The Effect of College Education on Mortality

Kasey Buckles, University of Notre Dame and NBER

Andreas Hagemann, University of Michigan

Ofer Malamud, University of Chicago and NBER

Melinda Morrill, North Carolina State University

Abigail Wozniak, University of Notre Dame, NBER and IZA

September 2015

ABSTRACT

We exploit exogenous variation in years of completed college induced by draft-avoidance behavior during the Vietnam War to examine the impact of college on adult mortality. Our estimates imply that increasing college attainment from the level of the state at the 25th percentile of the education distribution to that of the state at the 75th percentile would decrease cumulative mortality for cohorts in our sample by 6.5 percent relative to the mean. Most of the reduction in mortality is from deaths due to cancer and heart disease. We also explore potential mechanisms, including differential earnings, health insurance, and health behaviors, using data from the Census, ACS, and NHIS.

(JEL: I12, I23, J24)

Keywords: College education; mortality; health; Vietnam draft

---

Acknowledgments: The authors would like to thank William Evans, Bob Kaestner, Brigham Frandsen, and seminar participants at the NBER Education Program Meetings, Population American Association (PAA) Meetings, the Society of Labor Economists (SOLE) Meetings, University of Illinois at Chicago, NBER Education Group meetings, University of Illinois Urbana-Champaign, Georgia State University, University of Maryland, Brigham Young University, University of Calgary, University of Akron, University of Maryland-Baltimore County, the Federal Reserve Bank of New York, and the University of Notre Dame for helpful comments. Elizabeth Munnich provided valuable research assistance. All errors are our own.

I. Introduction

Schooling is highly correlated with subsequent health outcomes, including later life mortality. For example, in 2007, the age-adjusted mortality rate of high school graduates aged 25 to 64 was more than twice as large as the mortality rate of those with some college or a college degree (Xu et al., 2010). If these associations between health and education reflect a causal relationship, they would represent a significant non-pecuniary return to education. They would also imply that policies meant to increase educational attainment could serve as an important means for improving health. However, there is substantial debate about whether these associations actually represent causal effects (see the reviews by Grossman, 2006, Cutler and Lleras-Muney, 2010, and Mazumdar 2012). Previous work estimating the causal effect of education on mortality has exploited changes in education due to compulsory schooling requirements (e.g. Lleras-Muney (2005), Clark and Royer (2010), Meghir et al. (2012). To our knowledge, this paper is the first to provide a causal estimate of the effect of college education on mortality. In doing so, we contribute to knowledge about the impacts of education on health at the higher end of the schooling distribution, where the observed health gradient in education is steeper (Montez et al. 2012).

We use variation in college attainment induced by draft-avoidance behavior during the Vietnam War in an instrumental variables strategy. This enables us to identify the effect of increased higher education on mortality for men who were eligible to be drafted into the Vietnam War. This strategy builds on Card and Lemieux (2000, 2001) who document the excess educational attainment among cohorts induced to enter college in order to defer conscription. While Card and Lemieux focus on differences in induction risk *across* birth cohorts, we also exploit state level variation in induction risk *within* cohorts—an approach developed by Malamud and Wozniak (2012) in their study of the effect of college on mobility. The existence of state level variation allows us to decompose national induction risk into two constituent parts: induction risk faced by a young man's own state co-

hort and induction risk faced by young men of that cohort in the rest of the country. Our decomposition yields two instruments, which we use to identify the impact of the two endogenous variables—education and veteran status—in our empirical framework. This approach is an advance over studies that used the Card and Lemieux measures to identify the effect of college-going on health outcomes using only year-to-year variation in induction risk (e.g., De Walque, 2007; Grimard and Parent, 2007; MacInnis 2006).

We merge our data on national and state-level induction risk with the Vital Statistics Mortality Files from 1981 to 2007 and the 1980 U.S. Census to construct birth state-by-cohort level mortality rates, both cumulative and by cause. We first establish that the well-known gradient between education and mortality is present and statistically significant in our Vital Statistics data using OLS specifications. OLS shows that a one-year increase in a birth state-birth year cohort's average years of college education is associated with 22 fewer deaths per 1,000 persons by 2007. Given that the average mortality rate for these cohorts over this period is 138.6 per 1,000 persons, this represents a decrease of sixteen percent.

Our instrumental variables estimates indicate a causal effect that ranges from 17 to 35 fewer deaths per 1,000 persons. None of the 2SLS estimates are significantly different from the OLS estimates. For the birth state-birth year cohorts in our sample, our estimates imply that increasing college attainment from the level of the state at the 25th percentile of the education distribution to that of the state at the 75th percentile would decrease cumulative mortality by 6.5 percent relative to the mean. The largest effects are found for the impact of college education in lowering deaths due to cancer and heart disease, which represent the leading causes of mortality in our sample. Moving from the 25th to the 75th percentile in the cohort education distribution would decrease the cancer mortality rate by 16.6 percent, with lung cancer accounting for over half of the reduction.

Before concluding, we use the Census, the American Community Survey (ACS) and the Na-

tional Health Interview Survey (NHIS) to explore mechanisms that might explain the documented relationship between college education and mortality. We examine the causal effect of college attainment on auxiliary outcomes such as health insurance and wages, as well as the cross-sectional relationship between college attainment and health behaviors such as smoking, exercise, and obesity.

This paper helps to fill an important gap in the literature on the relationship between education and health. As mentioned earlier, previous analyses of the causal impacts of education on health outcomes, such as mortality, have relied on variation at the lower part of the schooling distribution. For example, Lleras-Muney (2005), Clark and Royer (2010), and Meghir et al. (2012) all exploit changes in compulsory schooling requirements to examine whether increased schooling improved the health of students on the margin of dropping out before 12th grade. Lleras-Muney (2005) finds large and significant effects of increased education on declines in mortality in the United States, whereas Clark and Royer (2010) find no evidence for an impact of education on mortality in England.[1] Meghir et al. (2012) find improvements in mortality and other health measures for affected cohorts following a reform in Sweden. However, regardless of the causal impact of schooling on health at the margin of dropping out of high school, the causal relationship may be different at the margin between high school and college. Montez et al. (2014) fit various functional forms to the education-health relationship in the United States and conclude that the causal relationship during the post-secondary range is likely different from that during compulsory years of schooling. Moreover, estimating the effect of education on health at the college margin may be of particular interest given that the largest increase in educational attainment in recent years has occurred among students entering college (Turner, 2004), as well as the fact that health disparities across education groups have widened in recent decades (Meara et al. 2008; Jemal et al. 2008).

The findings in this paper have important implications for both health and education policy.

[1] Arendt (2005) and Albouy and Lequien (2009) also find no statistically significant impact of compulsory school reforms on health outcomes in Denmark and France, respectively.

People value health, and the health returns to education may represent a substantial fraction of the pecuniary returns. Indeed, Cutler and Lleras-Muney (2006) calculate that their estimates of the health benefits from education increase the total returns to education by 15 to 55 percent. Because higher education policy in recent years has been focused on increasing college attainment, this represents an important margin of analysis. If individual investments in college education are suboptimal because of credit constraints, externalities, or lack of information, the presence of additional health returns to college strengthens the case for subsidizing education. This is particularly relevant given recent discussions about the rising cost of college and the decline in federal financial aid for college students. On the other hand, a positive causal impact of higher education on health may pose a dilemma for health policy. Health improvements, like smoking cessation, may reduce health care costs in the short run only to increase them in the long run as individuals live longer or as other health issues arise (Bearman, et al., 2011). Our analysis will help inform policymakers interested in the link between education policy and national healthcare spending.

The effect of postsecondary education on certain health behaviors—smoking in particular—has been examined previously. De Walque (2007) and Grimard and Parent (2007) exploit year-to-year variation in induction risk faced by cohorts of young men during the Vietnam War to identify the impact of education on smoking. Using different datasets (NHIS and the CPS Tobacco Supplements, respectively) and different specifications, they find that additional education has a negative and significant effect on the likelihood of smoking. MacInnis (2006) uses a similar identification strategy to document the effect of education in reducing obesity and its co-morbidities such as hypertension and adult-onset diabetes. Our paper extends this identification strategy by incorporating within-cohort variation in induction risk to account for veteran status. We also examine a wide range of potential mechanisms in order to provide a more complete picture of the channels through which higher education may affect health and mortality.

Finally, although it is not our main focus, this paper contributes to research examining the impact of military service during the Vietnam War on health outcomes and behaviors (Angrist, Chen, and Frandsen, 2010; Dobkin and Shabani, 2007; Conley and Heerwig, 2012; Hearst, Newman, and Hulley, 1986).[2] We find that veteran status has a statistically significant protective effect on mortality conditional on survival to 1980. This is potentially explained by higher rates of health insurance access among veterans in our sample.

II. Background on the Vietnam Draft

Our instrumental variables strategy exploits variation in the risk of induction (also referred to colloquially as the risk of "being drafted") to which young men in the US were exposed during the Vietnam conflict. This section provides a brief overview of the sources of this variation. A more detailed discussion can be found in Malamud and Wozniak (2012); henceforth MW.

Approximately 2 million American men were drafted during the Vietnam War. The Selective Service System, which comprised over 4,000 local draft boards across the nation at that time, was responsible for registering recruits and classifying them for either deferment or selection. Responsibility for devising and meeting the national target number of conscriptions rested with the federal Department of Defense (DoD). To achieve this target, the DoD issued monthly "draft calls" that divided the national number into quotas assigned to state draft boards, which did the active work of ordering men to be inducted.

Faced with an excess of eligible draft-age men, draft boards adopted generous deferral policies toward large categories of men. Enrollment in a four-year college was the second most common deferral category, after the exemption for dependents (Semiannual Reports of the Director of the

[2] These studies all exploit variation in veteran status induced by the Vietnam draft lottery, which is a different source of identification than our own. None reject the hypothesis that the impact of veteran status on health outcomes is zero.

Selective Service System, 1967-1973).[3] The Military Service Act of 1967 codified the existing de facto arrangement by stating that college students in good standing could defer induction until receipt of an undergraduate degree or age 24, whichever occurred first. Over 1.7 million college deferments had been granted by 1967. Although men who received college deferments were technically eligible for induction until age 35, very few men between the ages of 26 and 35 were ever drafted. Card and Lemieux (2000) estimate that, among men born between 1945 and 1947, those with a college degree were only one-third as likely to serve in Vietnam as compared to those without a college degree. Thus, the incentive to enroll in college to avoid the draft during these years was large.

Our identification strategy relies on two sources of variation in induction risk: over time and across states. The existence of intertemporal variation in induction risk is well-known (Card and Lemieux, 2000) and has been used in previous research (e.g., De Walque, 2007; Grimard and Parent, 2007). From 1960 to 1963, inductions were fairly low at approximately 8,000 per month. However, following the Gulf of Tonkin incident in August 2, 1964, Congress authorized an expanded role for the U.S. military in Vietnam. Inductions more than doubled from 1964 to 1965 and again from 1965 to 1966. By the spring of 1968, in the midst of raging student protests, the rate of inductions reached a peak of almost 42,000 a month.

The introduction of the draft lottery in 1969 led to a substantial change in the induction process. However, college deferments continued to be issued until September 1971, and men who were already enrolled were allowed to retain their deferment until the end of the school year. Risk of induction during this period was also much lower since men were at risk of induction for only a single year and the overall rate of inductions was substantially lower, falling from about 20,000 per month in late 1969 to 2,000 per month in late 1971. In February of 1973, the draft was suspended and no more inductions took place.

[3] See also Tatum and Tuchinsky, *Guide to the Draft*, Ch. 3. By contrast, enrollment in a two-year college was not considered grounds for automatic deferment (Rothenberg 1968).

The existence of state level variation in induction risk was less well known at the time, and remains so today. This type of variation arose through two channels. The first was uneven and idiosyncratic application of formal procedures across the thousands of local draft boards. In their influential study of the draft, Davis and Dolbeare write, "The conclusion seems inescapable: local board autonomy implies both within- and between-state variability, even among socioeconomically similar board jurisdictions." (Davis and Dolbeare, 1968, p. 84) Similar idiosyncrasies were described in the report of the U.S. National Advisory Commission on Selective Service, in 1967. A second source of state and year variation in induction risk was communication delays between federal, state, and local officials. These delays meant that the DoD assigned quotas using registrant numbers that were several months old. Thus, draft risk for an eligible man at a point in time was not only a function of the number of men in his state currently eligible for the draft but also of the number available several months ago. The current pool could be much larger than the past pool if, for example, a large number of local men graduated high school thus becoming draft eligible or much smaller if a large number married or aged out of the draft pool in the intervening months.

III. Empirical Strategy

We use variation in college attainment caused by draft-avoidance behavior during the Vietnam conflict to isolate the causal effect of education on mortality. An artifact of this identification strategy is that the likelihood an individual is a veteran also varies systematically across cohorts in our sample. Given that veteran status is a plausible determinant of health, it is important to control for this variable in our estimation strategy. However, selection into military service during the Vietnam War was likely based on characteristics that are unobserved in our data, which would confound our estimates of veteran status and potentially other covariates as well. To deal with this, we exploit changes in both national and state-level induction risk to generate exogenous variation in

both college attainment and veteran status.[4]

This strategy is similar to the one described in MW. In Appendix A, we provide formal econometric evidence of the relevance of our instruments by developing and implementing a test for first-stage power under the identifying assumptions appropriate to our model. In Appendix B, we provide a detailed discussion of the interpretation of our estimates when both direct and indirect effects are possible. To summarize, the estimated parameters identify "local average treatment effects" from two different interventions. Within the constant effects framework, the estimated coefficient on years of education measures the causal effect of education for individuals whose educational decisions are affected by their draft risk. Intuitively, as long as the correlation structure of the instruments and the endogenous variables is sufficiently rich, our instruments mimic a situation in which variation in college access and variation in veteran status come from two independent randomized experiments.[5] We are able to test whether the correlation structure meets this criterion, but we emphasize that our causal interpretation crucially depends on our parametric assumptions. We do not claim to *non-parametrically* identify a local average treatment or any other (direct, indirect, or total) causal effect.

A. Instruments for College Education and Veteran Status

To identify the effects of college education on health, we employ the same strategy as MW, who extend an instrumental variables strategy inspired by Card and Lemieux (2000, 2001); henceforth CL. Like CL, we assume that draft avoidance was proportional to the risk of induction. To account for the mechanical relationship between inductions and veteran status, we exploit state level variation within the cohort level variation identified by CL. The existence of state-cohort level varia-

[4] A number of related papers propose alternative solutions. De Walque (2007) includes veteran status but treats it as an exogenous variable. MacInnis (2006) instruments for veteran status using a quartic polynomial in age, while Grimard and Parent (2007) consider a specification that instruments for veteran status using information about early health problems.
[5] In other words, this is analogous to the thought experiment in which names are randomly selected from an urn to be treated with college access without affecting veteran status. For estimates of the effect of veteran status, all names are then placed back in the urn, and a second round of names is randomly drawn to be treated with veteran access without changing college status.

tion allows us to break national induction risk into its constituent parts and obtain separate instruments that can be used to identify both college attainment and veteran status. Thus, young men faced state-level cohort risk that is analogous to the CL measure, where *s* indexes state of residence and *c* indexes one-year birth cohorts:

$$(1) \qquad staterisk_{sc} = \frac{\left(\sum_{t \mid c=19\ to\ 22} I_{sct}\right)/4}{N_{sc}}$$

I_{sct} is the number of inductions from birth state–birth year cohort *sc* in year *t*. Like CL, we construct an average draft risk for the years a man was 19 to 22 since draft risk was non-trivial for men ages 20 to 22. N_{sc} is birth state-cohort size. This measure of state cohort risk is our first instrument. We then use our state level data on *I* and *N* to construct a second instrument in the following manner:

$$(2) \qquad nationalrisk_{sc} = \frac{\left(\sum_{t \mid c=19\ to\ 22} \sum_{-s} I_{-sct}\right)/4}{\sum_{-s} N_{-sc}}$$

This measure defines national cohort risk for a man born in state *s* and in birth cohort *c* as the number of inductees from the set of all other states, denoted –*s*, and birth cohort *c*, divided by the total number of such men at age 17. In other words, the numerator and denominator in (2) are national level inductions and cohort size for a cohort *c* minus the birth state inductions and cohort size for the same cohort, respectively.

To construct the measures in equations (1) and (2), we obtained data on the number of inductees from 1961 to 1972 in each state from reports of the Selective Service. We estimate state cohort size using enrollment numbers spanning 1959 to 1970, the academic years in which our cohorts of interest were in 11th grade. Thus state-cohort level risk (henceforth state risk) for a young man born in Alabama in 1950 equals the number of inductees from Alabama in 1969 (the year he turned 19) divided by the number of students enrolled in 11th grade in Alabama in 1967. National level risk

for the same young man roughly equals the number of men inducted nationally in 1969 divided by the size of his birth cohort; more precisely, we subtract own state inductions from the numerator and own state cohort size from the denominator. We then construct an average national draft risk for the years a man was 19 to 22.

B. Validity of Our Instruments

Given the novelty of our identification strategy, some of the IV assumptions bear more discussion. Our instruments could fail if young men attempted to exploit local variation in induction risk by moving between localities. In this case, risk would not be truly randomly assigned. Our risk measures would only bind for men who were unwilling or unable to move to low risk jurisdictions, which might in turn be correlated with other unobservable characteristics related to health. MW document that this type of "local board shopping" was prohibited by draft board regulations.

Identification further requires the assumption that induction risk only affected health through either education or veteran status. There are two ways in which this might fail. First, the health screenings required to determine draft eligibility might uncover an important health condition earlier than it might otherwise have been detected, thereby encouraging individuals to treat the condition and improving future health. Our reading of the historical literature suggests that this was unlikely. The required exams were cursory and did not involve testing for health conditions not easily observed in a brief physical exam. It is therefore unlikely these exams provided young men with information they did not already have.[6]

Second, young men may have viewed marriage or fertility as a means to reduce their chances of being drafted, and these choices may have connections to health. Marriage alone was never

[6] In fact, Tatum and Tuchinsky describe the exams this way, "Since the [Army] examines large numbers of men each day, and since the doctors, orderlies, and clerks assigned there must process them in assembly-line fashion as quickly as possible, the examinations are often careless. If you have a medical or other condition which should disqualify you, bring letters and other evidence from your own doctors…" (Ch. 6). A 1964 report to the president noted that "The current published medical standards are roughly the same as those which were in effect at the close of World War II," (U.S. Department of Health, Education, and Welfare, 1964, Ch. 3).

grounds for deferment, although for a brief period (1963-1965), married, childless men were placed in a lower priority category for induction (U.S. Selective Service 2008). Before and after that period, such men were treated equivalently to single men for the purposes of induction.[7] Fatherhood, on the other hand, was grounds for deferment throughout the draft era (U.S. Selective Service 2008). There is some evidence that contemporaneous fertility increased immediately around the time that treatment of childless, married fathers changed (Bailey 2011, Kutinova 2009). Bitler and Schmidt (2012), on the other hand, examine fertility changes among women who were likely to have been affected by an absence of men throughout the entire Vietnam War. Their findings indicate that men in our cohorts were modestly but statistically significantly *less* likely to become fathers during the Vietnam years. [8] Overall, the available evidence suggests that the impact of draft risk on fertility among draft-eligible men was modest, and because these studies focus on contemporaneous fertility, the effect on completed fertility is unknown.[9] Moreover, we know of no evidence linking fertility timing to later health for men. We conclude that there is little cause for concern that marriage or fertility among draft-age men had important health impacts in our cohorts.

Finally, there may be concern that our instruments are correlated with unobservable underlying health status. Our instruments are highly non-linear so any confounding variation in health status would also have to be quite non-linear and vary in such a way that is highly correlated with induction risk at age 18. We view this as unlikely. There is evidence that the health of potential inductees varied widely across states (President's Taskforce on Manpower Conservation, 1964), with high levels of health-related rejections in the population as a whole. Such differences in the levels of health by

[7] The availability of the marriage exemption for the first few of our cohorts is unlikely to substantially impact our results. Card and Lemieux (2000) show that enrollments in college were increasing commensurate with induction risk over the 1960 to 1969 period. There is no inflection in enrollments with the elimination of the marriage deferment in 1965.
[8] Fertility effects were largest among blacks and low skilled whites (Bitler and Schmidt, 2012). Blacks are omitted from our sample.
[9] Kutinova (2009) focuses on timing of the first birth in response to changes in the treatment of childless, married men in 1965 and notes that available Census data make studying the impacts of these policy changes on completed fertility difficult.

state are likely to be persistent and therefore will be absorbed by the state fixed effects and region-year trends in our models. Moreover, the same task force report forecasts levels of likely recruit health for the US male population over the period 1964-1970 (President's Taskforce on Manpower Conservation, 1964). The forecasts are linear and stable over time, further suggesting that although rejection on the basis of poor health was common, there is little reason to suspect its prevalence varied significantly across cohorts during the course of the war. We discuss this issue again when we present our robustness checks at the end of Section V.C.

IV. Data Sources and Estimating Equations

We perform our main analysis on data aggregated to the birth state-birth year cohort level, and we refer to these cells as birth state-cohorts. Our main measure of mortality at the birth state-cohort level is the cumulative mortality rate per 1000 persons between 1981 and 2007, constructed as follows:

$$MR_{sc} = \frac{deaths_{sc}\ between\ 1981\ and\ 2007}{cohort\ size_{sc}^{1980}} * 1{,}000 \tag{3}$$

where *s* is state of birth and *c* is year of birth. Thus the mortality rate gives the fraction of the cohort that died by 2007, conditional on having been alive in 1980. We also construct cumulative mortality rates by cause. These are defined as above, but where MR_{sc}^{j} is mortality due to cause *j,* and the numerator is all deaths due to cause *j* over the period.

The data for our analysis come from two sources. First, we use data from the IPUMS microdata 5% samples of the 1980 Census (Ruggles et al., 2004) to construct the denominators in the mortality rate measures as well as the birth state-cohort levels of college education and veteran status. We restrict our sample to men born between 1942 and 1953. These are the years for which both inductions and enrollments are available at the state level, which are the two components of our in-

duction risk measures. Finally, we omit non-white men from our sample because they may have been less able to avoid the draft by enrolling in college (Kuziemko 2010). Our second source is the Vital Statistics mortality data for the period 1981 to 2007, which contains observations on all deaths in the United States, at the annual level. We use these data to construct the numerators in our mortality rate measures, after applying the same sample restrictions applied to the Census data. We match numbers of deaths to the appropriate state-cohort information using year and state of birth. We also have information on the primary (or underlying) cause of death, and we use this to construct cause-specific mortality rates.

Our measure of educational attainment is years of education above high school; in the Census this variable ranges from zero to eight. All of the findings in this paper are robust to instead using college completion as our measure of educational attainment, as can be seen in the specifications presented in Buckles, et al. (2013). Our measure of veteran status is based on veteran information in the Census. Specifically, we define a veteran in our cohorts as someone who answered affirmatively that he was a Vietnam veteran.[10] We exclude anyone from our sample who continues to be on active duty in the military at the time of the Census, although this is a very small fraction of our sample. We also exclude observations with imputed values for a number of key variables.[11]

We estimate our main regressions at the birth state by birth cohort level. We use the aggregated data for several reasons. First, our source of exogenous variation occurs at this level. Second, since we cannot observe veteran status, education, and mortality for a representative set of individuals in a single data set, it allows for construction of a birth state-cohort panel from which we can estimate our main econometric model. Finally, an individual's risk of death in a given year is low, so

[10] Census respondents are asked whether they are veterans, and if they answer yes, they are asked to identify a specific period of conflict. A small number of men in our cohorts report that they are veterans but not Vietnam veterans. We also omit those living in group quarters.

[11] Specifically, we drop observations with imputed values for age, education, birth place, and veteran status. Our results are not sensitive to including the imputed values and those living in group quarters or on active duty in the military.

the fit of our model is likely better at the aggregate level than in a model estimating rare outcomes in individual level data. Our main econometric model is as follows:

$$MR_{sc} = a_1 C_{sc} + a_2 V_{sc} + \alpha_4 X_{sc} + \tau trend_{sc} + \delta_s + \varepsilon_{sc} \quad (4)$$

where *s* indexes state of birth and *c* indexes birth year, and the dependent variable MR_{sc} is the mortality rate per 1,000 persons as defined in (3)—either in total or by cause. The variable V_{sc} is the fraction of veterans, and C_{sc} is average years of college education for the cohort. The evidence in Montez et al. (2012) suggests that the educational gradient in health differs between years of post-secondary education and years of K-12 education. Their preferred specification for the relationship between education and health status models this as a linear relationship in years of education over the range of post-secondary schooling. By entering college attainment as years of schooling, our model captures both these features.

In addition, *trend* is a region-specific linear trend in birth cohort, X_{sc} is a set of state-cohort level controls, and δ_s represents a full set of state-of-birth dummies. Allowing for state-of-birth fixed effects removes variation arising from states with persistently higher or lower than average induction rates, which may be associated with other state characteristics (e.g., industrial composition) that are correlated with mortality rates. X_{sc} includes two variables to capture labor market conditions facing a cohort at the time of the college enrollment decision: (a) the employment-to-population ratio in the individual's state of birth the year his cohort turned 19, and (b) the log of the number of respondents from a birth state and year cohort in the 1960 Census. Together, these approximate the changes in labor demand and labor supply which may have occurred alongside changes in state-level induction risk.[12]

[12] The literature tends to find no consistent, significant relationship between local labor market conditions and college attendance (Wozniak, 2010; Card and Lemieux, 2001). However, early labor market conditions affect longer-run labor market outcomes (Wozniak, 2010; Kahn, 2010; Oreopoulos, et al., 2012) and labor market outcomes have been linked to

Our endogenous variables, C and V, are predicted from first stage equations that include the remaining right hand side covariates in equation (4) plus functions of *staterisk* and *nationalrisk* as defined in (1) and (2), respectively. De Walque (2007) uses a non-linear specification of induction risk to instrument for both education and veteran status. Consequently, our main results explore the robustness of estimates from three different first stage specifications in which *staterisk* and *nationalrisk* are entered as linear, quadratic, or cubic functions. Estimation is implemented via standard linear 2SLS as well as LIML for the higher order specifications, weighted by the number of observations in each state-cohort cell.[13] Standard errors are clustered at the birth-state level.[14]

Descriptive statistics are shown in Table 1, which summarizes the variables used in our analysis for the sample of white men born between 1942 and 1953. The average cumulative mortality rate per 1,000 persons as defined in (3) is 138.58. Table 1 also gives mortality rates by cause-of-death for ten important causes for these cohorts, in order of prevalence. The most common causes of death were cancers (36.5 per 1,000 white men) and heart disease (35.7 per 1,000). External causes (accidental injury, suicide, and homicide) accounted for 23.0 deaths per 1,000 men. For younger men, external causes like accidental injury and suicide are the leading causes of death, but by age 42 cancer and heart disease are the leading causes.

later adult health (Sullivan and von Wachter, 2009). We do not include fertility and marriage rates as controls since these may be endogenous to education and may therefore lead to biased estimates of education's total effect on mortality.

[13] See Wooldridge (2002) pp. 622-624 concerning 2SLS versus an approach with a probit first stage when the endogenous variable is a dummy variable. In some cases, the latter is more efficient but may tend to produce larger point estimates. Given our concerns about possible upward bias, we implement 2SLS estimation.

[14] Because our mortality measures combine Census data with data on the universe of mortality outcomes, we do not further correct our standard errors to account for multiple data sources, as for example, in two-sample IV.

Table 1: Summary Statistics

VARIABLES	Mean	SD
Census Data		
Years of College	1.99	0.38
College Graduate	0.3508	0.0666
Veteran	0.3092	0.1231
Individual Observations	14,392,122	
Mortality Data		
1981-2007 Mortality Rates by Cause of Death:		
Total	138.58	38.03
Cancers	36.52	15.18
Heart Disease	35.73	13.99
Accidental Injury	13.67	3.49
Suicide	7.31	1.37
Infectious and Parasitic Diseases	7.20	2.41
Liver Disease	5.75	1.93
Diabetes	3.82	1.44
Cerebrovascular Disease (Stroke)	3.61	1.47
Chronic Low Respiratory Disease	3.21	2.16
Homicide	1.98	0.97
Total Deaths	1,994,481	
State/Birth Year Cells	600	

Notes: Census data are from the 5% sample of the 1980 U. S. Census, available from IPUMS. Mortality data are from the Vital Statistics Multiple Cause of Death files from 1980 to 2007. The sample is restricted to white men born between 1942 and 1953. Veterans include any respondent that served in active duty in the Vietnam War. Respondents currently in active duty are excluded. Means are weighted by cell size. Mortality rates are deaths over the period per 1,000 population, where population is the cohort size in 1980.

V. Results

A. First Stage Results

Table 2 presents results from estimating the first stage with linear, quadratic, and cubic specifications of the risk measures. For transparency, we estimate two first stage equations for

Table 2: First Stage Estimates of Effect of Induction Risk on College Education and Veteran Rates

	Dependent Variable:					
	Years of College	Veteran	Years of College	Veteran	Years of College	Veteran
National Induction	4.682***	2.349***	9.046***	2.865***	9.560***	1.237*
	(-0.557)	(0.185)	(1.294)	(0.354)	(2.214)	(0.728)
National Risk ^2			-35.069***	-3.897*	-87.697***	25.194***
			(7.412)	(2.068)	(32.303)	(9.238)
National Risk ^3					407.644**	-154.080***
					(160.136)	(42.970)
State Induction Risk	-1.029*	0.768***	-1.199	0.653**	5.687***	1.171*
	(0.560)	(0.162)	(1.159)	(0.311)	(1.995)	(0.702)
State Risk ^ 2			1.040	0.627	-90.952***	-5.907
			(5.627)	(1.731)	(27.384)	(8.736)
State Risk ^ 3					382.226***	25.502
					(121.520)	(36.988)
Cragg-Donald F-stat	8.20		9.33		13.55	
[p-value]	[0.00]		[0.00]		[0.00]	
Angrist-Pischke F-Stat	10.71	96.89	14.89	239.27	23.81	259.39
F-Stat	239.49	2154.74	172.27	1137.29	145.68	851.75

Notes: Standard errors are in parenthesis and are clustered by birth state. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level respectively. *Years of College* and *Veteran* measure average years of higher education and the fraction of veterans, respectively, in the birth state-birth year cohort. *State induction risk* and *national induction risk* are defined as in equations (1) and (2) respectively. Number of observations at the birth state-birth cohort level is 600 in each OLS regression and all regressions are weighted by cell size. Additional controls include birth state fixed effects, the employment-to-population ratio, the cohort size (derived from the 1960 Census and defined at the birth-year level), and birth-region trends. See the text for a discussion of the null hypotheses of the three F-statistics. See Table 1 notes for data sources and sample restrictions.

each specification—predicting years of college and veteran status separately—although 2SLS estimates these equations jointly. Consistent with the manner in which 2SLS identifies endogenous variables, both equations include national and state cohort risk as identifying variables. We also report the Cragg and Donald (1993) F statistic for identifiability developed in 3.A., the Angrist-Pischke F statistics, and the traditional single equation first stage F statistics.

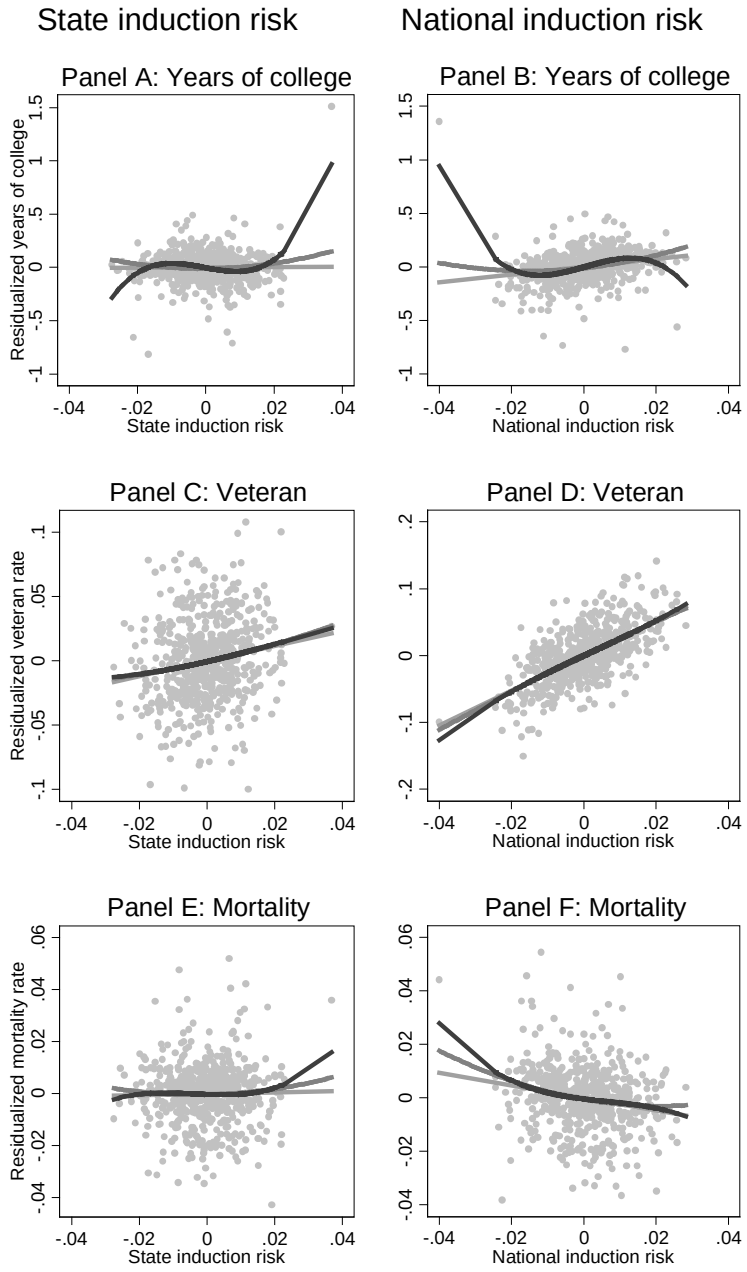
The first stage estimates with veteran status as the dependent variable show that veteran status varies positively with both national and state cohort risk. These patterns are also evident in Panels C and D of Figure 1. This is reassuring since higher rates of induction risk at both the state and national level should lead more young men to go to war. Across specifications, the F statistics suggest both that we meet the identification assumptions in our model (implied by the Cragg-Donald statistics) and that this first stage has substantial power (implied by the Angrist-Pischke F statistics).

We do not present the reduced form regression estimates in table form, but they are available upon request. The final two panels of Figure 1 display the reduced form relationships of mortality with state and national risk. Panel E does not reveal a strong relationship between later adult mortality and state risk while Panel F indicates a clear negative relationship between mortality and national risk. Together with the prior panels, these graphs show that state risk has a large positive effect on veteran status but not on college attainment or on mortality. On the other hand, national risk has a large positive effect on both veteran status and college attainment, as well as a corresponding negative effect on mortality. Thus, these patterns suggest that it is college attainment, more than veteran status, that is associated with the reduction in mortality since both instruments affect veteran status but only national induction risk affects college attainment. This is consistent with our 2SLS estimates described below.

B. OLS and 2SLS Effects of College Education on Total Mortality

Table 3 presents our OLS and 2SLS estimates of the effect of college attainment on

Figure 1: Residualized Scatterplots of First-Stages and Reduced-Form Relationships



Notes: State induction risk and national induction risk are defined as in equations (1) and (2) respectively. All variables plotted are residuals adjusted for birth state fixed effects, the employment-to-population ratio, cohort size (derived from the 1960 Census and defined at the birth-year level), birth-region trends and the corresponding alternative risk measure (i.e. state risk when plotting against national risk, and vice versa). Each point represents a birth state-birth year cohort. Linear, quadratic, and cubic fits are shown in the solid lines. For clarity of exposition, the fitted lines do not use the one outlier point at the extreme right and left of the state and national induction risk distribution respectively (corresponding to cohorts born in Alaska in 1942).

cumulative mortality, based on equation (4) (see Section V.D. below for a discussion of the estimates of the effect of veteran status). For the IV results, we show specifications in which the induction risk variables are included as linear, quadratic, and cubic polynomials. For the quadratic and cubic specifications the model is over-identified, so we can estimate the model using both 2SLS and limited-information maximum likelihood (LIML). We do this to further investigate the validity of our instruments—LIML is less precise than 2SLS but is also less biased in the presence of confounding variables. If the coefficients from the two approaches are similar, this is evidence that the bias in 2SLS is small (Angrist and Pischke 2009).

The OLS coefficient for years of college is -22.31. This indicates that increasing college attainment from the level of the state at the 25th percentile of the education distribution to that at the 75th percentile (a 0.52 year increase) is associated with 11.6 fewer deaths per 1,000 men. This is consistent with the well-documented educational gradient in most health outcomes.

Table 3, Column [2] shows the point estimates of the causal effect of college attainment and veteran status on mortality using a linear specification for the instruments. The results are remarkably similar to the OLS estimates, though they are less precise. With the quadratic specification, the estimated effect of college attainment is substantially larger and is statistically significant; with the cubic specification the point estimates are again very close to OLS and are statistically significant. In columns [3] through [6], the 2SLS and LIML results are nearly identical, further supporting our identification strategy. Taken together, the 2SLS results imply a large causal role for college attainment in reducing cumulative mortality for these cohorts. The magnitude of our estimated impacts is also economically significant. For example, the estimate from Table 3, Column [5], implies that increasing college attainment levels from those of the state at the 25th percentile in the distribution to those at the 75th percentile leads to nine fewer deaths per 1,000 men.

Table 3 shows that the impact of college attainment on later mortality is robust to three

Table 3: OLS and IV Estimates for the Impact of College Education on Cumulative Mortality, 1980-2007

		IV Specification:				
	OLS	Linear-2SLS	Quadratic-2SLS	Quadratic-LIML	Cubic-2SLS	Cubic-LIML
	[1]	[2]	[3]	[4]	[5]	[6]
Years of College	-22.308***	-21.152	-34.421***	-34.761***	-17.311***	-17.030***
	(3.804)	(16.983)	(8.499)	(8.727)	(5.420)	(5.669)
Veteran Status	-42.816***	-40.267**	-24.9017**	-24.490**	-45.0654***	-45.3507***
	(5.385)	(18.896)	(11.111)	(11.371)	(7.340)	(7.597)
Observations	600	600	600	600	600	600
R-squared	0.9516	0.9519	0.9511	0.9511	0.9518	0.9518

Notes: Standard errors are in parenthesis and are clustered by birth state. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level respectively. The dependent variable is the mortality rate from 1981-2007 per 1,000 persons. The IV specifications use the national and state-level induction risk to instrument for the average years of higher education (*Years of College*) and the fraction of veterans (*Veteran*) at the birth state-birth cohort level. The column headings indicate the functional form of the instruments in the first stage equation (linear, quadratic, or cubic) and the choice of model (two-staged least squares or limited-information maximum likelihood). Number of observations is 600 in each regression and regressions are weighted by cell size. Additional controls include birth state fixed effects, the employment-to-population ratio, the cohort size (derived from the 1960 Census), and birth region trends. See Table 1 notes for data sources and sample restrictions.

specifications of the first stage equation. We have examined robustness to other alternative specifications. In results not shown but available upon request, we substitute a common quadratic time trend for the linear birth region trends with similar results. We have also verified that our estimatesare robust to excluding region-specific trends and non-linear state-cohort varying controls (employment-to-population ratio and log cohort size).

A final check is shown in Appendix Table 1. Here we estimate our baseline OLS specification on subgroups of the sample defined by birth year. The "late" cohorts, those born 1951-1953, are somewhat different from the "early" cohorts, those born 1942-1944, and middle cohorts in that they were partially exposed to the draft lottery and had access to more generous GI Bill benefits upon release. One might be concerned that the late cohorts drive our results. If so, this might raise questions about our identification strategy. We do not have the first stage power to estimate our 2SLS specifications for a subset of only three cohorts.[15] However, the late cohorts are unlikely to contribute much to our estimates of the negative impact of college on mortality because the OLS results show no health gradient in college attainment for them. This is probably because these cohorts have not yet reached the ages where mortality is due to causes for which college attainment has a greater effect (something we show in more detail later). We therefore view our main results as robust to concerns about differential policy treatment across cohorts.

One might be concerned that our measures of induction risk are correlated with the state-cohort underlying health status. As discussed in Section III.B, we view this as unlikely, as we believe any state-cohort level health differences are likely controlled for either via birth state fixed effects or birth region trends.[16] Nevertheless, in robustness checks not shown here, we have confirmed that our main results are robust to the addition of proxies for population health in each state for each

[15] We generally lose first stage power when we drop more than two cohorts from the analysis.
[16] Results not shown, but available upon request, confirm that the estimated effects are not sensitive to the inclusion of birth state trends.

cohort's birth year (the state by birth year annual birth rate and the Korean War casualty rate). In addition, results are robust to the inclusion of a 1950's dummy, which is a general indicator of the baby boom generation. These checks provide strong evidence that our findings are not due to correlation between our instruments and underlying state-cohort health differences.[17]

C. OLS and 2SLS Effects of College Education on Mortality by Cause of Death

Table 4 shows the effect of state-cohort college attainment on leading causes of death for men in these cohorts. Causes are shown in descending order of prevalence between 1980 and 2007. OLS results are in Panel A and 2SLS results are in Panel B. The specification is as in column [5] of Table 3. For cancer and heart disease—the two leading causes of death—the point estimates on years of college are negative and statistically significant for both OLS and IV and are larger for IV. Moving from the birth state-birth year cohort at the 25th percentile of the education distribution to the 75th would reduce cancer deaths by about 6 per 1,000 (16.6% relative to the mean) and heart disease deaths by 4.4 per 1,000 (12.3% relative to the mean). For comparison, active treatment with statins reduces deaths by 13 per 1000 patients (LaRosa, He, Vupputuri, 1999).

In Appendix Table 2, we show results for three finer categories of deaths from disease. Lung cancer accounts for nearly one-third of the cancer deaths in our sample; the next leading specific cause is colon cancer, which accounts for about 10% of cancer deaths. The IV estimate of the effect of an additional year of college on lung cancer is -6.87 (s.e. = 1.09), suggesting that 59% of the decline in total cancer deaths is due to lung cancer. This is greater than the decline that would be observed if college education had an equal effect on deaths from all cancers. That college education has a particularly big effect on death from lung cancer is not surprising, as De Walque (2007) and Gri-

[17] Note that we restrict our analysis to men and do not attempt a falsification test using women. While Card and Lemieux (2001) use female college attainment as a counterfactual for male college attainment in the absence of the Vietnam War, female college attendance may have been affected by male college-going during the Vietnam years. In particular, the large inflows of men into college could have crowded out women who would otherwise have attended or encouraged more women to attend to take advantage of marriage market prospects. Moreover, to the extent that there are spillovers in health behaviors and outcomes between married men and women of a similar age, women would not offer an appropriate counterfactual for men.

Table 4: OLS and IV Estimates for the Impact of College Education on Mortality, by Leading Causes of Death

	Cancers	Heart Disease	Accidental Injury	Suicide	Infectious/ Parasitic	Liver Disease	Diabetes	Stroke	Chronic Low Resp.	Homicide
Panel A: OLS										
Years College	-8.05***	-6.67***	-1.13**	-0.01	-0.86	-0.39	-0.63**	-0.52***	-1.27***	-0.34***
	(1.24)	(1.21)	(0.44)	(0.31)	(0.53)	(0.32)	(0.30)	(0.19)	(0.28)	(0.13)
Veteran Status	-14.94***	-9.49***	-3.03***	-1.85***	-0.56	-1.06**	-0.53	-1.83***	-3.70***	0.08
	(1.63)	(1.84)	(0.68)	(0.51)	(0.53)	(0.43)	(0.38)	(0.29)	(0.32)	(0.17)
Panel B: IV										
Years College	-11.59***	-8.37***	-0.22	0.96	2.92***	2.41***	-1.31***	-1.28***	-1.92***	-0.04
	(1.98)	(2.36)	(0.96)	(0.60)	(1.05)	(0.83)	(0.44)	(0.40)	(0.54)	(0.29)
Veteran Status	-9.92***	-6.86**	-3.70***	-2.91***	-4.70***	-4.15***	0.32	-0.85*	-2.83***	-0.2
	(2.70)	(3.38)	(1.18)	(0.82)	(1.26)	(0.99)	(0.49)	(0.49)	(0.60)	(0.36)

Notes: Standard errors are in parenthesis and are clustered by birth state. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level respectively. The dependent variable is the mortality rate by cause from 1981-2007 per 1,000 persons. The IV specifications use a third-order polynomial in the national and state-level induction risks to instrument for average years of higher education (*Years College*) and fraction of veterans (*Veteran*) using a 2SLS model. Number of observations is 600 in each regression and regressions are weighted by cell size. Additional controls include birth state fixed effects, the employment-to-population ratio, the cohort size (derived from the 1960 Census), and birth region trends. See Table 1 notes for data sources and sample restrictions.

mard and Parent (2007) both find that education has a large and statistically significant negative effect on smoking. Cigarette smoking is known to be a leading cause of both lung cancer and cardiovascular (heart) disease (Chaloupka and Warner, 2000).

College education has a negative and statistically significant effect on mortality from three other causes: diabetes, stroke, and chronic low respiratory conditions (CLRCs). In the IV specifications, an additional year of college decreases deaths per 1,000 white men from both diabetes and stroke by about 1.3 and from CLRCs by 1.9. The latter effect is especially large and is also likely related to the effect of college on the likelihood of smoking, since smoking is a known cause of CLRCs like chronic bronchitis and emphysema (Chaloupka and Warner, 2000).

The OLS results show a negative association between a college education and death by two external causes—accidental injury and homicide. But for both of these, the IV coefficient is small and statistically insignificant. There is also no evidence that college education decreases deaths by suicide.

Finally, for infectious and parasitic diseases and liver diseases, the point estimate for years of college is negative in the OLS specification but positive and statistically significant in the 2SLS specifications. For these cohorts, 63% of deaths due to infectious and parasitic diseases were due to HIV, and as Appendix Table 2 shows, the estimated coefficient for HIV deaths specifically is 1.35—about half of the size of the coefficient for all infections and parasitic diseases. There are several ways in which college attendance might have increased one's risk of contracting HIV during this period. For example, college attendance is associated with living in an urban area and engaging in same-sex relationships, which are significant risk factors for contracting HIV (Shilts 1987).[18] Liver diseases, on the

[18] About half of all HIV cases in the U.S. are for men who have sex with men (Center for Disease Control 1997). Lambert et al. (2006) find that upper-class college students had more positive attitudes toward gay and lesbians than underclassmen, suggesting that college increases acceptance of homosexuality. Smith (1991) found that among respondents to the General Social Survey in 1989, college graduates were more likely to have engaged in risky sexual behaviors than less educated groups, although shares in the highest risk group were similar across education levels. Smith also found that risky behavior is higher among younger and more urban individuals, who are more likely to be college-educated.

other hand, include cirrhosis which can be caused by excessive alcohol consumption—48% of deaths to cirrhosis in 2007 were alcohol-related (Yoon and Yi, 2010). If a college degree increases this behavior, graduates may be more at risk for liver diseases.[19] We investigate the relationship between college education and alcohol consumption in the next section.

The results in Tables 3 and 4 indicate that college education has a negative effect on total mortality and that deaths due to heart disease and cancer are disproportionately affected. College-going has a particularly strong negative effect on deaths from causes related to smoking.

D. OLS and 2SLS Effects of Veteran Status on Mortality

As noted above, existing studies find mixed evidence on the relationship between veteran status and mortality for these cohorts. Our OLS and IV estimates in Table 3 show a negative relationship. In our preferred specification, a one-standard deviation increase in the rate of veteran status for a cohort would decrease deaths per 1,000 men in the cohort by about 5.5, or 4% relative to the mean. The IV results by cause of death in Table 4 show that veteran status has a statistically significant negative effect on deaths to all causes except diabetes and homicide. This may seem contrary to the perception that returning Vietnam veterans experienced high mortality rates, and elevated suicide rates specifically. In fact, a 1987 study by the Centers for Disease Control found that Vietnam veterans experienced higher rates of deaths due to external causes—including suicide—than other conflict veterans. However, this effect was limited to the first five years after service. After five years, the rates of death for all causes except drug-related deaths were comparable to other veterans and to the general population. Because our results are conditional on survival to 1980, veterans and non-veterans in our sample should have similar baseline mortality rates. In the next section, we consider potential channels for a positive effect of veteran status on health, including earnings differ-

[19] Competing risks may also explain these positive coefficients (as well as the coefficients on other specific causes of death). Since the first two columns in the table indicate that college prevented some men from dying of cancer and heart disease, these men may now be more susceptible to death by other causes (Honoré and Lleras-Muney, 2006).

ences and insurance access. We also discuss these results in the context of the existing literature in the conclusion.

VI. Potential Mechanisms

A number of potential mechanisms might explain the negative effect of college education on cumulative mortality and the heterogeneity in the effects across specific causes of death. For example, if the college-educated are less likely to smoke, this may explain the negative coefficients on cancer and respiratory disease. Mokdad et al. (2004) find that behavioral factors such as smoking, obesity, and excessive alcohol consumption represent almost half of the "actual" causes of death in the United States in 2000. Cutler and Lleras-Muney (2010) describe the role of cognition, resources, prices, and health behaviors in mediating the relationship between education and health for nationally-representative populations in the US and UK.

In this section, we explore potential mechanisms for our mortality results. First, using data from the Census, we examine the causal relationship between education and earnings or wages for our cohort. Second, we assess the impact on access to health insurance in the 2008-2010 American Community Survey (ACS) and 1980 National Health Interview Survey (NHIS) Health Insurance Supplement. Finally, using data from the 1985 NHIS, Health Promotion and Disease Prevention (HPDP) Supplement, we consider health outcomes, including diabetes, hypertension, and BMI, and health behaviors, including smoking, exercise, and alcohol consumption. Of course, there may be additional mechanisms, including effects on adult migration as described by MW, which could explain our reduced-form results.

A. Earnings and Wages

An extensive literature demonstrates that increased schooling is causally related to higher wages and earnings (Card, 1999). To the extent that higher income enables individuals to purchase

Table 5: OLS and IV Estimates for the Impact of College Education on Wages and Earnings, 1980

	Log Earnings		Log Wages	
VARIABLES	OLS	IV	OLS	IV
Years of College	0.101***	0.133***	0.0533***	0.0522***
	-0.0141	-0.0271	-0.0112	-0.0195
Veteran	0.119***	0.0618	0.0895***	0.0847**
	-0.0198	-0.0429	-0.0156	-0.0358
Mean Earnings/Wages	9.543		1.979	

Notes: Standard errors are in parenthesis and are clustered by birth state. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level respectively. Data are birth state-cohort cell averages from white men in the 1980 Census IPUMS extracts, born between 1942 and 1953. Dependent variable is mean log real earnings or log real wages. The IV specifications use a third-order polynomial in the national and state-level induction risks to instrument for average years of higher education (*Years of College*) and the fraction of veterans (*Veteran*). Underlying microdata sample is therefore restricted to those with valid earnings in 1980. Wage calculations exclude those with missing hours or weeks worked. Earnings of top-coded observations are multiplied by 1.5. Hourly wages are calculated as total wage and salary income divided by hours of labor supply. We truncate the bottom 1 percent of hourly earners and those above 1.5 times the maximum annual income amount divided by 1,750 (35 hours per week for 50 hours per year). Specifications are otherwise identical to those in the cubic specifications of Table 3. Number of observations is 600 in each regression and regressions are weighted by cell size.

better health care or acquire a healthier lifestyle, it represents a potentially important mechanism in explaining the relationship between education and mortality. Cutler and Lleras-Muney (2010) demonstrate that family income is an important mediating factor between education and health. We examine this potential mechanism by estimating the effect of college attainment on wages and earn ings for our cohorts. Using data from the 1980 Census, Table 5 presents OLS and IV estimates based on the cubic first-stage specification we used in Table 4.

The OLS estimates for log earnings indicate that increasing average years of college education in a state by one raises earnings by 10 percent and hourly wages by 5.3 percent. For a similar

comparison with our health outcomes, this implies that increasing college attainment from the level of the state at the 25th percentile of the education distribution to that of the state at the 75th percentile is associated with a 5.25 and 2.77 percent increase in earnings and wages respectively. The IV estimates for earnings are larger, suggesting that an additional year of college increases cohort earnings by 13 percent while the IV estimates for wages are similar to OLS at 5.2 percent. As with the mortality results, the IV estimates are not significantly different from the OLS estimates.

These results in Table 5 are in line with previous estimates in the literature of the causal impact of college education on earnings.[20] However, we caution that our IV estimates for the impact of college attainment on earnings and wages are sensitive to the choice of specification. In Appendix Table 3, we show that the quadratic specifications yield smaller impacts for earnings and insignificant impacts on wages, while none of the coefficients are significant in the linear specification. We now turn to an exploration of other potential mechanisms.[21]

B. Health Insurance Measures in the American Community Survey

We consider the role of health insurance as an additional potential mechanism through which education might reduce mortality. Higher education qualifies individuals for jobs that are more likely to include benefits such as employer-provided health insurance (Hipple and Stewart, 1996). Health insurance has been linked to better access to preventative care, which should lead to better health and lower mortality, all else equal (Miller, 2012). However, Finkelstein and McKnight (2008) find that the introduction of Medicare, i.e., nearly universal access to health insurance after age 65, had very little impact on mortality rates. When considering health insurance status among those under age 65, those with poorer health or who are at higher risk of needing expensive medical

[20] Oreopoulos and Petronijevic (2013) review several studies measuring the return to a college education and conclude that the average premium to one year of college is between 7 and 15 percent. Using an RD design on Florida data, Zimmerman (2014) estimates that an additional year of college leads to an 11 percent increase in long-run earnings.
[21] We have also examined impacts on labor force participation. Our cubic specifications show that an additional year of college is associated with an approximately 1.4 percentage point increase in participation, for both OLS and IV, although the estimates are insignificant in the linear and quadratic specifications.

services might be more inclined to purchase health insurance, complicating the measurement of a causal effect of health insurance on mortality. Here we explore whether college attainment is associated with a higher probability of being covered by health insurance in order to assess whether access to health insurance is a channel through which education might improve health.

Our primary data on insurance comes from the 2008-2010 ACS. In Section VI.C, below, we present parallel results using the 1980 NHIS Health Insurance Supplement which confirm these main findings. We first explore whether the individual has any health insurance; we then consider private and public health insurance coverage separately. Private health insurance includes both employer-provided health insurance and plans that individuals purchase in the private market. Our public health insurance measure includes Medicaid, Medicare and disability, and the veteran's health insurance program CHAMPUS. We therefore anticipate that veterans are much more likely to have public health insurance coverage. Note that in the ACS individuals can be classified as having both private and public health insurance.

The main results are reported in Table 6. The IV estimates suggest that an additional year of college raises the overall rate of health insurance coverage by 2.6 percentage points (which is about 2.8 percent of the mean of 92.4). For private coverage, the college graduate coefficient is only significant in the OLS specification, but the magnitude is similar in both OLS and IV suggesting that college graduates are more likely to have private health insurance coverage. Finally, we see that rates of college going are unrelated to public coverage rates, although the OLS relationship is negative. Veterans have higher rates of public coverage in both specifications, which appears to drive their higher rates of coverage overall.

C. Health Insurance Measures in the National Health Interview Survey

We next present results for access to health insurance from the more detailed 1980 NHIS Health Insurance Supplement. The 1980 NHIS is chosen because this is the baseline year from our

Table 6: College Education and Health Insurance Access

	Any Insurance	Private Coverage	Public Coverage
Panel A: OLS			
Years of College	0.022***	0.044***	-0.026***
	(0.005)	(0.006)	(0.008)
Veteran Status	0.048***	-0.04**	0.242***
	(0.01)	(0.02)	(0.028)
Panel B: IV			
Years of College	0.026*	0.024	0.014
	(0.015)	(0.021)	(0.025)
Veteran Status	0.074***	-0.028	0.307***
	(0.017)	(0.019)	(0.023)
Mean of Dep. variable	0.924	0.837	0.179

Notes: Standard errors are in parentheses and are clustered by birth state. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level, respectively. Data are from ACS survey years 2008-2010. Samples restricted to white, male respondents born between 1942 and 1953, and aged 64 or younger at the time of the survey. Dependent variable is the fraction of individuals reporting having any insurance coverage (*Any Insurance*), private insurance coverage (*Private Coverage*), or public insurance coverage (*Public Coverage*). The IV specifications use a third-order polynomial in the national and state-level induction risks to instrument for average years of higher education (*Years of College*) and the fraction of veterans (*Veteran*). C-D F statistic is 9.28 (p-value of 0.00). First stage standard F statistics are 13.6 and 11.2 for years of college and veteran, respectively; A-P F statistics are 13.3 and 146.1 for college and veteran, respectively.

mortality analysis. Estimation using the NHIS data necessitates some changes in our specifications since we cannot observe state of birth or state of residence (but we do observe the Census region of residence) in the public use versions of the NHIS. We therefore estimate our insurance models using individual-level data, rather than data aggregated to the birth state-birth year cohort level. Our econometric model for the NHIS data is the following:

$$(5) \qquad insurance_{irc} = a_1 C_{irc} + a_2 V_{irc} + \tau trend_{rc} + \delta_r + \varepsilon_{irc}$$

where *i* indexes individuals, *r* indexes region of residence, and *c* indexes birth year. Here the dependent variable is an indicator for either any insurance coverage, private coverage, or public coverage.

The variable V is an indicator for whether the individual is a veteran, and C is years of college.[22] In addition, *trend* is a region-specific linear trend in birth cohort, and δ_r represents region of residence dummies. Note that because of the lack of state identifiers in the NHIS, we cannot use the set of state-cohort level controls described above (the employment-to-population ratio and the cohort size).

We do not present results using instrumental variables with the NHIS data because of power issues. The 1980 Census sample is over 200 times larger than the 1980 wave of the NHIS. Although one could merge together data from three decades of the NHIS using restricted-access data that includes state of birth, even then the sample size is not sufficient to have a strong first stage.[23] We therefore emphasize that these results are from descriptive regressions with minimal controls, and cannot be used to make the same type of causal interpretation as with our 2SLS results.

The results from linear probability model regressions, with standard errors clustered by region of residence, are reported in the top of Table 7. We also report the sample size for each specification and the mean of the dependent variable. We see that the OLS estimates for health insurance coverage using the 1980 NHIS are similar to those from the more recent ACS data, with the exception of the probability of having public health insurance coverage. We find that an additional year of college education is associated with a 1.9 percentage point higher probability of having any health insurance and a 2.2 percentage point higher probability of having private health insurance coverage.

When we consider public health insurance coverage, we see some differences. In the NHIS, individuals can be classified as having public coverage only if they do not have any private health insurance coverage. Those who do not have private insurance or Medicare are asked to state the rea-

[22] In the 1980 NHIS Health Insurance Supplement, education is measured in categories. To create years of college, we make the following conversions: (1) 13-14 years completed = 1.5 years of college; (2) 15 years completed = 3 years of college; (3) 16 years completed (college graduate) = 4 years of college; (4) 17+ years completed (graduate school) = 7 years of college. Anyone with 12 or fewer years of school is assigned a value of 0 years of college.

[23] Results using pooled data from the 1997-2009 NHIS are available upon request, but show a weak first-stage and no statistically significant coefficients in 2SLS. The power issues arise because we are limited to looking at men from only 11 birth year cohorts. Other studies using the NHIS (e.g., DeWalque 2007) use larger samples.

Table 7: Health Outcomes and Health Behaviors

Dependent Variable	N	Mean	Years of College *Coefficient (SE)*		Veteran *Coefficient (SE)*	
1980 Health Insurance Supplement:						
Any Health Insurance	6,292	0.890	0.019***	(0.002)	0.036**	(0.008)
Private Coverage	6,292	0.873	0.022***	(0.003)	0.034***	(0.005)
Public Coverage	6,292	0.020	-0.004***	(0.000)	-0.001	(0.004)
1985 NHIS HPDP Supplement:						
Body Mass Index (BMI)	3,069	25.82	-0.236**	(0.064)	0.095	(0.099)
Obese (BMI ≥30)	3,069	0.109	-0.015**	(0.004)	-0.011	(0.008)
Ever told had:						
Hypertension	3,031	0.187	-0.002	(0.002)	0.006	(0.007)
High cholesterol	3,035	0.045	0.003	(0.003)	-0.015**	(0.003)
Heart Condition	3,033	0.032	-0.000	(0.001)	0.002	(0.011)
Stroke	3,035	0.005	-0.001*	(0.000)	0.003	(0.002)
Has Diabetes	3,032	0.013	-0.000	(0.001)	-0.007	(0.003)
Any Restricted Days in Past 2 Wks	3,069	0.080	-0.002	(0.001)	0.013	(0.013)
Any Work Loss Days in Past 2 Wks	3,069	0.047	-0.001	(0.001)	0.015	(0.015)
Any Bed Days in Past 2 Wks	3,069	0.042	-0.000	(0.002)	0.010	(0.009)
A Lot of Stress in Past 2 Weeks	3,018	0.227	0.015**	(0.004)	-0.007	(0.019)
Ever Smoked 100 Cigarettes	3,030	0.672	-0.043***	(0.002)	0.076***	(0.011)
Smoke Now	3,006	0.376	-0.048***	(0.002)	0.063**	(0.012)
Any days in Past Year 5+ Drinks	2,977	0.441	-0.010	(0.010)	0.070**	(0.015)
25+ days in Past Year 5+ Drinks	2,977	0.140	-0.020**	(0.005)	0.005	(0.009)
Any Drunk Driving in Past Year	2,996	0.209	-0.000	(0.003)	0.015	(0.024)
Exercise or play sports regularly	3,036	0.450	0.057***	(0.005)	0.010	(0.015)

Notes: The sample is restricted to white, male respondents born between 1942 and 1953. Each "Years of College" and "Veteran" pair are coefficients from a single regression. Specifications are estimated using microdata at the individual-level and include region-specific linear trend in birth cohort and region of residence fixed effects. Standard errors are clustered by the four Census regions of residence. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level respectively. Health insurance information is from the 1980 NHIS Health Insurance Supplement. For this sample, the average years of college is 2.0, and 41.7% are veterans. For public coverage, individuals are reported as having Medicare, or could report receiving services through Medicaid/Welfare or CHAMPUS. The remaining rows report data from the 1985 NHIS HPDP Supplement. With the exception of BMI, all dependent variables are dichotomous and estimates are from a linear probability model. For this sample, the average years of college is 2.0 and 38.1% are veterans.

son, to which they may respond, among other things, that they receive services through Medicaid, welfare, or CHAMPUS. As a result, only 2 percent of the sample report having any public insurance

coverage in the NHIS. This definition contributes to the differences between this figure and the one based on the ACS, where 17.9 percent have public coverage. Our cohort is also younger in 1980, so these men are much less likely to be disabled and qualify for Medicare than they will be at ages 55-64. Although the estimated coefficient on years of college is much larger in the ACS data, the effect size relative to the sample mean is quite similar between the two datasets. We view this as further evidence that access to health insurance is another potential mechanism through which college attainment could improve health and reduce mortality. In the NHIS data, we do not see any association between veteran status and public health insurance coverage, while the effect is positive and significant in the ACS. Because of the restrictive definition of public coverage in the NHIS and the lower rates of public coverage take-up in our cohorts in 1980 as compared to the ACS years, we prefer the ACS estimates and conclude that veteran status increases access to health insurance by providing higher rates of public coverage.

D. Health Outcomes and Health Behaviors in the National Health Interview Survey

Our final set of results considers how health conditions and health behaviors are associated with college attainment and veteran status for our cohort of men. Here we use the 1985 NHIS HPDP Supplement, which allows us to measure outcomes for our particular set of cohorts as close to the 1980 baseline year as possible. We again estimate equation (5) at the individual-level, but where the dependent variable is a measure of either a health outcome or a health behavior. We report estimated coefficients on years of college and veteran status in the bottom of Table 7.[24]

We first consider several health measures that could help to explain the association between education and mortality. First, we find that college education is associated with significantly lower body mass index (BMI) and a reduction in the probability of being obese. We find that each addi-

[24] In the 1985 NHIS HPDP Supplement, years of college is coded as 1-5 years or 6+. To be most consistent with the other datasets, we have classified individuals with 6+ years of college as having 6 years of college. This results in an average number of years of college of 2.0, which is similar to that found in the other datasets.

tional year of college is associated with a 1.5 percentage point higher probability of reporting having been under a lot of stress in the past two weeks. Added stress would work against the mortality advantage of the more educated. However, it may be that higher stress levels among those with college education are due to reporting differences across education groups, rather than true underlying stress levels. We do not find any relationship between college education and the probability of missing work or having restricted activity or bed days in the past two weeks.

The bottom of Table 7 explores a series of health behaviors that Cutler and Lleras-Muney (2010) identify as important mediating factors between education and health. Consistent with prior literature, we find that in our cohorts the probability of smoking decreases with years of college, while veterans are more likely to smoke. Similarly, we see that individuals with higher levels of college education are more likely to exercise. The evidence on heavy drinking is mixed. According to Table 7, rates of heavy drinking are not significantly related to years of college in two of the three categories, but college education is associated with lower levels of heavy drinking in the third category. Thus, drinking behavior alone is unlikely to explain the increase in mortality from liver disease seen in Table 4.[25] Health behaviors are also unlikely to explain the protective effect of veteran status on mortality.

VII. Conclusion

This paper examines the causal impact of college education on early adult mortality. We exploit changes in national and state-level risk of induction into military service during the Vietnam conflict as a source of exogenous variation in college attainment for white men who reached ages 38 to 49 by 1980. Using Census data from 1980 to 2000 and Vital Statistics data from 1981 to 2007, we

[25] If more educated respondents underreport heavy drinking relative to less educated respondents, perhaps because they perceive greater stigma to that behavior, then the results in Table 7 may understate the relationship between education and heavy drinking.

show that the effect of college education on mortality as estimated by 2SLS is quite similar in magnitude to the OLS gradient. In our cohorts, these estimates indicate that increasing college attainment from the level of the state at the 25th percentile of the education distribution to that of the state at the 75th percentile would decrease deaths by about nine per 1,000 men. This is equivalent to 6.5 percent of the mean of total mortality over our period, or nearly one-fourth of a standard deviation in birth state-cohort mortality. Alternatively, if we multiply our IV coefficient by four to approximate the effect of a college degree, we find that mortality rates over the period are about 1.7 times greater for non-college graduates than for college graduates. We find large negative effects of college education on deaths from cancer and heart disease, the leading causes of mortality among older adults. Increasing years of college from levels at the 25th percentile of the cohort education distribution those at the 75th would reduce cancer deaths by about 6 per 1,000 (16.6% relative to the mean) and heart disease deaths by 4.4 per 1,000 (12.3% relative to the mean).

For the purposes of comparison, 10-year mortality rates for 50-year old men in the U.S. are 2.6 times higher for smokers than for non-smokers (Woloshin et al. 2008).[26] We find that each additional year of college is associated with a 4.8 percentage point drop in the probability of smoking. Thus, if the difference in smoking rates between college graduates and others is causal, as argued by De Walque (2007) and Grimard and Parent (2007), smoking alone could explain a large portion of the mortality differential we find. We also shed some light on other potential mechanisms for the effect of college education on mortality by examining the causal effect of college on auxiliary outcomes such as health insurance and earnings, as well as the cross-sectional relationship between college education and health behaviors such as smoking, exercise, and obesity. Using data from the Census and the American Community Survey and employing the same IV strategy used to establish the mortality results, we show that increased college education leads to both higher earnings and

[26] A study of Norwegian men found that mortality by age 70 (conditional on living to age 40) was 2.9 times higher for smokers than non-smokers (Vollset et al. 2006).

higher rates of health insurance. Using data from the National Health Interview Study, we show that college education is also associated with less smoking and more exercise for our cohorts of interest.[27] These results suggest that the impact of college education on overall mortality may partially operate through greater financial and health resources, and that the impacts on cancer (especially lung cancer) and heart disease may be partially explained by the differences in behavior.

While our estimates of the impacts of college education on health and earnings are in line with what others have found using a variety of alternative identification approaches, our estimates of the causal impacts of veteran status on both long-run mortality and wages differ from what has been reported in the literature. Conley and Heerwig (2012) find no overall impact of veteran status on longer-run mortality, although they do find a small reduction in mortality for less-educated men. Our estimates of the impact of veteran status on mortality are generally negative and statistically significant. Angrist and Chen (2011) find no long-run impact of veteran status on earnings whereas we find very modest but statistically significant positive impacts. What do these differences imply about the plausibility of our findings? In answering this question, it is important to keep in mind that the most influential studies on the impact of veteran status (including those cited) use the Vietnam draft lottery as an instrument. As such, the marginal veteran in these studies is likely different from that in our study. The lottery was designed to improve the representativeness of inductees, specifically by making them more educated compared to the population of inductees that arrived under the draft board exemptions system. It is therefore possible that the positive effects on health that we find for veterans in our study are attributable to the fact that our marginal veteran is less educated than in a lottery IV design. The Conley and Heerwig (2012) results for less educated lottery inductees are consistent with this. It is possible that the positive earnings impacts we find can also be explained by

[27] See Cutler and Lleras-Muney (2010) for similar results in the NHIS for a broader set of cohorts.

this, but there are no relevant impacts in the literature to which we can compare ours.[28]

As a result of our identification strategy and sample, our results do have some limitations. First, the treatment effect we identify is the effect of college education on mortality for those who decide to get additional education in response to draft risk. The effect of going to college for other reasons may be different. Second, our population consists of white men who were born between 1942 and 1953. Results could be different for other demographic groups (minorities, women) or for more recent cohorts. The findings in this paper indicate that there is a large return to schooling on health for changes at the top part of the education distribution. Of course, whether these findings would point towards education subsidies depends on the presence of market failures, either in the form of externalities, credit constraints, or lack of knowledge about the health benefits of college education. To the extent that this paper provides useful information about the benefits of college in reducing adult mortality, we hope that it might help address one possible source of inefficient investments in human capital.

[28] Angrist and Chen (2011) provide separate estimates of the earnings impacts of lottery-based induction for blacks but not for less skilled individuals. The estimates for blacks are positive but insignificant.

References

Albouy, Valerie, and Lequien, Laurent. 2009. "Does Compulsory Education Lower Mortality?" *Journal of Health Economics*, 28(1): 155-168.

Angrist, Joshua, Chen, Stacey, and Brigham Frandsen. 2010. "Did Vietnam Veterans Get Sicker in the 1990s? The Complicated Effects of Military Service on Self-Reported Health," *Journal of Public Economics,* 94 (11-12): 824-837.

Angrist, Joshua D., and Stacey H. Chen. 2011. "Schooling and the Vietnam-Era GI Bill: Evidence from the Draft Lottery." *American Economic Journal: Applied Economics* 3, no. 2 (2011): 96-118.

Angrist, Joshua and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, New Jersey: Princeton University Press.

Arendt, Jacob Nielsen. 2005. "Does Education Cause Better Health? A Panel Data Analysis Using School Reforms for Identification." *Economics of Education Review*, 24(2): 149-160.

Bailey, M.J. 2011. "How America Avoided the Draft: The Demographic Legacy of Vietnam." Mimeo

Bearman, Peter; Kathryn M. Neckerman; and Leslie Wright. 2011. *After Tobacco: What Would Happen if Americans Stopped Smoking?* New York: Columbia University Press.

Bitler, Marianne and Lucie Schmidt. 2012. "Birth Rates and the Vietnam Draft." *American Economic Review Papers and Proceedings,* 102(3): 566-569.

Buckles, Kasey; Andreas Hagemann; Ofer Malamud; Melinda S. Morrill; and Abigail K. Wozniak, 2013. "The Effect of College Education on Health." *NBER Working Paper #19222*.

Card, David. 1999. "The Causal Effect of Education on Earnings." *Handbook of Labor Economics*. Vol 3. Ed. Orley Ashenfelter and David Card. Elsevier Science B.V.

Card, David, and Lemieux, Thomas. 2000. "Going to College to Avoid the Draft: The Unintended Legacy of the Vietnam War." *Manuscript*. Vancouver, Canada: University of British Columbia (2000).

Card, David, and Lemieux, Thomas. 2001. "Going to College to Avoid the Draft: The Unintended Legacy of the Vietnam War," *American Economic Review*, 91(2001): 97-102.

Center for Disease Control. 1987. "Postservice Mortality Among Vietnam Veterans." *Journal of the American Medical Association*, 257: 790-795.

Center for Disease Control. 1997. "Update: Trends in AIDS Incidence -- United States, 1996," Morbidity and Mortality Weekly Report, 46(37): 861-867.

Chaloupka, Frank, and Kenneth Warner. 2000. "The Economics of Smoking." In Culyer and Newhouse (Ed.), *Handbook of Health Economics*, Vol. 1, Part B, Elsevier.

Clark, Damon and Heather Royer. 2010. "The Effect of Education on Adult Health and Mortality: Evidence from Britain" NBER Working Paper No. 16013.

Conley, Dalton and Jennifer Heerwig. 2012. "The Long-Term Effects of Military Conscription on Mortality: Estimates from the Vietnam-Era Draft Lottery." *Demography,* 49: 841-855.

Cragg, John G. and Stephen G. Donald. 1993. "Testing Identifiability and Specification in Instrumental Variable Models." *Econometric Theory*, 9: 222-240.

Cutler, David M. and Adriana Lleras-Muney. 2006. "Education and Health: Evaluating Theories and Evidence." *NBER Working Paper #12352*.

Cutler, David M. and Adriana Lleras-Muney. 2010. "Understanding Differences in Health Behaviors by Education." *Journal of Health Economics*, 29(1): 1-28.

Davis, Jr., James W., and Dolbeare, Kenneth M. *Little Groups of Neighbors: The Selective Service System*. Chicago: Markham Publishing Company (1968).

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(5): 877-895.

Dobkin, Carlos and Shabani, Reza. 2007. "The Health Effects of Military Service: Evidence from the Vietnam Draft" *Economic Inquiry*, 47(1): 69–80.

Finkelstein, Amy, and Robin McKnight. 2008. "What did Medicare do? The initial impact of Medicare on mortality and out of pocked medical spending," *Journal of Public Economics*, 92(7): 1644-1668.

Grimard, F. and D. Parent. 2007. "Education and Smoking: Were Vietnam War Draft Avoiders Also More Likely to Avoid Smoking?" *Journal of Health Economics*, 26(5): 896-926.

Grossman, Michael, 2006. "Education and Nonmarket Outcomes," *Handbook of the Economics of Education*, Elsevier.

Hearst, Norman, Thomas Newman and Stephen Hulley. 1986. "Delayed Effects of the Military Draft on Mortality." *New England Journal of Medicine,* 314: 620-624.

Hipple, Steven, and Jay Stewart. 1996. "Earnings and Benefits of Workers in Alternative Work Arrangements," *Monthly Labor Review,* October 1996, 119(10): 46-54.

Honoré, Bo and Adriana Lleras-Muney. 2006. "Bounds in Competing Risks Models and the War on Cancer." *Econometrica*, 74(6): 1675-1698.

Jemal, Ahmedin; Elizabeth Ward; Robert N. Anderson; Taylor Murray; and Michael J. Thun. 2008. "Widening of Socioeconomic Inequalitys in U.S. Death Rates, 1993-2201." *PLoS ONE*. 3(50: e2181, 3, 1-8.

Kahn, Lisa. 2010. "The Long-Term Labor Market Consequences of Graduating from College in a Bad Economy." *Labour Economics* 17(2).

Kutinova, A. 2009. "Paternity Deferments and the Timing of Births: U.S. Natality During the Vietnam War." *Economic Inquiry* 47(2): 351-65

Kuziemko, Ilyana. 2010. "Did the Vietnam Draft Increase Human Capital Dispersion? Draft-

Avoidance Behavior by Race and Class." *Manuscript*. Columbia University.

Lambert, Eric, Lois Ventura, Daniel Hall, and Terry Cluse-Tolar. 2006. "College Students' Views on Gay and Lesbian Issues: Does Education Make a Difference?" *Journal of Homosexuality*, 50(4): 1-30.

LaRosa, John, Jiang He, and Suma Vupputuri. 1999. "Effect of Statins on Risk of Coronary Disease A Meta-analysis of Randomized Controlled Trials" JAMA 282(24): 2340-2346

Lleras-Muney, Adriana. 2005. "The Relationship between Education and Adult Mortality in the United States." *Review of Economic Studies*, 72, 189-221.

MacInnis, Bo. 2006. "The Long-Term Effects of College Education on Morbidities: New Evidence from the Pre-Lottery Vietnam Draft," *mimeo*, University of Michigan.

Mazumder, Bhashkar. 2012. "The Effects of Education on Health and Mortality" *Nordic Economic Policy Review*, 2012, 261-301

Malamud, Ofer and Wozniak, Abigail. 2012. "The Impact of College on Migration: Evidence from the Vietnam Generation." *Journal of Human Resources,* 47(4): 913-950.

Meara, Ellen R.; Seth Richards; and David M. Cutler. 2008. "The Gap Gets Bigger: Changes in Mortality and Life Expectancy, by Education, 1981-2000." *Health Affairs*. 27:350-60.

Meghir, Costas; Marten Palme; and Emilia Simeonova. 2012. "Education, Health, and Mortality: Evidence from a Social Experiment." *IZA Discussion Paper #6462*.

Miller, Sarah. 2012. "The Impact of the Massachusetts Health Care Reform on Health Care Use among Children." *American Economic Review Papers and Proceedings,* 102(3): 502-07.

Mokdad, A.H., Marks, J.S., Stroup, D.F, Gerberding, J.L., 2004. "Actual causes of death in the United States 2000." *Journal of the American Medical Association*, 291(10): 1238-1245.

Montez, Jennifer Karas; Robert A. Hummer; and Mark D. Hayward. 2012. "Educational Attainment and Adult Mortality in the United States: A Systematic Analysis of Functional Form." *Demography*. 49(1): 315-336.

Oreopoulos, Philip, and Uros Petronijevic. 2013. "Making College Worth It: A Review of the Returns to Higher Education." *The Future of Children*, 23(1): 41-65.

Oreopoulos, Philip; Till von Wachter, and Andrew Heisz. 2012. "Short- and Long-Term Career Effects of Graduating in a Recession." *American Economic Journal: Applied*. 4(1): 1-29.

President's Task Force on Manpower Conservation. 1964. *One Third of a Nation: A Report on Young Men Found Unqualified for Service.* U.S. Department of Health, Education and Welfare, Office of Education.

Rothenberg, Leslie S. *The Draft and You: A Handbook on Selective Service*. New York: Doubleday & Company, Inc. Anchor Books (1968).

Ruggles, Steven; and Matthew Sobek, Trent Alexander, Catherine A. Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, and Chad Ronnander. *Integrated Public Use Microdata Series: Version 3.0* [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor], 2004. URL: www.ipums.org.

Shilts, Randy. 1987. *And the Band Played On*. New York: St. Martin's Press.

Smith, Tom W. 1991. "Adult Sexual Behavior in 1989: Number of Partners, Frequency of Inter course and Risk of AIDS," *Family Planning Perspectives*, 23(3): 102-107.

Sullivan, Daniel and Till von Wachter. 2009. "Job Displacement and Mortality: An Analysis Using Administrative Data." *Quarterly Journal of Economics.* 124(3): 1265-1306.

Tatum, Arlo and Tuchinsky, Joseph S. *Guide to the Draft*. 2nd Edition. Boston: Beacon Press (1969).

Turner, Sarah. 2004. "Going to College and Finishing College: Explaining Different Educational Outcomes." In *College Decisions: How Students Actually Make Them and How They Could*, ed. Caroline Hoxby. University of Chicago Press for NBER.

United States National Advisory Commission on Selective Service. 1967. "In Pursuit of Equity: Who Serves When Not All Serve?" Report. Washington D.C. : U.S. Government Printing Office.

U.S. Selective Service. "Semiannual Report of the Director of Selective Service." Washington, DC: U.S. Government Printing Office. Various issues from 1967 to 1973.

U.S. Selective Service. 2008. "Effects of Marriage and Fatherhood on Draft Eligibility." *Fast Facts.* August 6, 2008.

Vollset, S.E., A. Tverdal, H.K. Giessing. 2006. "Smoking and Deaths between 40 and 70 Years of Age in Women and Men." *Annals of Internal Medicine,* 144(6): 381-389.

Woloshin, Steven, Lisa Schwartz and H. Gilbert Welch. 2008. "The Risk of Death by Age, Sex, and Smoking Status in the United States: Putting Health Risks in Context." *Journal of the National Cancer Institute*, 100(12): 845-853.

Wooldridge, Jeffrey M., 2002. *Econometric Analysis of Cross-section and Panel Data.* MIT Press: Cambridge, Massachusetts.

Wozniak, Abigail. (2010) "Are College Graduates More Responsive to Distant Labor Market Opportunities?" *Journal of Human Resources.*

Xu, J.Q., Kochanek K.D., Murphy S.L., Tejada-Vera B. (2010). Deaths: Final Data for 2007. National Vital Statistics Reports, 58(19). National Center for Health Statistics: Hyattsville, MD.

Yoon, Young-Hee, and Hsiao-ye Yi. 2010. "Surveillance Report #88: Liver Cirrhosis Mortality in the United States, 1970–2007." Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism, Division of Epidemiology and Prevention Research.

Zimmerman, Seth. 2014. "The Returns to College Admissions for Academically Marginal Students." *Journal of Labor Economics* 32:711–54.

Appendix Table 1: OLS Estimates for the Impact of College Education on Cumulative Mortality, 1980-2007, for Early and Late Cohorts

	Baseline OLS [1]	No Controls [2]	Early Cohorts [3]	Early-NC [4]	Late Cohorts [5]	Late-NC [6]
Years of College	-22.31***	-23.58***	-19.10	-25.32*	-7.49	-8.89
	(3.80)	(3.61)	(12.46)	(13.49)	(5.50)	(5.43)
Veteran Status	-42.82***	-47.58***	-66.09	-90.83	-49.95	-71.17**
	(5.38)	(5.32)	(63.70)	(66.93)	(31.92)	(31.30)
Controls?	Y	N	Y	N	Y	N
Birth Cohorts	42-53	42-53	42-44	42-44	51-53	51-53
Observations	600	600	150	150	150	150
R-squared	0.952	0.9493	0.9213	0.9107	0.9572	0.9565

Notes: Standard errors are in parenthesis and are clustered by birth state. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level respectively. The dependent variable is the mortality rate from 1981-2007 per 1,000 persons. *Years of College* and *Veteran* measure the average years of higher education and fraction of veterans respectively. "Early cohorts" include men born 1942-1944, while "late" cohorts include men born 1951-1953. Regressions are weighted by cell size. Additional controls include birth state fixed effects, the employment-to-population ratio, the cohort size (derived from the 1960 Census), and birth region trends. See Table 1 notes for data sources and sample restrictions.

Appendix Table 2: OLS and IV Estimates for the Impact of College Education on Mortality, Additional Causes of Death

	Lung Cancer	Colon Cancer	HIV
Panel A: OLS			
Years College	-4.04***	-0.74***	-0.77*
	(0.84)	(0.19)	(0.40)
Veteran Status	-6.97***	-1.00***	0.08
	(1.02)	(0.29)	(0.39)
Panel B: IV			
Years College	-6.87***	-0.88***	1.35*
	(1.09)	(0.33)	(0.75)
Veteran Status	-3.33**	-0.77*	-2.22**
	(1.45)	(0.43)	(1.06)

Notes: Standard errors are in parenthesis and are clustered by birth state. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level respectively. The dependent variable is the mortality rate by cause from 1981-2007 per 1,000 persons. The IV specifications use a third-order polynomial in the national and state-level induction risks to instrument for average years of higher education (*Years College*) and fraction of veterans (*Veteran*) using a 2SLS model. Number of observations is 600 in each regression and regressions are weighted by cell size. Additional controls include birth state fixed effects, the employment-to-population ratio, the cohort size (derived from the 1960 Census), and birth region trends. See Table 1 notes for data sources and sample restrictions.

Appendix Table 3: OLS and IV Estimates for the Impact of College Education on Wages and Earnings, 1980

		IV Specification:				
	OLS	Linear-2SLS	Quadratic-2SLS	Quadratic-LIML	Cubic-2SLS	Cubic-LIML
	[1]	[2]	[3]	[4]	[5]	[6]
Panel A: Log earnings						
Years of College	0.101***	-0.0277	0.0899***	0.0884**	0.133***	0.137***
	-0.0141	-0.0627	-0.0329	-0.0364	-0.0271	-0.0299
Veteran	0.119***	0.339***	0.139**	0.141**	0.0618	0.055
	-0.0198	-0.114	-0.0584	-0.0644	-0.0429	-0.0474
Mean log earnings	9.543	9.543	9.543	9.543	9.543	9.543
Panel B: Log wages						
Years of College	0.0533***	-0.00834	0.0137	0.0129	0.0522***	0.0521**
	-0.0112	-0.0483	-0.0291	-0.0297	-0.0195	-0.0205
Veteran	0.0895***	0.190**	0.152***	0.154***	0.0847**	0.0847**
	-0.0156	-0.0842	-0.0523	-0.0532	-0.0358	-0.0374
Mean log wages	1.979	1.979	1.979	1.979	1.979	1.979

Notes: Standard errors are in parenthesis and are clustered by birth state. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level respectively. Data are birth state-cohort cell averages from white men in the 1980 Census IPUMS extracts, born between 1942 and 1953. Dependent variable is mean log real earnings or log real wages. The IV specifications use the national and state-level induction risk to instrument for the for average years of higher education (*Years of College*) and the fraction of veterans (*Veteran*) Underlying microdata sample is therefore restricted to those with valid earnings in 1980. Earnings of top-coded observations are multiplied by 1.5. Hourly wages are calculated as total wage and salary income divided by hours of labor supply. Wage calculations exclude those with missing hours or weeks worked. We truncate the bottom 1 percent of hourly earners and those above 1.5 times the maximum annual income amount divided by 1,750 (35 hours per week for 50 hours per year). Specifications are otherwise identical to those in Table 3. Number of observations is 600 in each regression and regressions are weighted by cell size.

Appendix A: Identification in a Structural Model with Direct and Indirect Effects

We begin by presenting a simple structural model that relates veteran status and college attainment to health through both direct and indirect channels. The parameters of interest for our analysis of a health outcome *Y* are the coefficients on the years of college variable *C* and the veteran status indicator *V* in the first equation of the structural equation system[29]

(A1) $$Y = \alpha_1 C + \alpha_2 V + X'\alpha_4 + \varepsilon$$

(A2) $$C = \gamma_2 V + Z'\gamma_3 + X'\gamma_4 + \nu$$

(A3) $$V = Z'\beta_3 + X'\beta_4 + \eta.$$

Suppose for simplicity that the vector of instruments *Z* has two entries Z_1 and Z_2. The *p*-dimensional vector *X* contains additional covariates, including a constant. Both *Z* and *X* are exogenous in the sense that the unobserved error components ε, ν, and η satisfy E(ε | *Z*,*X*)=E(ε), E(ν | *Z*,*X*)=E(ν), and E(η | *Z*,*X*)=E(η). Since we have constants in the system, we can without loss of generality take E(ε)=E(ν)=E(η)=0. The variables *C* and *V* are assumed to be endogenous in (1) so that Cov(ε, *C*) ≠ 0 and Cov(ε, *V*) ≠ 0.

For our purposes, identification of the structural parameter of interest $\alpha_0 := (\alpha_1, \alpha_2)'$ is best thought of in terms of reduced-form restrictions. We can write the reduced form of the structural system (A1)-(A3) as

(A4) $$(Y \quad C \quad V) = (Z' \quad X') \begin{pmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{pmatrix} + (v_1 \quad v_2 \quad v_3).$$

Here (v_1, v_2, v_3) is a row vector of reduced-form errors and $(\Pi_{ij})_{ij=1,2}$ is the matrix of reduced-form parameters. The dimensions of Π_{11} and Π_{12} are 2×1 and 2×2, respectively; Π_{21} and Π_{22} are *p*×1 and

[29] Our model assumes that education only has a direct effect on health at the state-cohort level (that is, there is no indirect effect through veteran status). We believe this reflects the historical record. Certainly, college students could delay or avoid conscription by staying in school, but induction rates for a birth state-birth cohort were unlikely to be affected by individual draft-avoidance behavior. This is because local draft boards needed to fulfill specific manpower requirements set by the Department of Defense. This assumption is also supported by evidence (available upon request) showing that veteran status is not predicted by graduation rates at the birth-state birth-cohort level in our data. Ultimately this assumption simplifies our econometric analysis, but it is not required for identification.

$p \times 2$. The first stage for (1) (i.e., the reduced form of structural equations (A2) and (A3)) is therefore

$$(C \quad V) = (Z' \quad X')\begin{pmatrix}\Pi_{12}\\ \Pi_{22}\end{pmatrix} + (v_2 \quad v_3).$$

Plug the preceding display into (A1) to see

$$Y = (Z' \quad X')\begin{pmatrix}\Pi_{12}\alpha_0\\ \Pi_{22}\alpha_0 + \alpha_4\end{pmatrix} + v_1.$$

We conclude from (A4) that $\Pi_{12}\alpha_0 = \Pi_{11}$, and therefore $\alpha_0 = \Pi_{12}^{-1}\Pi_{11}$. The sample equivalent of this is just the IV estimator of α_0. Since the reduced-form parameters are identified as long as $\mathrm{E}(Z', X')'(Z', X')$ is invertible, it follows that α_0 is identified if and only if

(5) $$\mathrm{rank}(\Pi_{12}) = 2,$$

which requires the correlation structure of Z and (C,V) to be rich enough to disentangle the effects coming from C and V.

We now briefly outline a test of this rank condition based on the Cragg and Donald (1993) test for identifiability. In the standard case with a single endogenous variable, this test is identical to the usual first-stage F test. In our two-variable setting, the standard first-stage F test and the Cragg-Donald F test complement one another. We can use standard F statistics to separately test, in each first stage equation, the null hypothesis of no correlation between the instruments and the endogenous variables against the alternative of correlation. Rejecting the null in each case provides evidence that there is correlation, but does not necessarily imply that the correlation structure is rich enough to identify our structural parameter α_0. The presence of correlation with insufficient structure is precisely the null hypothesis of the Cragg-Donald F test and can be expressed as

(A6) $$H_0\colon \mathrm{rank}(\Pi_{12}) = 1.$$

The alternative is identification of α_0 in the sense of equation (A5). Hence, if we reject the null in both the first-stage F tests and the Cragg-Donald F test, we have evidence that the structural parameters are identified. The Cragg-Donald test can be interpreted like F statistics with $2(k-1)$ numer-

ator degrees of freedom, where $k \geq 2$ is the number of instruments.[30] The null (A6) and alternative (A5) do not change for different values of k.

Another test that is sometimes proposed in settings with multiple endogenous variables is the adjusted F statistic developed in Angrist and Pischke (2009). It tests, separately for each first stage model, whether the correlation between the instruments and a given endogenous variable is weak while accounting for the fact that some of the variation in the instruments is used in the remaining first stages. Although they do not represent a direct test of our structural model and identifying assumptions, the Angrist-Pischke F statistics are potentially of interest as a diagnostic test of weak instruments. Thus, we report Angrist-Pischke F statistics as well.

[30]We compute the minimum distance version of the Cragg-Donald statistic because, as opposed to the minimum eigenvalue version routinely reported in statistical software, it remains valid when cluster-robust covariance matrices are used; see Buckles et al. (2013) for details on the minimum distance statistic in our context.

Appendix B: Interpretation of the Structural Parameters

We now show that $\alpha_0 = (\alpha_1, \alpha_2)'$ has a causal interpretation by using the instruments Z_1 and Z_2 to mimic interventions that change *C* while holding *V* fixed (or vice versa) without affecting the error term ε. To avoid obscuring the problem with unnecessary notation, we do not explicitly include the covariates in the discussion, although all of the results below remain valid conditional on *X*. To interpret α_1, it follows from (A5) in Appendix A that we can find values (z_{11}, z_{21}) and (z_{12}, z_{22}) such that

(A7) $$\mathrm{E}(C \mid Z_1 = z_{11}, Z_2 = z_{21}) - \mathrm{E}(C \mid Z_1 = z_{12}, Z_2 = z_{22}) \neq 0$$ and

(A8) $$\mathrm{E}(V \mid Z_1 = z_{11}, Z_2 = z_{21}) - \mathrm{E}(V \mid Z_1 = z_{12}, Z_2 = z_{22}) = 0,$$

i.e., switching from (z_{11}, z_{21}) to (z_{12}, z_{22}) changes college attainment without changing the probability of veteran status. From these two relations, equation (A1), and the exogeneity assumption, we can conclude

$$\alpha_1 = \frac{\mathrm{E}(Y \mid Z_1 = z_{11}, Z_2 = z_{21}) - \mathrm{E}(Y \mid Z_1 = z_{12}, Z_2 = z_{22})}{\mathrm{E}(C \mid Z_1 = z_{11}, Z_2 = z_{21}) - \mathrm{E}(C \mid Z_1 = z_{12}, Z_2 = z_{22})}.$$

The parameter α_1 therefore measures the average change in the outcome variable for individuals where an intervention changed the average level of college attainment without changing the probability of veteran status, adjusted for the fact that a change in the average college attainment does not necessarily induce an effect on the college attainment of an individual. A similar argument can be made for α_2 with different values for (Z_1, Z_2) and reversed equality signs in (A7) and (A8).

It can also be seen that the parameters $\alpha_1$and α_2 are "local average treatment effects" from two different interventions. Within the constant effects framework, α_1 measures the causal effect of an intervention described by (A7) and (A8) on individuals that comply with the intervention. Intuitively, as long as condition (A5) holds, our instruments vary enough to mimic a situation in which variation in college access and variation in veteran status come from two independent randomized

experiments. This breaks the dependency between outcomes in equations (A1)-(A3). It may be that some individuals would be "randomly selected" to receive both interventions, but their receipt of both interventions is unrelated to the relationships in the structural model. This interpretation crucially depends on our parametric assumptions; we do not claim to *non-parametrically* identify local average treatment or any other (direct, indirect, or total) causal effect.