Published in final edited form as:

Educ Econ. 2015 ; 23(6): 735–750. doi:10.1080/09645292.2014.944858.

EXAMINING THE EDUCATION GRADIENT IN CHRONIC ILLNESS

PINKA CHATTERJI, HEESOO JOO, and KAJAL LAHIRI^a

^aDepartment of Economics, University at Albany: SUNY, 1400 Washington Avenue, Albany NY 12222

Abstract

We examine the education gradient in diabetes, hypertension, and high cholesterol. We take into account diagnosed as well as undiagnosed cases, and use methods accounting for the possibility of unmeasured factors that are correlated with education and drive both the likelihood of having illness and the propensity to be diagnosed. Data come from the National Health and Nutrition Examination Survey (NHANES) 1999–2012. The education gradient in chronic disease varies by whether self-reported or objective disease measures are used. Education is negatively associated with having undiagnosed disease in some cases, but findings vary by how we define undiagnosed disease.

Keywords

chronic disease; education; diabetes; hypertension; cholesterol

1. Introduction

Among adults in the US, there are important differences across education groups in the prevalence, morbidity, and mortality associated with chronic disease.¹ Based on data from NHANES 1999–2012, the age-adjusted prevalence rates for diabetes, hypertension, and high cholesterol were 7.8 percent, 27.8 percent, and 26.1 percent respectively among adults aged 20 and older with more than 12 years of education, and 13.5, 32.3, 26.8 percent respectively among adults with less than 12 years of education. When estimating education-related differences in chronic disease, economists typically measure chronic disease using self-reported information, despite the limitations in these measures (Baker, Stabile & Deri, 2004). This approach may lead to biased estimates of education-related health disparities if individuals have unmeasured characteristics that are correlated with education, associated with true disease prevalence, and associated with the propensity to be diagnosed with and aware of having disease (Johnston, Propper & Shields, 2009). Chronic diseases are frequently asymptomatic in their early stages, and as a result education may play a

Corresponding author: Pinka Chatterji, 518-442-4746 (p), 518-442-4736 (f), pchatterji@albany.edu.

¹Mortality differences by education group have been widening in recent years, and part of the explanation for these worsening disparities is education-related differences in chronic disease. Meara, Richards & Cutler (2008), for example, report that about 13 percent of the growth in education-related differences in mortality among US adults between 1990 and 2000 can be attributed to heart disease (Meara et al., 2008).

particularly important role in increasing the likelihood that individuals with disease are diagnosed and aware of their condition.

This empirical issue has important policy implications. The recently released Healthy People 2020 has a new topic area specifically addressing education and other social determinants of health. To measure progress in this domain, we need estimates of the education gradient in health that account for the fact that for many chronic illnesses, the less educated are both more likely to have disease as well as less likely to be diagnosed given that a disease exists (<http://healthypeople.gov/2020/>).

This study examines the education gradient in three chronic diseases (diabetes, hypertension, and high cholesterol) using data from the National Health and Nutrition Examination Survey (NHANES) 1999-2012 which includes both self-reported and objective measures of chronic disease. The paper has two objectives. The first objective is to estimate: (1) education-related disparities in self-reported, diagnosed disease; and (2) education-related disparities in the total prevalence of disease, which includes both diagnosed and undiagnosed cases. To estimate these education-related disparities in chronic disease, we use a bivariate probit model which accounts for the possibility that unmeasured factors exist which underlie both total prevalence of disease and self-reports of diagnosed illness.

The second objective is to use these estimates to calculate the education gradient in having undiagnosed disease using a standard and an alternative definition of being undiagnosed. The standard definition of being undiagnosed is the probability an individual self-reports not having the disease, conditional on actually having the disease. That is, the standard definition captures people who are unaware of having disease among those who have the disease according to medical criteria. The alternative definition of being undiagnosed is the probability an individual actually has the disease, conditional on self-reporting not having the disease. That is, the alternative definition captures people who have disease according to medical criteria among those individuals who state they do not have the disease. The standard definition is widely used in the literature, making it possible to compare our estimates with other published estimates of the education gradient. The alternative definition, however, may be more relevant for policymaking in cases where improving disease diagnosis among untreated individuals is of primary concern.

Our findings show that the education gradient in chronic disease varies by whether self-reported or total prevalence (self-report and/or objective measures) of disease are used. In the case of hypertension, there is no education gradient in self-reported hypertension, but college education is negatively associated with total prevalence of hypertension. For diabetes, education is negatively associated with self-reporting disease, and college education is negatively associated with total prevalence disease. Education is positively associated with self-reports of having high cholesterol, and high school education is associated with higher total prevalence of high cholesterol, but there is no association between college education and the total prevalence of high cholesterol. Education is negatively associated with the probability of having undiagnosed disease in some cases, but the findings vary by the definition of being undiagnosed that is used.

2. The education gradient in chronic disease

In the case of diabetes, there is some mixed evidence that education is negatively associated with diabetes prevalence (having diagnosed or undiagnosed disease) and positively associated with having diagnosed illness, given the disease exists. For example, Wilder et al. (2005), based on NHANES III do not find an association between education and the likelihood of having undiagnosed diabetes. Smith (2007), on the other hand, use data from NHANES II, III, and IV and report a mixed pattern of findings – college education (16+ years) is associated with a reduction in the total prevalence of diabetes among males in the NHANES III, and having at least some college education (more than a high school degree) is associated with lower likelihood of being undiagnosed with an existing disease in NHANES IV, but otherwise education is not associated with diabetes prevalence and diagnosis.

Johnston, Propper & Shields (2009), in a study of the income gradient in hypertension based on the Health Survey for England, estimate a censored probit model which accounts for the following: (1) the distribution of being diagnosed with hypertension is censored (because we do not observe what diagnosis status would have been for individuals who do not have hypertension if they actually had hypertension); and (2) this censoring may be driven by unmeasured factors that are correlated with income and also affect the likelihood of having hypertension. These authors find a strong income gradient exists in objective, but not in self-reported, measures of hypertension. Estimates from the censored probit model indicate that income and education are negatively associated with having undiagnosed hypertension among those with the disease. These findings indicate that (1) relying solely on self-reports of chronic health conditions; and (2) ignoring selection into disease prevalence and diagnosis can yield misleading estimates of the SES gradient in chronic disease.

In our own previous work, we study racial/ethnic disparities in undiagnosed chronic diseases using biomarker data from the 2006 Health & Retirement Study (Chatterji et al., 2012). We estimate a trivariate probit model with selection which accounts for common, unmeasured factors underlying: (1) self-reporting chronic disease; (2) participating in biomarker collection; and (3) having disease, conditional on participating in biomarker collection. We find that African-Americans are less likely to have undiagnosed hypertension than non-Latino whites, but the magnitude of this effect falls appreciably after we account for selection. Accounting for selection, we find that African-Americans and Latinos are more likely to have undiagnosed diabetes compared to non-Latino whites. These findings are based on a widely used definition of being diagnosed – the likelihood of self-reporting disease among those who have disease. When we use an alternative definition of being undiagnosed, which considers an individual to be undiagnosed if s/he actually has the disease conditional on self-reporting not having it, we find higher levels of undiagnosed disease among racial/ethnic minorities vs. non-Latino whites for both hypertension and diabetes.

In the present study, we apply a similar approach, but we focus on estimating the education gradient in chronic disease, and we use a nationally representative data set which includes people of all ages (NHANES 1999-2012). This is important because the education gradient

in chronic illness may begin in early middle age, when many individuals have onset of asymptomatic chronic illness. The main contribution of our study to existing literature is twofold. First, we account for the possibility that unmeasured factors correlated with SES, disease prevalence, and disease diagnosis may confound estimates of the education-health gradient. Second, our estimates allow us to calculate the effect of education on having undiagnosed disease using two, alternative definitions of having undiagnosed disease. This information is highly policy-relevant, as it quantifies the role education plays in timely diagnosis, a critical factor in reducing the disability burden of chronic disease.

3. Data, Definitions, and Sample Statistics

A. Data

This study uses data on adults aged 25 to 75 from NHANES 1999-2012. The NHANES is a nationally representative survey designed to collect detailed health and nutritional information from adults and children in the United States. In 1999-2012, NHANES interviewed 29,991 individuals aged 25 to 75. Among interviewed individuals, 28,731 (95.8%) participated in the medical examination portion of the NHANES survey, which includes a blood draw and three readings of blood pressure.² NHANES respondents who participated in the medical exam were randomly assigned to a morning or an afternoon/evening exam. Of the 28,731 participants, 14,055 (48.9%) were assigned to the morning session and the remaining 14,676 (51.1%) participated in the afternoon/evening session. As a result, the 8-h fasting plasma glucose test results, which we use to objectively measure diabetes (described in the next section), are only available for those examined during the morning session, while objective blood pressure and total cholesterol measurements are available for both morning and afternoon/evening session participants. Thus, 12,334 respondents have objective measures of diabetes, while 23,197 and 26,953 have objective measures for hypertension and high cholesterol. Our final analytic samples, which are limited to respondents with non-missing information for all variables used in the analysis, include 10,780, 20,170, and 23,412 respondents for the analyses of diabetes, hypertension and high cholesterol, respectively.³

²We exclude from our hypertension analysis sample respondents who do not provide both a 2nd and a 3rd reading of blood pressure since we use the average of the 2nd and 3rd reading as an objective measure of hypertension. We also exclude respondents who report having alcohol, coffee, or cigarettes in the 30 minutes before measuring blood pressure.

³Of the sample of 12,334 individuals with objective measures for diabetes, we dropped observations with missing values for income to poverty ratio (951), high cholesterol self-report (274), marital status (140), obesity (129), hypertension self-report (42), smoking (8), education (6), and diabetes self-report (4), yielding a sample size of 10,780 for the diabetes analytic sample. Of the sample of 23,197 individuals with objective measures for hypertension, we dropped observations with missing values for income to poverty ratio (1856), high cholesterol self-report (550), marital status (290), obesity (228), hypertension self-report (70), smoking (11), education (12), and diabetes self-report (10), yielding a sample size of 20,170 for the hypertension analytic sample. Of the sample of 26,953 individuals with objective measures for high cholesterol, we dropped observations with missing values for income to poverty ratio (2099), high cholesterol self-report (659), marital status (323), obesity (334), hypertension self-report (92), smoking (13), education (11), diabetes self-report (10), yielding a sample size of 23,412 for the diabetes analytic sample. Since we dropped large numbers of observations from the three samples due to missing information on poverty status, we re-estimated all models with an imputed version of this variable, and with an indicator included in the models for whether an imputed version of the income to poverty ratio is being used. The results were very similar to those presented in the paper and the estimated coefficients on the missing income to poverty ratio indicator were not statistically significant except the models for total prevalence and self-report of hypertension. These results are available upon request.

B. Definitions

To define the total prevalence of diabetes, hypertension, and high cholesterol, we include both diagnosed and undiagnosed individuals. That is, we draw on both self-reported information on whether respondents have ever been diagnosed by a doctor with these conditions as well as results from the NHANES blood draw or blood pressure measures conducted during the medical exam portion of the survey. Note that if we were to rely on self-reports of chronic conditions only, we would not capture those who actually have chronic conditions but are undiagnosed. If we were only to use the medical examination results, we would only capture individuals with uncontrolled disease and would miss individuals who have a chronic condition but are controlled by medication.

In the case of hypertension, the total prevalence includes respondents whose medical examination findings indicate systolic blood pressure over 140 mmHg or diastolic blood pressure over 90mmHg, as well as respondents who self-report that they are currently taking antihypertensive medication (Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure, 2003). Similarly, the total prevalence of high cholesterol includes respondents who self-report currently using cholesterol lowering medication, as well as respondents whose medical examination results show a total cholesterol level over 240 mg/dl (ATP III, 2001). The total prevalence of diabetes includes those who self-report a previous diagnosis of diabetes from a health professional and use of diabetes medication, and/or whose medical examination results show a fasting plasma glucose (FPG) over 7.0mmol/l (126mg/dl) (American Diabetes Association, 2009).

We define being undiagnosed with an existing disease in two alternative ways. The standard definition of being undiagnosed is the following: conditional on having the disease (based on the total prevalence definition described above, which includes people with untreated and treated disease), the respondent responds negatively to the NHANES question “Have you ever been told by a doctor or other health professional that you had (diabetes, hypertension, blood cholesterol level that was high)?”⁴ This standard definition of being undiagnosed is widely used in the public health literature – it is simply, among those individuals with (treated or untreated) disease, the proportion of people who are not aware of having it. An alternative definition of being undiagnosed is: conditional on self-reporting NOT having the disease, the respondent, based on the medical exam, is found to actually have the disease. The alternative definition of being undiagnosed captures, among those who self-report NOT having disease, the proportion of people who actually do have disease. The alternative definition is useful because it can highlight education-related disparities in chronic illness among individuals who believe, correctly or incorrectly, that they do not have the disease. Often, individuals who are not in treatment are the target of public health interventions for highly prevalent diseases. Thus, it may be useful to know the characteristics of individuals in this group who actually have disease.

⁴For high cholesterol, the NHANES respondents are initially asked “Have you ever had your blood cholesterol checked?” Those respondents who report “no” to this question but have blood test results indicating high cholesterol are considered to have undiagnosed high cholesterol.

In Figure 1, we show a concordance table between self-reported and total prevalence of disease in order to contrast the standard and alternative definitions of being undiagnosed. The standard definition of being undiagnosed is the proportion of negative self-reports among those who have disease ($c/(a+c)$), while by the alternative definition the undiagnosed are those who have the disease among those who give a negative self-report ($c/(c+d)$). The difference between the definitions lies in whether the denominator includes individuals diagnosed with a disease that exists (group a in Figure 1) versus individuals not diagnosed with a disease that does not exist (group d in Figure 1). From an epidemiological perspective, the standard definition is intuitive because typically there is a gold standard measurement that can be used to appropriately identify individuals with disease. From a policy perspective, particularly for diseases with high prevalence such as hypertension and high cholesterol, it may be more cost-effective to first use survey data to determine who is not currently in treatment (those who self-report not having disease) and then determine, among those individuals, who actually has disease. In this case, the alternative definition may be more useful.⁵

We measure education using three dummy variables: (1) less than high school graduate (<12 years, baseline category); (2) high school graduate (12 years, including GED); and (3) at least some college (>12 years). The models also include controls for the income to poverty ratio, race/ethnicity, co-morbid health conditions, obesity, and smoking. The income to poverty ratio is a continuous variable generated by NHANES by dividing family income by the federal poverty threshold from the Department of Health and Human Services (DHHS).⁶ We include four dummy variables for race/ethnicity: (1) non-Latino white (baseline category); (2) African-American; (3) Latino; and (4) Other race/ethnicity.

To measure co-morbid conditions, we include in the models dummy indicators for whether or not the respondent self-reports diabetes, hypertension, and/or high cholesterol. (We do not include the outcome of interest. For example, the diabetes models include self-reported hypertension and high cholesterol only as covariates.) We cannot use objective measures of co-morbid conditions because these measures are not available for the whole sample. The models also include dummy indicators for whether the respondent is overweight ($25 \leq$ body mass index (BMI) <30) or obese (BMI ≥ 30), with BMI less than 25 are defined as the baseline group. The BMI is calculated by using actual measurements of height and weight of respondents. We include in the models two dummy variables for current smoker and former smoker, with never smokers as the baseline. The models also include a dummy variable indicating the respondent is married (single/widowed/divorced is the baseline), age, and age squared.⁷ In addition, the models include six dummy indicators for survey year with the years 1999-2000 as the baseline.

⁵Note that although we focus on the problem of undiagnosed individuals, misreporting may occur in either direction – that is, it is also possible that respondents self-report having an illness, and their medical examinations do not indicate the existence of an illness (group b in Figure 1). In this paper, we focus on the education gradient in undiagnosed disease rather than the education gradient in over-diagnosed disease since in the case of diabetes, hypertension, and high cholesterol, undiagnosed disease is considered to be a much more important public health and public policy problem than over-diagnosis. Moreover, in our data, it is difficult to separate true false positives from cases in which an individual is controlling a disease so well that s/he appears to be a false positive but actually the self-report is accurate. For these reasons, we leave an analysis of education and over-diagnosis of disease to future work.

⁶The federal poverty threshold depends on survey year, family size, and state of residence. In cases in which the income to poverty ratio is bigger than 5.0, it is recorded as 5.0 in the data.

C. Sample Statistics

Table 1 shows weighted sample characteristics by education group for the diabetes, hypertension, and high cholesterol samples. There is a clear educational gradient in self-reported illness, obesity, and smoking. For example, in the sample used to examine diabetes outcomes, about 6 percent of those in the high education group report having diabetes while about 12 percent of those in the low education group report having diabetes. Rates of hypertension (but not high cholesterol) are higher in the low education group vs. the high group. Note that these chronic illness rates are based on self-reported information and thus do not include undiagnosed individuals. In the diabetes analysis sample, about 36 percent of those with less than 12 years of education are current smokers vs. 16 percent in group with more than 12 years of education.

Table 2 shows the weighted total prevalence rates (which include diagnosed and undiagnosed cases, as described above) by education group. The high educated (>12 years) group has significantly lower total prevalence of chronic disease compared to the low educated (<12 years) group, particularly in the case of diabetes. Only 7.5 percent of the high education group has diabetes (diagnosed and undiagnosed) while 15.2 percent of the low education group has diabetes. Table 2 also shows weighted percentages of disease diagnosis by education group based on the standard definition, which is the proportion of those with the disease who self-report being diagnosed with it. The only significant educational gradient in diagnosed illness occurs for high cholesterol – 68.0 percent of those who have high cholesterol and are in the low education group have been diagnosed with high cholesterol, while 73.2 percent of those who have high cholesterol and are in high education group have been diagnosed with the disease. For diabetes, the lowest diagnosed group is the mid education group (12 years) and diagnosis rates between the high and low education group are not significantly different. For hypertension, diagnosis rates do not vary much by education group (78.1-79.4 percent).

We emphasize that the descriptive relationships between disease diagnosis and education that we observe in Table 2 may be confounded by the other factors, such as the severity of health conditions. If less educated people are more likely to be in later stages of chronic conditions than more educated people, this greater severity may make them more likely to be diagnosed with illness. This type of bias from unmeasured severity should be considered when estimating the education gradient in disease diagnosis. In fact, the data support the idea that lower educated people have more severe conditions, and that people with severe conditions are more likely to be diagnosed with and aware of having an existing chronic disease (results not shown).

4. Methods

Our first goal is to estimate models in which we examine: (1) the education gradient in self-reported diagnosis of diabetes, high blood pressure, and high cholesterol; and (2) the education gradient in the total prevalence of each of these three diseases (diagnosed and

⁷We re-estimated all models with a measure of self-reported health as an additional covariate. Results were very similar to those presented in the paper and are available upon request.

undiagnosed). The dependent variables are a binary indicator of self-reported disease, and a binary indicator of whether the respondent either self-reports medication for disease and/or has the disease based on the NHANES medical examination results (total prevalence). For estimation, we use a bivariate probit model, which accounts for the possibility of unmeasured characteristics that affect both the propensity to have disease (total prevalence) and the propensity to self-report disease, such as unobserved health knowledge and unmeasured aspects of SES. Although the bivariate probit model does not require any exclusion restrictions for identification, researchers sometimes impose such restrictions to sharpen the identification. In our case, however, it is difficult to find plausible identifying exclusion restrictions, given that the determinants of disease prevalence and disease self-report are likely to overlap considerably. Thus, we rely on the non-linearity of the model for identification in our case.

The model includes controls for education, income to poverty ratio, race/ethnicity, gender, marital status, age, age squared, and survey year. We additionally include controls for self-reported co-morbid chronic conditions, whether the respondent is overweight or obese, and smoking status. However, we caution that these additional variables are potentially endogenous – for example, individuals with diabetes may quit smoking or lose weight to better control their disease. We estimated the models without these potentially endogenous variables, and the findings were qualitatively similar to those presented here (results not shown).

Our second goal is to use the results from these models to estimate the education gradient in being undiagnosed according to the standard and the alternative definitions of being undiagnosed. We have four possible outcomes:

$$\begin{aligned} \Pr(A) &= \Pr(\text{self-report}=\text{Yes}, \text{Has disease (total prevalence)}=\text{Yes}) \\ \Pr(B) &= \Pr(\text{self-report}=\text{Yes}, \text{Has disease (total prevalence)}=\text{No}) \\ \Pr(C) &= \Pr(\text{self-report}=\text{No}, \text{Has disease (total prevalence)}=\text{Yes}) \\ \Pr(D) &= \Pr(\text{self-report}=\text{No}, \text{Has disease (total prevalence)}=\text{No}) \end{aligned}$$

Individuals in Group A have chronic disease, and they have been diagnosed with disease; presumably these individuals are receiving some level of treatment. Individuals in Group D are healthy – they do not have disease and they do not report disease. From a policy perspective, the two interesting groups are Groups B and C. Individuals in Group B may be reporting disease that does not exist, but using a data set like the NHANES, we cannot definitively make this argument. It is possible that Group B individuals have been successfully treated without medication, but they still consider themselves to have the disease. Thus, our analysis does not focus on estimating the education gradient associated with being this group. Instead, we focus on the role of education for individuals in Group C, who have undiagnosed chronic disease.

We define the binary variable h_i for $i=1$, and 2. We call $i=1$ the “self-report disease” step, and $i=2$ the “total prevalence” step. The underlying latent variables h_{ij}^* can be expressed as follows for individual j and $i=1$, and 2:

$$h_{ij}^* = \beta_i' x_{ij} + \varepsilon_{ij}$$

For $i=1$ and 2 , $h_{ij} = 1$ if $h_{ij}^* \geq 0$ and $h_{ij} = 0$ otherwise. x_{ij} is a vector of individual j 's characteristics at step i . The probabilities of each outcome are written as

$$\begin{aligned} \Pr(A) &= \Pr(h_{1j}^* \geq 0, \text{ and } h_{2j}^* \geq 0) = \Phi_2(\beta_1' x_{1j}, \beta_2' x_{2j}, \Omega) \\ \Pr(B) &= \Pr(h_{1j}^* \geq 0, \text{ and } h_{2j}^* < 0) = \Phi_2(\beta_1' x_{1j}, -\beta_2' x_{2j}, \Omega) \\ \Pr(C) &= \Pr(h_{1j}^* < 0, \text{ and } h_{2j}^* \geq 0) = \Phi_2(-\beta_1' x_{1j}, \beta_2' x_{2j}, \Omega) \\ \Pr(D) &= \Pr(h_{1j}^* < 0, \text{ and } h_{2j}^* < 0) = \Phi_2(-\beta_1' x_{1j}, -\beta_2' x_{2j}, \Omega) \end{aligned}$$

Where the errors ($\varepsilon_1, \varepsilon_2$) are assumed to be normally distributed with mean zero and

correlation matrix $\Omega = \begin{pmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{pmatrix}$. Φ_2 is the bivariate joint normal distribution function.

Thus, the probability of having undiagnosed chronic disease can be written as follows:

Standard definition:

$$\begin{aligned} \Pr(\text{Self-report=No} \mid \text{Has disease (total prevalence) = Yes}) \\ &= \Pr(h_{1j}=0 \mid h_{2j}=1) \\ &= \frac{\Pr(h_{1j}^* < 0, \text{ and } h_{2j}^* \geq 0 \mid x_{ij})}{\Pr(h_{2j}^* \geq 0 \mid x_{ij})} \end{aligned}$$

Alternative definition:

$$\begin{aligned} \Pr(\text{Has disease (total prevalence) = Yes} \mid \text{Self-report=No}) \\ &= \Pr(h_{2j}=1 \mid h_{1j}=0) \\ &= \frac{\Pr(h_{1j}^* < 0, \text{ and } h_{2j}^* \geq 0 \mid x_{ij})}{\Pr(h_{1j}^* \geq 0 \mid x_{ij})} \end{aligned}$$

To evaluate the education gradient in having undiagnosed chronic disease according to each of these two definitions, we use these conditional probabilities to calculate the marginal effects of education on having undiagnosed chronic disease.

5. Results

Table 3 summarizes findings from bivariate probit models that estimate the education gradient in self-reported disease (Column 1) and the total prevalence of disease (Column 2). In the case of diabetes, education is negatively associated with both self-reported disease and

the total prevalence of disease, although the association between high school education and the total prevalence of diabetes is not statistically significant. The magnitudes of these effects are policy-relevant. Having at least some college education, for example, reduces the probability of having diagnosed or undiagnosed diabetes by 2.5 percentage points (Column 2, Table 3), which is a 26 percent reduction at the sample mean total prevalence rate of 9.8 percent (Table 2). Poverty is also strongly associated with both self-reported diabetes and the total prevalence of diabetes, with higher income individuals having lower risk for this disease.

For hypertension, education is not associated with self-reported disease, but college education is negatively associated with the total prevalence of hypertension. Having at least some college education reduces the risk of having diagnosed or undiagnosed hypertension by about 4 percentage points (Column 2, Table 3) which is a 13 percent reduction at the sample mean total prevalence rate of 30 percent (Table 2). High school education is not associated with the total prevalence of hypertension. Higher income individuals are less likely to self-report having hypertension, but there is no association between the income to poverty ratio and the total prevalence of hypertension.

The pattern of results for high cholesterol in Table 3 differs from those for hypertension and diabetes. Higher education is associated with higher risk of self-reporting high cholesterol, which is the case for higher income as well. When the total prevalence of high cholesterol is considered, the positive effect of having high school education on high cholesterol persists, but the association between college education and high cholesterol, as well as the association between higher income and high cholesterol, both become statistically insignificant. Thus, higher education and income increase the risk that respondents report having high cholesterol, but only high school education is associated with actually having the disease.

In sum, the findings in Tables 3 highlight the education gradient in chronic disease, and how it varies across chronic illnesses. Across all models, the estimated rho is large in magnitude and statistically significant, suggesting that there are advantages in estimating the two equations jointly. It is clear that the education gradient differs depending on whether self-reported or true prevalence of illness is the outcome of interest, particularly for hypertension and high cholesterol. This finding suggests that there may be a gradient in undiagnosed illness.

Table 4 shows estimates of the education gradient in undiagnosed illness for diabetes, hypertension, and high cholesterol. Panel A shows the marginal effects of education on undiagnosed illness calculated according to the standard definition of being undiagnosed, while Panel B shows marginal effects of education calculated according to the alternative definition of being undiagnosed. These estimates are calculated based on the bivariate probit models shown in Table 3, which account for the possibility of shared and unmeasured determinants of self-reporting disease and having disease. The standard definition includes in the denominator all individuals with disease, including those being treated for disease who, by definition, have been diagnosed with disease. The alternative definition focuses on identifying individuals with disease among individuals who are not being treated for disease.

For diabetes, we see that education is not associated with having undiagnosed diabetes based on the standard definition (Panel A, Table 4) but college education is associated about a 1.3 percent reduction in having undiagnosed diabetes based on the alternative definition (Panel B, Table 4). The interpretation of this finding is the following – among individuals who actually have diabetes, education is not associated with the probability of being diagnosed. However, among those who self-report not having diabetes, those with at least some college education are less likely than others to be reporting their true disease state inaccurately. We see the same relationship between the income to poverty ratio and undiagnosed diabetes, with higher income individuals being less likely to have undiagnosed diabetes according to the alternative definition, but not by the standard definition. In fact, by the standard definition, higher income individuals are actually at higher risk of being undiagnosed compared to lower income individuals. The findings are not consistent with our other findings related to SES and chronic disease (Panel A, Table 4).

In Table 4, we see that there is no education gradient in undiagnosed hypertension according to the standard definition. Using the alternative definition, however, we find that college educated individuals are about 2.4 percentage points less likely than those with less than a high school education to have undiagnosed diabetes (Panel B, Table 4). That is, if a policy or program is targeted at individuals who report being disease-free, college education will be positively associated with truly being disease-free. The difference in the findings across the standard and alternative definitions is due to the fact that many individuals are being treated for hypertension with medication, even in the lowest education group. These individuals clearly have diagnosed illness. But among individuals who are not being treated (and thus report not having disease), college education is still an important factor associated with having undiagnosed hypertension. We do not find an association between the income to poverty ratio and undiagnosed hypertension according to the alternative definition. In the case of the standard definition, there is a positive association between income and undiagnosed hypertension, but the magnitude is close to zero.

Finally, we consider the education gradient in undiagnosed high cholesterol in Table 4. By the standard definition, high school and college education are associated with 3.0 and 5.5 percentage point reductions, respectively, in the likelihood of having undiagnosed high cholesterol (Panel A, Table 4). Using the alternative definition, there is no association between high school education and undiagnosed high cholesterol, but college education remains a protective factor for having undiagnosed high cholesterol. According to both the alternative and standard definitions of being undiagnosed, higher income individuals are less likely than lower income individuals to have undiagnosed high cholesterol.

Although we focus on differences in chronic disease by education and income group in this paper, demographic and health characteristics also affect the risk of having chronic disease. All minority groups are at higher risk for diabetes (both self-reported and total prevalence) compared to non-Latino whites, and African-American race also is positively associated with both self-reported and the total prevalence of hypertension. Latinos, however, have lower risk of hypertension compared to non-Latino whites, and African-Americans and Latinos both have lower risk of high cholesterol compared to non-Latino whites (Results available upon request). As expected, older people, and those with other health conditions

and poor health behaviors, have higher risk of all chronic diseases (Results available upon request). Females have lower risk of diabetes than males, but they are at higher risk of having high cholesterol than males when we consider the total prevalence (diagnosed and undiagnosed cases) of this disease (Results available upon request).

6. Conclusions

As a whole, these findings indicate that relying on self-reported information to estimate the education-chronic disease gradient potentially distorts the gradient because education is correlated with the probability of having undiagnosed disease. Including undiagnosed cases, and allowing for shared, unmeasured factors that drive prevalence and diagnosis, reveals that education plays an important role not only in having disease but also in being diagnosed with an existing disease. Also, our findings indicate that the role of education in timely diagnosis varies by how undiagnosed cases are defined. For all three diseases, we find a consistent, negative association between having at least some college education and having undiagnosed disease according to the alternative definition. By the standard definition, however, we only find a negative association between education and having undiagnosed disease for high cholesterol.

Our findings for undiagnosed diabetes differ from those of Smith (2007), who uses earlier NHANES data (1999-2002) and a standard probit model to estimate the effect of education on the probability of being undiagnosed, among those who have diabetes. He reports an 8 percentage point reduction associated with having at least some college education – we find no association between diabetes and undiagnosed diabetes, using the same definition of undiagnosed but more recent data and empirical methods accounting for correlation between the unmeasured determinants of having diabetes and being undiagnosed given that diabetes exists. Like us, Smith (2007) also finds a counter-intuitive positive relationship between income and undiagnosed diabetes, although the estimated coefficient is marginally statistically significant.

Our findings may indicate that disease diagnosis is an important mechanism through which education ultimately affects health. However, to examine this idea further, longitudinal data are needed to examine whether compared to higher educated individuals, less educated people experience earlier onset of disease, later diagnosis of existing disease, or both. If both effects occur, it is important to understand which effect is stronger, and also the mechanisms at work. If the main issue is that the less educated experience earlier onset and higher prevalence, this may be due to factors such as lifestyle and environment. If the primary issue is that the less educated are less likely to be diagnosed, this may result from lack of accessible screening and primary care services. Understanding these linkages will have important implications for public policies targeted at reducing education-related disparities in chronic disease.

Acknowledgments

We gratefully acknowledge research support from the National Institute on Minority Health and Health Disparities, National Institutes of Health (grant number 1 P20 MD003373). The content is solely the responsibility of the

authors and does not represent the official views of the National Institute on Minority Health and Health Disparities or the National Institutes of Health.

References

- American Diabetes Association. Standards of medical care in diabetes. *Diabetes Care*. 2009; 32(S1):S13–S61. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL Jr., [PubMed: 19118286]
- Baker M, Stabile M, Deri C. What do Self-Reported, Objective, Measures of Health Measure? *Journal of Human Resources*. 2004; 39:1067–1093.
- Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention; Hyattsville, MD: 1999–2012. (http://www.cdc.gov/nchs/nhanes/search/nhanes_continuous.aspx)
- Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Questionnaire (or Examination Protocol, or Laboratory Protocol). U.S. Department of Health and Human Services, Centers for Disease Control and Prevention; Hyattsville, MD: 1999–2012. (http://www.cdc.gov/nchs/nhanes/search/nhanes_continuous.aspx)
- Chatterji P, Joo H, Lahiri K. Beware of being unaware: Racial/ethnic disparities in chronic illness in the USA. *Health Economics*. 2012; 21(9):1040–1060. [PubMed: 22764038]
- Johnston, David W.; Propper, Carol; Shield, Michael A. Comparing subjective and objective measure of health: Evidence from hypertension for the income/health gradient. *Journal of Health Economics*. 2009; 28:540–552. [PubMed: 19406496]
- Jones DW, Materson BJ, Oparil S, Wright JT Jr. Roccella EJ, the National High Blood Pressure Education Program Coordinating Committee. The seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure (JNC 7). *Hypertension*. 2003; 42:1206–1252. Available online at <http://hyper.ahajournals.org/cgi/content/full/42/5/1206>. [PubMed: 14656957]
- Meara, Ellen R.; Richards, Seth; Cutler, David M. The gap gets bigger: changes in mortality and life expectancy, by education, 1981–2000. *Health Affairs*. 2008; 27(2):350–360. [PubMed: 18332489]
- Smith, James P. Nature and causes of trends in male diabetes prevalence, undiagnosed diabetes, and the socioeconomic status health gradient. *Proceedings of the National Academy of Science*. 2007; 104(33):13225–13231.
- Third Report of the Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). National Institutes of Health; Washington, DC: 2001.
- Wilder, Ronald P.; Majumdar, Sumit R.; Klarenbach, Scott W.; Jacobs, Philip. Socio-economic status and undiagnosed diabetes. *Diabetes Research and Clinical Practice*. 2005; 70:26–30. [PubMed: 16126120]

		Has Disease (Total Prevalence)	
		1: yes	0: no
Self-Reported Disease	1: yes	<i>a</i>	<i>b</i>
	0: no	<i>c</i>	<i>d</i>

Figure 1.

Standard vs. Alternative Definition of Being Undiagnosed with Chronic Disease

Note: The standard definition of being undiagnosed is $c/(a+c)$ The alternative definition of being undiagnosed is $c/(c+d)$.

Table 1

Weighted sample characteristics by education

	(1) Diabetes Analysis Sample				(2) Hypertension Analysis Sample				(3) High Cholesterol Analysis Sample			
	Total	Low	Mid	High	Total	Low	Mid	High	Total	Low	Mid	High
Non-Latino white	71.4	47.6	74.4 [-14.3]	77.1 [-12.2]	71.8	47.7	74.0 [-13.9]	77.9 [-13.1]	71.7	47.0	74.3 [-15.4]	77.7 [-13.7]
African-American	10.8	15.1	11.4 [3.2]	9.3 [5.2]	10.2	15.0	11.0 [4.1]	8.6 [6.2]	10.2	15.1	10.8 [4.7]	8.6 [6.7]
Latino	12.0	32.0	9.9 [12.3]	6.9 [11.5]	12.4	32.1	11.1 [12.7]	7.3 [12.3]	12.5	32.2	10.8 [13.8]	7.4 [12.7]
Other race/ethnicity	5.8	5.3	4.3 [1.5]	6.6 [-1.8]	5.5	5.2	3.9 [2.3]	6.2 [-2.1]	5.6	5.6	4.1 [2.6]	6.2 [-1.0]
Female	51.5	50.5	49.4 [0.7]	52.7 [-1.6]	51.4	50.2	49.3 [0.8]	52.5 [-2.3]	51.3	49.9	49.3 [0.6]	52.6 [-2.8]
Married	63.8	55.3	62.6 [-3.7]	66.8 [-6.7]	63.3	56.2	62.0 [-3.9]	65.8 [-7.6]	63.4	55.7	62.0 [-4.8]	66.2 [-9.1]
Age (years)	46.7	48.0	47.6 [0.9]	46.0 [4.6]	46.9	48.5	47.8 [1.8]	46.1 [6.5]	46.8	48.3	47.6 [1.9]	46.0 [7.3]
Self-reported												
Diabetes	7.7	12.3	8.5 [3.5]	6.1 [6.8]	7.7	13.1	8.3 [5.7]	6.0 [11.3]	7.8	12.9	8.0 [7.2]	6.2 [12.5]
Hypertension	30.9	35.5	34.7 [0.6]	28.1 [5.8]	29.8	35.1	32.5 [2.6]	27.2 [7.9]	29.6	34.8	32.6 [2.3]	26.8 [8.2]
High cholesterol	31.8	29.5	33.8 [-2.7]	31.7 [-1.7]	31.4	30.3	32.7 [-1.9]	31.2 [-0.8]	31.3	29.9	32.5 [-2.2]	31.3 [-1.5]
BMI												
BMI<25	30.6	27.8	26.6 [0.7]	33.0 [-2.8]	30.0	25.9	27.0 [-0.8]	32.2 [-4.7]	30.0	26.3	26.9 [-0.5]	32.2 [-4.8]
25 BMI<30 (Overweight)	34.0	32.9	35.2 [-1.4]	33.9 [-0.6]	34.7	35.5	33.8 [1.6]	34.8 [0.6]	34.7	35.4	34.3 [1.0]	34.7 [0.7]
BMI 30 (Obese)	35.4	39.3	38.2 [0.7]	33.3 [4.1]	35.3	38.6	39.2 [-0.5]	32.9 [4.8]	35.3	38.2	38.8 [-0.4]	33.0 [5.1]

	(1) Diabetes Analysis Sample				(2) Hypertension Analysis Sample				(3) High Cholesterol Analysis Sample			
	Total	Low	Mid	High	Total	Low	Mid	High	Total	Low	Mid	High
Non-smoker	51.2	40.0	44.8 [-2.5]	57.1 [-9.7]	52.4	42.0	45.8 [-3.0]	58.0 [-13.2]	51.5	40.9	44.6 [-3.0]	57.3 [-13.2]
Ever smoker	25.9	24.5	24.9 [-0.3]	26.7 [-1.3]	25.7	23.6	25.7 [-1.8]	26.3 [-2.8]	25.4	23.4	25.0 [-1.5]	26.1 [-2.8]
Current smoker	22.9	35.5	30.3 [2.8]	16.2 [12.9]	21.9	34.4	28.6 [4.8]	15.8 [18.6]	23.1	35.7	30.4 [4.6]	16.6 [18.7]
Sample size	10,780	2,896	2,426	5,458	20,170	5,374	4,587	10,209	23,412	6,320	5,327	11,765

Note: All numbers are percentages except age. T-test statistics are shown in square brackets. T-tests were conducted to test for differences between the mid and high education groups and the low education group. The low education group is less than 12 years of education. The mid education group is 12 years of education. The high education group is more than 12 years of education.

Table 2

Total prevalence and diagnosis of chronic conditions by education group (%)

		Total	Education		
			Low	Mid	High
Diabetes	Prevalence				
	Objective or medication *	9.8 (9.0-10.6)	15.2 (13.5-16.8)	11.6 (10.1-13.1) [3.1]	7.5 (6.7-8.4) [7.9]
	Self-report	7.7 (7.0-8.4)	12.3 (10.6-13.9)	8.5 (7.1-9.9) [3.5]	6.1 (5.3-6.8) [6.8]
	Objective	7.6 (6.9-8.3)	12.1 (10.7-13.6)	9.0 (7.6-10.3) [3.1]	5.7 (4.9-6.5) [7.8]
	Medication	6.4 (5.7-7.0)	10.3 (8.9-11.7)	7.0 (5.7-8.3) [3.4]	5.0 (4.2-5.7) [6.7]
	Within Prevalence (n=1,434)				
	Diagnosed	70.0 (66.7-73.3)	72.6 (68.3-76.8)	64.8 (58.0-71.7) [1.8]	71.7 (66.3-77.1) [0.3]
	Taking Medication	65.0 (61.4-68.6)	67.8 (63.3-72.3)	60.4 (53.4-67.4) [1.7]	66.1 (60.3-72.0) [0.5]
Hypertension	Prevalence				
	Objective or medication *	30.1 (29.1-31.1)	36.2 (34.4-38.1)	33.8 (32.1-35.4) [2.0]	26.9 (25.7-28.1) [9.1]
	Self-report	29.8 (28.8-30.8)	35.1 (33.3-36.8)	32.5 (31.0-34.1) [2.6]	27.2 (26.0-28.5) [7.9]
	Objective	15.4 (14.7-16.1)	20.6 (19.1-22.1)	17.4 (16.2-18.7) [3.4]	13.2 (12.3-14.1) [8.5]
	Medication	21.1 (20.2-22.0)	25.2 (23.5-26.8)	23.1 (21.8-24.5) [2.0]	19.2 (18.0-20.4) [6.6]
	Within Prevalence (n=6,976)				
	Diagnosed	79.0 (77.5-80.4)	79.4 (77.1-81.6)	78.1 (75.6-80.6) [0.9]	79.2 (77.1-81.3) [0.1]
	Taking Medication	70.2 (68.4-72.0)	69.5 (66.6-72.4)	68.5 (66.0-71.0) [0.6]	71.3 (68.8-73.8) [-1.0]
High Cholesterol	Prevalence				
	Objective or medication *	28.6 (27.7-29.4)	30.5 (28.8-32.2)	31.4 (29.8-33.0) [-0.8]	26.9 (25.7-28.1) [3.6]
	Self-report	31.3 (30.4-32.3)	29.9 (28.3-31.6)	32.5 (30.7-34.2) [-2.2]	31.3 (30.1-32.5) [-1.5]
	Objective	16.4 (15.7-17.2)	17.7 (16.3-19.0)	18.2 (16.9-19.5) [-0.6]	15.4 (14.5-16.4) [2.6]
	Medication	14.0 (13.2-14.7)	15.8 (14.3-17.2)	15.2 (14.1-16.3) [0.7]	12.9 (12.0-13.9) [3.4]
	Within Prevalence (n=6,987)				
	Diagnosed	71.6 (70.0-73.2)	68.0 (65.3-70.8)	70.6 (67.9-73.3) [-1.4]	73.2 (70.9-75.6) [-3.0]
	Taking Medication	48.9 (47.0-50.8)	51.7 (48.4-54.9)	48.3 (45.6-51.1) [1.8]	48.2 (45.5-50.9) [1.6]

Notes:

* Total prevalence. 95% CI is in parentheses. T-test statistics are shown in squared brackets. T-tests conducted to test for differences between mid education group and low education group, and for differences between high education group and low education group. Statistics are weighted. "Within prevalence" indicates that the sample is limited to respondents who are in the total prevalence group, meaning that they have the disease based on objective measurement or self-report medication for the disease.

Table 3

The education gradient in self-reported and total prevalence of chronic disease

		Self-report disease Column 1		Has disease (Tot. Prev.) Column 2	
		Coeff.	ME	Coeff.	ME
Diabetes (N=10,780)	High school graduate	-0.127 [*] (0.051)	-0.014 ^{**} (0.005)	-0.079 (0.048)	-0.011 (0.007)
	College and more	-0.110 [*] (0.048)	-0.013 [*] (0.006)	-0.168 ^{**} (0.046)	-0.025 ^{**} (0.007)
	Income to poverty ratio	-0.101 ^{**} (0.014)	-0.012 ^{**} (0.002)	-0.087 ^{**} (0.013)	-0.013 ^{**} (0.002)
	Rho	0.952 ^{**}	(0.005)		
Hypertension (N=20,170)	High school graduate	0.008 (0.030)	0.003 (0.010)	0.016 (0.030)	0.006 (0.011)
	College and more	-0.050 (0.028)	-0.018 (0.010)	-0.115 ^{**} (0.029)	-0.040 ^{**} (0.010)
	Income to poverty ratio	-0.024 ^{**} (0.007)	-0.008 ^{**} (0.003)	-0.014 (0.007)	-0.005 (0.003)
	Rho	0.851 ^{**}	(0.005)		
High Cholesterol (N=23,412)	High school graduate	0.099 ^{**} (0.027)	0.035 ^{**} (0.010)	0.064 [*] (0.027)	0.022 [*] (0.009)
	College and more	0.117 ^{**} (0.026)	0.040 ^{**} (0.009)	0.004 (0.025)	0.001 (0.009)
	Income to poverty ratio	0.041 ^{**} (0.007)	0.014 ^{**} (0.002)	0.004 (0.007)	0.001 (0.002)
	Rho	0.703 ^{**}	(0.007)		

Notes:

Table shows estimated coefficients and marginal effects (ME) from bivariate probit models. Standard errors are in parentheses. Race, gender, marital status, age, age squared, self-reported diabetes, high cholesterol, hypertension, obesity, smoking status, and survey year dummies are included in the models, but the estimated coefficients are not shown.

* denotes statistical significance at the 0.05 level and

** at the 0.01 level.

Table 4

The education gradient in undiagnosed disease

		Panel A: Standard definition Pr (Self-report=0 Has disease=1)		Panel B: Alternative definition Pr (Has disease=1 Self-report=0)	
Diabetes	High school graduate	0.054	(0.032)	−0.001	(0.004)
	College and more	−0.029	(0.028)	−0.013 **	(0.004)
	Income to poverty ratio	0.024 **	(0.008)	−0.003 **	(0.001)
Hypertension	High school graduate	0.002	(0.012)	0.003	(0.007)
	College and more	−0.015	(0.012)	−0.024 **	(0.006)
	Income to poverty ratio	0.007 **	(0.003)	−0.00001	(0.002)
High Cholesterol	High school graduate	−0.030 **	(0.011)	0.006	(0.007)
	College and more	−0.055 **	(0.011)	−0.013 *	(0.006)
	Income to poverty ratio	−0.019 **	(0.003)	−0.004 *	(0.002)

Notes:

Marginal effects shown. Standard errors are in parentheses. Marginal effects of race, gender, marital status, age, age square, self-reported diabetes, cholesterol, hypertension, obesity, smoking status, and survey year dummy variables are not shown.

* denotes statistical significance at the 0.05 level and

** at the 0.01 level.