2016

The Effect of Education on Health: New Evidence from the Elimination of the Social Security Student Benefit Program

Taylor Cranor
tcranor@wellesley.edu

The Effect of Education on Health: New Evidence from the Elimination of the Social Security Student Benefit Program

Taylor J. Cranor

Submitted in Partial Fulfillment
of the
Prerequisite for Honors
in Economics

April 2016

Acknowledgements

I would like to thank all of the faculty, staff, friends, and family that supported me throughout the thesis process. First and foremost, I am extremely appreciative of all the guidance from my advisor, Courtney Coile, whose thoughtful comments and suggestions were integral to the development of my thesis. Thanks also go to Robin McKnight, who kindly reminded me of deadlines, taught me what "belies" actually means, and provided key advice during several thesis crises. I am also indebted to David Lindauer, Akila Weerapana, Sue Skeath, Casey Rothschild, Phil Levine, Kartini Shastry, Pinar Keskin, Kyung Park, Kristin Butcher, Dan Fetter, and Sheila Datz. They shaped my choice of major, future career plans, and Wellesley experience in an overwhelmingly positive way. Thanks also go to Tim Peltason and Sari Kerr for their willingness to serve on my reading committee.

I also give utmost thanks to my friends, Jackie Li, Amy Wickett, Alice Liang, Priyanka Fouda, and Nancy Zhang, who supported me throughout my thesis and Real Analysis. Finally, thank you to my family for their constant love and support.

Abstract

An individual's education and health status are closely correlated, but the causal pathway between an increase in education and health outcomes has remained largely elusive in empirical work. I explore how an increase in education affects long-term health outcomes by exploiting the removal of the Social Security Student Benefit program in 1982 as an instrument for college attendance and completion in a two-stage least squares model. The outcomes of interest include likelihood of poor self-reported health, pain that interferes with work, arthritis, mental health, hypertension, heart problems, diabetes, smoking, and exercising. I am unable to reject increases or decreases in one's health outcomes or behaviors. My findings are inconclusive due to insufficient statistical power, which stems from a small sample size.

Table of Contents

I. Introduction

Policy makers and public intellectuals often stress the importance of a college education, arguing that education leads to higher expected earnings and expanded job prospects. Yet there may be other important benefits of college attendance that are less well understood. Recent reviews of the economic literature on the effects of education cited substantial evidence that education is linked not only to increases in an individual's earnings potential, but also to decreases in criminal behavior, increases in voting and democratic participation, and improved health outcomes (Grossman 2015; Lochner 2011). Indeed, one of the best-documented correlates of education is health: "The one social factor that researchers agree is consistently linked to longer lives in every country where it has been studied is education. It is more important than race; it obliterates any effects of income" (Kolata cited in Grossman 2015).

Despite empirical evidence of this association, it has been difficult to establish a causal link between education and health, due to the issue of dual causality. Health in childhood and young adulthood may impact one's ability to get an education. Yet educational attainment clearly affects income and perhaps other aspects of one's life that facilitate access to preventive and medical care; more educated individuals may also practice safer behaviors, such as choosing not to smoke, always using a seatbelt, or exercising regularly, more often than less educated people. In addition, individual-level characteristics like patience that make one more likely to both invest in education and in long-term health may also contribute to the strong relationship between health and education. It is thus difficult to figure out which way the causation goes and which mechanisms drive the effects. Many of the rigorous empirical analyses that do address the issue of causality focus on compulsory education law changes in the 1960s or earlier that required students to attend various levels of secondary school. While these studies are empirically sound, their focus on secondary schooling means that the findings cannot be extrapolated to the effect of college attendance on health outcomes. In addition, these studies rely on data from the early to mid-1900s, which means their findings may not apply to more recent cohorts of students.

This paper aims to shed light on the causal effect of education on the health outcomes of more recent cohorts by investigating how a decrease in college attendance caused by a Social Security Administration policy change affected long-term health outcomes. From 1965 to 1982, the Social Security Administration paid monthly benefits to full-time college students ages 18-22 with an eligible, deceased parent. Congress passed a law in 1981 that ended the program, mandating that students (18 or older) entering college after May 1982 were unable to receive any Social Security benefits. According to Susan Dynarski's (2003) paper "Does Aid Matter?", students who received benefits were 22 percentage points more likely to attend college by age 23 than those who would have been eligible for the benefits but entered college just after May 1982. These findings indicate that the removal of the program significantly affected college-going; if college attendance actually impacts health, then we would expect to see different health outcomes between students eligible for benefits and ineligible students, controlling for trends in health outcomes over time. Importantly, since eligibility for the student benefit was not tied to the students' health, the change in educational attainment was unrelated to the health status of beneficiaries.

Using data from the National Longitudinal Survey of Youth 1979 (NLSY79), which has followed the same group of respondents since 1979, this paper will examine the health outcomes of people who were directly affected by the policy change and who are now in their early-to-mid 50s, ages when health problems become more common. The health outcomes of interest include self-reported health, pain that interferes with work, arthritis, mental health, hypertension, heart problems, and diabetes. I will also use two health behaviors, smoking and exercising, as dependent variables, as they may be mechanisms through which education effects health. To estimate the causal effects of education on health, I will first estimate the variation in college attendance due solely to the policy change. Then I will estimate how that variation in college attendance affected long-term health outcomes, which captures the effect of an exogenous change in college attendance on health outcomes.

In the following section, I will discuss the findings from previously published research on the relationship between education and health. I will then briefly summarize the policy change in question. Next, I will explain how Susan Dynarksi's "Does Aid Matter?" is a "proof of concept" for this project. The following four sections will detail the empirical strategy, data, results, and discussion, respectively. A conclusion ends the paper.

II. Literature Review

Given the importance of both education and health, it is no surprise that many studies in the past have examined the relationship between the two. A large portion of research has used cross-sectional data to document the strong correlation between health and education.[1] Cutler and Lleras-Muney (2006) document the basic correlations between health and education using the National Health Interview Survey, a cross-sectional data set. They find that individuals with higher levels of education are less likely to die within 5 years and have lower morbidity from heart conditions, stroke, diabetes, and other chronic conditions. Interestingly, when controlling for exercise, smoking, drinking, seat belt use, and use of preventive care, the effect of education

[1] For additional papers that study the correlation between health and education, please see Michael Grossman's 2015 paper, "The Relationship between Health and Schooling: What's New?" (2015) and Brian Goesling's "The Rising Significance of Education for Health?" (2007).

on mortality falls by 30 percent. This indicates that increases in education may affect one's likelihood of practicing safe habits. As Cutler and Lleras-Muney point out, the correlations they find do not prove that a causal relationship between education and health exists. Cross-sectional analysis is helpful for establishing the descriptive relationship between education and health, but other empirical analyses better isolate how changes in education affect health.

Researchers have typically used changes in compulsory schooling laws as natural experiments to test the effects of education on health, but the results vary substantially by study. Lleras-Muney's (2006) landmark paper analyzed the effects of increases and decreases in compulsory schooling laws in the United States between 1915 and 1939 on mortality. Lleras-Muney uses US census data from 1960, 1970, and 1980 to create "synthetic cohorts" so that she can follow age groups over time. Her preferred model isolates the variation in educational attainment due solely to compulsory schooling laws and uses that variation to estimate how education directly affected mortality rates. Lleras-Muney finds that increasing education by 1 additional year significantly increases life expectancy at age 35 by as much as 1.7 years. One year of additional compulsory schooling decreased mortality after age 35 by about 3 percent. As most of the compulsory law changes in question increased the school leaving age to 12-14, depending on the state, the relevance of this study to today's educational landscape, in which 81 percent of students graduate high school, is limited (Institute of Education Sciences 2015).

A number of studies investigating compulsory schooling law changes in other countries have also contributed to the existing literature. Meghir, Palme, and Simeonova (2012) estimate the impact of an increase in compulsory schooling in Sweden from 7 or 8 years (depending on municipality) to 9 years between 1949 and 1962. Because the timing of adoption of this new law varied across municipalities, Meghir et. al can compare people who were born in the same year and are in the same local labor markets (i.e. geographically close municipalities) in a given year, which ensures that differences in outcomes between individuals are not due to differential trends across cohorts or labor markets. Their estimates suggest that the reform had no significant effect on overall mortality for women. Men who attended school after the reform was implemented have a 6.8 percent lower mortality risk than the control group, but this change is only marginally significant. Oreopoulos (2003) employs a similar approach as Meghir et al., using data extracts from Census data from the United States (1950 - 2000) and Canada (1971-2001) and from the U.K.'s and Ireland's General Household Surveys (1983-1998) to examine the effects of compulsory schooling law changes on earnings and health. In the United States, most laws changes increased the compulsory schooling age from 14 to 16. The estimates for the US imply that an additional year of compulsory schooling lowers the likelihood of having a disability that one's ability to care for their basic needs by 1.7 percentage points. In the UK and Ireland, the compulsory schooling age rose from 14 to 15. For the UK, a one-year increase in schooling increases the probability of self-reporting being in good health by 6 percentage points and lowers the likelihood of reporting being in poor health by 3.2 percentage points.

Though causal estimates of the effect of college attendance on health outcomes are scarce, Buckles et. al's (2013) paper is a notable exception for investigating the effects of college attendance on mortality. The authors instrument for an increase in postsecondary education using state variation in draft avoidance behavior from the Vietnam War Era and investigate its effects on the cumulative mortality rate per 1000 men between 1981 and 2007 (the fraction of the cohort that died by 2007 conditional on being alive in 1981). The Department of Defense assigned draft quotas to each state, which local boards within the state were responsible for fulfilling. Uneven application of formal procedures and communication delays between federal, state, and local governments created wide variation in the number of men in each state selected for the draft each year. This created substantial differences in the number of men who avoided the draft by attending college in each state. Decomposing their sample into birth-state-by-cohort groups, Buckles et. al estimate that the mortality rate for non-college graduates is 2.2 times higher than the mortality rate for college graduates. The authors cite suggestive evidence based on correlations that reduced smoking, increased exercise, greater access to health care, and increased wealth are potential mechanisms for the reduction in mortality rates among college graduates. Buckles et. al provide evidence of a causal link between a college education and mortality rates.

The literature on the effect of education on health documents causal estimates of the effect of education on health, but these estimates may not be directly applicable to today's policy environment. First of all, all of the discussed papers that estimate causal models use instruments based on policy decisions made before 1972, so their findings may not apply in a modern context. Second, since Meghir et. al and Oreopulous primarily use international data, the applicability of their estimates to American policy is limited. Sweden and the UK provide universal healthcare to their citizens while the U.S. does not, so the causal estimates might have been very different if the law changes they used had occurred within the U.S. where education may be correlated with health care access and quality. Finally, Buckles et. al use draft avoidance as an instrument for education; since the U.S. no longer has a draft, their conclusion is less informative for shaping modern policies. Direct evidence on how changes to education due to different financial incentives affected health would be more relevant to the policy options the government currently has, such as tax incentives and tuition subsidies.

III. Brief History of the Social Security Student Benefit Program

The Social Security Administration's policy change regarding survivors benefits for young adults created an ideal "quasi-experimental" set-up to further investigate the effect of college attendance on long-term health outcomes. The Social Security Administration (hereafter referred to as the SSA) is a government agency that administers several well-known public programs in the United States, including Social Security, Disability Insurance, and Supplemental Security Income. These programs are funded by payroll taxes paid by employers, employees, and the self-employed. The Social Security Administration uses these funds to pay benefits to retired workers, their spouses, and their children, and to survivors of deceased insured workers. Children and widows or widowers receive survivors benefits if a working parent or spouse, respectively, dies. Children and widows (or widowers) who have dependent children or are age 60 or older are generally eligible if the deceased worker made 10 years of contributions into Social Security. [2] Survivors benefits for children who are younger than 18 or are still in secondary school equal 75 percent of what their deceased parent would receive if (s)he retired at normal retirement age (called the Primary Insurance Amount, or PIA), which, in turn, depends on the parent's average lifetime earnings. Widows or widowers under 60 with a child younger than 16 get 75 percent of the worker's benefit amount; those over 60 but under full retirement age get 71-99 percent of the worker's PIA. However, the SSA caps the total benefits paid to a

---

[2] To be eligible for survivors' benefits, a worker must accumulate a number of credits to equal the number of years elapsing between age 21 and the year in which he/she dies. If a worker dies before achieving fully insured status, the worker must be at least currently insured for benefits to be paid to family members, which means they must have worked for at least 3 quarters of the past 3 years (12 quarters) preceding the death. If a worker has already earned 40 credits, the amount necessary to receive retirement benefits, they are eligible for all other Social Security benefits. Please see http://www.socialsecuritymatters.org/Get_the_Facts_files/D488WomenandSS.pdf

family at 150 percent to 180 percent of the basic benefit rate.[3] If the sum of the benefits payable to the family exceeds the limit, each family member's payment is reduced proportionally.

In 1965, Congress expanded all children's benefits for those with a deceased parent or a retired parent to unmarried, full-time college students ages 18 to 22, reasoning that those students were still dependent on their parents, despite their legal "adult" status. From 1965 to 1981, benefits paid to students 18-22 ballooned from $1.2 billion per year to $6.3 billion per year (adjusted for inflation), with the number of beneficiaries increasing with similar rapidity from 206,000 in 1965 to 733,000 in 1980 (U.S. Social Security Administration 2015). In 1980, 10 percent of full-time college students aged 18 to 24 received Social Security student benefits (U.S. Social Security Administration 2015) (Institute of Education Sciences 2015). The average annual benefit paid to full-time college students with a deceased parent in 1980 was $9,150 (U.S. Social Security Administration 2015).[4] For comparison, in 1980, the average Pell Grant, a federal aid program for low-income, college-going students, was $2,835 and the average guaranteed student loan was $5,800 (Dynarski 2003; "Federal Pell Grant Program Data Books 1980-81 - 1989-90" 2011). Average college tuition for a public four-year institution in 1980 was $2,400 and average tuition at a private non-profit institution was $10,511.[5] Tuition costs for the average student receiving benefits were thus substantially more affordable than they would have been without the benefits.

---

[3]Today, the maximum family benefit cannot exceed "(a) 150 percent of the first $1,056 of the worker's PIA, plus (b) 272 percent of the worker's PIA over $1,056 through $1,524, plus (c) 134 percent of the worker's PIA over $1,524 through $1,987, plus (d) 175 percent of the worker's PIA over $1,987." Please see http://www.ssa.gov/oact/cola/familymax.html for more detailed information.

[4] The average benefit paid to all full-time students receiving Social Security benefits due to a deceased, disabled, or retired parent in 1980 was $7,856 (U. S. Social Security Administration 2015).

[5] http://trends.collegeboard.org/college-pricing/figures-tables/tuition-fees-room-board-time-1974-75-2014-15-selected-years

In the mid-1970s, a number of problems with the student beneficiary program began cropping up. Eligible children only received benefits if they were a full-time student. If at any point the child dropped out of college or became a part-time student, their benefits were supposed to stop, but the SSA relied on self-reporting from students and their parents to ensure beneficiaries remained full-time students. In 1978, an internal study revealed that overpayments totaled as much as $544 million a year because of underreporting changes in full-time status. The General Accounting Office estimated the number was closer to $978 million (DeWitt 2001). After several years of economic distress that left the SSA concerned about the program's solvency, the Student Benefit program became a target for budget cuts. High inflation rates increased program expenditures substantially and rampant unemployment reduced payroll taxes, leaving the Social Security trust fund underfunded. The slowing birth rate and increase in life-expectancy also threatened the long-term viability of the program, with the government estimating that the trust fund would not be able to pay benefits on time beginning some time in the 1980s. In addition, the public had come to see the student benefits program as a form of student-aid instead of a social insurance program, which was the program's original intention. Combined with the overpayments and concerns about the trust fund's solvency, this prompted Congress to cut the SSA's Student Benefit program, as other forms of federal educational assistance were available to students in need of financial assistance to attend college. Students who were not enrolled in college by May 1982 were no longer eligible to receive any benefits, while students who were enrolled in college had their benefits reduced by 25 percent each year for 3 years. The last student benefits were paid out in April of 1985 (DeWitt 2001).

IV. Framing Paper

Estimating a causal effect of education on health requires an exogenous source of variation in college attendance. Susan Dynarski's "Does Aid Matter?" illustrates that the removal of the Social Security Student Benefit Program significantly decreased college attendance. Using data from the National Longitudinal Survey of Youth from 1979 to 1983, Dynarski employs a difference-in-differences framework to estimate how the removal of the Social Security Student Benefit program affected college attendance. She compares changes in college-going before and after the program ended between the treatment group, college-age children with deceased fathers and the control group, college-age children with living fathers. Her preferred estimates suggest that aid eligibility prompted about 24.3 percent of eligible children to attend college who would not otherwise have gone to college and increased average schooling by age 28 by 0.679 years. These findings indicate that the removal of the Social Security Student Benefit Program created an exogenous decrease in education for children with a deceased parent. Her findings suggest that if education indeed affects one's long-term health outcomes, this policy change would provide an ideal quasi-experimental framework to measure that causal effect.

V. Empirical Framework

I will use a two-stage least squares empirical framework that instruments for a change in education using the removal of the Student Benefit Program. Perhaps unusually, the first stage is a difference-in-differences regression that will essentially replicate Dynarski's findings from "Does Aid Matter?", estimating variation in college attendance due to the removal of the Social Security Student Benefit program. To capture the policy's effect on college attendance, I will run a difference-in-differences regression that uses individuals with a deceased father, who were eligible for student benefits when the program existed, as the treatment group and those whose father was living as the control group. I will run the following regression:

$$College_i = \tag{1}$$

$$\beta_o + \beta_1 Deceased\ Father_i + \beta_2 After_i + \beta_3 (Deceased\ Father\ x\ After)_i + \beta_4 V_i + \varepsilon_i$$

in which the dependent variable, $College$, is a dummy variable for whether or not a student attended college full-time by 23 or for whether or not a student completed at least 1 year of college by 23. $Deceased\ Father_i$ is a dummy variable that equals 1 if a student's father died before the student was 18 and $After_i$ is a dummy variable set to 1 if a person graduated from high school after student benefits were eliminated. V_i is a vector of controls for individual characteristics, including race/ethnicity, sex, and age dummies, as well as an indicator for having a deceased mother. β_3 is our coefficient of interest, indicating the effect of the policy change on college attendance. Only students who were eligible for student benefits due to the death, but not the retirement, of a parent are in the treatment group; a parent's death is exogenous but a parent may retire or apply for disability insurance to trigger benefit eligibility for their child. The model only counts students who have a deceased *father* as treated because 90 percent of student beneficiaries were entitled to the benefits because of the death of their father (U. S. Social Security Administration 2015). This empirical strategy controls for changes over time in average college attendance rates and time invariant differences in the college attendance rates of those with a deceased father and those with a living father.

The second stage of my 2SLS model will estimate how college attendance affected long-term health outcomes. Using the previous equation, I will calculate the predicted values of college attendance and use those values as the main variable of interest in the following equation:

$$Y_i = \beta_o + \beta_1 \widehat{College}_i + \beta_2 Deceased\ Father_i + \beta_3 After_i + \beta_4 V_i + \varepsilon_i \qquad (2)$$

in which the dependent variable, Y_i, is one of the health outcomes of interest, including self-reported health, incidence of pain that interferes with work, arthritis, mental health, hypertension, heart problems, and diabetes, or a health behavior (smoking or exercising). Health behaviors are of interest as they may be the mechanisms through which education affects health. To the extent that health behaviors vary by education level, these estimates may also be informative for explaining those differences. $Deceased\ Father_i$ remains in the second stage, since having a deceased father may be linked to a genetic predisposition for poor health.[6] $After_i$ also stays in the regression, controlling for any cohort effects on health. Neither $Deceased\ Father_i$ nor $After_i$ are valid instruments because they may have independent effects on health.

A number of assumptions underpin the two steps of my empirical strategy. The first stage implicitly assumes that the only reason for any differential changes over time in college-going between the treatment and control groups is that the Social Security Student Benefit program changed in 1982. In the second stage of my regression, I assume that the removal of policy benefits was exogenous to individual health characteristics. Since Congress terminated the program due to budget concerns, this assumption holds. However, instrumenting for a change in education with the policy change may violate the exclusion restriction condition in other ways. For example, the removal of a substantial source of income may reduce a family's ability to pay for medication and buy healthy food or it may increase stress so much so that it prompts individuals to begin smoking or take up other unhealthy behaviors. This scenario would only be an issue if we thought that not receiving $36,000 over four years when one is 18-22 caused one's

[6] "Deceased mother" is also included as a control variable in both the first and second stage, as having a deceased parent (whether mother or father) could affect one's ability to go to college and long-term health outcomes.

long-term health outcomes to deteriorate. If the assumptions hold, this model will accurately capture the effect of an exogenous decrease in education on individuals' long-term health.

VI. Data

I will use data from the National Longitudinal Survey of Youth 1979 (NLSY79) to complete this study. The NLSY79 interviewed 12,686 people ages 14-22 in 1979 and subsequently interviewed them annually through 1994. Respondents are now interviewed every other year. When each participant turned 40, interviewers conducted a detailed health survey of the respondent.[7] These special health interviews provide the main data on outcomes of interest for this study - self-reported health, pain that interferes with work, arthritis, mental health, hypertension, heart problems, and diabetes.[8] Other waves of the NLSY79 collected information on health behaviors, such as smoking and exercising, which are also relevant for this paper. Those who graduated high school before 1982 form the "before" cohort and those who were seniors in 1982 or 1983 make up the "after" cohort.[9] In 1988, interviewers asked participants to describe any periods before age 18 when they did not live with both biological parents, which allows me to determine which respondents' fathers died before their 18th birthday. My sample includes 5,942 people.[10]

---

7 NLSY also collected health data when each participant turned 50. I do not include these variables in my analysis because not all of the people in the sample are 50 years old yet.

8 I was also interested in mortality as a health outcome, but I could not pursue it due to small sample sizes. Very few people die by the age of 40 and, in some of the groups I'm comparing, no one had died by 40.

9 After 1981, children already receiving benefits due to a deceased or working parent while in college continued to receive payments from SSA, but the amounts were reduced for ¼ each year. I chose to code my treatment variable as an indicator variable for eligibility. Coding eligibility as a continuous variable that measure what fraction of years in college were paid did not change my results and are thus not included in this paper.

10 Please see the data appendix for additional information on coding methods. The sample size listed is for the IV model in which "self-reported poor health" is the outcome and "attend college full-time by 23" is the instrumented variable.

VII. Results

A. Replication Results

Since the first stage of my 2SLS model relies on Dynarski's difference-in-differences regression model from "Does Aid Matter?" I will first detail the results of my replication attempt before I explain more fully the ways in which my analysis differs from hers. I coded the treatment variable and control variables according to Dynarski's description in the data appendix of "Does Aid Matter?" Since Dynarski drops those who never completed their junior year of high school, I also drop that group of people from my sample.[11] Dynarski uses two outcome measures in her paper; different numbers of people answered the requisite questions that I used to code "attend college full-time by 23" and "complete at least one year of college by 23," so the two outcome variables have slightly different samples sizes.[12]

Despite my best efforts, I could not exactly replicate the assignment to treatment that Dynarski uses in "Does Aid Matter?" For simplicity's sake, I will focus on the changes to the "attend college full-time by 23" sample, shown in Table 1. The replication sample is about 1 percent larger than Dynarski's. The group of those who graduated between 1979 and 1981 ("Before") and whose fathers were deceased ("Deceased Father") is the same size as the parallel group in Dynarski's analysis. The "Before & Living Father" and "After & Deceased Father" groups are about 1 percent larger than Dynarski's parallel groups. Most striking is the 69 person (5.6 percent) difference between the replication "After & Living Father" category and

---

[11] Children who did not finish their junior year of high school would not go to college, regardless of whether or not they were eligible for the Student Benefit program.

[12] I coded the "Attend college by 23" as a 1 if the observation ever attended college full-time before the age of 23 and a 0 if the observation never attended college full-time before the age of 23. I coded "complete any college" as a 1 if the observation reported that they their highest grade completed was at least the first year of college by the time they turned 23 and a 0 if they only completed 12th grade. Respondents who left the sample by age 23 and did not complete a year of college before they left the sample were coded as missing.

Dynarski's category. These differences complicate the replication exercise, as variation in who composes each group could limit my ability to estimate results that exactly match Dynarski's.[13]

As expected, the replication regression results in columns (3) and (4) in Table 2 are similar to but not exactly the same as Dynarski's original results. For example, comparing the coefficient on the interaction term in the model without controls, we see that the replication coefficient is highly similar in size in and significance (own: 0.18 vs. Dynarksi's 0.182); the replication coefficient indicates that having access to the Student Benefit program increased the likelihood of attending college full-time by 18 percentage points. Dynarski used restricted state of residence data for state fixed effects. Unfortunately, I could not access this data, which means my replication results can never perfectly match her results, since I cannot include state fixed effects. The replication coefficient on "Deceased Father*Before" in the full model indicates that eligibility for the Student Benefit Program increased full-time college attendance 17.6 percentage points. As the mean probability of attending college full-time by age 23 is 49.4 percent in this sample, the coefficient suggests that the removal of the Student Benefit program reduced full-time college attendance by 36 percent for those with deceased fathers. A substantial difference, this finding suggests that the removal of the program could have had long-term impacts on the health of individuals. Next, I describe in more detail the first stage in my analysis, which is modeled on Dynarski's work but differs from it in several ways.

B. The Effect of Social Security Benefits on College Attendance

For the first stage of my analysis, I make several modifications to Dynarski's sample. Employing Dynarski's definition of the "Before" category would limit the power of my ultimate

---

[13] I requested the original coding of the treatment and outcome variables from Sue Dynarski so that I could more closely replicate her findings. I did not receive a response from Professor Dynarski.

2SLS regression method. Dynarski only includes those who were high school seniors in 1979, 1980, or 1981 in her "Before" group, even though there are 19-22 year olds in the 1979 NLSY survey who have graduated high school before the Social Security Student Benefit program ended. To increase my treatment sample size, reduce my standard errors, and thus increase the likelihood of a statistically significant coefficient of interest in my ultimate IV regression model, I count all people who were high school seniors before 1981, including those who graduated high school before 1979, as part of the "Before" group. This change more than doubles my treatment group size and the "Before & Living Father" group size (see Table 1). It also introduces some measurement error into two of the control variables Dynarski uses in her first stage.[14] Family size and income were calculated for the year that the individual was a high school senior. Since the people who graduated before 1979 did not complete a survey during their senior year of high school, they did not provide comparable data. I assigned these people the mean values of their "category," as defined by whether or not they were in the "before" period and whether or not they had a deceased father by age 18. While this change introduces classical measurement error into the estimation of family size and family income, the benefits from increasing my sample size outweigh this small cost. My second change was to restrict this expanded sample only to those people who answered the health survey questions collected when they turned 40 years old. This modification leaves my final sample for "attend college full-time by 23" with 244 people in my treatment group and 5,942 people total.[15]

For a closer inspection of the expanded sample and the controls that I will use moving forward, consider the summary statistics in Table 3 Panel A. I control only for those variables

[14] Please see the data appendix for additional information on how the variables were coded.
[15] As some people did not answer every question, each regression has a slightly different sample size. The numbers reported here and in Table 1 are for the regressions of "likelihood of reporting poor health" on the instrumented "complete at least one year of college by 23" variable.

that may affect health through a mechanism other than educational attainment, since my instrumental variable captures the variation in health due to variation in educational attainment. Thus race/ethnicity, sex, and age dummies, as well as indicators for having a having a deceased father or mother by age 18 and for being in the "before" cohort, comprise my controls. The deceased father/mother indicators aim to capture a genetic component to health outcomes; if a parent died at a relatively young age, his or her child may have a higher risk of health problems. Comparing the means of these variables across the different categories, we see that more women graduated in the "Before" period than men and fewer women than men graduated high school in the "After" period, which reflects the distribution of women by graduation date in the overall NLSY79 data. Those with a deceased father were more likely to have deceased mothers as well. In both the "Before" and "After" periods, African-Americans are more likely to have a deceased father than their peers who graduated high school in the same time period; African Americans comprise 30.2 percent of the overall sample but 35.7 and 45.8 percent of the sample for those with deceased fathers in the before and after periods, respectively. The fraction Hispanic remains relatively stable across groups, though the "After & Living Father" category contains slightly more Hispanic people than other categories. The variation in the covariates between categories is partially due to differential attrition between the categories, since this study can only use data from participants who completed the 40+ Health Survey questions. Given these differences in covariates across categories, I control for demographics in my final IV specification.

Returning to Table 2, the final two columns show the first stage results that are used for the IV analysis in this paper.[16] The results are quite similar to Dynarksi's and my replication results. In my preferred specification, which includes covariates and is shown in column (8) of

[16] Again, since a slightly different sample of people answered each health question, this regression uses the sample for "likelihood of reporting poor health" because it is the largest sample.

Table 2, eligibility for the Social Security Student Benefit program significantly increased one's likelihood of attending any college full-time by age 23 by 16.8 percentage points.

C. The Effect of Education on Health

Now we focus on the ultimate objective of the paper – estimating the effect of education on health. In the summary statistics shown in Table 3 Panel C, we see some striking differences in the likelihood of developing adverse health conditions across the categories.[17] Those with deceased fathers in the "Before" category are 9 percentage points more likely to have pain interfere with their work at home or on the job than those with deceased fathers in the "After" category (13.2 percent vs. 4.17 percent). They are also 11 percentage points more likely to have hypertension and 8 percentage points more likely to have arthritis. Only the difference in the likelihood of having heart problems has the expected sign: those with a deceased father in the "Before" group are slightly less likely to develop heart problems than those with a deceased father in the "After" group (3.7 percent vs. 4.1). The rest of the health variables are fairly similar across the categories. These results are the opposite of what we would expect if education leads to improved health, since people in the "Before" category had access to survivors benefits and people in the "After" group did not. On the other hand, there could be trends across cohorts in health outcomes (that is, people in the after period could be healthier for reasons other than the policy) that are not accounted for in this simple difference. The full IV analysis will control for these possible trends.

[17] Note that these summary statistics are only for the "complete at least one year of college" sample, for brevity's and consistency's sake. The summary statistics for the "attend college by age 23" is not substantially different.

The results of the IV regressions without covariates are presented in Table 4. Given the large literature documenting positive correlations between education and good health, we would expect the coefficients of interest in my IV regressions to be negative. In general the coefficients are wrong-signed but insignificant. Only the coefficient on "likelihood of having a diagnosed heart problem" is negative, indicating that attending college full-time by 23 is associated with a decrease in one's chances of reporting diagnosed heart problems by a statistically insignificant 5.3 percentage points. The coefficient in the "probability of having diagnosed hypertension" regression suggests that attending college full-time by 23 increases the likelihood of having diagnosed hypertension by 55.1 percentage points. The coefficient of interest is unexpectedly positive, large, and insignificant, for the rest of the outcomes as well. These puzzling results depart from what previous literature and correlations between health and education would suggest. The results from the 2SLS model that uses "complete at least one year of college by 23" are largely similar in size, sign, and significance to the results using the other definition of college-going. Only the coefficient on "likelihood of reporting hypertension" changes substantially between the two models; the coefficient is 5 percentage points larger in the "complete at least one year of college by 23" model than in the "attend college full-time by 23" model and marginally significant. The addition of covariates (see Table 5) has little effect on the results.

Before exploring the source of these unexpected results, we examine the effect of college-going on health behaviors. Buckles et. al's paper finds suggestive evidence of a link between education and reduced smoking and increased exercising, but causal evidence has remained elusive. I investigate how increases in education affect smoking and exercising (defined as engaging in vigorous activity at least once a week).[18] These serve as dependent variables in my IV model. In Table 6, we see that, according to my preferred specification (column (2)), attending college full-time by 23 makes one 20 percentage points more likely to have smoked more than 100 cigarettes in one's lifetime. Attending college full-time by 23 also makes one less likely to engage in vigorous physical activity at least once week by 9.5 percentage points (column (4)). Like the coefficients for the regressions estimating effects of education on health outcomes, these estimates are insignificant (except for the estimate in column (3)) and their confidence intervals all include 0. Thus I cannot reject large positive or negative changes in health behaviors due to increases in education.

D. Discussion of Results

As the coefficients on all of the health outcomes (except hypertension) and the health behaviors are insignificant and have confidence intervals that include zero, the paper's central result is that we do not find evidence of an effect of education on health or health behaviors. This does not, however, mean that there is no effect of education on health. It is possible that the findings reflect statistical error. An explanation of the unexpected coefficient sign and size requires further investigation.

The imprecision and unexpected sign and size of these estimates stem in part from the weakness of the instruments. An F-test of significance on the "Before*Deceased Father" variable as an instrument for "Attend College Full-Time by 23" shows that the instrument is fairly weak, with an F-statistic of 5.8-5.9, depending on the specific regression (see Table 4). The F-statistic

[18] I use the questions "have you ever smoked more than 100 cigarettes in your lifetime" (from NLSY 1998) and "how often do you engage in vigorous physical activity?" (NLSY 2000) to code the two indicator variables. "Vigorous activities" include running, aerobics, swimming, etc.

for "Complete At Least One Year of College" is similarly small, ranging from 5.0 to 5.1. An F-statistic of 10 or higher is the typical standard for a "strong" instrument. A weak instrument tends to have large standard errors, making it difficult to detect a significant coefficient (John Bound, David A. Jaeger, and Regina M. Baker 1995). Thus, the IV model for health outcomes falls far short of the statistical ideal. The F-statistics in the health behaviors regressions are above the cutoff of 10 because the compositions of the samples in those models are different from those in the previous model.[19] Even with the stronger instrument, the results for the health behaviors were surprising, which indicates that a larger issue is at the root of the unexpected results.

Digging deeper into the root cause of these unexpected findings may provide additional insights. The simple correlations between education and health in the NLSY79 data have the expected size and sign, which leaves the reduced form as the main reason why my IV results were negative.[20] The reduced form of an IV regression is a regression of the dependent variable, in this case, a health outcome, on the instrument, which is the interaction of "Before" and "deceased father," as well as on the other controls. Intuitively, the reduced form is a difference in difference estimating the effect of the Social Security Student Benefit program on health. A simple comparison of health outcome means across the categories indicates the source of the odd results from the IV regressions. Consider Table 7, which diagrammatically shows the difference-in-differences analysis. If the program removal affected people as we expected, in that it increased people's overall health by increasing their education, (A-B) would be negative,

[19] E.g., 300 people in the "poor self-reported health" sample had missing data for "ever smoked," while 281 people in the "ever smoked" sample had missing data for the "poor self-reported health" sample.

[20] In Appendix Table 1, I estimate OLS models that regress health on education, as measured by "complete at least one year of college by 23," to confirm the expected association between health and education. This analysis cannot be interpreted as representing the causal effect of education on health for reasons detailed above, but is intended to show the direction of correlation. The results are negative, as expected: increasing one's education is associated with less adverse health outcomes.

indicating that the share of people reporting fair or poor health in the "After" period is larger than in the "Before" period. We would also predict that (C-D) would be small in magnitude, given that cohort effects are probably small across this narrow time span. Using "likelihood of reporting poor health" as our health outcome, we see that (A-B) is positive (.0317) and (C-D) is negative (-.0291). The positive sign on (A-B) already reveals the issues we saw in the IV regression, as eligibility for the Student Benefit Program seemingly *increased* one's likelihood of having poor self-reported health. We then calculate (A-B)-(C-D), which only augments the magnitude of the seemingly positive effect of the removal of the Student Benefit program on health outcomes (.0317-(-.0291)=.061). In Table 8 the results of the reduced form model using "attend college full-time by 23" as the independent variable yield the same positive sign on the coefficient of interest.[21] This simple exercise partially explains why we see such large and positive results in the IV regressions.

This discussion can also help to explain why the magnitudes of the estimate effects are so large. In the first stage (Table 2), the Social Security Benefit program induced roughly an extra 1 out of 6 people to attend college full-time by age 23. Multiplying the reduced form of effect of the program on the likelihood of reporting poor health by 6 approximates the effect of increased education for those who went to college because of the Student Benefit program on health (6 X .061= .366); as expected, this number approximates the magnitude of the IV results.

While the preceding discussion illustrated mechanically why we obtain these large and negative estimates, it may be helpful to consider how sampling error may contribute to this situation. In particular, the small sample size of the "After & Deceased Father" category, which contains only 48 people, likely contributed to the unexpected results. Whenever you sample a

[21] This sample is slightly different because there are some people for whom I do not have education data but for whom I do have health data.

population, the sample mean is likely to differ from the true population mean. Drawing a sample of a fixed size multiple times will yield a different sample mean each time, but having a large sample size improves the chances that your sample mean will more closely approximate the population mean in any given sample. With a sample size as small as 48, there is a relatively high chance that any single random sample will have a sample mean vastly different than the population mean. In Figure 1 we see a hypothetical representation of the distributions of "likelihood of having poor self-reported health" for the "Before & Deceased Father" and "After & Deceased Father" groups. As the values on the x-axis move towards the right, one's health is *worse*. From Table 3, we know that the likelihood of poor self-reported health is higher in the "Before & Deceased Father" group than in the "After & Deceased Father" group, which is the opposite of what we would expect. Figure 1 shows how this result may stem from how the sample means differ from the true population means. The *increase* in health for those who were *not* eligible for the Student Benefit program and were therefore less likely to attend college suggests that the "After & Deceased Father" sample is healthier than the average person in that group. In other words, there may be a spurious correlation between the "Deceased Father*Before" interaction term and the health outcomes.

Finally, it is worth noting that all of the confidence intervals for the coefficients of interest are large and include 0, which means I cannot reject improvements in health due to increased education. Using poor self-reported health as the dependent variable in the IV specification with controls, the confidence interval around "attend college full-time by 23" suggests anything from a 30 percentage point decline to a 105 percentage point increase in the likelihood of reporting poor health. Fundamentally, my results show that I lack the statistical power to draw conclusive findings about the effect of education on health.

VIII. Conclusion

Estimating the effect of education on health outcomes continually poses a challenge to researchers. The majority of studies discussing the link of education on health rely on correlations, which cannot get at the ultimate question of causality. The few research papers that have estimated causal effects of education on health rely on data on international data or U.S. data from 1972 or before, making their conclusions less relevant for today's policy context.

The approach used in this paper offers a hypothetically useful way to obtain a causal effect of education on health. It uses the removal of a benefit program that caused an exogenous decrease in education as an instrument for education. This theoretically enables us to estimate how people's health outcomes are affected by variation in college attendance alone, without bias from unobserved characteristics of individuals. Unfortunately, given the small sample sizes in my data, I am unable to reject large declines or large improvements in health outcomes due to college attendance. Despite the inconclusiveness of my estimates, the empirical strategy in this paper could be used in future research with a larger, representative sample to continue the important work of determining the effect of education on health.

Fully understanding how education affects one's long-term life outcomes would enable policy-makers to craft education policy that maximizes positive long-term outcomes for society as a whole. The United States already spends many billions on incentives for college education; in 2013, the federal government lost $14.4 billion in foregone revenue due to the American Opportunity Tax Credit, which gives tax filers up to a $2,500 credit for spending money on college tuition or fees (McCann 2016). In the current primary campaign season, two candidates support expanding financial aid for community college to make it tuition free for low-income individuals (or, in the case of Bernie Sanders, all individuals). The costs and benefits of such proposals are typically evaluated by looking at how increases in education would affect wages; however, the discussion in this paper reminds us there could be other important impacts of education that should be considered in policy conversations.

IX. References

Bound, John, David A. Jaeger, and Regina M. Baker. 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): 443–50.

Buckles, Kasey, Andreas Hagemann, Ofer Malamud, Melinda S. Morrill, and Abigail K. Wozniak. 2013. “The Effect of College Education on Health.” Working Paper 19222. National Bureau of Economic Research. http://www.nber.org/papers/w19222.

DeWitt, Larry. 2001. “The History of Social Security ‘Student’ Benefits.” January. http://www.ssa.gov/history/studentbenefit.html.

Dynarski, Susan. 2003. “Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion.” *The American Economic Review* 93 (1): 279–88.

“Federal Pell Grant Program Data Books 1980-81 - 1989-90.” 2011. Reports; Datasets; Statistical Reports. April 11. http://www2.ed.gov/finaid/prof/resources/data/pell-historical/hist-2.html.

Goesling, Brian. 2007. “The Rising Significance of Education for Health?” *Social Forces* 85 (4): 1621–44.

Grossman, Michael. 2015. “The Relationship between Health and Schooling: What’s New?” Working Paper 21609. National Bureau of Economic Research. http://www.nber.org/papers/w21609.

Institute of Education Sciences. 2015. “Digest of Education Statistics, 2013.” Accessed September 29. http://nces.ed.gov/programs/digest/d13/tables/dt13_303.40.asp.

Lleras-Muney, Adriana. 2006. “The Relationship between Education and Adult Mortality in the United States: Erratum.” *Review of Economic Studies* 73 (3): 847–847. doi:http://restud.oxfordjournals.org/content/by/year.

Lochner, Lance. 2011. “Non-Production Benefits of Education: Crime, Health, and Good Citizenship.” Working Paper 16722. National Bureau of Economic Research. http://www.nber.org/papers/w16722.

McCann, Clare. 2016. “Tax Benefits for Education.” *EdCentral*. Accessed April 14. http://www.edcentral.org/edcyclopedia/federal-education-tax-benefits/.

Meghir, Costas, Mårten Palme, and Emilia Simeonova. 2012. “Education, Health and Mortality: Evidence from a Social Experiment.” Working Paper 17932. National Bureau of Economic Research. http://www.nber.org/papers/w17932.

Oreopoulos, Philip. 2003. “Do Dropouts Drop Out Too Soon? International Evidence From Changes in School-Leaving Laws.” Working Paper 10155. National Bureau of Economic Research. http://www.nber.org/papers/w10155.

U. S. Social Security Administration. 2015. “Annual Statistical Supplement, 2014.” *Social Security*. Accessed September 29. https://www.socialsecurity.gov/policy/docs/statcomps/supplement/.

X. Tables and Figures

Table 1: Category Breakdown by Sample

Attend College Full-Time by 23				
	Dynarski (Published Paper)	Dynarski Replication (1988 Definition)	Dynarski Definition + Expanded "Before"	Final Sample
Before & Living Father	2745	2802	6027	4776
Before & Deceased Father	137	137	324	244
After & Living Father	1050	1109	1109	874
After & Deceased Father	54	56	56	48
Total	3986	4104	7516	5942
Complete At Least 1 Year of College by 23				
	Dynarski (Published Paper)	Dynarski Replication (1988 Definition)	Dynarski Definition + Expanded "Before"	Final Sample
Before & Living Father	2745	2791	5950	4705
Before & Deceased Father	137	136	314	238
After & Living Father	1050	1107	1107	872
After & Deceased Father	54	56	56	48
Total	3986	4090	7427	5863

Note: Please see “Replication Results” for a description of the different samples.

Table 2: First Stage Results by Sample (Y= Likelihood of attending any college full time)

	Dynarski		Own		Expanded Sample		Final Sample	
	Difference -in- difference s (1)	Add covariates (2)	Difference- in- differences (3)	Add covariates (4)	Difference- in- differences (5)	Add covariates (6)	Difference- in- differences (7)	Add covariates (8)
Deceased father X Before	0.182*	0.219**	0.180*	0.176*	0.186**	0.138	0.172**	0.168**
	(0.096)	(0.102)	(0.0941)	(0.0960)	(0.0870)	(0.0905)	(0.0763)	(0.0764)
Deceased father	-0.123	Y	-0.133	0.869	-0.133	-0.289	-0.0747	-0.0643
	(0.083)		(0.0812)	(0.775)	(0.0812)	(0.441)	(0.0719)	(0.0717)
Before	0.026	Y	0.0409**	-0.436	0.0405**	-1.301***	0.0717***	0.0666***
	(0.021)		(0.0206)	(0.512)	(0.0194)	(0.467)	(0.0181)	(0.0181)
Senior-year family income		Y		Y		Y		N
AFQT score		Y		Y		Y		N
Black		Y		Y		Y		Y
Hispanic		Y		Y		Y		Y
Father attended college		Y		Y		Y		N
Mother attended college		Y		Y		Y		N
Single-parent household		Y		Y		Y		N
Family size		Y		Y		Y		N
Female		Y		Y		Y		Y
Age in 1988		Y		Y		Y		N
State dummies		Y		N		N		N
All covariates X before		Y		Y		Y		N
All covariates X deceased father		Y		Y		Y		N
Deceased mother		N		N		N		Y
Age dummies		N		N		N		Y
R-Squared	0.002	0.339	0.00309	0.300	0.00212	0.294	0.00504	0.00978
N	3,986	3,986	4104	4104	7526	7516	5942	5942

Table 3: Summary Statistics					
	Before & Living Father	Before & Deceased Father	After & Living Father	After & Deceased Father	Overall Sample
Panel A: Covariates					
Hispanic	0.153	0.172	0.182	0.146	0.158
	(0.360)	(0.378)	(0.386)	(0.357)	(0.365)
Black	0.295	0.357	0.316	0.458	0.302
	(0.456)	(0.480)	(0.465)	(0.504)	(0.459)
Female	0.526	0.520	0.466	0.458	0.517
	(0.499)	(0.501)	(0.499)	(0.504)	(0.500)
Deceased Father	0	1	0	1	0.0491
	(0)	(0)	(0)	(0)	(0.216)
Deceased Mother	0.0220	0.0697	0.0206	0.125	0.0246
	(0.147)	(0.255)	(0.142)	(0.334)	(0.155)
Panel B: First Stage (Education) Outcomes					
Attend College Full-Time by 23	0.521	0.619	0.450	0.375	0.514
	(0.500)	(0.487)	(0.498)	(0.489)	(0.500)
Complete At Least One Year of College by 23	0.458	0.530	0.394	0.292	0.450
	(0.498)	(0.500)	(0.489)	(0.459)	(0.498)
Panel C: Second Stage (Health) Outcomes					
Poor Self-Reported Health	0.0959	0.115	0.125	0.0833	0.101
	(0.294)	(0.319)	(0.331)	(0.279)	(0.301)
Pain Interferes with Work	0.102	0.132	0.0953	0.0417	0.102
	(0.303)	(0.339)	(0.294)	(0.202)	(0.303)
Arthritis	0.108	0.123	0.110	0.0417	0.108
	(0.310)	(0.329)	(0.313)	(0.202)	(0.311)
Mental Health Problems	0.113	0.127	0.131	0.0833	0.116
	(0.317)	(0.334)	(0.338)	(0.279)	(0.320)
Hypertension	0.170	0.172	0.156	0.0625	0.167
	(0.376)	(0.378)	(0.363)	(0.245)	(0.373)
Heart Problems	0.0249	0.0369	0.0206	0.0417	0.0249
	(0.156)	(0.189)	(0.142)	(0.202)	(0.156)
Diabetes	0.0522	0.0574	0.0481	0.0417	0.0517
	(0.222)	(0.233)	(0.214)	(0.202)	(0.221)
N	4776	244	874	48	5942

Note: “Poor self-reported health” means the respondent either reported “fair” or “poor” health.

Table 4: IV Regressions without Controls

	(1) Poor Self-Reported Health	(2) Pain Interferes with Work	(3) Arthritis	(4) Mental Health Problems	(5) Hypertension	(6) Heart Problems	(7) Diabetes
Complete At Least One Year of College by 23	0.380	0.487	0.458	0.391	0.600*	-0.0362	0.0841
	(0.309)	(0.306)	(0.300)	(0.314)	(0.331)	(0.181)	(0.195)
N	5863	5860	5857	5853	5857	5860	5861
F-statistic	5.898	5.967	5.856	5.842	5.932	5.879	5.833
Attend College Full-Time by 23	0.350	0.474	0.485	0.360	0.551	-0.0530	0.0669
	(0.337)	(0.321)	(0.327)	(0.327)	(0.361)	(0.191)	(0.201)
N	5942	5939	5936	5933	5937	5940	5941
F-statistic	5.086	5.191	5.052	5.051	5.152	5.073	5.048

Standard errors in parentheses

* p<0.10 ** p<0.05 *** p<0.01

Note: “Poor self-reported health” means the respondent either reported “fair” or “poor” health.

All y-variables are dummies. All models are estimated as linear probability models.

Table 5: IV Regressions with Controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Poor Self-Reported Health	Pain Interferes with Work	Arthritis	Mental Health Problems	Hypertension	Heart Problems	Diabetes
Complete At Least One Year of College by 23	0.406	0.505	0.468	0.385	0.659*	-0.0279	0.0673
	(0.321)	(0.320)	(0.304)	(0.318)	(0.365)	(0.184)	(0.197)
N	5863	5860	5857	5853	5857	5860	5861
F-statistic	5.685	5.764	5.686	5.629	5.767	5.686	5.636
Attend College Full-Time by 23	0.378	0.494	0.490	0.350	0.612	-0.0451	0.0550
	(0.347)	(0.329)	(0.321)	(0.326)	(0.388)	(0.191)	(0.201)
N	5942	5939	5936	5933	5937	5940	5941
F-statistic	5.274	5.384	5.281	5.228	5.393	5.278	5.239

Standard errors in parentheses

* p<0.10 ** p<0.05 ***p<0.01

Note: Controls include dummies for deceased mother/deceased father/age/sex/race/ethnicity/“before”

“Poor self-reported health” means the respondent either reported “fair” or “poor” health.

All y-variables are dummies. All models are estimated as linear probability models.

Table 6: IV Regressions for Health Behaviors

	(1)	(2)	(3)	(4)
	Smoked More than 100 Cigarettes	Smoked More than 100 Cigarettes	Vigorous Activity At Least Once Per Week	Vigorous Activity At Least Once Per Week
Attend College Full-Time by 23	0.474	0.202	-0.603*	-0.0954
	(0.368)	(0.335)	(0.321)	(0.311)
Controls	No	Yes	No	Yes
N	5923	5923	5704	5704
F-statistic	10.65	11.66	12.61	12.26

Standard errors in parentheses

* p<0.10 ** p<0.05 *** p<0.01

Note: Controls include dummies for deceased mother/deceased father/age/sex/race/ethnicity/“before”.

“Vigorous activity” includes running, swimming, aerobics, cycling, etc.

Table 7: Reduced Form							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Poor Self-Reported Health	Pain Interferes with Work	Arthritis	Mental Health Problems	Hypertension	Heart Problems	Diabetes
Deceased Father*Before	0.0639	0.0849**	0.0830**	0.0589	0.105**	-0.00762	0.00947
	(0.0468)	(0.0373)	(0.0374)	(0.0475)	(0.0455)	(0.0315)	(0.0331)
N	5942	5936	5931	5929	5933	5936	5936

Standard errors in parentheses

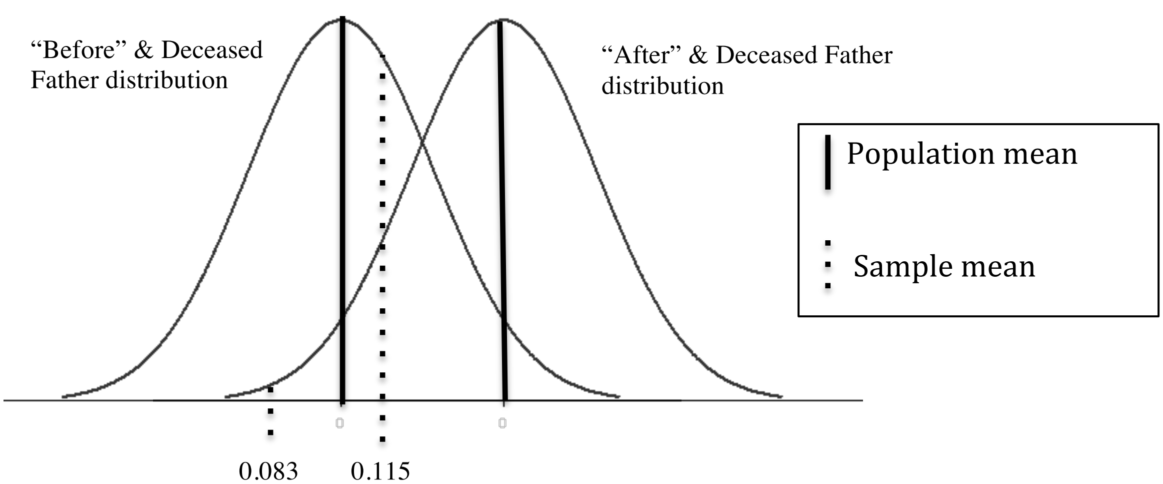
* p<0.10 ** p<0.05 *** p<0.01

Note: Controls include dummies for deceased mother/deceased father/age/sex/race/ethnicity.

"Poor self-reported health" means the respondent either reported "fair" or "poor" health.

Table 8: Difference-in-differences for likelihood of reporting bad health			
	Before	After	Difference
Deceased Father	A = 0.115	B = 0.0833	(A-B) = .0317
Living Father	C= 0.0959	D= 0.125	(C-D) = -.0291
Difference-in-Differences			(A-B)-(C-D) = .061

Figure 1: Likelihood of reporting poor health distributions by category



XI. Data Appendix

Like Dynarksi, I dropped everyone who did not finish their junior year of high school. The "Before" indicator variable was created by assigning a 1 to everyone who was attending 12th grade in 1979, 1980, or 1981, and to those who had already graduated high school by 1979 (their highest grade completed in 1979 was 12th grade or higher). Those who attended their senior year of high school in 1982 or 1982 were assigned a 0 for the "Before" variable. Dynarski's "Before" variable was coded based only on if one had graduated between 1979 and 1981 or after 1981, as she did not include those who were seniors in high school before the NLSY survey began. I used the 1988 NLSY survey to create the "deceased father" dummy, since that survey included a section in which respondents detailed any periods during their childhood when they did not live with both biological parents and the reason why they did not live with those parents. Individuals who lived with both parents until 18 were assigned a 0. Those who did not live with both parents because their mother, stepfather, or stepmother no longer lived with them were also assigned a 0. Those who did not live with their biological father until they were 18 because of a reason other than the father's death were assigned a 0. Only those who did not live with both biological parents until 18 because their father died were assigned a 1. Dynarski used the same method, to the best of my knowledge.

To code "complete at least one year of college by age 23" I used the question "what was the highest grade that you completed?" in the year that the respondent turned 23. Dynarski's comparable variable was coded in the same way. Coding the "attend college full-time by 23" proved more challenging, particularly when I expanded my sample. I used the "did you attend college full-time" question in each of the survey years before an individual turned 23 to create the indicator for "attend college full-time by 23." When I expanded the sample, I included 18-22 year olds who had graduated high school before 1979 and who may have attended college full-time *before* 1979. In 1984 (and later years), NLSY collected information on the three most recent colleges that an observation attended in their lifetime and whether or not they attended that college full-time or part-time. For those who attended college before 1979, I assigned them a 1 for "attended college full-time by age 23" if they were enrolled in a college before 1979 and attended that college full-time. Those who only ever attended college part-time or never attended college before the age of 23 were assigned a 0. Dynarski only used the "in the last year, did your college consider you part-time or full-time?" question from the survey years before the respondent turned 23. She did not need to use the information on the three most recent colleges attended since she did not include those who had graduated high school before 1979 in her sample.