Impact of low health literacy on adverse health behaviors in the state of North Carolina, 2016

Manan Roy*

Appalachian State University

Adam Hege†

Appalachian State University

Erin Bouldin‡

Appalachian State University

Abstract

The rural Appalachia region of the U.S. is replete with health disparities rooted in social determinants. One of the effects of social determinants is limited health literacy (LHL). Healthy People 2030 states attaining HL is required to achieve health and well-being. This study examines the causal effect of LHL on adverse health behaviors among North Carolina residents using a novel bias-adjusted treatment effect estimator (Oster, 2017). Data come from NC's 2016 Behavioral Risk Factor Surveillance System and the three-question HL module, which asks respondents to rate how difficult it is for them to get health-related advice or to understand verbal or written medical information. Respondents were classified as having limited HL if they find it somewhat or very difficult, or say they do not look for information on at least one question. Outcomes include tobacco use, alcohol consumption, exercise, sleep, oral health, and medical and dental check-ups. We find LHL has a direct causal effect on adverse health behaviors such as not exercising, sleeping less than 6 or 7 hours a day, on average, and not keeping up with routine medical and dental checkups. We cannot, however, assess the relationship between LHL and health outcomes such as obesity and poor oral health. This study aims to inform interventions that will improve health and behavior outcomes in Appalachia.

JEL: I12, I18
Keywords: Health literacy, omitted variable bias, treatment effect, health behavior, BRFSS

Acknowledgement: The authors would like to thank Dr. Daniel L. Millimet for suggestions and James Cassell and the NC BRFSS team in the NC Department of Health and Human Services for providing data that enabled us to explore health in Appalachia.

*Corresponding author. Department of Nutrition and Health Care Management, Appalachian State University. Email: roym1@appstate.edu. Phone: 828-262-8641.

†Department of Health and Exercise Science, Appalachian State University.

‡Department of Health and Exercise Science, Appalachian State University.

1 Introduction

Health literacy remains an urgent public health challenge facing the United States (USDHHS, 2018; Koh et al., 2012; Rudd, 2010; and Rasu et l., 2015). The concept of health literacy, which is quite complex and has been defined in a multitude of ways, generally refers to the ways that citizens access, understand, perceive, process and apply health information (Pleasant, 2014; Sorensen et al., 2012).[1] More than one-third of American adults have limited health literacy and just slightly more than 10 percent are considered proficient (Services USDoHaH, 2008). Low or inadequate health literacy has been associated with poorer health behaviors and outcomes (Batterman et al., 2016; Berkman et al., 2011; and Lee et al., 2012). Some have, however, posited that health literacy could serve as a mediator between social determinants of health (i.e., income and educational attainment) and health behaviors and outcomes (Lee et al., 2012; Bennett et al., 2009; and Schillinger et al., 2006).

With health disparities prevalent across the United States, geographic location, and often rurality, plays a significant role (Bolin et al., 2015). In particular, the largely rural southeastern United States consistently is the unhealthiest region of the country (Appalachian Regional Commission, 2017). Many of these disparities are rooted in social determinants, but the prevalence of risky health behaviors (e.g., smoking, alcohol abuse) cannot be overlooked (Singh et al., 2017; Behringer and Friedell, 2006; and Wewers et al., 2006). The literature, however, is yet to assess the causal effect of low health literacy (LHL) on health behavior in this region of the country.

This study examines the causal effect of LHL on adverse health behaviors among North Carolina residents, using Behavioral Risk Factor Surveillance System (BRFSS) data. Using a novel bias-adjusted treatment effect estimator (Oster, 2017), we find LHL has a direct causal effect on adverse health behaviors such as not exercising, sleeping less than 6 or 7 hours a day, on average, and not keeping up with routine medical and dental checkups. We cannot, however, assess the relationship between LHL and health outcomes such as obesity and poor

[1] Also see: https://www.cdc.gov/healthliteracy/learn/index.html

oral health. The findings can help to inform intervention efforts related to improving both health literacy and health behavior.

The rest of the paper is organized as follows. Section 2 motivates the study and Section 3 discusses the data. Section 4 presents the empirical methodology. Section 5 discusses the main results (with binary LHL) while section 6 presents the additional results using a health literacy index. Section 7 concludes.

2 Background

The findings and research gaps in the *Health Disparities in Appalachia* report[2] and the overarching goal of the draft summary of *Healthy People 2030* to identify regions and groups with poor health motivate this study. The *Healthy People 2030* states attaining health literacy, among others, is required to achieve health and well-being.[3] One of the overarching goals is to attain health literacy, health equity, and eliminate disparities to improve health and well-being of all populations. The draft framework also recommends identifying regions and groups with poor health or at risk of poor health in the future to alleviate their poor health status. At the same time, the *Health Disparities in Appalachia* report recently published by the Appalachian Regional Commission, Robert Wood Johnson Foundation, and others shows that the Appalachian region lags behind the nation in mortality rates for major chronic conditions, higher depression prevalence, and physically and mentally unhealthy days; heart disease, cancer, and stroke mortality; infant mortality; primary care physician supply; poverty; education; and years of potential life lost. However, the state level data used in the *Health Disparities in Appalachia* report cannot be used for individual level analyses.[4] We propose to fill this research gap by focusing on a variety of adverse health behavior and health outcomes at the individual level in one Appalachian state: North Carolina (NC).

[2] https://www.arc.gov/assets/research_reports/Health_Disparities_in_Appalachia_August_2017.pdf

[3] https://www.healthypeople.gov/2020/About-Healthy-People/Development-Healthy-People-2030/Draft-Framework

[4] https://www.arc.gov/assets/research_reports/Health_Disparities_in_Appalachia_August_2017.pdf

Such analyses are crucial to better understand the challenges and needs of the population to devise interventions that will improve health and behavior outcomes in Appalachian NC.

3 Data

We used data from the 2016 NC BRFSS to conduct our study. The BRFSS is a telephone survey - including both cellular and landline numbers - of adults age 18 and older who live in the community (i.e., not in institutional settings like long-term care facilities or prisons) (Mokdad, 2009). It is conducted annually in all US states and territories by state health departments in collaboration with the Centers for Disease Control and Prevention (CDC). The purpose is to assess a variety of health-related topics including diet, physical activity, health care access, health conditions, and disability among a representative sample. All data collected through the BRFSS are self-reported. The health literacy questions were included as an optional module for NC in 2016. The NC sample included 6,537 adults. We restricted the sample to 6,396 respondents between 18 - 89 years old without missing information on age and gender.

3.1 Measures

The treatment variable of interest is LHL, which was measured by three questions. The three questions include: "How difficult is it for you to get advice about health or medical topics if you need it? Would you say it is…"; "How difficult is it for you to understand information that doctors, nurses and other health professionals tell you? Would you say it is…"; and, "You can find written information about health on the Internet, in newspapers and magazines, and in brochures in the doctor's office and clinic. In general, how difficult is it for you to understand written health information? Would you say it is...." For each of the questions, response options included: "very easy", "somewhat easy", "somewhat difficult", "very difficult", or "I don't look for health information." Respondents were classified as having LHL if they find it somewhat or very difficult, or say they do not look for information on at least one question. This definition is different from the trichotomized health literacy variable used in the literature using, for example, the Kansas BRFSS data (Chesser et al., 2018). The health literacy response categories in NC BRFSS are in reverse order compared to the Kansas BRFSS data.

We also created an HL index using these three questions. We created binary variables for responses to each question so that respondents who rated the task as very easy or somewhat easy were classified as having adequate HL in that domain and respondents who said the task was somewhat or very difficult or who said they did not engage in the task were classified as having inadequate HL in that domain. We combined these binary variables to create a HL index, and we classified respondents as having LHL overall if they had inadequate HL in at least one domain.

We considered several binary measures (yes, no) of adverse health behavior as dependent variables to include: exercising in the past month; a medical checkup more than 2 years ago or never having one; inadequate sleep (less than 6 hours and less than 7 hours), on average, in a 24 hour period; visiting a dentist or dental clinic in the last two years, or last 5 years, or more than 5 years ago; having 5 permanent teeth removed or 6 or more teeth removed due to tooth decay or gum disease; smoking at least 100 cigarettes over the lifetime; smoking every day; and obesity. We controlled for several socioeconomic and demographic characteristics in the regression models. These include dummy variables for male, white and non-white, married, education dummies for less than high school, high school, some college, college degree, employed, income dummy variables for annual income less than \$15,000, less than \$25,000, less than \$50,000, less than \$75,000, greater than \$75,000. Homeownership dummy variables included are for homeowners and renters. We also included an indicator for Appalachian counties.

4 Empirics

Ordinary least squares (OLS) regressions are not sufficient to analyze the causal effect of a treatment variable (LHL, here) on an outcome. This is because empirical research using OLS (linear or non-linear models) suffer from omitted variable bias - whether they omit controls or improper proxies of controls available in the data (observed characteristics) from the regression models or whether they cannot control for unobservable attributes that influence both the treatment variable and the outcomes that are not available in the data. Moreover, OLS and its non-linear counterparts assume selection on observed characteristics (SOO) is random and assume selection on unobserved characteristics (SOU) is zero. In the absence of a randomized experiment, neither of these assumptions hold. Accordingly, the OLS estimates cannot be interpreted as causal effects.

Oster (2017) notes that empirical studies often assess the sensitivity of the estimated treatment effect to the inclusion of observed controls as a sign of the extent of the omitted variable bias. Greater stability of the coefficient across baseline to augmented or full regression models usually imply limited omitted variable bias. And this link is often direct. However, this assumes that the bias arising from the included observed controls contains information about the bias arising from the full set of controls, including unobserved attributes. However, observable factors can be used to identify bias from the unobserved factors only if we impose assumptions about the covariance characteristics of the two sets of factors. Additionally, coefficient movements alone cannot be used to calculate the bias. The quality of the included covariates also depends on how much of the variance of the outcome it can explain - that is, the change in the R-squared when the covariates are introduced in the model.

Oster (2017) proposes an omitted variable bias-adjusted estimator that allows the omitted variable bias (OVB) to be proportional to coefficient movements scaled by the change in R-squared when the covariates are introduced.

The approach begins with a model with some observed confounders and some unobserved confounders. Specifically, consider the following regression model:

$$Y = \beta LHL + \gamma\omega^0 + W_2 + \varepsilon, \qquad (1)$$

where Y is the outcome of interest, LHL is the LHL indicator, and ω^0 is a vector of observed control variables ω_1^0, ..., ω_J^0. The index W_2 is not observed.[5] We define $W_1 = \gamma\omega^0$ and assume all elements in ω^0 are orthogonal to W_2, so W_1 and W_2 are orthogonal.

We then define the proportional selection relationship between the observed characteristics and unobserved attributes as follows:

$$\delta\frac{\sigma_{1LHL}}{\sigma_1^2} = \frac{\sigma_{2LHL}}{\sigma_2^2},$$

where

$$\sigma_{iLHL} = cov(W_i, LHL), \sigma_i^2 = var(W_i^2) \text{ for } i\epsilon\{1, 2\},$$

and δ is the coefficient of proportionality. Since we do not make any assumptions about δ now, this relationship always holds for some δ.

Let the coefficient on LHL from the uncontrolled regression of Y on LHL be β_0 and the R-squared from that regression be R_0. Let the coefficient from the controlled regression of Y on LHL and ω^0 be β_1 and the R-squared R_1. Finally, let $R_{\max}$ be the R-squared from a hypothetical regression of Y on LHL, ω^0, and W_2 (full model).

The omitted variable bias-adjusted treatment effect can be approximated as follows:

$$\beta^* \approx \beta_1 - \delta\left[\beta_0 - \beta_1\right]\frac{R_{\max} - R_1}{R_1 - R_0}.$$

Oster (2017) suggests two ways to assess the robustness of coefficient stability using the above bias-adjusted estimator:

One approach is to calculate the value of δ (the coefficient of proportionality) for which $\beta = 0$ by assuming a value for $R_{\max}$. That is, the degree of SOU relative to SOO that

[5] The treatment variable can be either binary or continuous.

would be needed to explain away the estimated effect under exogeneity. For example, a value of $\delta = 3$ suggests that unobservable attributes would have to be three times as important as the observed characteristics to produce a treatment effect of zero.

The second approach is to set bounds on $R_{\max}$ and δ to develop a set of bounds for β. Following the partial identification literature, the estimator can be denoted as $\beta^*(R_{\max}, \delta)$. Now, without any additional assumptions, $R_{\max}$ is bounded between R_1 (the R-squared in the controlled regression) and 1. Assuming $\delta > 0$, that is the covariance between LHL and observed characteristics is of the same sign as the correlation between LHL and the unobserved covariates, bounds δ between 0 and an arbitrary upper bound $\overline{\delta}$.

The bounding set for β is:

$$\Delta_s = [\beta_1, \beta^*(\overline{R_{\max}}, 1)].$$

If the bounds of the set are outside the confidence interval on β_1 then the conclusions based on the controlled regression's coefficient is not robust. That is, if the identified set includes zero, the estimated treatment effect from the intermediate regression is not robust.

Finally, Oster (2017) uses empirical evidence from randomized control trial studies to determine a plausible value for $R_{\max}$.

She suggests using

$$R_{\max} = 1.3 * R_1.$$

In other words, if the identified set $[\beta_1, \beta^*(\{1.3 * R_1, 1\}, 1)]$ excludes zero or the estimated δ which produces $\beta = 0$ with $R_{\max} = 1.3 * R_1$ exceeds 1, then we can conclude the estimated coefficient is robust. The cutoff value of $\delta = 1$ is appropriate since it suggests the observables are at least as important as the unobservable attributes. This is because researchers typically choose regression controls which they believe ex ante are the most important (Angrist and Pischke, 2010).

The Oster (2017) method is an extension of the theory that links bias explicitly to coefficient stability. It shows the importance of and develops a tractable strategy to account for both coefficient and R-squared movements in addressing omitted variable bias and generates bounds on the treatment effect. One of the drawbacks of the approach is that it is applicable to only linear regressions. However, as illustrated in this study, linear probability models can be used to assess the treatment effect on binary outcomes. Since this method bounds the treatment effect, it does not provide a point estimate of the treatment effect. Another challenge in using this method pertains to the choice of appropriate controls in the models. These results do not hold if unobserved attributes related to the treatment contain no information about the relationship between the treatment and observables included in the model. Broadly, this challenge points towards improving the set of controls in any empirical analysis.

5 Main Results

5.1 Descriptive statistics

Compared to their non-Appalachian neighbors, NC residents living in Appalachian counties were more likely to have low health literacy. People in Appalachian counties were significantly older, and more often white (Table 1). People in Appalachian counties tended to have lower levels of education and household income and were less likely to be employed and rent than people in non-Appalachian counties. In unadjusted comparisons, people in Appalachian counties were less likely to exercise, more likely to have visited a dentist more than 5 years ago, and more likely to have all teeth removed. In sum, there are limited significant differences across Appalachian and non-Appalachian counties in terms of the outcomes of interest and most covariates. Accordingly, we focus on estimating the statewide effects of LHL.

Among community-dwelling North Carolina adults aged 18-89, 17% had low health literacy. Compared to people with higher health literacy, people with low health literacy were significantly older, more frequently men, less often white and married or partnered (Table 1).

People with low health literacy also tended to have lower levels of education and household income and were less likely to be employed and own their home than people with higher health literacy. In unadjusted comparisons, people with low health literacy had statistically significantly lower likelihood of exercising and getting adequate sleep; were more likely to be obese, to be current smokers or other tobacco users, or to have more than 6 permanent teeth removed; and less likely to have had a recent dental visit or health check-up or to have had recommended medical checkup than people with higher health literacy.

5.2 OLS results

Table 2 illustrates how OLS estimates and R-squared values change as we introduce controls.[6] For all panels, Column (1) does not include any controls. Column (2) includes only two controls - age and an male indicator. We refer to this model as the Uncontrolled Model following Oster (2017). Column (3) augments Column (2) by adding education dummies for less than high school, high school, some college, college degree, employed, income dummy variables for annual income less than \$15,000, less than \$25,000, less than \$50,000, less than \$75,000, greater than \$75,000, and dummy variables for homeowners and renters. Column (4) further augments the model in Column (3) by adding indicators for white and non-white race categories, indicator for being married, an interaction terms between the male indicator and age, and squared and cubed age. It also includes an indicator for Appalachian counties.[7] We refer to this model as the Controlled Model following Oster (2017).

As we move from Column (1) to Column (4) for all outcomes, we observe changes in the magnitudes of R-squared values and coefficient estimates of LHL. The coefficient estimates also change in statistical significance. For example, in Panel I, the coefficient estimates of LHL changes from 14.2% without any controls to 13.7% after controlling for age and gender.

[6] The Oster (2017) approach is applicable only for linear regressions. We compared marginal effects from linear probability models (LPM) and probit models. Since the marginal effects from the latter are identical to the estimates from the LPM, we implemented the approach with LPM. The comparison results are available upon request.

[7] Columns (3) and (4) in Panel II include an additional indicator of whether the respondent has any health insurance coverage.

The estimate further changes to 7.5% in Column (3) and finally becomes 7.2% after including all observed controls. The estimates are statistically significant at the 1% confidence level across all models. R-squared values also change from 0.016 in Column (1) to 0.037 in Column (2) to 0.082 in Column (3) to 0.084 in Column (4). The movements in both the LHL estimate and R-squared implies there is ample selection on observed characteristics and most likely on unobserved attributes too, and thus omitted variable bias. For example, the data does not have information on the attitude of respondents towards exercising. Moreover, since the BRFSS is not a longitudinal survey we do not have information about the respondents' history of exercising, employment, income, health status, an other health behaviors.

For Panels II - IV, the results are similar. The coefficient estimate of LHL changes as we introduce more controls along with movements in the R-squared. The LHL estimates also continue to be statistically significant across all models indicating a strong positive association between LHL and probability of inadequate sleep and irregular health check-up. But, as before, these estimates do not indicate a causal relationship.

For outcomes related to oral check-up, panels V - VII present a similar story but in some cases the LHL estimate loses statistical significance as more controls re introduced. For example, in Panel V (likelihood of dental visit in the last 2 years), the LHL estimate is statistically significant at the 5% level in the first two models with none or few controls. The LHL estimate becomes statistically insignificant as more controls are introduced in Columns (3) and (4). Similarly, the LHL estimate for the likelihood of a dental check-up in the past 5 years (Panel VI) is statistically significant at the 1% level with none or few controls in Columns (1) and (2). Introduction of more covariates in Columns (3) and (4) reduces the magnitude of the LHL estimate as well as its statistical significance. Finally, in Panel VII, the LHL estimate changes from 11.4% with no controls to 10.8% with two controls (statistically significant at the 1% level). As more controls are introduced, the magnitude of the LHL estimate drops to around 4% and is statistically significant at only the 5% level.

For the two outcomes related to oral health in Panels VIII and IX, only the likelihood of having six permanent teeth removed (Panel IX) has a statistically significant (1% confidence level) and positive association with LHL even after introducing all the controls. The magnitude, however, changes from 10.4% without controls to 4.6% with all the observed controls.

Similar to the dental check up outcomes, the probability of having smoked at least 100 cigarettes in life (Panel X) has a positive and statistically significant association with LHL without controls (9.8% at 1% confidence level) and with controls (4% at 10% confidence level). However, the association is positive but imprecise in the full model. Similar results hold for the probability of smoking everyday (Panel XI).

Finally, in Panel XII, the coefficient estimate of LHL is statistically significant at the 5% level with none or two controls in Columns (1) and (2). However, the association between LHL and probability of being obese becomes imprecise. although positive, as more controls are introduced in Columns (3) and (4).

In sum, OLS estimates cannot be interpreted as causal since the assumptions cannot address omitted variable bias. In the next section, we turn to the Oster (2017) estimates to identify the causal effect of LHL on the various outcomes.

5.3 Bias-adjusted treatment effect

Table 3 presents the results for the omitted variable bias-adjusted treatment effect. For each of the 12 outcomes, Column (1) presents the estimated coefficient, robust standard errors (s.e.), and R-squared from the Uncontrolled Model - it is Column (2) in Table 2. Column (2) presents the same information for the Controlled Model - it is Column (4) in Table 2. Column (3) presents the estimated bounds on the bias-adjusted treatment effect. Column (4) presents the estimated coefficient of proportionality between the SOO and SOU (δ) so that the estimated effect under the controlled model is zero and $R_{\max}$ is 1.3 times the R-squared from the controlled model.[8]

[8] All estimations are done in Stata using the user written command –*psacalc*-.

Exercise. For the probability of not exercising in the past month, there seems to be a direct causal effect of LHL on not exercising. The uncontrolled model indicates individuals with LHL are 13.7% more likely to not exercise in the past month. This estimate remains positive and statistically significant and drops to 7.2% in Column (2) as we include more control variables. Column (3) indicates the estimated bias-adjusted treatment effect of LHL lies between 3.5% and 7.2%. Finally, the estimated δ in Column (4) is 1.8, that is, the unobservable attributes are almost twice as important as the observed characteristics. That seems unlikely since the full model includes almost all available and important covariates. So, the estimated direct relationship between LHL and likelihood of not exercising is robust.

Medical checkup. The uncontrolled model in Panel II shows that respondents with LHL are 8.4% more likely to have had a health checkup more than 2 years ago or never. Including covariates reduces that likelihood to 5.4% in Column (2). Column (3) shows the bias-adjusted treatment effect of LHL lies between 3.6% and 5.4%. Finally, the estimated δ in Column (4) is almost 2.7, that is, the unobservable attributes are almost thrice as important as the observed characteristics. Since that is unlikely to be true, the estimated direct relationship between LHL and likelihood of medical checkup more than 2 years ago is robust.

Inadequate sleep. For the probability of sleeping less than 6 hours in Panel III, the uncontrolled model indicates individuals with LHL are 7.3% more likely to sleep less than 6 hours on average in the past 24 hours. This estimate drops to 4.6% in Column (2). Column (3) indicates the estimated bias-adjusted treatment effect of LHL lies between 3.5% and 4.6%. Finally, the estimated δ in Column (4) is 3.2, that is, the unobservable attributes are more than thrice as important as the observed characteristics. Since that is unlikely, the estimated direct relationship between LHL and likelihood of sleeping less than 6 hours is robust. In Panel IV, the estimated treatment effect of LHL on the probability of sleeping less than 7 hours changes from 8.5% to 6.9% as we move from Column (1) to Column (2). The identified set shows the bias-adjusted treatment effect is bounded by about 6% and almost 7%. Additionally, since the estimated δ in Column (4) exceeds 1, the effect of LHL on this measure of inadequate sleep is robust as well.

Oral health checkup. Column (1), Panel VII shows respondents with LHL have a 10.8% likelihood of visiting a dentist or dental clinic more than 5 years ago. The likelihood drops to 3.8% (although at the 5% confidence level) after including the covariates. The identified set in Column (3) shows the bias-adjusted treatment effect of LHL varies from 0.9% to 3.8%. In Panels V and VI for the probabilities of visiting a dentist or dental clinic in the past 2 years and 5 years, the estimated coefficients in the controlled model are positive (as in the uncontrolled model in Column (1)), but are not statistically significant. Thus it is infeasible to assess the robustness of the relationship in our sample using this empirical approach.

Oral health. Panels VIII and IX present the results for probabilities of having 5 permanent teeth and 6 or more permanent teeth removed for tooth decay or gum infection. The relationship in Panel VIII cannot be assessed with this approach since none of the estimated treatment effects in Columns (1) and (2) are statistically significant. However, in Panel IX, respondents with LHL are 9.3% more likely to have at least 6 teeth permanent teeth removed for tooth decay or gum infection in Column (1) and that drops to 4.6% after including covariates. However, the bounds on the estimated bias-adjusted treatment effect includes zero (see Column (3)). Moreover, estimated coefficient of proportionality in Column (4) is less than one. Columns (3) and (4) indicate the estimated direct relationship between LHL and the probability of having at least 6 teeth removed is not robust.

Tobacco use. In Panel X, for the probability of having smoked at least 100 cigarettes in entire life, the estimated treatment effect drops from 8.1% in Column (1) to 3.3% in Column (2) although the latter is imprecisely estimated. Thus it is infeasible to assess the robustness of the relationship in our sample using this empirical approach. Similarly, we cannot assess the omitted variable bias-adjusted treatment effect for the probability of smoking every day in Panel XI.

Obesity. Similar to Panels X and XI, we cannot assess the omitted variable bias-adjusted treatment effect for the likelihood of being obese in Panel XII.

6 Sensitivity analysis

In this section, we assess the causal effect of a health literacy (HL) index on the above outcomes. First, we code each health literacy question such that 0 indicates adequacy and 1 indicates inadequacy. For the question on understanding written health information, the written indicator takes a value of 1 if the response is somewhat or very difficult or if the respondent does not pay attention to written health information, and zero otherwise. For the question on understanding information shared by health professionals, the understanding indicator takes a value of 1 if the response is somewhat or very difficult, and zero otherwise. Finally, the advice indicator takes a value of 1 if the respondent finds it somewhat or very difficult to follow health advice or information or does not look for health information, and zero otherwise.

Second, we add these binary variables to generate the overall HL index. A value of zero for the HL index denotes adequate health literacy since the respondent demonstrates adequate HL on all three health literacy questions. A value of 3 indicates lowest level of HL since the respondent demonstrates low or inadequate HL on all three health literacy questions. Values of 1 or 2 indicate inadequate HL on one or two of the health literacy questions. The coefficient on the LHL index thus should be interpreted as the effect of moving from adequate HL to decreasing levels of health literacy.

6.1 OLS estimates

Similar to Table 2, we estimated how the OLS coefficients and R-squared values change as we introduce controls using the health literacy index. The movements in both magnitude and often statistical significance of the OLS estimate of the health literacy index along with R-squared values imply omitted variable bias. As before, we cannot interpret these findings as causal effects. The results are available upon request.

6.2 Bias-adjusted treatment effect

Table 3 presents the results for the omitted variable bias-adjusted treatment effect of the health literacy index. For each of the 12 outcomes, Column (1) presents the estimated coefficient, robust standard errors (s.e.), and R-squared from the uncontrolled regression. Column (2) presents the same information for the controlled regression. Column (3) presents the estimated bounds on the bias-adjusted treatment effect. Column (4) presents the estimated coefficient of proportionality between the SOO and SOU (δ) so that the estimated effect under the controlled model is zero and $R_{\max}$ is 1.3 times the R-squared from the controlled model.

Exercise. For the probability of not exercising in the past month, there appears to be a robust relationship between health literacy and not exercising. The uncontrolled model indicates as respondents' HL decreases by one unit, they are 8.2 percentage points more likely to not exercise in the past month. This estimate remains positive and statistically significant and drops to 3.9 percentage points in Column (2) as we include more control variables. Column (3) indicates the estimated bias-adjusted treatment effect of HL lies between 1.4 and 3.9 percentage points. That is, as the health literacy level falls, people are more likely to not exercise. Finally, the estimated δ in Column (4) is 1.5, that is, the unobservable attributes are almost twice as important as the observed characteristics. That seems unlikely since the full model includes almost all available and important observed covariates. So, there exists a direct causal relationship between HL and likelihood of not exercising.

Medical checkup. The uncontrolled model in Panel II shows that as HL decreases by one unit, the probability of having had a health checkup more than 2 years ago or never increases by 5.6 percentage points. Including covariates reduces that likelihood to 3.4 percentage points in Column (2). Column (3) shows the bias-adjusted treatment effect of HL lies between 2 and 3.4 percentage points. Finally, the estimated δ in Column (4) is around 2.2, that is, the unobservable attributes are more than twice as important as the observed characteristics.

Since that is unlikely to be true, the estimated direct relationship between HL and likelihood of medical checkup more than 2 years ago is robust.

Inadequate sleep. For the probability of sleeping less than 6 hours in Panel III, the uncontrolled model indicates as HL decreases by one unit, the likelihood of sleeping less than 6 hours on average increases by 4.9 percentage points. This estimate drops to 3 percentage points in Column (2). Column (3) indicates the estimated bias-adjusted treatment effect of HL lies between 2.2 and 3 percentage points. Finally, the estimated δ in Column (4) is 2.7. Thus, the estimated direct relationship between HL and likelihood of sleeping less than 6 hours is robust. In Panel IV, the estimated treatment effect of HL on the probability of sleeping less than 7 hours changes from 5.4 percentage points to 4.3 percentage points as HL decreases by one unit as we move from Column (1) to Column (2). The identified set shows the bias-adjusted treatment effect is bounded by about 3.8 to 4.3 percentage points. Additionally, since the estimated δ in Column (4) exceeds 1, the effect of HL on this measure of inadequate sleep is robust as well.

Oral health checkup. Column (1), Panel VII shows respondents with lower HL have a 8 percentage point likelihood of visiting a dentist or dental clinic more than 5 years ago. The likelihood drops to 3.4 percentage points after including covariates in Column (2). The identified set in Column (3) shows the bias-adjusted treatment effect of HL varies from 1.3 to 3.4 percentage points. Moreover, the estimated δ in Column (4) exceeds 1 so that the direct relationship with lower levels of HL is robust. The estimated coefficient for the HL index in Panel V (the probability of visiting a dentist or dental clinic in the past 2 years) is imprecise. Thus it is infeasible to assess the robustness of the relationship using this empirical approach in this sample of NC respondents. In Panel VI, respondents with lower HL have a 3.6 percentage point likelihood of visiting a dentist or dental clinic in the past 5 years. The likelihood drops to 1.9 percentage points after including covariates in Column (2). The identified set in Column (3) shows the bias-adjusted treatment effect of HL varies from 1.2 to 1.9 percentage points. Moreover, the estimated δ in Column (4) exceeds 1 so that the direct relationship with lower levels of HL is robust. This result is different from the results in Table 2 where the same relationship was imprecisely estimated.

Oral health. Panels VIII and IX present the results for probabilities of having 5 permanent teeth and 6 or more permanent teeth removed for tooth decay or gum infection. The relationship in Panel VIII cannot be assessed with this approach since none of the estimated treatment effects in Columns (1) and (2) are statistically significant. However, in Panel IX, respondents with lower HL are 5.2 percentage points more likely to have at least 6 permanent teeth removed for tooth decay or gum infection in Column (1) and that estimate drops to 2 percentage points after including covariates. However, the bounds on the estimated bias-adjusted treatment effect includes zero (see Column (3)). Moreover, the estimated coefficient of proportionality in Column (4) is less than one. Columns (3) and (4) indicate the estimated direct relationship between HL and the probability of having at least 6 teeth removed is not robust.

Tobacco use. In Panel X, for the probability of having smoked at least 100 cigarettes in entire life, the estimated treatment effect drops from 4.7 percentage points in Column (1) to 1.5 percentage points in Column (2) although the latter is imprecisely estimated. Thus it is infeasible to assess the robustness of the relationship in our sample using this empirical approach. Similarly, we cannot assess the omitted variable bias-adjusted treatment effect for the probability of smoking every day in Panel XI.

Obesity. Similar to Panels X and XI, we cannot assess the omitted variable bias-adjusted treatment effect for the likelihood of being obese in Panel XII.

In sum, the sensitivity analyses produce identical results as with the binary indicator of limited health literacy for all outcomes but one. Using an index for varying levels of health literacy, we conclude there is a direct robust relationship between likelihood of visiting a dentist or dental clinic in the past 5 years and health literacy. Using a binary limited health literacy indicator in Section 5.3, we could not ascertain a relationship since the estimated coefficients were imprecise. For all other outcomes, we arrive at the same conclusion: there exists a direct causal effect of lower levels of health literacy or limited health literacy on adverse health behaviors such as not exercising, sleeping less than 6 or 7 hours a day, on average, and not keeping up with routine medical and dental checkups. Additionally, we cannot assess the relationship between HL and health outcomes such as obesity and poor oral health.

7 Conclusion

Limited health literacy is a public health concern since it adversely affects a population's ability to follow and utilize necessary and often crucial health advice and recommended behavioral changes. Our findings suggest that low health literacy is relatively common in North Carolina - more than one in six adults experiences it. Furthermore, we find that low health literacy has a causal relationship with several important health behaviors that impact quality of life and influence the risk of health outcomes like chronic diseases. Specifically, people with low health literacy are less likely to exercise and have adequate sleep. They also are less likely to utilize recommended health care services, namely regular check-ups.

However, it is worth noting all data in this study are self-reported and are therefore subject to measurement error and recall biases. Also, this study used data from only a single state and therefore may not represent the experiences of other adults in the US, and also may not represent the experience of all people living across the 13-state Appalachian region.

Nonetheless, this study contributes to the growing literature on the importance of health literacy and the dire consequences of its inadequacy in the general public. The first contribution constitutes defining limited health literacy based on three newly introduced questions in the optional module in the NC chapter of the BRFSS. The questions are in reverse order compared to other states that introduced these questions earlier. The second contribution pertains to estimating the causal effect of limited health literacy on several adverse behavioral outcomes instead of mere associations. To that end, we applied a novel omitted variable bias-adjusted treatment effect estimator following Oster (2017). Our findings are interesting. First, limited health literacy has a direct causal impact on not exercising, inadequate sleep, and irregular health and dental checkup. This finding corroborates existing literature on associations between limited health literacy and the outcomes. Surprisingly, in our sample of NC residents along with the estimator of choice, we do not find any causal relationship between smoking and alcohol consumption and limited health literacy. Second, and somewhat unsurprisingly, we do not find any causal link between limited health literacy and health outcomes such as obesity and teeth removal due to tooth decay or gum infection. Further research connecting the causal relationship between limited health literacy and adverse health behaviors such as not exercising, inadequate sleep, and irregular health and dental checkups to adverse health outcomes such as obesity and teeth removal due to tooth decay or gum infection is warranted.

8 Compliance with Ethical Standards

Funding: The study received no external funding.

Conflict of Interest: The authors declare that they have no conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.We conducted secondary data analyses using the 2016 NC BRFSS. This study was reviewed and considered exempt by the authors' University's Institutional Review Board (IRB #18-0071).

References

[1] Appalachian Regional Commission. Health Disparities in Appalachia. Washington, DC, USA: 2017.

[2] Angrist, J. D., and Pischke, J.-S. (2010), "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics," Journal of Economic Perspectives, 24, 3–30. [1,3,5,10,17]

[3] Batterman R, Hawkins M, Collins P, Buchbinder R, Osborne R. Health literacy: applying current concepts to improve health services and reduce health inequalities. Public Health. 2016;132:3-12.

[4] Bauer U, Briss P, Goodman R, Bowman B. Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA. Lancet. 2014;384:45-52.

[5] Behringer B, Friedell G. Appalachia: where place matters to health. Preventing Chronic Disease:. 2006;3(4).

[6] Bennett I, Chen J, Soroui J, White S. The contribution of health literacy to disparities in self-rated health status and preventive health behaviors in older adults. Annals of Family Medicine. 20009;7(3):204-11.

[7] Berkman N, Sheridan S, Donahue K, Halpern D, Crotty K. Low Health Literacy and Health Outcomes: An Updated Systematic Review. Annals of Internal Medicine. 2011;155(2):97-107.

[8] Bolin J, Bellamy G, Ferdinand A, Vuong A, Kash B, Schulze A, et al. Rural healthy people 2020: new decade, same challenges. The Journal of Rural Health. 2015;31(3):326-33.

[9] Koh H, Berwick D, Clancy C, Baur C, Brach C, Harris L, et al. New Federal Policy Initiatives To Boost Health Literacy Can Help the Nation Move Beyond The Cycle of Costly 'Crisis Care'. Health Affairs. 2012;31(2):434-43.

[10] Lee J, Divaris K, Baker A, Rozier R, WF Vann J. The Relationship of Oral Health Literacy and Self-Efficacy With Oral Health Status and Dental Neglect. American Journal of Public Health. 2012;102:923-9.

[11] Mokdad AH. The Behavioral Risk Factors Surveillance System: past, present, and future. Annu Rev Public Health. 2009;30:43-54. doi:10.1146/annurev.publhealth.031308.100226.

[12] Pleasant A. Advancing health literacy measurement: a pathway to better health and health system performance. Journal of health communication. 2014;19(12):1481-96.

[13] Rasu R, Bawa W, Suminski R, Snella K, Warady B. Health literacy impact on national healthcare utilization and expenditure. International journal of health policy and management. 2015;4(11).

[14] Rudd R. Improving Americans' health literacy. New England Journal of Medicine. 2010;363(24):2283-5.

[15] Schillinger D, Barton L, Karter A, Wang F, Adler N. Does literacy mediate the relationship between education and health outcomes? A study of a low-income population with diabetes. Public Health Reports. 2006;121(3):245-54.

[16] Singh G, Kogan M, Slifkin R. Widening Disparities in Infant Mortality and Life Expectancy Between Appalachia and the rest of the United States, 1990-2013. Health Affairs. 2017;36(8):1423-32.

[17] Services USDoHaH. Healthy People 2030 Framework 2018 [cited 2018 July 25]. Available from: https://www.healthypeople.gov/2020/About-Healthy-People/Development-Healthy-People-2030/Proposed-Framework.

[18] Services USDoHaH. America's Health Literacy: Why We Need Accessible Health Information. 2008.

[19] Sorensen K, Broucke SVd, Fullam J, Doyle G, Pelikan J, Slonska Z, et al. Health literacy and public health: A systematic review and integration of definitions and models. BMC Public Health. 2012;12.

[20] Oster, E. (2017). Unobservable selection and coefficient stability: Theory and evidence. Journal of Business & Economic Statistics. Published online https://doi.org/10.1080/07350015.2016.1227711

[21] Wewers M, Fickle D, Paskett E. Risky Behaviors Among Ohio Appalachian Adults. Preventing Chronic Disease:. 2006;3(4).

Table 1. Summary statistics

Variable	Obs	Mean	Std. Dev.	By LHL Difference in means	By Appalachian County Difference in means
Treatment					
Limited health literacy	6,396	0.168	0.374		0.031†
Behavior and health outcomes					
No exercise (1= no physical activity past month)	6,396	0.231	0.421	0.142*	0.033†
Medical checkup (1=check up before 2 yrs/never)	6,325	0.149	0.356	0.079*	0.009
Sleep < 6 (1 = sleep < 6 hours)	6,396	0.120	0.325	0.071*	0.006
Sleep <7 (1 = sleep < 7 hours)	6,396	0.335	0.472	0.083*	0.013
Oral health					
Dental visit 1 (1=dental visit within 1 year)	6,396	0.631	0.483	-0.223*	-0.022
Dental visit 2 (1=dental visit within 2 years)	6,396	0.110	0.313	0.036*	-0.012
Dental visit 5 (1=dental visit within 5 years)	6,396	0.113	0.317	0.044*	0.000
Dental visit 5+ (1=dental visit more than 5 years ago)	6,396	0.127	0.333	0.114*	0.038*
Teeth Removed 5 (1=1-5 permanent teeth removed)	6,396	0.288	0.453	0.029‡	-0.021
Teeth Removed 6 (1= > 6 permanent teeth removed)	6,396	0.120	0.326	0.104*	0.018‡
Teeth Removed All (1=all permanent teeth removed)	6,396	0.059	0.235	0.068*	0.024*
Tobacco use					
Smoke 100 (1=smoked >=100 cigarettes in life)	6,396	0.420	0.494	0.098*	0.062*
Smoke everyday (1=now smoke every day)	6,396	0.124	0.330	0.067*	0.017
Smoke somedays (1=now smoke some days)	6,396	0.051	0.220	0.017†	-0.002
No smoking (1=now not smoking at all)	6,396	0.245	0.430	0.014	0.046*
Cigar (1=smoke cigar)	6,396	0.039	0.193	0.028*	0.009
Hookah (1=smoke hookah)	6,396	0.015	0.123	0.017*	0.005
Alcohol consumption					
Drinks per week (# drinks/wk past 30 days)	6,396	0.471	1.373	-0.146*	-0.129*
Drinks per month (# drinks/mnth past 30 days)	6,396	2.657	6.541	-0.475†	0.232
Binge drinker (1=binge drinker)	6,396	0.139	0.346	0.006	-0.021‡
Health outcome					
Obese (1 = obese)	6,396	0.294	0.456	0.042*	0.017
Covariates					
Demographics					
Age (reported age in years)	6,396	47.139	18.019	2.419*	2.785*
Male (1=male)	6,396	0.482	0.500	0.057*	0.012
White (1=white)	6,396	0.658	0.474	-0.072*	0.194*
Non-white (1=non-white)	6,396	0.291	0.454	0.072*	-0.160*
Missing race (1= race is missing)	6,396	0.009	0.092	-0.001	-0.005
Married (1=married)	6,380	0.553	0.497	-0.096*	-0.010
Education					
Less than HS (1=education less than high school)	6,396	0.146	0.353	0.227*	0.054*
High school (1=high school degree)	6,396	0.272	0.445	0.061*	0.013
Some college (1=some years in college)	6,396	0.326	0.469	-0.086*	-0.010
College (1=college degree)	6,396	0.255	0.436	-0.203*	-0.055*
Missing education (1=education is missing)	6,396	0.002	0.043	0.001	-0.002
Economic factors					
Employed (1=employed)	6,373	0.565	0.496	-0.131*	-0.030‡
Health insurance (1=has health insurance)	6,376	0.862	0.345	-0.106*	0.010
Inc < $15,000 (1=annual income less than $15,000)	6,396	0.089	0.285	0.086*	0.041*
Inc < $25,000 (1=annual income less than $25,000)	6,396	0.159	0.366	0.100*	0.041*
Inc < $50,000 (1=annual income less than $50,000)	6,396	0.200	0.400	-0.012	-0.021
Inc < $75,000 (1=annual income less than $75,000)	6,396	0.133	0.340	-0.080*	0.010
Inc > $75,000 (1=annual income exceeds $75,000)	6,396	0.236	0.425	-0.140*	-0.076*
Missing income (1=income is missing)	6,396	0.182	0.386	0.045*	0.004
Household characteristics					
Number of children in household	1,713	2.897	0.997	0.096	0.022
Own home (1=home owner)	6,396	0.669	0.471	-0.075*	0.037†
Rent (1=renter)	6,396	0.247	0.431	0.047*	-0.051*
Other home (1=other home type -mobile trailer etc.)	6,396	0.079	0.269	0.025*	0.018†
Regional characteristics					
Appalachian County (1=Appalachian county)	6,396	1.822	0.382	0.032†	

Note: Test of equality of means across respondents with LHL and those wtihout are conducted by Chi-square tests.Statistical significance: * p<0.01; † p<0.05; ‡ p<0.10. BRFSS survey weights are utilized. Outcomes and covariates with N < 6369 observations have missing information.

Table 2. Sensitivity of OLS coefficient and R-squared to controls

Outcomes		(1) No controls	(2) Uncontrolled Model: 2 controls	(3) Additional controls	(4) Full/Controlled Model
I. P(No exercise)					
Limited Health Literacy	*Estimate*	0.142*	0.137*	0.075*	0.072*
	Robust S.E.	(0.020)	(0.019)	(0.020)	(0.020)
	R^2	0.016	0.037	0.082	0.084
II. P(Medical check up more than 2 years ago)					
Limited Health Literacy	*Estimate*	0.079*	0.084*	0.055*	0.054*
	Robust S.E.	(0.018)	(0.017)	(0.017)	(0.017)
	R^2	0.0067	0.058	0.1205	0.133
III. P(Sleep < 6 hours)					
Limited Health Literacy	*Estimate*	0.071*	0.073*	0.049*	0.046*
	Robust S.E.	(0.016)	(0.016)	(0.017)	(0.016)
	R^2	0.0067	0.008	0.0244	0.0304
IV. P(Sleep < 7 hours)					
Limited Health Literacy	*Estimate*	0.083*	0.085*	0.075*	0.069*
	Robust S.E.	(0.021)	(0.021)	(0.022)	(0.021)
	R^2	0.0043	0.0083	0.0206	0.0345
V. P(Dental visit in past 2 years)					
Limited Health Literacy	*Estimate*	0.036†	0.038†	0.025	0.025
	Robust S.E.	(0.015)	(0.015)	(0.016)	(0.016)
	R^2	0.0018	0.0037	0.0118	0.014
VI. P(Dental visit in past 5 years)					
Limited Health Literacy	*Estimate*	0.044*	0.046*	0.021	0.019
	Robust S.E.	(0.015)	(0.015)	(0.016)	(0.015)
	R^2	0.0027	0.0073	0.0254	0.028
VII. P(Dental visit more than 5 years ago)					
Limited Health Literacy	*Estimate*	0.114*	0.108*	0.041†	0.038†
	Robust S.E.	(0.017)	(0.017)	(0.016)	(0.016)
	R^2	0.0165	0.029	0.0997	0.104
VIII. P(Five teeth removed)					
Limited Health Literacy	*Estimate*	0.029	0.021	0.004	-0.000
	Robust S.E.	(0.020)	(0.020)	(0.020)	(0.020)
	R^2	0.0006	0.016	0.0325	0.057
IX. P(Six teeth removed)					
Limited Health Literacy	*Estimate*	0.104*	0.093*	0.047*	0.046*
	Robust S.E.	(0.016)	(0.015)	(0.015)	(0.015)
	R^2	0.0143	0.086	0.1236	0.127
X. P(Smoked at least 100 cigarettes in life)					
Limited Health Literacy	*Estimate*	0.098*	0.081*	0.040‡	0.033
	Robust S.E.	(0.021)	(0.021)	(0.021)	(0.021)
	R^2	0.0055	0.041	0.0736	0.112
XI. P(Smoking everyday)					
Limited Health Literacy	*Estimate*	0.067*	0.068*	0.022	0.016
	Robust S.E.	(0.016)	(0.016)	(0.017)	(0.016)
	R^2	0.0057	0.011	0.0544	0.088
XII. P(Obese)					
Limited Health Literacy	*Estimate*	0.042†	0.039†	0.021	0.015
	Robust S.E.	(0.020)	(0.020)	(0.020)	(0.020)
	R^2	0.0012	0.004	0.019	0.046

*Notes: Column (1) has no controls. Column (2) includes indicators for male and age. Column (3) augments Column (2) with education indicators for less than HS, some college, college degree, income indicators for less than $15,000, $25,000, $50,000, $75,000, and income exceeding $75,000. The model also includes indicators for owning and renting home. Column (4) additionally includes age squared, age cubed, interaction term between male and age, indicators for white and non-white, an indicator for married, and an indicator for Appalachian counties. Columns (3) and (4) for medical check up more than 2 years ago includes an indicator for any health insurance coverage. All analyses are weighted using BRFSS survey weights. Statistical significance: * p<0.01; † p<0.05; ‡ p<0.10.

Table 3. Bias-adjusted effect of limited health literacy on adverse health behaviors/outcomes

Outcomes		(1) Uncontrolled Model	(2) Controlled Model	(3) Identified Set	(4) δ for β=0 given R_{max}=1.3*$R_{controlled}$
I. P(No exercise)					
Limited Health Literacy	*Estimate*	0.137*	0.072*	[0.035, 0.072]	1.820
	Robust S.E.	(0.019)	(0.020)		
	R^2	0.037	0.084		
II. P(Medical check up more than 2 years ago)					
Limited Health Literacy	*Estimate*	0.084*	0.054*	[0.036, 0.054]	2.686
	Robust S.E.	(0.017)	(0.017)		
	R^2	0.058	0.133		
III. P(Sleep < 6 hours)					
Limited Health Literacy	*Estimate*	0.073*	0.046*	[0.035, 0.046]	3.201
	Robust S.E.	(0.016)	(0.016)		
	R^2	0.008	0.030		
IV. P(Sleep < 7 hours)					
Limited Health Literacy	*Estimate*	0.085*	0.069*	[0.061, 0.0689]	5.465
	Robust S.E.	(0.021)	(0.021)		
	R^2	0.008	0.035		
V. P(Dental visit in past 2 years)					
Limited Health Literacy	*Estimate*	0.038†	0.025	[0.019, 0.025]	3.675
	Robust S.E.	(0.015)	(0.016)		
	R^2	0.004	0.014		
VI. P(Dental visit in past 5 years)					
Limited Health Literacy	*Estimate*	0.046*	0.019	[0.008, 0.019]	1.653
	Robust S.E.	(0.015)	(0.015)		
	R^2	0.007	0.028		
VII. P(Dental visit more than 5 years ago)					
Limited Health Literacy	*Estimate*	0.108*	0.038†	[0.009, 0.038]	1.293
	Robust S.E.	(0.017)	(0.016)		
	R^2	0.029	0.104		
VIII. P(Five teeth removed)					
Limited Health Literacy	*Estimate*	0.021	-0.000	[-0.010, -0.000]	-0.020
	Robust S.E.	(0.020)	(0.020)		
	R^2	0.016	0.057		
IX. P(Six teeth removed)					
Limited Health Literacy	*Estimate*	0.093*	0.046*	[-0.000, 0.046]	0.991
	Robust S.E.	(0.015)	(0.015)		
	R^2	0.086	0.127		
X. P(Smoked at least 100 cigarettes in life)					
Limited Health Literacy	*Estimate*	0.081*	0.033	[0.008, 0.0330]	1.286
	Robust S.E.	(0.021)	(0.021)		
	R^2	0.041	0.112		
XI. P(Smoking everyday)					
Limited Health Literacy	*Estimate*	0.068*	0.016	[-0.003, 0.0161]	0.861
	Robust S.E.	(0.016)	(0.016)		
	R^2	0.011	0.088		
XII. P(Obese)					
Limited Health Literacy	*Estimate*	0.039†	0.015	[0.007, 0.015]	1.796
	Robust S.E.	(0.020)	(0.020)		
	R^2	0.004	0.046		

*Notes: Uncontrolled model includes age and an indicator for male. The controlled model additionally includes age squared, age cubed, intercation term between male and age, indicatrs for white and non-white, indicator for married, education indicators for less than HS, some college, college degree, income indicators for less than $15,000, $25,000, $50,000, $75,000, and income exceeding $75,000. The model also includes indicators for owning and renting home, and an indicator for Appalachian counties. The omitted categories for race is missing race, for education is HS degree, for income is missing income, and for home type is Other (mobile trailer etc.). The full model for medical check up more than 2 years ago includes an indiactor for any health insurance coverage. All analyses are weighted using BRFSS survey weights. Statistical significance: * p<0.01; † p<0.05; ‡ p<0.10. Model estimates are generated using user written Stata command -*psacalc*-.

Table 4. Bias-adjusted effect of health literacy on adverse health behaviors/outcomes

Treatment Variable: Health Literacy Index

Outcomes		(1) Uncontrolled Model	(2) Controlled Model	(3) Identified Set	(4) δ for β=0 given R_{max}=1.3*$R_{controlled}$
I. P(No exercise)					
Health Literacy	*Estimate*	0.082*	0.039*	[0.014,0.039]	1.50992
	Robust S.E.	(0.013)	(0.013)		
	R^2	0.036	0.083		
II. P(Medical check up more than 2 years ago)					
Health Literacy	*Estimate*	0.056*	0.034*	[0.020, 0.034]	2.20143
	Robust S.E.	(0.012)	(0.011)		
	R^2	0.061	0.133		
III. P(Sleep < 6 hours)					
Health Literacy	*Estimate*	0.049*	0.030*	[0.022, 0.030]	2.74321
	Robust S.E.	(0.011)	(0.011)		
	R^2	0.009	0.031		
IV. P(Sleep < 7 hours)					
Health Literacy	*Estimate*	0.054*	0.043*	[0.038, 0.043]	5.45476
	Robust S.E.	(0.013)	(0.014)		
	R^2	0.009	0.035		
V. P(dental visit in past 2 years)					
Health Literacy	*Estimate*	0.013	0.005	[0.001, 0.005]	1.31586
	Robust S.E.	(0.009)	(0.009)		
	R^2	0.002	0.013		
VI. P(dental visit in past 5 years)					
Health Literacy	*Estimate*	0.036*	0.019‡	[0.012, 0.019]	2.22004
	Robust S.E.	(0.010)	(0.011)		
	R^2	0.009	0.029		
VII. P(dental visit more than 5 years ago)					
Health Literacy	*Estimate*	0.080*	0.034*	[0.013, 0.034]	1.52804
	Robust S.E.	(0.012)	(0.012)		
	R^2	0.035	0.106		
VIII. P(Five teeth removed)					
Health Literacy	*Estimate*	0.007	-0.008	[-0.016, -0.008]	-1.20425
	Robust S.E.	(0.012)	(0.013)		
	R^2	0.015	0.058		
IX. P(Six teeth removed)					
Health Literacy	*Estimate*	0.052*	0.020‡	[-0.012, 0.020]	0.64356
	Robust S.E.	(0.010)	(0.010)		
	R^2	0.084	0.126		
X. P(Smoked at least 100 cigarettes in life)					
Health Literacy	*Estimate*	0.047*	0.015	[-0.003, 0.015]	0.84182
	Robust S.E.	(0.013)	(0.013)		
	R^2	0.040	0.112		
XI. P(Smoking everyday)					
Health Literacy	*Estimate*	0.038*	0.002	[-0.011, 0.002]	0.15897
	Robust S.E.	(0.010)	(0.010)		
	R^2	0.010	0.088		
XII. P(Obese)					
Health Literacy	*Estimate*	0.033†	0.018	[0.012, 0.018]	2.84492
	Robust S.E.	(0.013)	(0.013)		
	R^2	0.005	0.047		

*Notes: Uncontrolled model includes age and an indicator for male. The controlled model additionally includes age squared, age cubed, intercation term between male and age, indicatrs for white and non-white, indicator for married, education indicators for less than HS, some college, college degree, income indicators for less than $15,000, $25,000, $50,000, $75,000, and income exceeding $75,000. The model also includes indicators for owning and renting home, and an indicator for Appalachian counties. The omitted categories for race is missing race, for education is HS degree, for income is missing income, and for home type is Other (mobile trailer etc.). The full model for medical check up more than 2 years ago includes an indicator for any health insurance coverage. All analyses are weighted using BRFSS survey weights. *Statistical significance* : * p<0.01; † p<0.05; ‡ p<0.10. Model estimates are generated using user written Stata command -*psacalc* -.