1 = 1, h_2 = 1, and\ h_3 = 1)$. Although we observe binary outcome h_i, the underlying latent variable h_i^* can be expressed as follows:

$$h_{ij}^* = \beta_i' x_{ij} + \varepsilon_{ij} \qquad (2)$$

for i=1,2, and 3. The vector x_i represents individual characteristics at step i, and j are individuals. We call i=1 the "acceptance" step (i.e., agreeing to the medical examination), i=2 the "incidence of disease" step and i=3 the "awareness" step. For i=1, h_{1j} equals 1 if $h_{1j}^* \geq 0$ and h_{1j} equals 0 otherwise. For i=2 and3, h_{ij} equals 1 if $h_{ij}^* \geq 0, and\ h_{kj}^* \geq 0$ with all k<i and otherwise h_{ij} equals 0. The probabilities of each outcome are written as:

$$\Pr(D) = \Pr\left(h_{1j}^* < 0\right) = \Phi(-\beta_1' x_{1j}) \qquad (3)$$

$$\Pr(B) = \Pr\left(h_{1j}^* \geq 0, and\ h_{2j}^* < 0\right) = \Phi_2(\beta_1' x_{1j}, -\beta_2' x_{2j}, \Omega_2) \qquad (4)$$

$$\Pr(C) = \Pr\left(h_{1j}^* \geq 0, h_{2j}^* \geq 0, and\ h_{3j}^* < 0\right) = \Phi_3(\beta_1' x_{1j}, \beta_2' x_{2j}, -\beta_3' x_{3j}, \Omega_3) \qquad (5)$$

$$\Pr(A) = \Pr\left(h_{1j}^* \geq 0, h_{2j}^* \geq 0, and\ h_{3j}^* \geq 0\right) = \Phi_3(\beta_1' x_{1j}, \beta_2' x_{2j}, \beta_3' x_{3j}, \Omega_3) \qquad (6)$$

where the errors $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ are assumed to be normally distributed with mean zero and correlation matrix $\Omega_3 = \begin{pmatrix} 1 & & \\ \rho_{21} & 1 & \\ \rho_{31} & \rho_{32} & 1 \end{pmatrix}$. Ω_2 is the 2x2 sub-matrix involving ρ_{21}. The joint log-likelihood function is given by:

$$L = \sum_{j \in d} \ln P\{h_{1j}^* < 0\} + \sum_{j \in b} \ln P\{h_{1j}^* \geq 0, and\ h_{2j}^* < 0\}$$
$$+ \sum_{j \in c} \ln P\{h_{1j}^* \geq 0, h_{2j}^* \geq 0, and\ h_{3j}^* < 0\}$$
$$+ \sum_{j \in a} \ln P\{h_{1j}^* \geq 0, h_{2j}^* \geq 0, and\ h_{3j}^* \geq 0\} \quad (7)$$

Like many sequential response models, however, we have unequal observations across steps. This gives rise to a multivariate probit model with partial observability. As a result, the summation of the log-probability for each outcome is taken only over the sample cases that have attained that final outcome, as denoted by the sets *d, b, c,* and a. The evaluation of the log-likelihood function L in general involves higher-dimension multiple integrals of normal rectangle probabilities unless Ω assumes a simplified structure. We use GHK Monte Carlo simulator to simulate the probabilities and evaluate the log-likelihood function.[10] Note that in expression (7), each of the integration of the probabilities for different outcomes involves only a top-left subset of Ω, depending on the dimension of the integral. The advantage of the 3-step sequential probit model is that we can maximize a joint likelihood function.

Obviously, when Ω is an identity matrix, the log-likelihood function in (7) is equivalent to that of a 3-stage step-wise sequential probit model:

$$L = (\sum_{j \in d} \ln P\{h_{1j}^* < 0\} + \sum_{j \in b \cup c \cup a} \ln P\{h_{1j}^* \geq 0\})$$
$$+ (\sum_{j \in b} \ln P\{h_{2j}^* < 0\} + \sum_{j \in a \cup c} \ln P\{h_{2j}^* \geq 0\})$$
$$+ (\sum_{j \in c} \ln P\{h_{3j}^* < 0\} + \sum_{j \in a} \ln P\{h_{3j}^* \geq 0\}) \quad (8)$$

Note that the expressions inside each pair of parentheses are independent of each other and may be evaluated separately. However, it is reasonable to assume that common unmeasured

[10] We used the GAUSS code, available on the Internet, provided by Hajivassiliou, McFadden and Ruud (1996).

factors exist which affect the errors in all equations. Health knowledge, for instance, may be difficult to measure but it may affect both the 2nd and 3rd steps, by decreasing prevalence and increasing awareness of chronic illnesses. Also, this factor also may affect the 1st step by increasing the likelihood of participating in the medical examination. Thus, we need to maximize joint likelihood function instead of maximizing separate likelihood functions.

The equations also embody certain natural exclusion restrictions. Specifically, we include three binary variables related to the HRS respondent's behavior during the interview in the "acceptance" step, but were not significant in the "disease" and "awareness" steps. These three variables are the HRS interviewer's observations about: (1) whether the respondent appeared hostile during the interview; (2) whether the respondent was uncooperative during the interview; and (3) whether the respondent expressed concerns about time during the interview. The HRS respondent's attitude during the interview is likely to influence whether the respondent agrees to participate in the medical examination, a more invasive and time-intensive portion of the interview. However, the respondent's attitude during the HRS interview should not be directly related to disease and disease awareness, after controlling for other factors.

In addition, measures of obesity and overweight status are included only in the "has disease" equation, and not in the "acceptance" and "awareness" steps. This exclusion is reasonable since body weight has physiological influence on the likelihood of having disease but does not directly affect awareness of disease and participation in the interview, after controlling for other factors. Since obesity is a major risk factor for hypertension and diabetes, this restriction is reasonable. The second restriction that obesity does not affect the awareness of chronic conditions may seem less reasonable. Obesity could affect awareness of chronic conditions if obese individuals have more opportunities to have blood pressure and/or blood sugar level measurement because of their bodyweight. However, obesity was insignificant in the awareness level possibly because of other controls including self-assessed health and co-morbidity in the specification.

In order to establish the importance of self-selection and other unmeasured factors, we need to test for the significance of non-diagonal terms in the correlation matrix of disturbances in the hypertension and diabetes models. We amended an LM test statistic for a diagonal correlation matrix (Ω=I) in a multivariate Probit model originally proposed by Kiefer (1982) for our sequential model.[11]

5.0 Results

Tables 6 and 7 report the simulated maximum likelihood (SML) estimates using the GHK simulator for the three-step sequential probit model for hypertension and diabetes respectively.[12] In both tables, Panel 1 shows estimates from the first stage (agreed to participate in the medical exam), Panel 2 shows estimates from the second stage (has the disease, among those who agreed to participate), and Panel 3 shows estimates from the third stage (aware of disease, among those who agreed to participate and also have the disease). The final rows of each table show estimates of the ρs, the estimated correlations between unmeasured factors affecting the outcomes of interest. In both Tables 6 and 7, our primary interest is in the estimated coefficients

[11] For the 3-step sequential probit model, the LM statistic is as follows: LM=$\frac{g_{12}^2}{k_{12}}+\frac{g_{23}^2}{k_{23}}+\frac{g_{13}^2}{k_{13}} \sim \chi_3^2$,
where $g_{lm} = \sum_{i=1}^{n} q_{il}\, q_{im} \frac{\varphi(w_{il})\varphi(w_{im})}{\Phi(w_{il})\Phi(w_{im})}$, $k_{lm} = \sum_{i=1}^{n} \frac{(\varphi(w_{il})\varphi(w_{im}))^2}{\Phi(w_{il})\Phi(-w_{il})\Phi(w_{im})\Phi(-w_{im})}$,

$w_{im} = q_{im} z_{im}$, $q_{im} = 2h_{im} - 1$, and $z_{im} = \beta_m' x_{im}$, (i=1,…N and m=1,2 and3).

[12] The normalization that the diagonal elements of Ω are unity prevents us from directly using the convenient derivatives of the multivariate normal rectangle probabilities in the GHK simulator as derived by Hajivassiliou, McFadden and Ruud (1996) to estimate standard errors. Instead, we rely on the maximum likelihood algorithm to generate standard errors for all parameters. We first used a sampling size of 50 for the GHK simulator to obtain initial estimates and then fine-tuned the estimation finally with a sampling size of 500.

on the race/ethnicity indicators and their marginal effects in Panel 3 of each table, since these indicators provide information on the significance and direction of racial/ethnic disparities in health awareness. In addition, the estimated coefficients on education and income in Panel 3 of each table provide information regarding the SES gradient in health awareness.[13]

In Panel 1 of Table 6, the estimates indicate that older individuals and Latinos are more likely than others to participate in measurement of blood pressure, while those in worse self-reported health status and smokers are less likely than others to agree to have a blood pressure measurement. African-American race is not associated with participation. Individuals with self-reported diabetes are less likely to participate, perhaps because they are aware of having diabetes and regularly check their blood pressure as part of diabetes self-management. However, individuals with self-reported elevated cholesterol are more likely to participate than others. Notably, individuals who appeared to be non-cooperative, hostile, or concerned about the length of the interview (based on HRS interviewer observation) were much less likely than others to agree to have their blood pressure measured by the interviewer.

---

[13] The reported conditional marginal effects were calculated using the following formulas. (1st step is unconditional.)

x is continuous:

$$\frac{\partial E[h_1|x]}{\partial x} = \frac{\partial \Pr[h_1=1|x]}{\partial x} \; ; \; \frac{\partial E[h_2|h_1=1,x]}{\partial x} = \frac{\partial \frac{\Pr[h_1=1,h_2=1|x]}{\Pr[h_1=1|x]}}{\partial x} \; ; \; \frac{\partial E[h_3|h_1=1,h_2=1,x]}{\partial x} = \frac{\partial \frac{\Pr[h_1=1,h_2=1,h_3=1|x]}{\Pr[h_1=1,h_2=1|x]}}{\partial x}$$

x is binary:

$$\frac{\partial E[h_1|x]}{\partial x} = \Pr[h_1 = 1|x = 1] - \Pr[h_1 = 1|x = 0] \; ; \; \frac{\partial E[h_2|h_1=1,x]}{\partial x} = \frac{\Pr[h_1=1,h_2=1|x=1]}{\Pr[h_1=1|x=1]} - \frac{\Pr[h_1=1,h_2=1|x=0]}{\Pr[h_1=1|x=0]}$$

$$\frac{\partial E[h_3|h_1=1,h_2=1,x]}{\partial x} = \frac{\Pr[h_1=1,h_2=1,h_3=1|x=1]}{\Pr[h_1=1,h_2=1|x=1]} - \frac{\Pr[h_1=1,h_2=1,h_3=1|x=0]}{\Pr[h_1=1,h_2=1|x=0]}$$

Greene and Hensher (2010, pp.83-96) present corresponding formulas for bivariate probit models with selection.

Panel 2 of Table 6 shows correlates of having hypertension, among those who participated in hypertension measurement. African-American race is associated with a 14 percentage point higher probability of being hypertensive. However, Latinos are not statistically different from non-Latino whites in their likelihood of being hypertensive. College education is negatively associated with hypertension. As expected, older age, obesity, other health chronic conditions, and worse self-reported health all are associated with having hypertension.

Finally, in Panel 3 of Table 6, we examine awareness of hypertension among those who have the disease and agreed to participate in hypertension measurement. Recent work (Hertz et al., 2005; Ong et al., 2006; Howard et al., 2007) indicates that African-Americans are more likely to be aware of their hypertension than non-Latino whites. Our results, in contrast, show no statistically significant differences between racial/ethnic groups in awareness of hypertension (Table 6, panel 3). Surprisingly, there is no income gradient in awareness of hypertension either. However, we do see an awareness gradient in post-secondary education; individuals with more than 12 years of education are more likely to be aware of their hypertension compared to less educated individuals. In addition, chronic health conditions, worse self-reported health, and female gender all are associated with higher likelihood of awareness of hypertension.

Table 7 shows findings related to diabetes. Unlike hypertension, measurement of diabetes involves a blood draw, making the first stage of the model (agreed to participate in medical examination) potentially more important. Panel 1 in Table 7 indicates that African-Americans are 8 percentage points less likely than non-Latino whites to participate in the blood draw to measure diabetes. Similarly, individuals from the other race/ethnicity category are about 10 percentage points less likely to participate compared to non-Latino whites. These differences may be due to factors such as mistrust of the US health care and research system, culture, and language. However, Latinos are just as likely to participate as non-Latino whites. Age, employment, female gender, and income are all positively associated with participating, while worse self-reported health is negatively associated with participating. As in the case of hypertension, respondents who were hostile, uncooperative, or concerned about time during the HRS interview were appreciably less likely to participate in a blood draw to measure diabetes.

Panel 2 of Table 7 shows correlates of having diabetes. All minority individuals are more likely to have diabetes than non-Latino whites. Respectively, African-Americans are 13 percentage points more likely, Latinos are 18 percentage points more likely, and individuals in the other race/ethnicity category are 17 percentage points more likely to have diabetes compared to non-Latino whites. As expected, age, other chronic illnesses, worse self-reported health, and obesity are positively associated with diabetes. Higher income is associated with lower likelihood of having diabetes.

Finally, in Panel 3 of Table 7, we examine awareness of diabetes. African-Americans and Latinos are less likely than non-Latino whites to be aware of having diabetes. We find no education or income gradient in awareness of diabetes. Surprisingly, males are more likely than females, and employed individuals are less likely than non-employed individuals to be aware of having diabetes. As expected, individuals who self-report other chronic health conditions and worse overall health are more likely to be aware of having diabetes.

The estimated correlations of errors are shown in the bottom rows in Tables 6 and 7. The results indicate that for hypertension, there is a large, statistically significant, negative correlation between the error terms of the 2^{nd} and 3^{rd} equations (has the disease, aware of the disease). The negative correlation suggests that there are unmeasured factors which decrease the likelihood of having the disease and increase awareness of the disease if one has it. The same is true for diabetes.

For hypertension, there are no statistically significant correlations between the error terms in the first and second equations, and between the error terms in the first and third equations. In the case of diabetes, however, there is a large negative correlation between the error terms in the first and second equations. This finding suggests the existence of unmeasured factors that increase the likelihood of participating in the medical exam (which involves a blood draw in the case of diabetes), and decrease the likelihood of having the disease. The LM statistics show that we can reject null hypothesis that the errors across are uncorrelated at the 1% significance level for both hypertension (LM=168) and diabetes (LM=707).

Next, we examine the impact, if any, of the selection problem on estimates of racial/ethnic disparities in the awareness of hypertension and diabetes. In Table 8 we have consolidated relevant results for $\Omega \neq I$ from Tables 6 and 7, and report comparable estimates that we obtained when assumption $\Omega = I$ was imposed on the estimation inappropriately. We find that the estimation results under the zero correlation assumption ($\Omega = I$) and the non-zero correlation assumption ($\Omega \neq I$) are not significantly different from each other in the 1st and 2nd steps of both the hypertension and the diabetes models. This result is consistent with the estimated correlation matrix, which shows that there is zero correlation between the 1st and 2nd steps and the 1st and 3rd steps in the diabetes and hypertension models.

In the case of hypertension, Latinos are more likely to accept the medical exam than non-Latino whites, and African-Americans and other minorities are not significantly different from non-Latino whites in their acceptance of the medical exam, regardless of whether we assume

Ω=I or $\Omega \neq$I. In either model, African-Americans are more likely to have hypertension compared to non-Latino whites, but Latinos and other minorities are not statistically different from non-Latino whites in disease prevalence. Similarly, assumptions about Ω do not affect estimation of racial/ethnic differences in the acceptance of the medical examination for diabetes, or the likelihood of having diabetes. In both models, African-Americans are less likely than non-Latino whites to accept the medical exam, and all minorities are more likely than non-Latino whites to have diabetes.

However, the findings in Table 8 show that accounting for selection does appear to affect estimates of racial/ethnic disparities in awareness of chronic diseases. Under the assumption, $\Omega \neq$ I, we find no differences across racial/ethnic groups in the awareness of hypertension, but we find that African-Americans and Latinos are less likely to be aware of having diabetes than non-Latino whites. If we had assumed that Ω=I, however, the findings in Table 8 would instead suggest, as prior researchers have reported, African-Americans are significantly more aware of hypertension than non-Latino whites. This assumption of no correlation between the error terms of the equations also leads to the finding that there are no racial/ethnic differences in awareness of diabetes. In sum, the findings in Table 8 indicate that if we ignore the correlation between unmeasured determinants of accepting the medical exam, having the disease, and being aware of the disease, we tend to underestimate racial/ethnic disparities in awareness of hypertension and diabetes.

6.0 An alternative definition of awareness

These estimates of disease awareness are based on a definition of awareness that has been used extensively in prior work. In terms of Table 4, this definition of awareness is simply the *sensitivity* of self-reports, $a/(a+c)$, that is the proportion of individuals who self-report illness in a group that consists of individuals who meet objective criteria for that illness. The definition stems from an epidemiological framework, in which objective measurement of disease would be considered a clinical “gold standard.” From a clinician’s perspective, it makes sense to consider objective measurement to be the first step. Given that a clinician has determined that a person has the illness and is subsequently found to be ignorant of the illness, this suggests that the patient needs intervention since “unaware” patients cannot effectively manage their conditions. Note that this measure completely ignores the *specificity* of self reports, i.e., given that the medical examination has determined that the subject does not have the disease, the probability that the self report concords with the test result.

An alternative definition of awareness may be more appropriate in certain cases, and can help public health officials to target high risk populations for early testing. In particular, an alternate definition of awareness could be the following: conditional on self-reporting not having the disease, a respondent, based on objective measurement, actually does not have the disease, i.e., $b2/(b2+c)$ in Table 4. The proportion of unaware using this definition is $c/(b2+c)$. These two alternative measures of unawareness will give similar estimates if a and $b2$ are similar in magnitude. On the other hand, in cases where $b2$ is considerably bigger (smaller) than a, this alternative measure will yield a smaller (bigger) incidence of unawareness. The model structure underlying this alternate definition of (un)awareness is mapped out in Figure 4, and the incidence of unawareness for different race/ethnic groups based on our sample are reported in Table 9.

As we expect, the incidence of unawareness is uniformly higher for hypertension and lower for diabetes for all groups because the prevalence rate is higher for hypertension than for diabetes. For costly but relatively rare chronic diseases, this alternative measure inappropriately dilutes the unaware percentage by bringing in the much larger healthy group in the denominator. However, what is most remarkable in Table 9 is that with the alternative definition, the percentage unaware is relatively much higher for the minorities compared to non-Latino whites. For instance, the unaware percentage is only 11 percent for non-Latino whites but over 20 percent for African-Americans and Latinos. For hypertension, even though for Latinos it makes no difference, the rate is 23 percent for non-Latino whites but 34 percent for African-Americans.

The intuition behind this result can be derived from the way we slice the data in two alternative sequential patterns as depicted in Figures 3 and 4. Note that prevalence includes both objectively measured disease and/or medication use. In Figure 3, at step 3, a large portion of the aware group (viz. Group A) includes people who are under treatment and take medication to control their illness. By contrast, at step 3 of Figure 4, neither of the two groups is on medication. These are the patients who reported in the survey that they were never told by a doctor of the existence of either hypertension or blood pressure, yet the medical examination determined that a portion of them in fact had the disease. Thus, in the alternative definition of awareness, the people who are on medication are not included. However, in our sample we found (cf. Table 2) that a disproportionate number of African-Americans are on hypertension medication (65% African-Americans, 47% Latinos, and 48% non-Latino whites). With respect to diabetes patients, while only 14% non-Latino whites are on medication, the percentages are 25% for African-Americans and 27% for Latinos. Thus, even though a larger proportion of African-Americans are treated with hypertension medication than non-Latino whites, there are even larger number of undiagnosed African-American patients who are unaware of their disease, presumably because they do not come in contact with the medical establishment.

Regarding diabetes, we find that compared to non-Latino whites, a much larger percentage of both African-Americans and Latinos are on medication, but Table 9 reveals that even then the incidence of unawareness among the remaining minorities is almost double that of non-Latino whites. The fact that a higher percentage of minorities are on medication is simply reflective of the reality that these minorities are more extensively and severely impaired by the chronic diseases, possibly due to initial period unawareness, lack of early treatment, and the resultant progression to the latter stages of the disease. The alternative definition of unawareness leads us to the important realization that even though using the conventional definition of awareness, studies routinely report that African Americans are more aware about hypertension than non-Latino whites, we find this result is misleading about the true level of awareness in the untreated African American population.

From an alternate perspective, it may be more useful to employ a definition of awareness which begins with individuals who *self-report* not having the illness, rather than starting with a group that objectively meets criteria for illness. For example, a public health authority may not have objective measures of illness for a large population. For diabetes, it is well known that a mass unselective population based screening is not cost effective, and that it should be limited to high risk groups, Bishop et al. (1998). Thus, the authority may use a survey to identify individuals who report not having an illness, and then use further inference to determine whether there are individuals in this group who actually have the illness but are unaware of it.

However, these reported differences in unawareness are subject to selection biases. As we have emphasized before, there may be unmeasured factors that are correlated with race/ethnicity and that determine participation, self-reporting not having the illness, and actually having the illness. Thus, in order to establish statistically if minorities are indeed less aware, we re-estimated the sequential probit model based on this alternative structure. The maximum likelihood estimates for hypertension and diabetes are reported in Tables 10 and 11 respectively. The estimates of the first two equations are very similar to those of equations 1 and 3 of the main sequential model. Now the correlation in errors between the participation and awareness equations for hypertension and between medical examination and self reports are highly significant. Among those who participate in the survey and self-report not having hypertension, African-Americans are more likely to actually have the disease compared to non-Latino whites.

In the case of diabetes, all racial/ethnic minority groups are more likely to have the illness compared to non-Latino whites, conditional on self-reporting not having the illness and agreeing to participate in the medical examination. What is noteworthy is that the African American dummy in the awareness equation for hypertension, and all minority dummies in the awareness equation for diabetes are statistically significant, and suggests that these minorities are more unaware of their chronic illnesses compared to non-Latino whites. The marginal effects are substantial, around 10%. These results are consistent with Table 9 and indicate that despite a higher proportion of minorities are treated for hypertension and diabetes, there are even higher proportion minorities who are unaware of their chronic conditions, and needs to be identified and medically treated.

7.0 Conclusions

In this paper, we examine racial/ethnic disparities in the awareness of hypertension and diabetes among older individuals in the US taking advantage of the biomarker data from the 2006 round of HRS that included blood pressure measurement and a blood draw for diabetes. We characterize awareness in the context of a 3-step sequential probit model involving participation in the collection of biomarker data, having a chronic disease and finally being aware of it. The rich socio-economic data available in the core part of the HRS survey together with the results from an almost simultaneous medical examination help us to adjust for a range of individual-level factors including disease severity that may be correlated with both race/ethnicity and awareness. We account for the possibility that factors exist that affect both the likelihood of having a chronic disease and the likelihood of being aware of the illness if it exists, and the censoring resulting from respondents refusing to participate in the collection of health examination and biomarker data.

This paper has two broad messages for public health policy: First, based on a conventional definition of awareness, i.e., conditional on having the disease as determined by medical examination, respondent self-reports that s/he does indeed have the illness, we find that awareness is higher among African-Americans compared to non-Latino whites, which is consistent with previous studies. For diabetes, our raw counts indicate small differences in awareness between racial and ethnic groups, which are also consistent with previous studies. We supplement these estimates with two new tests of concordance and dependence between self report and medical examination in order to establish the statistical significance of the sample estimates.

We, however, show that accounting for multiple selections affect estimates of racial/ethnic disparities in awareness of chronic diseases. Allowing for a non-diagonal correlation matrix of errors in the 3-equation sequential probit model, we find no differences across racial/ethnic groups in the awareness of hypertension, but that African-Americans and Latinos are less likely to be aware of having diabetes than non-Latino whites. Imposing the assumption that the error correlation matrix is diagonal, we obtain the result that African-

Americans are significantly more aware of hypertension than non-Latino whites and that there are no racial/ethnic differences in awareness of diabetes, as prior researchers have reported. In sum, our analysis clearly indicates that if we ignore the correlation between unmeasured determinants of accepting the medical exam, having the disease, and being aware of the disease, we tend to underestimate racial/ethnic disparities in awareness of hypertension and diabetes.

Second, we explore an alternative definition of awareness under which the unaware respondents are the ones who report in the survey that they were never told by a doctor of the existence of a disease, yet the medical examination determined that they in fact have it. Interestingly, we find that with this alternative definition, the percentage unaware is relatively much higher for all minorities compared to non-Latino whites. Under this alternative definition, people who are on medication are logically vetted out while distinguishing between aware and unaware. In our sample, however, a disproportionate number of African-Americans and Latinos are on medication for hypertension and diabetes. Thus, even though a larger proportion of minorities are treated for hypertension and diabetes than non-Latino whites, there are even a larger number of undiagnosed African-American and Latino patients who are unaware of their chronic diseases. The fact that a higher percentage of minorities are on medication simply reflects the reality that these minorities are more extensively and severely impaired. The simulated maximum likelihood estimates of the alternative sequential probit model allowing for selection suggest that the African-Americans are almost 10% less likely to be as aware as non-Latino whites for hypertension, and that all minorities are significantly less aware for diabetes by a similar magnitude. Presumably these minorities do not come in contact with the medical establishment such that they can be aware of their existing medical conditions.

We conclude by noting that even though a higher proportion of minorities are treated for hypertension and diabetes, among those who self-report not having disease, minorities are more likely than non-Latino whites to be unaware of their chronic conditions. Under the conventional definition of awareness, which includes people being treated by medication, the relatively large number of treated minorities gives the impression that they as a group are more aware than non-Latino whites. However, when examining the untreated, we find that significant racial/ethnic disparities persist in the awareness of hypertension and diabetes.

References

Achter, Paul, Roxanne Parrott, and Kami Silk. 2005. "African Americans' opinions about human-genetics research." *Politics and the Life Sciences*, 23(1): 60-66.

Adler, Nancy E., and Judith Stewart. 2010. "Health disparities across the lifespan: Meaning, methods, and mechanisms." *Annals of the New York Academy of Sciences*, 1186: 5-23.

Aizer, Anna, and Laura Stroud. 2010. "Education, knowledge and the evolution of disparities in health." NBER Working Paper No.15840.

Alegria, Margarita, Glorisa Canino, Ruth Ríos, Mildred Vera, José Calderón, Dana Rusch, and Alexander N. Ortega. 2002. "Inequalities in use of specialty mental health services among Latinos, African Americans, and non-Latino whites." *Psychiatric Services*, 53(12): 1547-1555.

Angell, Sonia Y., Renu K. Garg, R. Charon Gwynn, Lori Bash, Lorna E. Thorpe, and Thomas R. Frieden. 2008. "Prevalence, awareness, treatment, and predictors of control of hypertension in New York City." *Circulation, Cardiovascular Quality and Outcomes*, 1: 46-53.

Bahr, Peter R. 2007. "Race and nutrition: an investigation of Black-White differences in health-repeated nutritional behaviours." *Sociology of Health & Illness*, 29(6): 831-856.

Balsa, Ana I., and Thomas G. McGuire. 2003. "Prejudice, clinical uncertainty and stereotyping as sources of health disparities." *Journal of Health Economics*, 22: 89-116.

Balsa, Ana I., Thomas G. McGuire, and Lisa S. Meredith. 2005. "Testing for statistical discrimination in health care." *Health Services Research*, 40(1): 227-252.

Benítez-Silva, Hugo, Moshe Buchinsky, Hiu Man Chan, Sofia Cheidvasser, and John Rust. 2004. "How large is the bias in self-reported disability?" *Journal of Applied Econometrics*, 19: 649-670.

Bishop, Donald B., Bruce R. Zimmerman, and Jon S. Roesler. 1998. "Chapter 14. Diabetes." In *Chronic Disease Epidemiology and Control*, ed. Ross C. Brownson, Patrick L. Remington, and James R. Davis, 421-464. 2nd ed. Washington, DC.: American Public Health Association.

Buell, Catherine, Duclie Kermah, and Mayer B. Davidson. 2007. "Utility of A1C for diabetes screening in the 1999-2004 NHANES Population." *Diabetes Care*, 30 (9): 2233-2235.

Caillier, James G. 2006. "Race, gender, and cardiovascular disease: Do disparities exist at hospitals that serve majority black populations when patients present with ischemic heart disease and myocardial infarction?" *Journal of Cultural Diversity*, 13(4): 202-207.

Centers for Disease Control and Prevention. 2003. "Prevalence of diabetes and impaired fasting glucose in adults-United States, 1999-2000." *Morbidity and Mortality Weekly Report*. 52 (September (35)), 833-837.

Chandra, Amitabh, and Jonathan Skinner. 2003. "Geography and racial health disparities." NBER Working Paper No.9513.

Chobanian, Aram V., George L. Bakris, Henry R. Black, William C. Cushman, Lee A. Green, Joseph L. Izzo, Jr., Daniel W. Jones, Barry J. Materson, Suzanne Oparil, Jackson T. Wright, Jr., Edward J. Roccella, and the National High Blood Pressure Education Program Coordinating Committee. 2003. "The seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure (JNC 7)" *Hypertension*, 42: 1206-1252.

Cutler, David M., and Adriana Lleras-Muney. 2010. "Understanding differences in health behaviors by education." *Journal of Health Economics*, 29: 1-28.

Danaei, Goodarz, Ari B. Friedman, Shefali Oza, Christopher JL. Murray, and Majid Ezzati. 2009. "Diabetes prevalence and diagnosis in US states: analysis of health surveys." *Population Health Metrics*, 7(16).

Ginde, Adit A., Enrico Cagliero, David M. Nathan, and Carlos A. Camarga, Jr. 2008. "Value of risk stratification to increase the predictive validity of HbA1C in screening for undiagnosed diabetes in the US population." *Journal of General Internal Medicine*, 23(9): 1346-1353.

Goldman, Dana P., and Darius N. Lakdawalla. 2005. "A theory of health disparities and medical technology." *Contributions to Economic Analysis & Policy*, 4(1). Article 8.

Gray, Bradford H., Mark Schlesinger, Shannon M. Siegfried, and Emily Horowitz. 2009. "Racial and ethnic disparities in the use of high-volume hospitals." *Inquiry*, 46: 322-338.

Greene, William H., and David A. Hensher. 2010. *Modeling ordered choices: A primer.* Cambridge, MA: Cambridge University Press.

Hajivassiliou,Vassilis, Daniel McFadden, and Paul Ruud. 1996. "Simulation of multivariate normal rectangle probabilities and their derivatives theoretical and computational results." *Journal of Econometrics*, 72(1-2): 85-134.

Halbert, Chanita H., Benita Weathers, Ernestine Delmoor, Brandon Mahler, James Coyne, Hayley S. Thompson, Thomas T. Have, David Vaughn, S. Bruce Malkowicz, and David Lee. 2009. "Racial differences in medical mistrust among men diagnosed with prostate cancer." *Cancer*, June: 2553-2561.

Harding, Don, and Adrian Pagan. 2002. "Dissecting the cycle: a methodological investigation." *Journal of Monetary Economics*, 49: 365-381.

Health and Retirement Study, 2006 Biomarker data. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI, (2006).

Hertz, Robin P., Alan N. Unger, Jeffrey A. Cornell, and Elijah Saunders. 2005. "Racial disparities in hypertension prevalence, awareness, and management." *Archives of Internal Medicine*, 165: 2098-2104.

Howard, George, Ron Prineas, Claudia Moy, Mary Cushman, Martha Kellum, Ella Temple, Andra Graham, and Virginia Howard. 2006. "Racial and geographic differences in awareness, treatment, and control of hypertension: The REasons for Geographic and racial differences in stroke study." *Stroke*, 37(5): 1171-1178.

Johnston, David W., Carol Propper, and Michael A. Shields. 2009. "Comparing subjective and objective measures of health: Evidence from hypertension for the income/health gradient." *Journal of Health Economics*, 28: 540-552.

Kiefer, Nicholas M. 1982. "Testing for dependence in multivariate probit models." *Biometrika*, 69 (1): 161-166.

LaVeist, Thomas A., Kim J. Nickerson, and Janice V. Bowie. 2000. "Attitudes about racism, medical mistrust, and satisfaction with care among African American and White cardiac patients." *Medical Care Research and Review*, 57 Suppl1: 146-161.

LêCook, Benjamin, Thomas G. McGuire, Kari Lock, and Alan M. Zaslavsky. 2010. "Comparing methods of racial and ethnic disparities measurement across different settings of mental health care." *Health Services Research*, 45(3): 825-847.

Meara, Ellen. 2001. "Why is health related to socioeconomic status? The case of pregnancy and low birth weight." NBER Working Paper No.8231.

Morenoff, Jeffrey D., James S. House, Ben B. Hansen, David R. Williams, George A. Kaplan, and Haslyn E. Hunte. 2007. "Understanding social disparities in hypertension prevalence, awareness, treatment, and control: The role of neighborhood context." *Social Science & Medicine* 65: 1853-1866.

National Healthcare Disparities Report, 2003. Full Report. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/qual/nhdr03/fullreport/

Ong, Kwok Leung, Bernard M.Y. Cheung, Yu Bun Man, Chu Pak Lau, and Karen S.L. Lam. 2007. "Prevalence, awareness, treatment and control of hypertension among United States adults 1999-2004." *Hypertension*, 49: 69-75. http://hyper.ahajournals.org/cgi/content/full/49/1/69.

Pierce, Mary B., Paola Zaninotto, Nicholas Steel, and Jennifer Mindell. 2009. "Undiagnosed diabetes - data from the English longitudinal study of ageing." *Diabetic Medicine*, 26: 679-685.

Primatesta, Paola, and Neil R. Poulter. 2006. "Improvement in hypertension management in England: results from the Health Survey for England 2003." *Journal of Hypertension*, 24(6): 1187-1192.

Figure 1: Prevalence of hypertension by race/ethnicity and age

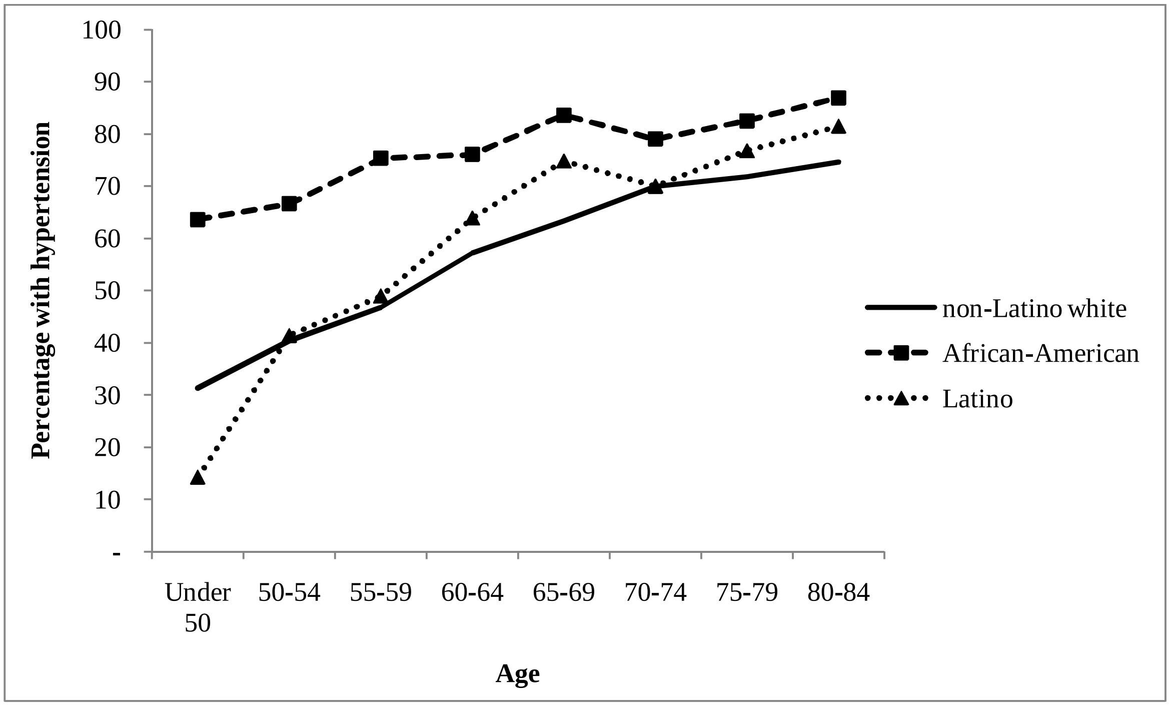

Figure 2: Prevalence of diabetes by race/ethnicity and age

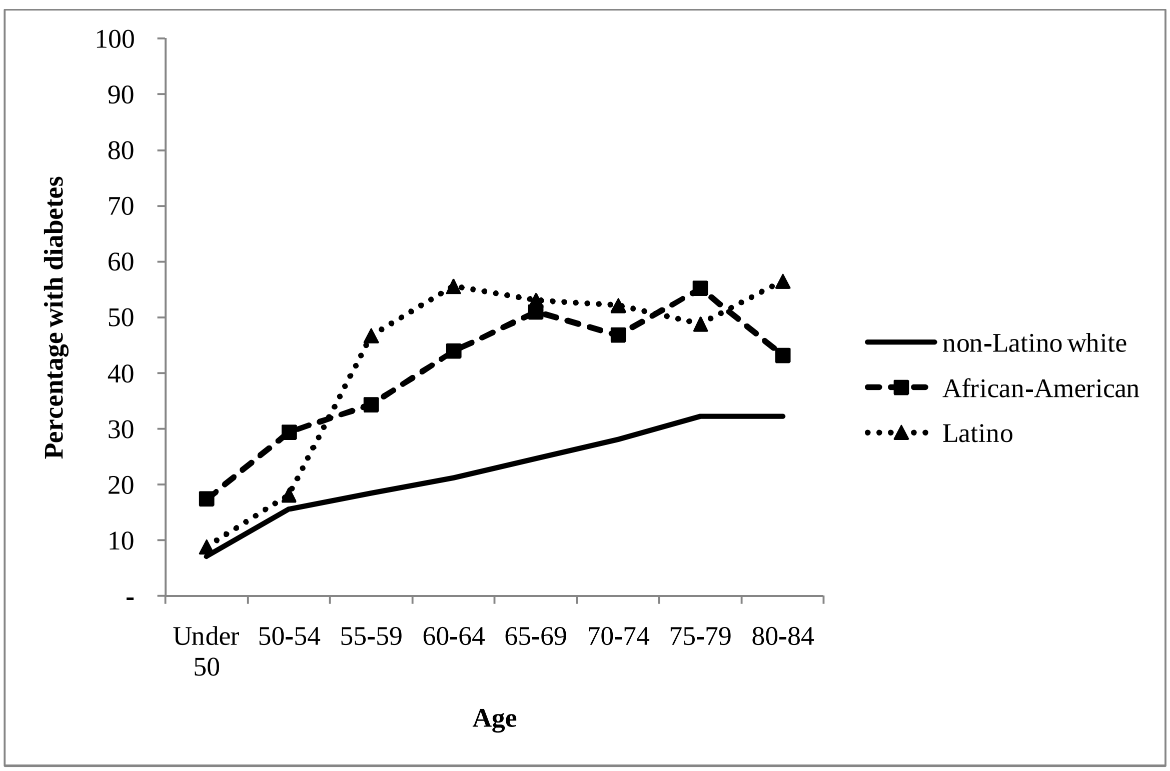

Figure 3: Flow chart depicting awareness of chronic diseases, 2006 HRS

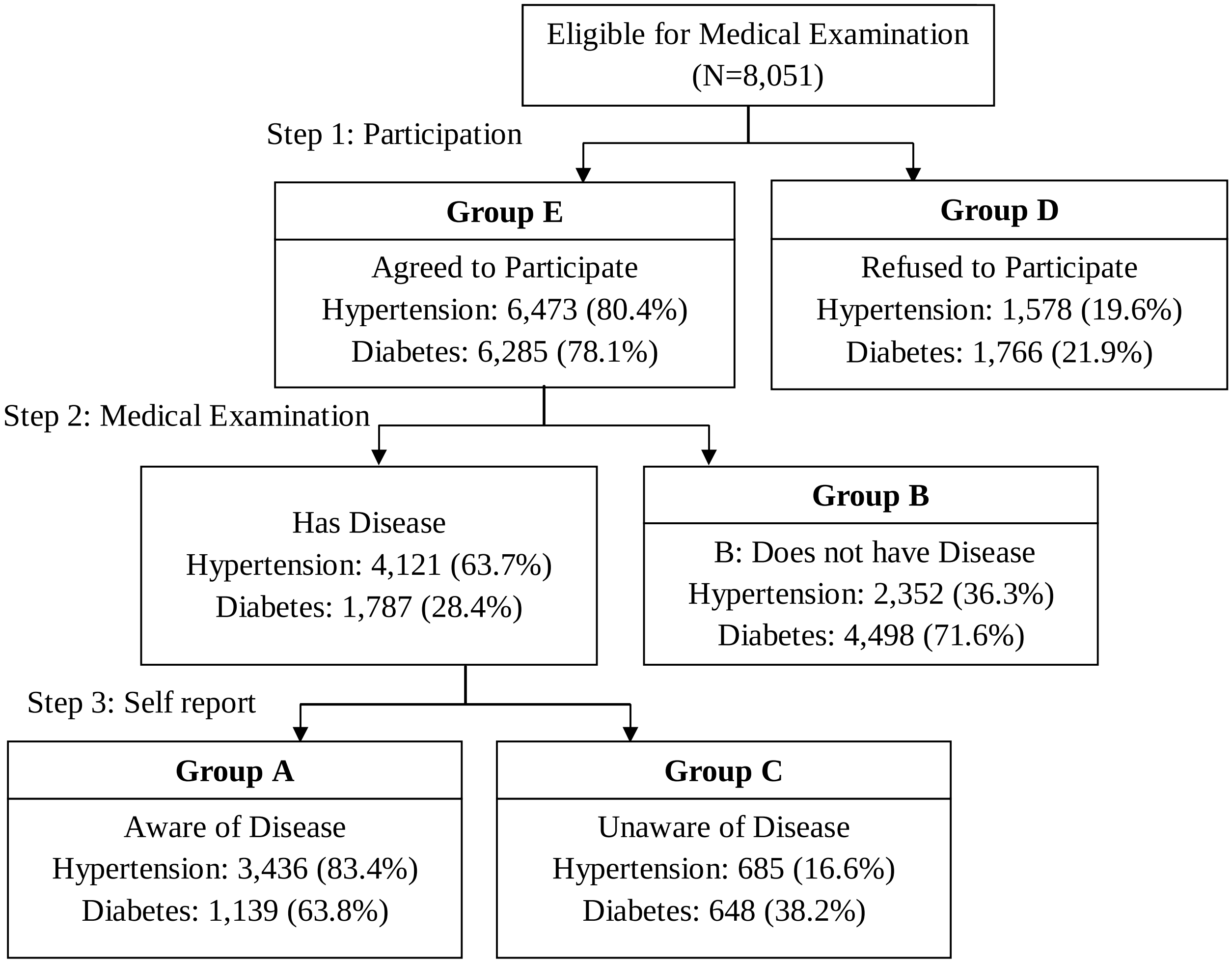

Figure 4: Flow chart for an alternative model of chronic disease awareness, 2006 HRS

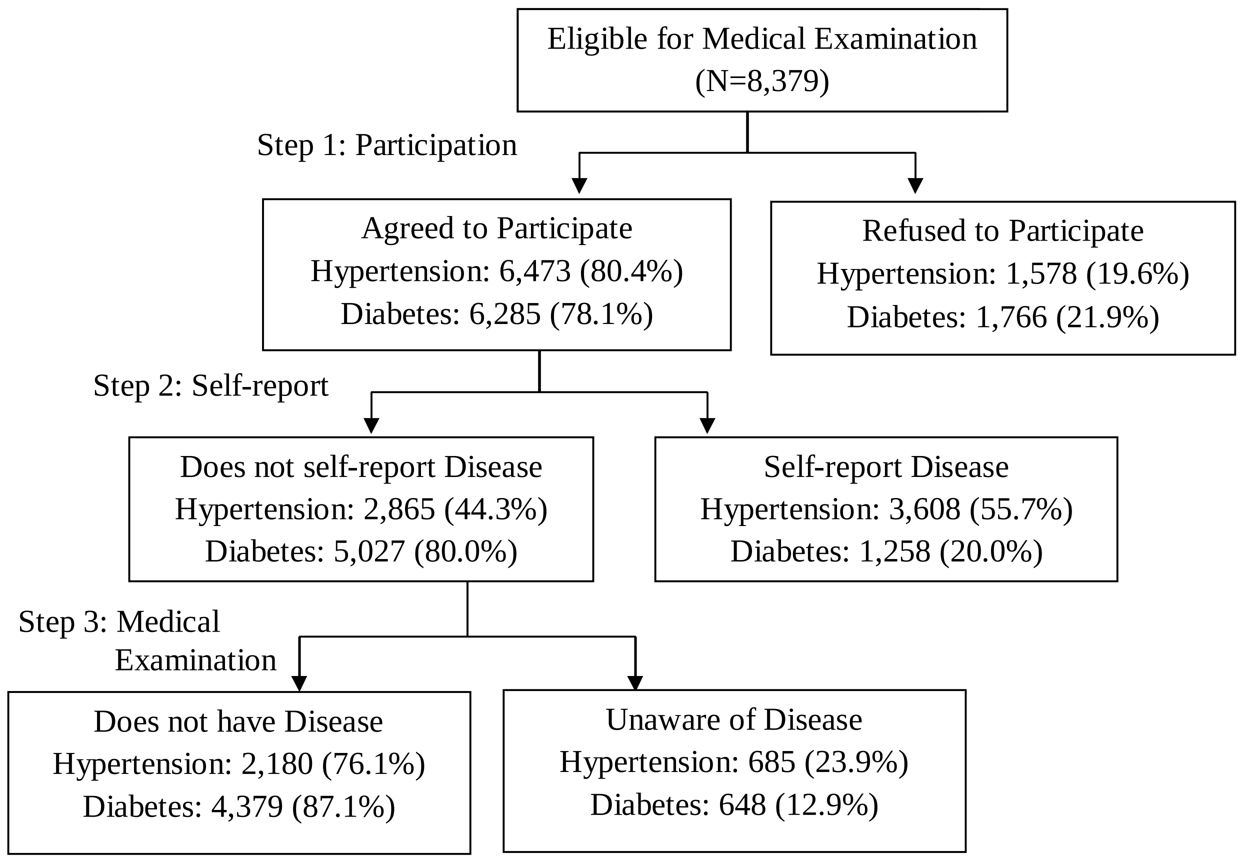

Table 1: Sample Characteristics (N = 8,051)

	(1) Total	Hypertension			Diabetes		
		(2) Agreed	(3) Refused	(4) t-test	(5) Agreed	(6) Refused	(7) t-test
Non-Latino white	75.7	76.3	73.3	**-2.48**	77.5	69.4	**-7.00**
African-American	13.8	13.0	17.1	**4.29**	12.2	19.5	**7.98**
Latino	8.2	8.5	7.0	**-1.94**	8.2	7.9	-0.43
Other race	2.3	2.3	2.6	0.77	2.1	3.1	**2.45**
Male	41.2	41.4	40.3	-0.83	40.7	43.1	**1.80**
Average Age	67.38	67.51	66.87	**-2.11**	67.12	68.32	**4.12**
High school dropout	22.0	21.2	25.2	**3.48**	21.1	25.3	**3.77**
High school graduate	33.8	33.8	33.5	-0.23	34.1	32.7	-1.11
Some college	22.1	22.4	20.9	-1.32	22.5	21.1	-1.24
College graduate or more	22.1	22.5	20.3	**-1.89**	22.4	21.0	-1.26
Married	64.2	65.6	58.5	**-5.30**	65.6	59.1	**-5.07**
Currently Employed	34.6	34.8	34.1	-0.51	36.1	29.6	**-5.09**
Household Income $	69,922	72,123	60,894	**-1.79**	72,913	59,276	**-2.27**
Self-report hypertension	56.4	55.7	58.9	**2.25**	55.8	58.5	**2.05**
Self-report diabetes	19.9	19.5	21.7	**1.98**	20.0	19.4	-0.55
Use cholesterol medication	39.0	39.6	36.6	**-2.16**	38.7	40.1	1.06
SAH=Excellent	11.5	12.1	8.9	**-3.64**	12.2	9.1	**-3.63**
SAH=Very Good	30.3	31.3	25.9	**-4.26**	31.3	26.4	**-3.96**
SAH=Good	30.7	30.7	30.8	0.07	30.8	30.4	-0.39
SAH=Fair	20.4	19.7	23.4	**3.30**	19.2	24.9	**5.21**
SAH=Poor	7.1	6.1	11.0	**6.86**	6.4	9.3	**4.12**
Overweight	33.5	34.4	29.5	**-3.75**	33.7	32.6	-0.85
Obese	28.2	26.8	33.7	**5.48**	28.4	27.1	-1.09
No BMI information	11.0	10.6	12.6	**2.35**	10.4	13.0	**3.07**
Private insurance only	29.9	29.9	30.0	0.08	30.5	27.9	**-2.05**
Medicare only	20.5	20.6	20.2	-0.36	19.6	23.7	**3.76**
Medicare & Private	32.3	32.9	29.7	**-2.43**	32.9	30.2	**-2.14**
Other insurance	16.1	15.4	19.0	**3.53**	15.9	16.7	0.80
Northeast	15.6	15.9	14.4	-1.44	14.8	18.6	**3.94**
Midwest	24.6	24.2	26.2	**1.66**	25.2	22.5	**-2.30**
South	38.5	38.8	37.4	0.72	38.4	38.9	-1.55
West	21.3	21.1	21.9	-1.00	21.6	19.9	0.40
Current smoker	14.0	11.6	24.3	**13.2**	13.6	14.7	0.92
Hostile	6.6	5.0	13.2	**11.9**	4.9	12.5	**11.32**
Uncooperative	3.1	1.7	9.2	**15.6**	1.5	9.1	**16.56**
Concerned about time	14.1	12.7	19.8	**7.3**	12.5	19.7	**7.69**

Note: Entries show mean percentages (except for household income and age).

Table 2: Prevalence, awareness and treatment of chronic diseases by race (%)

		Total	Non-Latino white	African-American	Latino	Other Race
Hypertension	**"Agreed to Participate" Sample (N=6,473)**					
	Self-Reported Hypertension	**55.7** [52.4] (0.0062)	**53.4** (0.0071)	**71.8**[10.0] (0.0156)	**54.4** [0.5] (0.0213)	**49.7**[-0.9] (0.0414)
	Objective Hypertension	**31.9** [30.8] (0.0058)	**30.5** (0.0066)	**40.2** [5.6] (0.0169)	**33.6** [1.5] (0.0202)	**26.5**[-1.0] (0.0365)
	Medication for Hypertension	**50.2** [46.7] (0.0062)	**48.1** (0.0071)	**65.4** [9.4] (0.0164)	**46.5**[-0.7] (0.0213)	**46.3**[-0.4] (0.0413)
	Objective or Medication	**63.7** [60.8] (0.0060)	**61.6** (0.0069)	**78.4** [9.5] (0.0142)	**61.1**[-0.2] (0.0208)	**57.8**[-0.9] (0.0409)
	"Has disease" (Objective measurement or Medication) Sample (N=4,121)					
	Awareness	**83.4** [81.7] (0.0058)	**82.5** (0.0069)	**87.8** [3.4] (0.0128)	**83.3** [0.4] (0.0204)	**82.3**[-0.0] (0.0416)
	Medication	**78.8** [76.9] (0.0064)	**78.0** (0.0075)	**83.4** [3.1] (0.0145)	**76.1** [-0.8] (0.0233)	**80.0** [0.4] (0.0436)
Diabetes	**"Agreed to Participate" Sample (N=6,285)**					
	Self-Reported Diabetes	**20.0** [18.2] (0.0050)	**17.4** (0.0054)	**28.8** [9.8] (0.0164)	**30.5** [8.1] (0.0202)	**24.8**[1.9] (0.0376)
	Objective measurement	**24.3** [22.3] (0.0054)	**20.4** (0.0058)	**37.3** [7.6] (0.0175)	**39.6** [9.6] (0.0215)	**32.3**[2.4] (0.0407)
	Medication for Diabetes	**16.5** [15.0] (0.0047)	**13.9** (0.0050)	**25.1**[10.6] (0.0157)	**27.4** [9.0] (0.0204)	**21.1**[2.2] (0.0355)
	Objective or Medication	**28.4** [26.0] (0.0057)	**24.4** (0.0061)	**42.7**[10.7] (0.0179)	**43.6** [9.6] (0.0218)	**36.8**[3.3] (0.0420)
	"Has disease" (Objective measurement or Medication) Sample (N=1,787)					
	Awareness	**63.7** [63.0] (0.0114)	**62.9** (0.0140)	**64.1** [0.4] (0.0266)	**67.3** [1.2] (0.0313)	**65.3**[0.3] (0.0687)
	Medication	**58.2** [57.8] (0.0117)	**57.2** (0.0144)	**58.9** [0.6] (0.0273)	**62.8** [1.6] (0.0322)	**57.1**[-0.0] (0.0714)

Notes: For total sample in the first column, weighted percentages are in square brackets. In the 3rd, 4th and 5th column, t-test statistics are in square brackets. Baseline for the t-statistics is non-Latino white group. Standard errors are in parentheses.

Table 3: Comparison of prevalence and awareness rates from HRS and NHANES

		2006 HRS*	NHANES**
Hypertension	Has Disease	60.8	66.3
	Aware of Disease	81.7	81.0
Diabetes	Has Disease***	26.0	23.6
	Aware of Disease	63.0	71.6

*We use Physical Measure weight for Hypertension and Biomarker weight for Diabetes which are provided by 2006 HRS tracker file.

**For hypertension, estimates come from Ong et al. (2007) which used 2003-2004 NHANES. For diabetes, Danaei et al. (2009) provide these estimates based on 2003-2006 NHANES. Both these studies use individuals over 60 years old. Thus, the survey window and the age distribution in HRS and NHANES do not exactly match.

***For diabetes, the definition of "has disease" is self-reported usage of insulin or other medication for controlling diabetes or A1C>6% in the 2006 HRS. Danaei et al. (2009), however, use an alternative definition that includes self-reported diagnosed diabetes or FPG(Fasting Plasma Glucose level)>126mg/dL.

Table 4: Concordance Table for "Has Disease" and "Self-Reported Disease"

		Medical Examination - Disease	
		1: yes	0: no
Self-Reported Disease	1: yes	*a*	*b1*
	0: no	*c*	*b2*

Note: The sum of b1 and b2 equals b in the text.

Table 5: Dependency tests between "has disease" and "self-reported disease" by race/ethnicity

		Sample Size	χ^2 test (χ_1^2)	Index of Concordance			OLS Regression Test	
				$\hat{I}$	t-statistics $H_0: \hat{I} = 1$ $H_1: \hat{I} \neq 1$	t-statistics $H_0: \hat{I}_w = \hat{I}_M$ $H_0: \hat{I}_w > \hat{I}_M$	Coeff.	z
Hyper-tension	Non-Latino white	4,939	2,700	0.8663 (0.0042)	**-27.60****	Reference	Reference	
	African American	839	390	0.8760 (0.0114)	**-10.89****	-0.7647	-0.0159 (0.0133)	-1.20
	Latino	548	290	0.8631 (0.0147)	**-9.31****	0.2106	-0.0151 (0.0160)	-0.94
	Other Race	147	-	-	-	-	-0.0010 (0.0297)	-0.03
	Constant	-	-	-	-	-	**0.0826**** (0.0050)	16.35
	Total	6,473	3,500	0.8676 (0.0042)	**-31.43****	-	6,473	
	R-Squared	-	-	-	-	-	0.0477	
Diabetes	Non-Latino white	4,870	2,300	0.8889 (0.0045)	**-24.66****	Reference	Reference	
	African-American	801	346	0.8324 (0.0135)	**-12.39****	**4.4918****	**0.0691**** (0.0131)	5.26
	Latino	518	255	0.8456 (0.0159)	**-9.72****	**2.9393****	**0.0617**** (0.0156)	3.95
	Other Race	196	-	-	-	-	0.0507 (0.0297)	1.71
	Constant	-	-	-	-	-	**0.0696**** (0.0048)	14.37
	Total	6,285	3,000	0.8780 (0.0041)	**-29.55****	-	6,285	
	R-Squared	-	-	-	-	-	0.0641	

Notes: * P<0.05, ** P<0.01. $\hat{I}_M$ is concordance index of minorities. Standard errors are in parentheses.

Table 6: Effect of socio-demographic factors and selected risk factors on participation, prevalence, and awareness of hypertension

	Panel 1 (N=8,051) Agreed to Participate		Panel 2 (N=6,473) Has Disease		Panel 3 (N=4,121) Aware of Disease	
African-American	-0.0545 (0.0488)	[-0.0148]	**0.4122**** (0.0585)	[0.1428]	0.1377 (0.0944)	[0.0435]
Latino	**0.1647**** (0.0679)	[0.0413]	-0.0339 (0.0700)	[-0.0158]	0.0085 (0.1026)	[0.0020]
Other Race	-0.0785 (0.1090)	[-0.0217]	-0.0262 (0.1106)	[-0.0082]	-0.0764 (0.1631)	[-0.0204]
Male	0.0272 (0.0351)	[0.0072]	0.0371 (0.0361)	[0.0132]	**-0.3387**** (0.0525)	[-0.0768]
Age	**0.3850*** (0.1728)	[0.0998]	**0.8141**** (0.2031)	[0.0828]	0.3763 (0.3201)	[0.1334]
Age squared	**-0.3886*** (0.1675)	[-0.1008]	**-0.5622**** (0.1978)	[-0.0858]	-0.4362 (0.3002)	[-0.1351]
Some college	0.0494 (0.0535)	[0.0130]	-0.0826 (0.0559)	[-0.0319]	**0.1383*** (0.0797)	[0.0268]
College or more	-0.0315 (0.0563)	[-0.0085]	**-0.2054**** (0.0595)	[-0.0771]	**0.1502*** (0.0829)	[0.0228]
Self-reported diabetes	**-0.0805*** (0.0438)	[-0.0220]	**0.2817**** (0.0489)	[0.1021]	**0.2661**** (0.0766)	[0.0649]
Uses cholesterol medication	**0.0624*** (0.0365)	[0.0166]	**0.4032**** (0.0374)	[0.1452]	**0.5278**** (0.0797)	[0.1378]
SAH=Very Good	-0.0541 (0.0602)	[-0.0146]	**0.2285**** (0.0567)	[0.0843]	**0.2282*** (0.0998)	[0.0573]
SAH=Good	**-0.1371*** (0.0608)	[-0.0375]	**0.2495**** (0.0590)	[0.0933]	**0.3909**** (0.1089)	[0.0896]
SAH=Fair	**-0.2060**** (0.0668)	[-0.0580]	**0.2617**** (0.0684)	[0.0977]	**0.5198**** (0.1206)	[0.1047]
SAH=Poor	**-0.4051**** (0.0821)	[-0.1241]	**0.3122**** (0.0979)	[0.1165]	**0.5809**** (0.1560)	[0.0990]
Obese	-	-	**0.5651**** (0.0483)	[0.1958]	-	-
Overweight			**0.2058**** (0.0424)	[0.0753]		
Hostile	**-0.3589**** (0.0653)	[-0.1086]	-		-	
Uncooperative	**-0.7833**** (0.0887)	[-0.2668]				
Concerned about time	**-0.2154**** (0.0456)	[-0.0614]				
ρ	ρ_{12}=0.1359 (0.1847)		ρ_{13}=-0.1686 (0.2972)		ρ_{23}= **-0.5010****(0.1938)	

Notes: *P<0.05, **P<0.01, Standard errors are in parentheses and marginal effects are in square bracket.
Coefficients on marital status, employment status, insurance status, region, smoking, and missing BMI are not shown in table.

Table 7: Effect of socio-demographic factors and selected risk factors on participation, prevalence, and awareness of diabetes

	Panel 1(N=8,051) Agreed to Participate		Panel 2 (N=6,285) Has Disease		Panel 3 (N=1,787) Aware of Disease	
African-American	-0.2718** (0.0479)	[-0.0843]	0.4035** (0.0542)	[0.1263]	-0.2793** (0.0946)	[-0.0399]
Latino	-0.0065 (0.0645)	[-0.0019]	0.4755** (0.0691)	[0.1798]	-0.1994* (0.1088)	[0.0034]
Other Race	-0.3012** (0.1015)	[-0.0963]	0.5069** (0.1183)	[0.1679]	-0.2718 (0.1896)	[-0.0181]
Male	-0.0883** (0.0338)	[-0.0256]	0.0594 (0.0379)	[0.0137]	0.1209* (0.0677)	[0.0673]
Age	0.3339* (0.1734)	[0.0933]	0.9096** (0.2391)	[0.2079]	0.0349 (0.4435)	[0.2246]
Age squared	-0.3891* (0.1675)	[-0.1087]	-0.7536** (0.2305)	[-0.2214]	-0.1851 (0.4239)	[-0.2481]
Currently Employed	0.1320** (0.0453)	[0.0374]	0.0086 (0.0495)	[0.0127]	-0.1734* (0.0860)	[-0.0791]
Self-reported Hypertension	0.0218 (0.0350)	[0.0063]	0.1952** (0.0398)	[0.0679]	0.1616* (0.0797)	[0.1168]
Use cholesterol medication	-0.0096 (0.0351)	[-0.0028]	0.3864** (0.0397)	[0.1347]	0.3077** (0.0930)	[0.2206]
SAH=Very Good	-0.0544 (0.0589)	[-0.0158]	0.2660** (0.0721)	[0.0896]	0.1583 (0.1551)	[0.1191]
SAH=Good	-0.1090* (0.0599)	[-0.0320]	0.5134** (0.0728)	[0.1764]	0.3217* (0.1802)	[0.2344]
SAH=Fair	-0.2272** (0.0651)	[-0.0689]	0.6252** (0.0775)	[0.2157]	0.4340* (0.2003)	[0.2840]
SAH=Poor	-0.2646** (0.0810)	[-0.0830]	0.6745** (0.0956)	[0.2395]	0.5246* (0.2264)	[0.2780]
Log of Income	0.0451* (0.0224)	[0.0126]	-0.0876** (0.0259)	[0.0137]	-0.0008 (0.0459)	[-0.0207]
Obese		-	0.6019** (0.0560)	[0.1507]		-
Overweight			0.2383** (0.0471)	[0.0275]		
Hostile	-0.2959** (0.0645)	[-0.0937]		-		-
Uncooperative	-0.8972** (0.0883)	[-0.3231]				
Concerned about time	-0.2575** (0.0445)	[-0.0796]				
ρ	ρ_{12}=-0.4708** (0.1582)		ρ_{13}=0.2650 (0.2356)		ρ_{23}=-0.6783** (0.1160)	

Notes: *P<0.05, **P<0.01, Standard errors are in parentheses and marginal effects are in square bracket.
Coefficients on education, region, marital status, smoking, insurance status and missing BMI are not shown.

Table 8: Racial/ethnic disparities in "Agreed to Participate", "Has Disease" and "Aware of Disease" outcomes under non-zero and zero error correlation assumption

Hypertension

	Non-zero correlation $\Omega \neq I$			Zero correlation $\Omega = I$		
	1st Step: Agreed to Participate	2nd Step: Has Disease	3rd Step: Aware of Disease	1st Step: Agreed to Participate	2nd Step: Has Disease	3rd Step: Aware of Disease
African-American	-0.0545 (0.0488)	**0.4122**** (0.0585)	0.1377 (0.0944)	-0.0531 (0.0488)	**0.4260**** (0.0576)	**0.2508**** (0.0838)
Latino	**0.1647**** (0.0679)	-0.0339 (0.0700)	0.0085 (0.1026)	**0.1649**** (0.0679)	-0.0415 (0.0683)	0.0190 (0.1056)
Other Race	-0.0785 (0.1090)	-0.0262 (0.1106)	-0.0764 (0.1631)	-0.0790 (0.1090)	-0.0273 (0.1110)	-0.1091 (0.1717)

Notes: *P<0.05, **P<0.01, Standard errors are in parentheses. All variables which are considered in Table 4 are controlled.

Diabetes

	Non-zero correlation $\Omega \neq I$			Zero correlation $\Omega = I$		
	1st Step: Agreed to Participate	2nd Step: Has Disease	3rd Step: Aware of Disease	1st Step: Agreed to Participate	2nd Step: Has Disease	3rd Step: Aware of Disease
African-American	**-0.2718**** (0.0479)	**0.4035**** (0.0542)	**-0.2793**** (0.0946)	**-0.2748**** (0.0479)	**0.3649**** (0.0552)	-0.0819 (0.0961)
Latino	-0.0065 (0.0645)	**0.4755**** (0.0691)	**-0.1994*** (0.1088)	-0.0080 (0.0645)	**0.5026**** (0.0689)	0.0100 (0.1150)
Other Race	**-0.3012**** (0.1015)	**0.5069**** (0.1183)	-0.2718 (0.1896)	**-0.3039**** (0.1015)	**0.4688**** (0.1228)	-0.0994 (0.2072)

Notes: *P<0.05, **P<0.01, Standard errors are in parentheses. All variables which are considered in Table 5 are controlled.

Table 9: Estimates of unawareness based on alternative definitions (%)

		Non-Latino whites	African-Americans	Latino	All
Main definition *c/(c+a)*	Hypertension	18	12	17	17
	Diabetes	37	36	33	36
Alternative Definition c/(*c+b2)*	Hypertension	23	34	22	24
	Diabetes	11	22	21	13

Note: See table 4 for the definitions of *a*, *b2* and *c*.

Table 10: Effect of socio-demographic factors and selected risk factors on participation, self-report, and unawareness of hypertension: Alternative Sequential Model

	Agreed to Participate (1), N=8,051		Doesn't self-report Disease (2), N=6,473		Unaware of Disease (3), N=2,865	
African-American	-0.0279 (0.0493)	[-0.0075]	**-0.4212**** (0.0576)	[-0.1589]	**0.3672**** (0.0958)	[0.0809]
Latino	**0.1685**** (0.0682)	[0.0420]	0.0383 (0.0696)	[0.0171]	-0.0494 (0.1032)	[-0.0171]
Other Race	-0.0854 (0.1097)	[-0.0235]	0.0793 (0.1108)	[0.0303]	-0.0861 (0.1560)	[-0.0170]
Male	0.0274 (0.0353)	[0.0073]	**0.1558**** (0.0360)	[0.0618]	**0.2205**** (0.0686)	[0.0930]
Age	**0.4021*** (0.1734)	[0.1037]	**-0.7644**** (0.2002)	[0.1046]	**0.5757*** (0.2862)	[0.0899]
Age squared	**-0.4285*** (0.1682)	[-0.1105]	**0.5975**** (0.1945)	[-0.1126]	-0.3162 (0.2790)	[-0.0229]
Some college	0.0468 (0.0538)	[0.0123]	-0.0115 (0.0556)	[-0.0040]	**-0.1444*** (0.0836)	[-0.0501]
College or more	-0.0523 (0.0569)	[-0.0141]	**0.1037*** (0.0587)	[0.0404]	**-0.2971**** (0.0858)	[-0.0829]
Self-reported diabetes	-0.0362 (0.0447)	[-0.0097]	**-0.3846**** (0.0471)	[-0.1471]	0.1365 (0.1059)	[0.0013]
Use cholesterol medication	**0.0699*** (0.0367)	[0.0185]	**-0.5547**** (0.0367)	[-0.2126]	0.0311 (0.1169)	[-0.0562]
SAH=Very Good	-0.0289 (0.0606)	[-0.0077]	**-0.3390**** (0.0585)	[-0.1312]	-	-
SAH=Good	-0.0984 (0.0612)	[-0.0266]	**-0.4559**** (0.0607)	[-0.1748]		
SAH=Fair	**-0.1652**** (0.0669)	[-0.0458]	**-0.5361**** (0.0691)	[-0.2006]		
SAH=Poor	**-0.3592**** (0.0827)	[-0.1082]	**-0.6205**** (0.0953)	[-0.2220]		
Current Smoker	**-0.5100**** (0.0471)	[-0.1565]	**0.1502*** (0.0679)	[0.0525]	-0.0443 (0.0963)	[0.0211]
Obese	**-0.2772**** (0.0479)	[-0.0776]	**-0.4630**** (0.0537)	[-0.1798]	**0.4934**** (0.0916)	[0.1330]
Overweight	**-0.0825**** (0.0455)	[-0.0222]	**-0.2006**** (0.0439)	[-0.0794]	**0.1640**** (0.0658)	[0.0356]
Hostile	**-0.3622**** (0.0651)	[-0.1094]	-	-	-	-
Uncooperative	**-0.7876**** (0.0890)	[-0.2678]				
Concerned about time	**-0.2184**** (0.0457)	[-0.0621]				
ρ	ρ_{12}=-0.0787 (0.1922) ρ_{13}=0.2846 (0.2769) ρ_{23}=**-0.5194**** (0.1840)					

Notes: *P<0.05, **P<0.01, Standard errors are in parentheses and marginal effects are in square bracket.
Coefficients on marital status, employment status, insurance status, regions and missing BMI are not shown in table.

Table 11: Effect of socio-demographic factors and selected risk factors on participation, self-report, and unawareness of diabetes: Alternative Sequential Model

	Agreed to Participate (1), N=8,051		Doesn't self-report Disease (2), N=6,285		Unaware of Disease (3), N=5,027	
African-American	**-0.2718**** (0.0482)	[-0.0864]	**-0.2033**** (0.0666)	[-0.0442]	**0.4513**** (0.0704)	[0.0980]
Latino	-0.0082 (0.0646)	[-0.0024]	**-0.3180**** (0.0747)	[-0.0865]	**0.4309**** (0.0965)	[0.1155]
Other Race	**-0.3011**** (0.1011)	[-0.0962]	**-0.3379**** (0.1344)	[-0.0838]	**0.4486**** (0.1480)	[0.1003]
Male	**-0.0880**** (0.0341)	[-0.0255]	**-0.1164**** (0.0429)	[-0.0260]	-0.0235 (0.0504)	[-0.0089]
Age	**0.3309*** (0.1737)	[0.0924]	**-0.9709**** (0.2585)	[0.1017]	**0.6471*** (0.2895)	[0.1647]
Age squared	**-0.3825*** (0.1680)	[-0.1068]	**0.9198**** (0.2776)	[-0.1142]	-0.4428 (0.2759)	[-0.1230]
Self-reported Hypertension	0.0142 (0.0351)	[0.0041]	**-0.3448**** (0.0457)	[-0.0815]	0.0600 (0.0580)	[0.0156]
Use cholesterol medication	-0.0128 (0.0351)	[-0.0037]	**-0.5449**** (0.0420)	[-0.1387]	0.0338 (0.0732)	[0.0103]
SAH=Very Good	-0.0387 (0.0591)	[-0.0112]	**-0.2751**** (0.0916)	[-0.0655]	-	-
SAH=Good	-0.0860 (0.0602)	[-0.0251]	**-0.6258**** (0.0903)	[-0.1648]		
SAH=Fair	**-0.2076**** (0.0654)	[-0.0627]	**-0.8242**** (0.0954)	[-0.2429]		
SAH=Poor	**-0.2524**** (0.0814)	[-0.0788]	**-0.9264**** (0.1122)	[-0.3008]		
No BMI Information	**-0.1191*** (0.0566)	[-0.0356]	**-0.2167**** (0.0757)	[-0.0531]	**0.1742*** (0.0822)	[0.0344]
Log of Income	**0.0456*** (0.0224)	[0.0127]	**0.0972**** (0.0293)	[0.0101]	**-0.0685*** (0.0365)	[-0.0132]
Obese	0.0518 (0.0455)	[0.0148]	**-0.6055**** (0.0608)	[-0.1660]	**0.3283**** (0.0910)	[0.0857]
Overweight	-0.0003 (0.0425)	[-0.0001]	**-0.2367**** (0.0562)	[-0.0591]	**0.1324**** (0.0632)	[0.0304]
Hostile	**-0.3032**** (0.0649)	[-0.0961]	-	-	-	-
Uncooperative	**-0.8949**** (0.0882)	[-0.3221]				
Concerned about time	**-0.2589**** (0.0445)	[-0.0800]				
ρ	ρ_{12}=0.2342 (0.1875)		ρ_{13}= **-0.5036**** (0.1631)		ρ_{23}=0.0304 (0.2192)	

Notes: *P<0.05, **P<0.01, Standard errors are in parentheses and marginal effects are in square bracket.
Coefficients on education, region, marital status, employment status, smoking and insurance status are not shown in table.