Spring 5-5-2018

The Effect of College Degree Attainment on Risky Health Behaviors: Smoking, Obesity, and Vaping.

Sal R. Ansari
CUNY Hunter College

The Effect of College Degree Attainment on Risky Health Behaviors: Smoking, Obesity, and Vaping.

by

Sal R. Ansari

Submitted in partial fulfillment
of the requirements for the degree of
Master of Arts Economics, Hunter College
The City University of New York

2018

Thesis Sponsor:

May 5, 2018	Purvi Sevak
Date	Signature
May 5, 2018	Kenneth J. McLaughlin
Date	Signature of Second Reader

Table of Contents

List of Figures

List of Tables

Abstract

This paper tests the hypothesis that completing college results in better health outcomes through the causal pathway of risky health behaviors. I analyze data from the 2016 Behavioral Risk Factor Surveillance System (BRFSS 2016) to explore the effect college education has on an individual's decision to smoke, e-smoke, and tendency to have a healthy body mass index (BMI). The results imply that completing college reduces the likelihood an individual is a smoker or e-smoker. The results do not show any causal relationship between college education and having a healthy BMI.

1 Introduction

Aside from the obvious advantages that completing higher education has, such as increased lifetime earnings, there may be many other benefits that arise from attaining more education. This paper focuses on aversion from risky health behaviors as a potential pathway by which health outcomes are affected. Specifically, I estimate the effect of college degree attainment on smoking, obesity, and vaping & e-cigarettes. The first two behaviors, smoking and obesity, are important as both are known to lead to heart disease, the number one cause of premature death in the United States (Center for Disease Control). Evaluation of a causal relationship between these health outcomes and education will aid policy makers when evaluating the importance of public education. The third health behavior, e-cigarette use and vaping, is a relatively new trend and has only recently been added to various datasets. Although the health effects of this new trend are unclear, finding a potential relationship between schooling and vaping will be useful looking forward as medical research in this area expands.

This study uses an instrumental variable approach that takes advantage of the variation of high school graduation requirements, known as Carnegie Units, across and within states. In states with high Carnegie Unit requirements, students are less likely to complete college; meanwhile, in states with low Carnegie Unit requirements, students are more likely to complete college. This variation in graduation requirements enables a two stage least squares model to identify the causal effect of college completion on various risky health behaviors. The results, although only borderline statistically significant, suggest that completing college reduces an individual's likelihood to smoke and e-smoke. The results do not show any significant causal relationship between college education and having a healthy BMI.

2 Literature Review

The effect of education on health is one of the most well-documented benefits of education[1]. Education's effect on health is often split into two categories, direct effects and indirect effects. As by Grossman explores, education could potentially have a direct effect on health and health behaviors through its effect on productive and allocative efficiency of health inputs (Grossman, 2005). As education increases, knowledge of general health information increases as well. This causes a shift in the production function of health and could also change the allocation of different health inputs.

The indirect effects of education on health are far broader. The least controversial indirect effect is related to the effect of education on labor market opportunities. Higher attainment of education results in higher rates of employment and improved earnings (Card, 1999). The increased earnings could affect health by potentially making health improving goods such as a gym membership more affordable or by increasing access to medical care through employer-based insurance. Another indirect effect proposed is that more educated people simply work in safer environments (Cutler & Lleras-Muney, 2006). Also, more educated individuals could potentially be exposed to healthier colleagues who affect their own health through peer effects (Gaviria & Raphael, 2001).

Although not a causal pathway, one explanation for a correlation between education and health outcomes is the patience levels of individuals. More educated individuals could naturally be more patient given their decision to pursue education and thus partake in healthier activities that don't supply individuals with instant gratification (Fuchs, 1982). For example, more patient

[1] For a more comprehensive literature review, see Education and Health: Evaluating Theories and Evidence by Cutler, David M, & Lleras-Muney, Adriana. 2006

individuals may be more likely to invest in education but also more patient when withstanding the withdrawal symptoms of smoking cessation. More patient individuals may also have more perseverance in exercise habits despite not seeing immediate results. Despite the present correlation between college completion and these two actions, college completion is not directly affecting an individual's decision to smoke or exercise.

Education's effect on risky health behaviors could also be indirect. Previous research shows that both high school completion and GED attainment results in decreased smoking rates, with GED attainment on a smaller magnitude (Kenkel et al., 2006). The research also shows that individuals with a college education are less likely to smoke and more likely to be former smokers (de Walque, 2004). There is however, very little evidence showing a relationship between high school completion and obesity (Kenkel et al., 2006).

I contribute the existing literature by evaluating the effect of college completion on cigarette use, e-cigarette use, and BMI. Although Kenkel, Lillard and Mathios address the effect of high school education and GED attainment on smoking and obesity, high school education is an expectation in the current society. College education is a far more popular topic when it comes to policy implications. Many recent political conversations in the 2016 election centered around the idea of tuition-free public higher education. Although the idea has traction with some policy makers, it's far from a unanimous ideology. If my research can demonstrate a causal relationship between college education and a healthier society through lower rates of risky health behaviors, it could potentially give policy makers another factor to evaluate for public higher education policy. This research also goes beyond de Walque's research by using a newer dataset. The updated dataset allows us to see this relationship in light of the changing cultural view of different risky health behaviors.

3 Research Questions

With these policy implications in mind, I evaluate risky health behaviors closely related to health outcomes. My analysis looks at the effect of college completion on cigarette use, obesity, and vaping. The first behavior, smoking, has many well documented adverse effects. Smoking causes more deaths per year than HIV, illegal drug use, alcohol use, motor vehicle injuries, and firearm related incidents combined (Center for Disease Control). The second dependent variable is overweight or obesity, as measured by body mass index (BMI) above a threshold value. Prior empirical literature does not show a significant causal relationship between education and a healthy BMI. The third health behavior is vaping and e-cigarette usage. E-cigarettes are a recent invention and its health effects have yet to be documented. Despite the ambiguous effect e-cigarettes have on health, finding a potential relationship between higher education and e-smoking will be useful looking forward as it receives more study in the medical sector.

There is a well-documented negative correlation between education and each of these risky health behaviors. This correlation is potentially explained by the following hypotheses:

1. Educational attainment directly causes individuals to be less likely to partake in risky health behaviors
2. Higher earnings from education causes individuals to be more likely to partake in risky health behaviors as part of an income effect
3. Individuals who attend college are naturally less likely to partake in risky health behaviors.

4 Methods

4.1 Data

To test these hypotheses I combine data from two sources. The first data source is from the 2016 Behavioral Risk Factor Surveillance System BRFSS, compiled by the Center for Disease Control (CDC). The BRFSS is a telephone survey that collects data on U.S. state residents and their risky health behaviors. The initial cross-sectional dataset has a total of 486,303 observations. The BRFSS contains observations of individuals older than 18 as well as individuals that live in U.S. territories. My instrument, high school graduation requirements, is only relevant to individuals between the ages of 25 to 50 that live in the United States. After I drop observations that are outside of the scope of my instrument the sample is reduced to 116,069 observations. I also drop observations with missing information for independent or key dependent variables depending on the outcome variable of the regression. My final samples of analysis consist of about 90,000 observations

The second key data are Carnegie Unit requirements across states and years[2]. The Carnegie Unit is the standardized measure of high school graduation requirements. A Carnegie Unit is defined as 120 hours of class or contact time with an instructor. Figure 1 in the appendix illustrates the difference in Carnegie Unit requirements throughout the United States and how it's evolved over time.

The complete list of Carnegie Unit requirement policies spans from 1980 to 2006. Within this time period, the Carnegie Unit requirements change six times; however, the changes are not uniform across states. Some states, like Rhode Island, has a constant 16 unit requirement from the

[2] Carnegie unit requirement data was taken from the National Center for Education Statistics website nces.ed.gov

years 1980 to 2004. Other states, such as Vermont, fluctuates between 12 units and 20 units in that same time period.

4.2 Challenges of Endogeneity

Estimating the relationship between risky health behaviors and schooling is unfortunately not as straightforward as an Ordinary Least Squares (OLS) regression. An OLS model attempting to explain the relationship between a risky health behavior and college education will suffer from endogeneity. Endogeneity occurs when correlation exists between the independent variable and the error term. Because education is not randomly assigned, an individual's decision to pursue college education and their decision to engage in a risky health behavior could be correlated.

Another threat to modeling this relationship is omitted variable bias. Omitted variable bias is a form of endogeneity that occurs when there is an important control variable missing in a regression. Even with a rich set of controls, an OLS estimation of this relationship will have several confounding variables. Because several these confounding variables are not easily measured, the model will suffer from omitted variable bias and our estimates would not be an accurate representation of college's effect on risky health behaviors. For example, more patient individuals could be more likely to forgo present income to attend college; meanwhile, more patient individuals could also be more likely to exercise despite not realizing a present benefit. The inability to control for patience would result in the estimated effect of college to account for an individual's patience level instead of purely the effect of college.

4.3 Two Stage Least Squares (2SLS)

To combat the endogeneity of college attainment, I use an instrumental variable (IV) in a two stage least squares (2SLS) regression model. The two requirements for a valid instrument are instrument exogeneity and instrument relevance (Bound et al., 1995). Instrument exogeneity requires that the instrument is not correlated to the to the second stage dependent variable of interest in any way other than the endogenous variable of interest. Failure to satisfy this requirement results in biased estimates in the second stage regression (Wooldridge, 2012). Thankfully, Carnegie Unit requirements satisfy this first requirement because they do not directly affect individual health behaviors except through individuals schooling. The second requirement, instrument relevance, requires that the instrument is highly correlated to the endogenous independent variable. Failure to satisfy this requirement results in larger than normal standard errors (Wooldridge, 2012). This can be tested through the instrument's significance in the first stage results (Staiger and Stock, 1997). In section 5.1, I show that the Carnegie Unit requirements significantly affect college attainment and satisfy this requirement

To take advantage of this instrument, I utilize the variation in Carnegie Unit requirements across states and within states of high school attendance. Across state variation is present because Carnegie Unit requirement policy is dictated at the state level. For example, two individuals who went to high school in the same year may not be subject to the same requirement if they attended high school in different states. In 2006, an individual in Alabama was required to complete 24 Carnegie Units; however, an individual in Rhode Island was required to complete 18 Carnegie Units. In state variation in Carnegie Units is present because of the policy changes that occurred between the years 1980 and 2006. For example, an individual who attended high school in 1980 West Virginia was required to complete 18 Carnegie Units; meanwhile, an individual who attended high school in 2006 West Virginia was required to complete 24 Carnegie Units.

The available data have a few limitations. The first limitation is the lack of Carnegie Unit data before 1980. The Carnegie Unit system has been around for about a century; however, because I only have a record of state requirements as far back as 1980, I drop observations of individuals who are older than 49 years old (because they attended high school prior to 1980).

The second limitation is the lack of information regarding the exact Carnegie Unit policy that influenced individuals. To combat this, I assume individuals are only affected by graduation policies that were in effect their first year of high school. For example, someone who is 29 years old in the 2016 BRFSS entered high school 15 years ago. Therefore, this individual entered high school in 2001; however, the two changes in graduation requirements around 2001 was in the years 1998 and 2002. Thus, the 29-year-old individual was only affected by the graduation requirements enacted in the year 1998.

The third limitation is the reported age grouping of the 2016 BRFSS. Because the 2016 BRFSS groups observations into reported age groups instead of reporting exact ages, I apply the Carnegie Unit policy associated with the oldest of the age group. Following the previous example, the 25-29 age group will be associated with the graduation requirement policy that affects the 29 year olds. The reasoning behind applying the older policy is driven by the observation that graduation requirements tended to increase as time went on, therefore applying the older policy to the younger individuals in an age group was a conservative decision. In practice, the difference of 4 years didn't affect much when analyzing changes in graduation requirements.

The last limitation is the lack of information on an individual's state of high school attendance. The BRFSS only reports current state of residence. Although some individuals may not be living in the same state they attended high school, I assume they did not migrate and apply the Carnegie Unit requirement associated with an individual's current state of residence.

Considering these limitations, I apply the Carnegie Unit requirements that were in place for the oldest person in the age group in the year they turned fourteen. Typically, this process introduces error in the independent variable and could lead to bias. Thankfully, the instrumental variable method solves this issue; however, it will impact the efficiency of the estimates.

Using the Carnegie Unit requirements as an instrument, I estimate two-stage least squares models of equation (1) and (2) below. While I use the standard term "two-stage least squares" I estimate these models in one step using Stata's "ivregress" command. In the first equation, I regress college on required Carnegie Units, age, race/ethnicity, and sex, with detailed census region fixed effects.

$$\mathrm{college_i} = \beta_0 + \beta_1 \mathrm{CarnegieUnits_i} + \beta_n \mathrm{X_i} + \sum_{r=1}^{9} R_r + \mathrm{U_i} \tag{1}$$

This approach assumes that the Carnegie Units are exogenous and do not reflect any unobservable state factors that may affect health behaviors. To address this, one could leverage within state variation in Carnegie Units over time. To see if this is feasible I estimate a version of equation (1) that replaces detailed census region fixed effects with state fixed effects. Unfortunately, it does not meet the test for a strong instrument.

In the second equation (2), the instrumented variable for college is the independent variable and the risky health behavior is the dependent variable. I also include demographic variables, X, to control for differences in age, race/ ethnicity, and sex and detailed Census region fixed effects, R.

$$\mathrm{RHB_i} = \beta_0 + \beta_1 \widehat{\mathrm{college}}_\mathrm{i} + \beta_n \mathrm{X_i} + \sum_{r=1}^{9} R_r + \mathrm{U_i} \tag{2}$$

RHB is one of three risky health behaviors: smoking behavior, e-smoking behavior, and overweight or obese weight. One major concern is that different areas in the United States have systematic trends in their health behaviors. For example, researchers have identified a 'stroke belt' in the south (Center for Disease Control). To control for this, I include control variables for detailed census regions in my second stage equation. Summary statistics containing the mean and standard deviation of my key dependent and independent variables can be found in Table 1.

5 Results

5.1 First Stage Results

My first stage results are presented in table 2. The first column reports the effect of Carnegie Units required on college completion without any covariates. The second column reports the effect of Carnegie Units required on college completion with all the covariates I use for my second stage regression, including control variables for detailed census regions. Column 3 reports the effect of Carnegie Units on college completion with added state fixed effects. Adding the covariates reduces the F-statistic of Carnegie Units required from 195.96 to 39.90. Although the decrease is great, 39.90 is still greater than 10 and therefore passes the test of a weak instrument (Staiger and Stock, 1997). The results in column 2 suggest that for each marginal increase in Carnegie Units required, an individual's probability to complete college decreases by 0.5%. Despite the seemingly small magnitude, the true impact of Carnegie Units required is greater when considering that some states have differences in requirements as great as 20 units. Differences of this magnitude result in a 10% reduction in college completion. The insignificant estimates in column 3 suggest that Carnegie Unit requirements within state lack the variation sufficient to identify causal relationships. While Carnegie Units do vary over time within state, these results suggest that they do not vary enough to allow an IV approach that leverages only within state variation[3].

5.2 Second Stage Results

I organize my second stage results using three separate tables, tables 3 – 5. Each table has results grouped according to health behavior. Columns 1 and 2 of each table report results from a linear probability OLS model. The model includes the various controls for age, race, and sex, shown below the key independent variable, college. The model also includes region fixed effects for the 9 different detailed census regions. The base race is white. The base age group is 25-29. And the base region is the northeast. Columns 3 and 4 of each table report results from a two-stage least squares model using high school graduation requirements as an instrument for college degree attainment.

The F-statistics for the significance test of my instrument in the first stage are reported on the bottom of columns 3 and 4 of tables 3 – 5. For the estimates with detailed census region fixed effects, the F-stats are greater than 10, and in fact always greater than 30. As was displayed by my first stage table, they are not suggestive of a weak IV problem (Staiger and Stock, 1997)[4].

The results indicate that college completion's effect on smoking behaviors are both statistically and economically significant. Table 3 displays that individuals with a college degree are less likely to smoke when compared to individuals without a college degree. This finding is

---

[3] IV estimates with added state fixed effects can be found in the appendix

[4] The differences in F-statistics between my first stage table and my second stage tables is caused by the differences in samples of analyses. In my second stage table, the samples are adjusted for missing responses in the dependent variables.

consistent with de Walque's finding that college educated individuals are less likely to smoke (de Walque, 2004). To illustrate the magnitude of my estimates table 3 suggests that individuals with a college degree are 37.3% less likely to have ever smoked and 25.6% less likely to be current smokers, while 40% of this sample has smoked at least one cigarette and 20% of this sample is currently a smoker. When compared to the OLS estimates, the coefficients have the same sign of coefficients; however, the magnitude and statistical significance differ. The IV estimates for Ever Smoked and Current Smoker are greater in magnitude than the OLS estimates. The IV estimates are less precisely estimated than the OLS estimates and therefore reduce statistical significance but are still significant at the 10% level for Current Smoker and the 5% level for Ever Smoked. The estimates in table 3 also suggests that older individuals are more likely to smoke. The results also suggest that Black and Hispanic individuals are less likely to be current smokers and to have ever smoked; meanwhile, Asians are less likely to have ever smoked but there is no statistically significant relationship between being Asian and being a current smoker. Females are less likely to have ever smoked and to be current smokers.

The results suggest that college degree attainment also affects e-smoking decisions. Individuals with a college education are less likely to have ever used an e-cigarette as well as less likely to be current e-smokers. Table 4 displays that individuals with college degrees are 37% less likely to have ever used an e-cigarette meanwhile 26% of this sample has used an e-cigarette at least once in their life. Individuals with college degrees are also estimated to be 18.5% less likely to be current e-smokers while only 5.2% of the sample currently e-smoke. Compared to the OLS results, the signs of the coefficients are the same, but the magnitude and statistical significance differ. For both outcome variables, Ever E-Smoked and Current E-Smoker, the coefficients are greater in magnitude. The IV estimate of Current E-Smoker loses significance and is only significant at the 5% level. The IV estimate of Ever E-Smoked is still significant at the 1% level. The estimates in table 4 suggest older individuals are less likely to currently smoke and have ever smoked. The results also suggest that Black and Hispanic individuals are less likely to have ever e-smoked and less likely to be current e-smokers. Females are less likely to have ever e-smoked and be current e-smokers.

Lastly, the results in table 5 imply that individuals with a college degree are not significantly different in their self-reported BMI category from those who do not have a college degree. These results are similar to previous research which showed that high school completion has no effect on BMI (Kenkel et al. 2006). The results for Overweight and Obese differ greatly from the OLS estimates. The OLS estimates suggest that individuals with college degrees are 2.2% more likely to be overweight and 11.7% less likely to be obese. The IV estimates of this relationship have no statistical significance. Given that about 31% of this sample is overweight and 30% of this sample is obese this null finding could potentially be explained by the different types of individuals who are categorized as overweight or obese. For some, education may lead individuals to engage in healthier eating and exercise habits; for others, pursuit of education could be stressful and the outcome of having a desk job could lead individuals to live sedentary lifestyles. These conflicting effects could result in a complex relationship between college and BMI with a null aggregate effect.

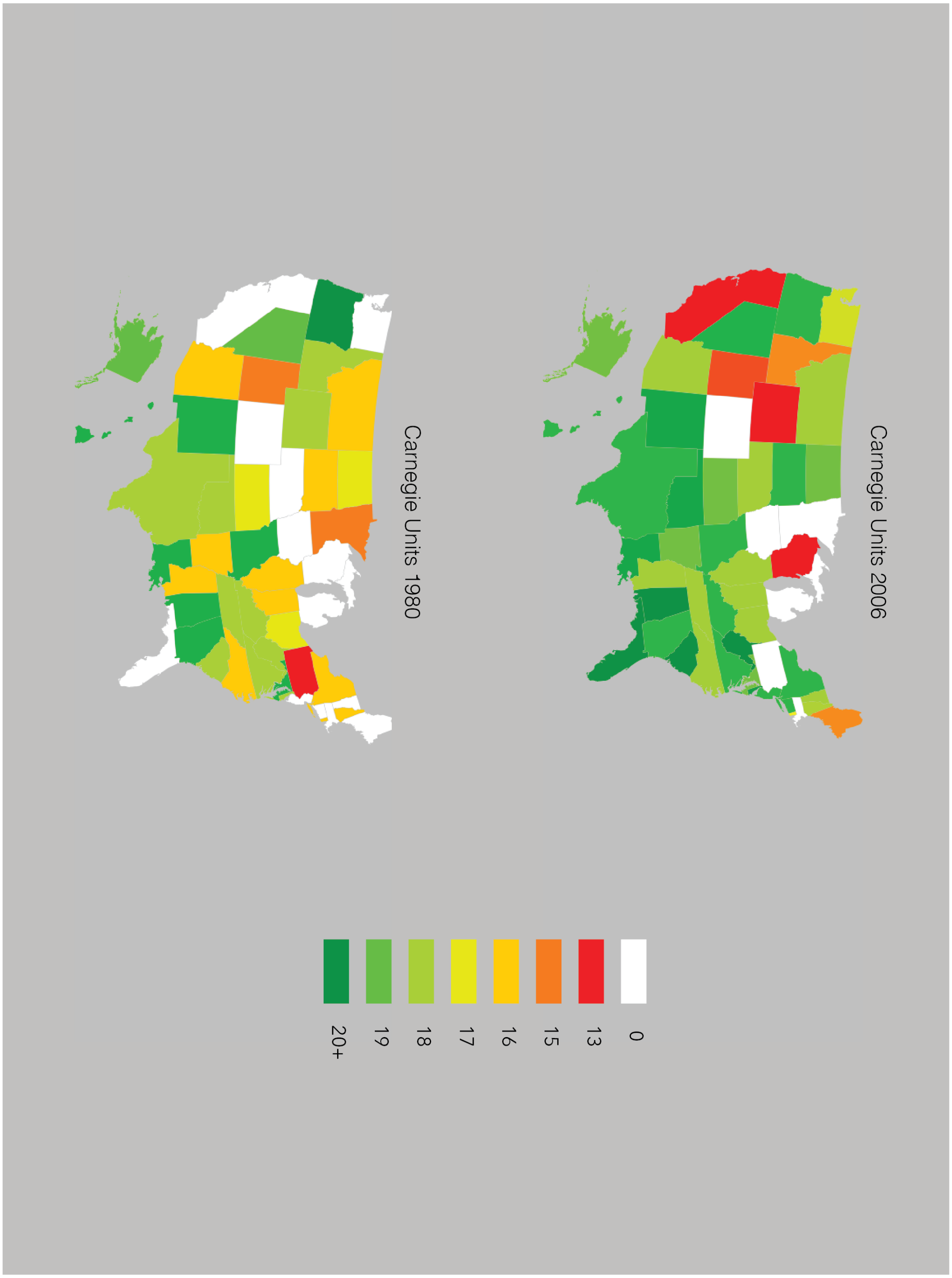
6 Conclusion

The correlation between health and education is well-documented. Showing where this correlation comes from, and ultimately providing results for causal pathways has been a far more challenging task. This paper used the variation in high school graduation requirements across years and states to isolate exogenous variation in college attainment. Doing so allowed me to analyze college completion's effect on risky health behaviors as a potential causal pathway that education effects health outcomes.

The main shortcoming of this paper was the inability to include observations of individuals older than 49 years old because of the limited data of the instrument. A future improvement on this study is to expand the sample size to include ages beyond 50 to take advantage of the within-state variation in Carnegie Unit policy. Inclusion of these observations could lead to enough variation to identify causal relationships with state fixed effects.

The results imply that completing college makes individuals are less likely to engage in risky health behaviors. Specifically speaking, the results report that completing college reduces the likelihood an individual is a smoker or e-smoker. The results suggest that education does not have an apparent causal impact on body mass index.

Figure 1



Author's tabulation of Carnegie Unit data available at the National Center for Education Statistics website: nces.ed.gov

Table 1
Sample Statistics

Variables	Mean	SD
Ever E-Smoked	0.265	0.441
Current E-Smoker	0.052	0.223
Current Smoker	0.204	0.403
Ever Smoked	0.397	0.489
Underweight	0.014	0.116
Overweight	0.309	0.462
Obese	0.299	0.458
High School Graduate	0.923	0.266
Some College	0.679	0.467
College Graduate	0.411	0.492
Age 25 - 29	0.190	0.392
Age 30 - 34	0.208	0.406
Age 35 - 39	0.189	0.392
Age 40 - 44	0.187	0.390
Age 45 - 49	0.226	0.418
Black	0.107	0.309
Asian	0.033	0.179
Hispanic	0.114	0.318
Other Race	0.057	0.232
Female	0.541	0.498
Carnegie Unit Requirements	18.528	2.714
N	94,056	

Table 2
Effect of Carnegie Units on College Completion

	1 College	2 College	(3) college
Carnegie Units	-0.00825***	-0.00503***	0.000228
	(-13.97)	(-6.32)	(0.14)
Age 30 - 34		0.0178***	0.0189***
		(3.57)	(3.81)
Age 35 - 39		0.0232***	0.0362***
		(4.03)	(4.78)
Age 40 - 44		0.0284***	0.0410***
		(4.93)	(5.42)
Age 45 - 49		0.00221	0.0141*
		(0.40)	(1.91)
Black		-0.118***	-0.141***
		(-22.16)	(-25.66)
Asian		0.267***	0.250***
		(29.86)	(27.49)
Hispanic		-0.230***	-0.243***
		(-45.05)	(-46.11)
Other Race		-0.157***	-0.158***
		(-22.57)	(-22.33)
Female		0.0515***	0.0521***
		(16.33)	(16.61)
Constant	0.564***	0.544***	0.324***
	(51.01)	(31.92)	(8.31)
N	94,056	94,041	94,041
F-Statistic	195.96	39.90	0.0196

Note: Columns 1 displays results from a first stage without the covariates used in my final method of estimation. Column 2 displays results including all covariates and detailed census region fixed effects. F-Statistics listed display instrument relevance

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3
Estimates of the Relationship Between College Completion and Smoking

	OLS Estimates		IV Estimates	
	1 Ever Smoked	2 Current Smoker	3 Ever Smoked	4 Current Smoker
College	-0.256***	-0.211***	-0.373**	-0.256*
	(-79.77)	(-78.70)	(-2.34)	(-1.93)
Age 30 - 34	0.0553***	0.0108***	0.0573***	0.0116**
	(11.21)	(2.63)	(10.06)	(2.46)
Age 35 - 39	0.0793***	0.00717*	0.0836***	0.00885
	(15.56)	(1.69)	(10.68)	(1.37)
Age 40 - 44	0.0428***	-0.0210***	0.0478***	-0.0191***
	(8.38)	(-4.95)	(5.58)	(-2.69)
Age 45 - 49	0.0437***	-0.0245***	0.0457***	-0.0237***
	(8.96)	(-6.03)	(8.17)	(-5.13)
Black	-0.183***	-0.0557***	-0.197***	-0.0611***
	(-34.34)	(-12.57)	(-10.04)	(-3.77)
Asian	-0.159***	-0.0654***	-0.128***	-0.0533
	(-17.67)	(-8.72)	(-2.96)	(-1.49)
Hispanic	-0.176***	-0.112***	-0.202***	-0.123***
	(-34.19)	(-26.26)	(-5.45)	(-3.99)
Other Race	0.0423***	0.0502***	0.0238	0.0430**
	(6.17)	(8.79)	(0.90)	(1.97)
Female	-0.0773***	-0.0250***	-0.0712***	-0.0227***
	(-24.79)	(-9.64)	(-7.99)	(-3.07)
Constant	0.525***	0.297***	0.577***	0.317***
	(77.82)	(52.92)	(8.04)	(5.35)
N	90,039	90,039	90,039	90,039
F-Statistic (first stage IV)			36.950	36.950

<u>Note</u>*: Columns 1 and 2 contain results from standard linear probability models. Columns 3 and 4 contain results using Carnegie Unit requirements as an Instrument for College. All Models include detailed census region fixed effects. F-Statistic for the first stage of the 2SLS is found at the bottom to display instrument relevance.*

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4
Estimates of the Relationship Between College Completion and E-Smoking

	OLS Estimates		IV Estimates	
	1	2	3	4
	Ever E-Smoked	Current E-Smoker	Ever E-Smoked	Current E-Smoker
College	-0.167***	-0.0418***	-0.374***	-0.185**
	(-56.90)	(-27.36)	(-2.59)	(-2.43)
Age 30 - 34	-0.0588***	-0.00682***	-0.0553***	-0.00442
	(-13.04)	(-2.91)	(-10.61)	(-1.60)
Age 35 - 39	-0.113***	-0.0152***	-0.106***	-0.0100***
	(-24.35)	(-6.29)	(-14.92)	(-2.67)
Age 40 - 44	-0.157***	-0.0241***	-0.149***	-0.0181***
	(-33.76)	(-9.94)	(-19.29)	(-4.43)
Age 45 - 49	-0.179***	-0.0275***	-0.176***	-0.0252***
	(-40.13)	(-11.83)	(-34.31)	(-9.25)
Black	-0.0991***	-0.0277***	-0.123***	-0.0445***
	(-20.34)	(-10.94)	(-7.01)	(-4.78)
Asian	-0.112***	-0.0122***	-0.0578	0.0253
	(-13.58)	(-2.84)	(-1.49)	(1.24)
Hispanic	-0.128***	-0.0358***	-0.176***	-0.0687***
	(-27.35)	(-14.64)	(-5.25)	(-3.88)
Other Race	0.0392***	0.00493	0.00616	-0.0179
	(6.24)	(1.51)	(0.26)	(-1.42)
Female	-0.0625***	-0.0201***	-0.0518***	-0.0127***
	(-21.94)	(-13.57)	(-6.44)	(-2.98)
Constant	0.521***	0.104***	0.614***	0.168***
	(84.46)	(32.46)	(9.46)	(4.91)
N	90,081	90,049	90,081	90,049
F-Statistic (first stage IV)			39.456	39.732

Note: *Columns 1 and 2 contain results from standard linear probability models. Columns 3 and 4 contain results using Carnegie Unit requirements as an Instrument for College. All Models include detailed census region fixed effects. F-Statistic for the first stage of the 2SLS is found at the bottom to display instrument relevance.*
t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5
Estimates of the Relationship Between College Completion and BMI

	OLS Estimates		IV Estimates	
	1	2	3	4
	Overweight	Obese	Overweight	Obese
College	0.0226***	-0.117***	0.0968	-0.120
	(6.83)	(-36.08)	(0.58)	(-0.73)
Age 30 - 34	0.0333***	0.0430***	0.0322***	0.0430***
	(6.52)	(8.56)	(5.66)	(7.71)
Age 35 - 39	0.0344***	0.0765***	0.0316***	0.0766***
	(6.53)	(14.82)	(3.88)	(9.60)
Age 40 - 44	0.0337***	0.0974***	0.0305***	0.0975***
	(6.38)	(18.80)	(3.39)	(11.09)
Age 45 - 49	0.0427***	0.115***	0.0415***	0.115***
	(8.48)	(23.30)	(7.20)	(20.43)
Black	-0.00184	0.114***	0.00700	0.113***
	(-0.34)	(21.12)	(0.34)	(5.59)
Asian	-0.0226**	-0.142***	-0.0419	-0.141***
	(-2.45)	(-15.66)	(-0.94)	(-3.24)
Hispanic	0.0394***	0.0130**	0.0554	0.0125
	(7.26)	(2.44)	(1.52)	(0.35)
Other Race	-0.0239***	0.0829***	-0.0118	0.0825***
	(-3.40)	(11.99)	(-0.42)	(2.99)
Female	-0.136***	0.00692**	-0.140***	0.00704
	(-42.28)	(2.19)	(-14.52)	(0.75)
Constant	0.365***	0.266***	0.332***	0.267***
	(52.82)	(39.29)	(4.41)	(3.63)
N	85,797	85,797	85,797	85,797
F-Statistic (first stage IV)			33.655	33.655

Note: *Columns 1 and 2 contain results from standard linear probability models. Columns 3 and 4 contain results using Carnegie Unit requirements as an Instrument for College. All Models include detailed census region fixed effects. F-Statistic for the first stage of the 2SLS is found at the bottom to display instrument relevance.*

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix

Table 6
Estimates of the Relationship Between College Completion and Smoking

	IV Estimates with State Fixed Effects	
	1	2
	Ever Smoked	Current Smoker
College	1.770	5.589
	(0.14)	(0.17)
Age 30 - 34	0.0188	-0.0932
	(0.08)	(-0.15)
Age 35 - 39	0.0118	-0.179
	(0.03)	(-0.17)
Age 40 - 44	-0.0376	-0.241
	(-0.08)	(-0.19)
Age 45 - 49	0.0168	-0.0897
	(0.12)	(-0.24)
Black	0.102	0.765
	(0.06)	(0.16)
Asian	-0.660	-1.496
	(-0.21)	(-0.18)
Hispanic	0.315	1.300
	(0.10)	(0.16)
Other Race	0.352	0.961
	(0.17)	(0.18)
Female	-0.185	-0.334
	(-0.27)	(-0.19)
Constant	-0.0905	-1.570
	(-0.02)	(-0.14)
N	90,039	90,039
F-Statistic (first stage IV)	0.030	0.030

Note: Columns 1 and 2 contain results from standard linear probability models. Columns 3 and 4 contain results using Carnegie Unit requirements as an Instrument for College. All Models include detailed census region fixed effects. F-Statistic for the first stage of the 2SLS is found at the bottom to display instrument relevance.

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7
Estimates of the Relationship Between College Completion and Smoking

	IV Estimates with State Fixed Effects	
	1	2
	Ever E-Smoked	Current E-Smoker
College	6.338	-1.688
	(0.10)	(-0.10)
Age 30 - 34	-0.171	0.0218
	(-0.16)	(0.08)
Age 35 - 39	-0.318	0.0357
	(-0.16)	(0.07)
Age 40 - 44	-0.398	0.0356
	(-0.18)	(0.06)
Age 45 - 49	-0.249	-0.0115
	(-0.39)	(-0.07)
Black	0.811	-0.258
	(0.09)	(-0.11)
Asian	-1.715	0.391
	(-0.11)	(0.10)
Hispanic	1.443	-0.437
	(0.10)	(-0.11)
Other Race	1.060	-0.255
	(0.11)	(-0.10)
Female	-0.407	0.0671
	(-0.12)	(0.08)
Constant	-1.630	0.660
	(-0.08)	(0.12)
N	90081	90049
F-Statistic (first stage IV)	0.011	0.011

Note: Columns 1 and 2 contain results from standard linear probability models. Columns 3 and 4 contain results using Carnegie Unit requirements as an Instrument for College. All Models include detailed census region fixed effects. F-Statistic for the first stage of the 2SLS is found at the bottom to display instrument relevance.
t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8
Estimates of the relationship between college completion and smoking

	IV Estimates with State Fixed Effects	
	1	2
	Overweight	Obese
College	-67.93	46.00
	(-0.01)	(0.01)
Age 30 - 34	1.090	-0.674
	(0.01)	(-0.01)
Age 35 - 39	2.279	-1.449
	(0.01)	(-0.01)
Age 40 - 44	2.683	-1.703
	(0.01)	(-0.01)
Age 45 - 49	0.840	-0.428
	(0.01)	(-0.01)
Black	-9.602	6.637
	(-0.01)	(0.01)
Asian	16.54	-11.38
	(0.01)	(-0.01)
Hispanic	-15.49	10.56
	(-0.01)	(0.01)
Other Race	-10.94	7.488
	(-0.01)	(0.01)
Female	3.621	-2.543
	(0.01)	(-0.01)
Constant	22.80	-14.90
	(0.01)	(-0.01)
N	85797	85797
F-Statistic (first stage IV)	0.000	0.000

Note: Columns 1 and 2 contain results from standard linear probability models. Columns 3 and 4 contain results using Carnegie Unit requirements as an Instrument for College. All Models include detailed census region fixed effects. F-Statistic for the first stage of the 2SLS is found at the bottom to display instrument relevance.

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

References

Bound, J., Jaeger, D., & Baker, R. (1995). Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak. Journal of the American Statistical Association*,* *90*(430), 443-450. doi:10.2307/2291055

Card, D. E. (1999). The Causal Effect of Education on Earnings. In: Ashenfelter, Orley, & Card, David E (eds), The Handbook of Labor Economics Vol 3a. Amsterdam: Elsevier/North Holland.

Centers for Disease Control and Prevention (CDC). (2017, March 17). National Center for Health Statistics. Retrieved from https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm

Cutler, D. M, & Lleras-Muney, A. (2006). Education and Health: Evaluating Theories and Evidence. NBER Working Paper 12352, Cambridge, MA.

de Walque, D. (2007). Does education affect smoking behaviors?: Evidence using the Vietnam draft as an instrument for college education, Journal of Health Economics, 26, issue 5, p. 877-895,

Fuchs, V. R. (1982). Time preference and health: an exploratory study. In: Fuchs, V R (ed), Economic Aspects of Health. New York: Russell Sage Foundation.

Gaviria, A., & Raphael, S. (2001). School-based peer effects and juvenile behavior. Review of Economics and Statistics, 83(2), 257–268.

Grossman, M.. (2005). Education and Non-market Outcomes. NBER Working Paper 11582, Cambridge, MA.

Kenkel, D., Lillard, D., & Mathios, A. (2006) The Roles of High School Completion and GED Receipt in Smoking and Obesity, Journal of Labor Economics 24, no. 3 (July 2006): 635-660.

Wooldridge, J. M., (2012). Introductory econometrics: a modern approach. Mason, Ohio: South-Western Cengage Learning,