The Effect of College Education on Mortality

Kasey Buckles, University of Notre Dame and NBER

Andreas Hagemann, University of Notre Dame

Ofer Malamud, University of Chicago and NBER

Melinda Morrill, North Carolina State University

Abigail Wozniak, University of Notre Dame, NBER and IZA

May 2014

ABSTRACT

We exploit exogenous variation in college completion induced by draft-avoidance behavior during the Vietnam War to examine the impact of college completion on adult mortality. Our preferred estimates imply that increasing college completion rates from the level of the state with the lowest induced rate to the highest would decrease cumulative mortality by 28 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)

Acknowledgments: The authors would like to thank William Evans, Bob Kaestner 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, University of Maryland-Baltimore County, 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 collegiate 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, and Cutler and Lleras-Muney, 2010). To our knowledge, this paper is the first to provide a causal estimate of the effect of college completion on mortality. In doing so, we contribute to knowledge about the impacts of education on health for a new part of the schooling distribution, where the observed health gradient in education is steeper than at lower parts of the schooling distribution examined in the prior literature.

We use variation in college attainment induced by draft-avoidance behavior during the Vietnam War with an instrumental variables strategy, as in Malamud and Wozniak (2012). 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. 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 cohort 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 endoge-

nous 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). This strategy was developed by Malamud and Wozniak (2012); here we provide a more detailed explanation of how identification works for a setting in which there are two linked endogenous variables. We also provide a test for identification grounded in the assumptions of our structural model.

We merge our data on national and state-level induction risk with the Vital Statistics Mortality Files from 1981 to 2007 and the U.S. Censuses for 1980, 1990, and 2000 to construct birth state-by-cohort level mortality rates. We also include information about basic demographic characteristics, cohort size, and labor market conditions at the time of entry from a variety of additional sources. Our instruments predict both veteran status and educational attainment for men in the affected cohorts, with the increase in education coming primarily from increased post-secondary schooling attainment. We therefore have a viable IV strategy for educational attainment at higher levels that can be purged of its correlation with veteran status for the Vietnam cohorts.

We focus on the cumulative mortality rate between 1981 and 2007 and perform our main analysis using data aggregated to the birth state-birth year cohort level. 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. Specifically, OLS shows that a birth state-birth year cohort with 100 percent college completion is associated with 102 fewer deaths by 2007 per 1,000 persons compared to a cohort with no one completing college. Given that the mortality rate for non-college graduates over this period is 171.5 per 1,000 persons, this represents a decrease of almost 60 percent.

Our instrumental variables estimates indicate an effect that ranges from 93 to 172 fewer deaths per 1,000 persons, with our preferred specification yielding a magnitude similar to the OLS estimate.

For the birth state-birth year cohorts in our sample, this translates into a reduction in mortality of 42.8 per 1,000 for states with the highest induced completion rates versus the lowest. Furthermore, none of these 2SLS estimates are significantly different from the OLS estimates. Results using 10-year mortality rates for the 1980s, 1990s and 2000s show that the effects of college completion on health are not confined to any particular decade. The largest effects are found for the impact of college completion in lowering deaths due to cancer and heart disease, which represent the leading causes of mortality in our sample of older adults. College completion decreases the cancer mortality rate by 72 percent, with lung cancer accounting for over half of the reduction.

Before concluding, we use the U.S. Census, the American Community Survey (ACS) and the National 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 completion on auxiliary outcomes such as health insurance and wages, as well as the cross-sectional relationship between college completion and various 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. 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

[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, but in both studies the estimated effects have large standard errors.

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. 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).

The findings in this paper have important implications for both health and education policy. 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 completion, this represents a particularly 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 provides an even stronger 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 may 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.[2] Our paper extends this identification strategy by incorporating within-cohort variation in induction risk to account for veteran status. Moreover, we examine a wide range of both health outcomes and behaviors across multiple decades, providing a much broader picture of higher education's potential health impacts.

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).[3] 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); hereafter 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

[2] 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.
[3] 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.

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 Selective Service System, 1967-1973).[4] 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 were granted in 1967 alone.[5] 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 is well-known (Card and Lemieux, 2000) and has been used in previous research (e.g., De Walque, 2007; Grimard and Parent, 2007). Inductions varied considerably over the course of the Vietnam War. 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

[4] 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. See Rothenberg (1968).
[5] The number of college deferments remained above 1.7 million in 1968 and 1969, and then fell to 1.5 million and 1.3 million in 1970 and 1971 respectively (Semi-annual Reports of the Director of the Selective Service System, 1967-1973).

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.

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 state 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 completion and veteran status.[6]

This strategy is identical to the one described in MW. However, we extend the MW analysis in two ways. First, 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. Second, we provide a detailed discussion of the interpretation of our estimates when both direct and indirect effects are possible.

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 completion 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 college completion indicator C and the veteran status indicator V in the first equation of the structural equation system[7]

[6] 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.
[7] 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

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

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

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

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($\varepsilon \mid Z,X$)=E(ε), E($\nu \mid Z,X$)=E(ν), and E($\eta \mid Z,X$)=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) $\neq$ 0 and Cov(ε, V) $\neq$ 0.

For our purposes, identification of $\alpha_0 := (\alpha_1, \alpha_2)'$ is best thought of in terms of reduced-form restrictions. We can write the reduced form of the structural system (1)-(3) as

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

Here Ψ 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 p×2. The first stage for (1) (i.e., the reduced form of structural equations (2) and (3)) is therefore

$$(C \quad V) = (Z' \quad X')\begin{pmatrix} \Pi_{12} \\ \Pi_{22} \end{pmatrix} + \Psi_2,$$

where Ψ_2 consists of the second and third entry of Ψ. Denote the first entry by Ψ_1 and plug the preceding display into (1) to see

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.

$$Y = (Z' \quad X')\begin{pmatrix} \Pi_{12}\alpha_0 \\ \Pi_{22}\alpha_0 + \alpha_4 \end{pmatrix} + \Psi_1.$$

We conclude from (4) 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 E[(*Z*',*X*')'(*Z*',*X*')] is invertible, it follows that α_0 is identified if and only if

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

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

(6) $$H_0\text{: rank}(\Pi_{12}) = 1.$$

The alternative is identification of α_0 in the sense of equation (5). 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 $2(k-1)$ numerator

degrees of freedom, where $k \geq 2$ is the number of instruments.[8] The null (6) and alternative (5) 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 so we report them as well.

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 otherwise impossible 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 (5) that we can find values (z_{11}, z_{21}) and (z_{12}, z_{22}) such that

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

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

[8]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.

i.e., switching from (z_{11}, z_{21}) to (z_{12}, z_{22}) changes the probability of college education without changing the probability of veteran status. From these two relations, equation (1), 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 probability of college education without changing the probability of veteran status, adjusted for the fact that a change in the probability does not necessarily induce an effect in the outcome. A similar argument can be made for α_2 with different values for (Z_1, Z_2) and reversed equality signs in (7) and (8).

We can also show that α_1 and α_2 measure the direct causal effect of a change in *C* (respectively *V*) status holding *V* (respectively *C*) status constant in a potential outcomes framework. Write the observed outcome *Y* as a function of the potential outcomes Y_{11}, Y_{01}, Y_{10}, or Y_{00} depending on realizations (*C*,*V*)=(1,1), (0,1), (1,0), or (0,0),

(9) $$Y = Y_{11}CV + Y_{01}(1 - C)V + Y_{10}C(1 - V) + Y_{00}(1 - C)(1 - V)$$

$$= \big(Y_{11} - Y_{01} - (Y_{10} - Y_{00})\big)CV + (Y_{10} - Y_{00})C + (Y_{01} - Y_{00})V + Y_{00}.$$

Under linear regression (constant effects) assumptions, we have $\alpha_1 = Y_{11} - Y_{01} = Y_{10} - Y_{00}$ and $\alpha_2 = Y_{11} - Y_{10} = Y_{01} - Y_{00}$. In particular, the marginal effect of a college education is the same for veterans and non-veterans and the marginal effect of veteran status is the same for college graduates and non-college graduates. Hence, the first term in the second line of the preceding display drops out and, if we take $\mathrm{E}(Y_{00})$ as the intercept and $\varepsilon = Y_{00} - \mathrm{E}(Y_{00})$, we are back in the framework of the structural system in equation (1). As such, α_1 is the "average treatment effect" because $\alpha_1 =$

$\mathrm{E}(Y_{11} - Y_{01}) = Y_{11} - Y_{01} = Y_{10} - Y_{00} = \mathrm{E}(Y_{10} - Y_{00})$. The same argument, with the roles of *C* and *V* reversed, yields the equivalent result for α_2.

Finally, it can also be seen that the parameters α_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 (7) and (8) on individuals that comply with the intervention. Intuitively, as long as condition (5) 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.[9] This breaks the dependency between outcomes in equations (1)-(3). It may be that some individuals would be "randomly selected" to receive both interventions, but their receipt of both interventions is random and unrelated to the relationships in the structural model. However, we emphasize that this interpretation crucially depends on our parametric assumptions. For a general nonparametric potential outcomes model, we cannot simplify (9). We therefore we do not claim to *non-parametrically* identify local average treatment or any other (direct, indirect, or total) causal effect.

C. Instruments for College Education and Veteran Status

Finding instrumental variables that meet the conditions on *Z* discussed in the previous section is obviously critical to our analysis. Fortunately, we are able to employ the same strategy as MW, who in turn 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

[9] 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 α_2, 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.

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

$$(10) \qquad staterisk_{sc} = \frac{\left(\sum_{t|c=19\,to\,22} I_{sct}\right)/4}{N_{sc}}$$

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:

$$(11) \qquad nationalrisk_{sc} = \frac{\left(\sum_{t|c=19\,to\,22} \sum_{-s} I_{-sct}\right)/4}{\sum_{-s} N_{-sc}}$$

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.[10] In other words, the numerator and denominator in (11) 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 (10) and (11), 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

[10] Note that MW use variation at the six month birth cohort level. Here we use yearly variation because month of birth is not available in the public use mortality data.

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.

D. 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 very standard and did not involve any invasive testing for health conditions not easily observed in a doctor's office. It is therefore unlikely these exams provided young men with information they did not already have. [11]

[11] 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

Second, young men may have viewed marriage or fertility choices as a means to reduce their chances of being drafted, and these choices may have connections to health. Marriage was not in itself grounds for deferment after 1965 (Selective Service 2008).[12] Fatherhood, on the other hand, was grounds for deferment and was the most common draft exemption. However, there is little evidence that male fertility among affected cohorts responded to this. Bitler and Schmidt (2012) examine fertility changes among women who were likely to have been affected by the absence of men in affected cohorts. Their findings indicate that men in our affected cohorts were modestly but statistically significantly less likely to become fathers during the Vietnam years. Whether there is an effect on completed fertility is unknown. Nevertheless, we know of no evidence linking fertility timing to later adult health for men. Since the known effects (for women) are small, and since there is no known channel to link any potential effects for men to health, we believe that the exclusion restriction is valid in our context.[13]

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 it 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 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

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).

[12] The availability of this 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.

[13] Also, fertility effects were largest among blacks and low skilled whites (Bitler and Schmidt, 2012). Since we omit blacks from our sample, this further reduces concerns about fertility changes in young adulthood driving later health for men.

the US male population over the period 1964-1970 (President's Taskforce on Manpower Conservation, 1964). There is no appreciable non-linearity in the forecasts, 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:

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

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 rates of college completion 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 induction 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 if surging applications from marginal white men crowded out

black college applicants, as discussed in Kuziemko (2010). This would violate the monotonicity assumption of our IV strategy—that the instrument should have a uniform direction of impact on affected individuals.

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 primary measure of educational attainment is a dummy variable for college completion. To inform this choice, we explored the impact of our instruments on educational attainment and confirmed that higher national and state-level induction risk increased male educational attainment at all post-secondary levels, including completion. We also replicate our main results using years of college as our measure of educational attainment, and our findings are robust to this change.[14]

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.[15] 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.[16]

---

[14] College completion has been shown to be the most accurately measured higher education outcome and to contain little measurement error overall (Black, et al., 2003). On the other hand, years of post-secondary schooling is unlikely to be plagued by error that is negatively correlated with the recorded value, as is the case with the dummy variable schooling measures in general (Kane, Rouse, and Staiger, 1999, and Black, et al., 2000).

[15] 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. For 1990, we define Reservists or National Guard members as non-veterans to match the veteran definition in the 1980 and 2000 Censuses. We also omit those living in group quarters.

[16] 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.

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 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:

$$MR_{sc} = a_1 C_{sc} + a_2 V_{sc} + \alpha_4 X_{sc} + \tau trend_{sc} + \delta_s + \varepsilon_{sc} \tag{13}$$

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 (12)—either in total or by cause. The variable V_{sc} is the fraction of veterans, and C_{sc} is the fraction with a college degree. 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.[17]

Our endogenous variables, *C* and *V*, are predicted from first stage equations that include the

[17] 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 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.

remaining right hand side covariates in equation (13) plus functions of *staterisk* and *nationalrisk* as defined in (10) and (11), 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 linearl, 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.[18] We follow Bertrand et al. (2004) and cluster standard errors at the birth state level after collapsing our data to birth state-cohort cells.[19]

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 (12) 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 (35.7/1,000). 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 their forties cancer and heart disease are the leading causes.

V. Results

A. First Stage Results

[18] 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.

[19] 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 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 each specification—predicting college graduation 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 two columns show results from the linear specification. Conditional on state cohort risk, a 10 percentage point increase in national cohort risk (roughly the entire range of this variable) increased the fraction of men with a college degree by 9.4 percentage points. In other words, cohorts with higher national risk were more likely to graduate college.[20] Coefficients from the quadratic and cubic specifications also indicate an overall positive relationship between higher national risk and college graduation over the range of the risk variable. In contrast, there does not appear to be a strong relationship between state cohort risk and college graduation. Conditional on national induction risk, college graduation is negatively related to state cohort risk in the linear specification, insignificant in the quadratic specification, and has mixed signs in the cubic specification.[21] These patterns are displayed in Panels A and B of Figure 1 which show that the relationship between college graduation and state risk is modest compared to the impact of national risk on college completion.

These results are consistent with the fact that men were unaware of how state relative induction risks fluctuated over time. Indeed, our review of the historical literature suggests that, while young men were aware that some states had average risk levels above others, they were unaware of the year

[20] The comparable estimate from Table 1 in Card and Lemieux (2001) shows a 4.6 percentage point increase. Our estimate is larger because we restrict the sample to whites. When we do not make this restriction, our estimates are very close to those in Card and Lemieux.

[21] Note, when national risk is excluded from the first-stage regression model (not shown), the relationship between state level risk and college graduation is positive and significant.

to year changes in this relative risk that we exploit as our identifying variation. The relatively small and significant negative coefficient in the linear specification could reflect a modest impact on college going working through some other channel.[22] Alternatively, the negative relationship may also be an artifact of the high correlation (i.e. collinearity) between national and state risk.

The first stage estimates with veteran status as the dependent variable show that veteran status positively varies with both national and state cohort risk (although more strongly with national risk). These patterns are also clearly 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).

Due to space limitations, the reduced form regression results are not shown here (but available upon request). However, 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 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 graduation or on mortality whereas national risk has a large effect on both veteran status and college graduation, as well as a corresponding negative effect on mortality. Thus, these patterns might suggest that it is college graduation, more than veteran status, which is associated

[22] Recall that the coefficient on *staterisk* reflects responses to changes in *staterisk* conditional on national level risk and other controlled factors like state fixed effects. It is possible that increases in *staterisk* above what men might have expected based on national level risk and state averages meant that any high ability young men who had not gone to college already were at higher risk of being drafted. Card and Lemieux (2000) document that conditional on not enrolling in college, high ability men were much more likely than others to be drafted. If some in this group would have gone to college eventually, then the unexpected rise in *staterisk* may have disrupted their college-going plans and ultimately lowered college completion rates.

with the reduction in mortality since both instruments affect veteran status but only national induction risk affects college graduation.

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

Table 3 presents our OLS and 2SLS estimates of the effect of college graduation on cumulative mortality (equation 13). For the IV results, we show specifications in which the induction risk variables are linear, quadratic, and cubic. 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 college graduation is -101.97. This indicates that increasing college completion rates from the level of the state with the lowest rates to that with the highest (a 42 percentage point change) is associated with 42.8 fewer deaths per 1,000 men. This is consistent with the well-documented educational gradient in most health outcomes. Veteran status is also negatively associated with mortality for these cohorts, but it is less clear what we should expect for the sign on this coefficient, even in our OLS estimates. This is because we observe mortality conditional on surviving to 1980. Veteran status may increase mortality at young ages, both via combat injuries and death but also through risky behavior while young (Hearst et al., 1986), but less is known about the relationship between veteran status and health in the longer term. As noted above, the existing studies find mixed evidence.

Table 3, Column [2] shows the point estimates of the causal effect of college graduation 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 graduation 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 both college graduation and veteran status 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 completion rates from the level of the state with the lowest rates to that with the highest (a 42 percentage point change) leads to 39.5 fewer deaths per 1,000 men.

Table 3 shows the impact of high college completion rates on later mortality is robust to three 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 estimates are robust to excluding the non-linear state-cohort varying controls (employment to population ratio and log cohort size). Finally, we have reproduced our main results using years of college as the measure of educational attainment. The results for cumulative mortality are shown in the first two columns of Appendix Table 1. Instrumental variables estimates of the effect of an additional year of college on mortality are negative and statistically significant. Multiplying the coefficient by 4 yields an estimate that is comparable to, but less than, our preferred estimate of the effect of college completion from Table 3 (-69.24 vs. -93.93). This suggests that there may be non-linearities in the effect of college education on health (for example, due to sheepskin effects).

A final check is shown in Appendix Table 2. 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.[23] 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 college completion health gradient for them. This is probably because these cohorts have not yet reached the ages where mortality is due to causes for which college completion 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.

Returning to our main analysis, one might wonder whether the effect of education on mortality changes over time, since the causes of death vary considerably over the period, as Figure 2 shows. In Table 4, we show OLS and IV estimates for the effect of college completion rates on mortality by decade. In these and most remaining results, we show results for the cubic 2SLS specification only.[24] For each decade, this measure is the total number of deaths over the decade (1981-1990, 1991-2000, 2001-2007) divided by the cohort size in the beginning of the decade (taken from the 1980, 1990, and 2000 Censuses, respectively). These shorter-term mortality rates allow us to determine whether the role of education changes as the likely reasons for mortality evolve.[25]

For each decade, college graduation and veteran status are negatively associated with mortality. For the 1980s (where the first stage is strongest), IV results are again very similar to OLS. The estimated effect of increased college attainment on mortality during the 1980s is -24.38; increasing college

[23] We generally lose first stage power when we drop more than two cohorts from the analysis.
[24] We choose to show the cubic specification because the estimates are generally more precise, but results using lower order polynomial functional forms for the first stage equations are qualitatively similar. The 2SLS and LIML results for decadal mortality are similar using the cubic functional form (though more so for 1980 and 2000 than for 1990), which again suggests that the bias from using 2SLS is small.
[25] Ideally, we would construct mortality rates by age, rather than by decade (for example, rates for men in their forties). However, to construct the denominators for the mortality rates, we would need to know the number alive at certain ages, by birth state. We would also need the average education level and veteran status for the cell. These cannot be constructed from the Census since it is only conducted every 10 years, and the SEER population counts are not available by state of birth and do not have information on education and veteran status.

education from the level of the least educated birth state-cohort to the highest would decrease mortality by 10.2 per 1,000 men (or 40% relative to the mean). For the 1990s and 2000s, the IV estimates are larger than the OLS estimates and indicate that moving from the lowest-education birth state-cohort to the highest would decrease mortality by 61% and 57% relative to the mean, respectively. Results for these decades are less precise, however, which may be driven by the weaker first stage for college graduation (first-stage *F*-statistics are reported in the table). So while we interpret these results with caution, there is little evidence that the negative effect of college education on total mortality is confined to a particular decade. Veteran status has a negative and statistically significant negative effect on mortality in all periods.

As discussed earlier, one might be concerned that our measures of induction risk are correlated with the state-cohort underlying health status. As discussed in Section III.D., 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.[26] 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 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.[27]

[26] Results not shown, but available upon request, confirm that the estimated effects are not sensitive to the inclusion of birth state trends.

[27] 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 may 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. Consequently, we prefer to control directly for secular trends in male college completion and use our birth state and birth cohort variation to identify changes in male college completion driven by induction risk.

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

Table 5 shows the effect of state-cohort college graduation rates 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. All specifications are the same as in Table 3. For cancer and heart disease—the two leading causes of death—the point estimates on college graduate are negative and statistically significant for both OLS and IV and are larger for IV. Moving from the lowest-education birth state-cohort to the highest would reduce cancer deaths by 25.9 per 1,000 (71% relative to the mean) and heart disease deaths by 18.9 (42% relative to the mean). The effect of veteran status on mortality due to these causes is also negative, though not statistically significant for heart disease.

In results not shown here, we estimated the model for deaths by specific type of cancer. 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 a college degree on lung cancer is -35.59 (s.e. = 5.09), suggesting that 57% 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 Grimard 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).[28]

College education has a negative and statistically significant effect on mortality from three

[28] Results for cancers and heart disease using years of college as the measure of educational attainment are available in Appendix 1.

other causes: diabetes, stroke, and chronic low respiratory conditions (CLRCs). In the IV specifications, a college degree decreases deaths per 1,000 white men from both diabetes and stroke by 7 and from CLRCs by 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). Veteran status decreases deaths due to chronic low respiratory conditions and stroke, while its effect on diabetes is small and statistically insignificant at the 5 percent level.

The OLS results show a negative association between a college degree and death by two external causes—accidental injury and homicide. But for both of these, the IV coefficient is smaller and statistically insignificant. There is also no evidence that college decreases deaths by suicide. Veteran status, however, does decrease deaths due to accidents and suicide, conditional on surviving to 1980.

Finally, for infectious and parasitic diseases and liver diseases, the point estimate for college graduate is negative in the OLS specification but positive and statistically significant in the 2SLS specifications. For infectious and parasitic diseases, 78% of deaths for these cohorts were due to HIV alone, and HIV deaths are driving the positive coefficient. There are several ways in which college attendance might increase 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).[29] Liver diseases, on the 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

[29] 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 finds that risky behavior is significantly higher among younger and more urban individuals, who are more likely to be college-educated.

be more at risk for liver diseases.[30] We investigate the relationship between college education and alcohol consumption in the next section. The effect of veteran status on deaths by infectious diseases and liver disease is negative.

The results in Tables 3, 4, and 5 indicate that college education has a negative effect on total mortality, that the effect was not confined to any particular decade, and that deaths due to heart disease and cancer are particularly affected. A college education has a particularly strong negative effect on deaths from causes related to smoking.

VI. Potential Mechanisms

A number of potential mechanisms might explain the negative effect of college completion 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 and more likely to drink, this may explain the negative coefficient on cancer and the positive coefficient on liver disease, respectively. 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 focus on our cohort in particular and explore whether we can find evidence supporting or refuting potential mechanisms that would explain the cause-of-death 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)

[30] 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).

Health Insurance Supplement. Finally, using data from the 1985 National Health Interview Survey (NHIS), Health Promotion and Disease Prevention (HPDP) Supplement, we consider acute health outcomes, including diabetes, hypertension, and BMI, and health behaviors, including smoking, exercise, and alcohol consumption.

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 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. However, health may directly affect earnings potential, so these results must be interpreted only as suggestive evidence of a possible mechanism. We examine this potential mechanism by estimating the effect of college completion on wages and earnings for cohorts who came of age during the Vietnam War. Using data from the 1980 Census, Table 6 presents OLS and IV estimates based on our preferred cubic first-stage specification.[31]

The OLS estimates for log earnings indicate that a cohort moving from a 0 to 100 percent college completion rate is associated with an increase of 46 percent in earnings for that cohort. Thus, increasing college completion rates from the level of the state with the lowest rates to that with the highest (a 42 percentage point change) is associated with 19 percent higher wages for the more educated cohort. The IV estimate suggests that college completion causes a significant increase in earnings of approximately 72 percent. The OLS estimate for hourly wages indicates that moving from a 0 to 100 percent college completion rate is associated with an increase in hourly wages of 23 percent. The

[31] All of the regressions are restricted to full-time workers. Hourly wages are further truncated to remove the bottom and top 0.5 percent in order to reduce measurement error.

IV specifications yield large and significant impacts indicating that universal college completion would raise wages by about 30 percent. These results are in line with other estimates of the causal impact of college education (or completion) on earnings.[32]As with the earnings results, we cannot reject that the IV estimates are significantly different from the OLS estimates.

Note that our IV estimates for the impact of college completion 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. So, while our preferred specifications suggest that wages and earnings may be an important mediating factor in the effect of education on health, we explore other potential mechanisms as well.[33]

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. Earning a college degree 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 services might be more inclined to purchase health insurance, complicating the measurement of a causal effect of

[32] 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. Similarly, Kane and Rouse (1995) find that the premium to a bachelor's degree for men is about 32 percent on average. Using an RD design on Florida data, Zimmerman (forthcoming) finds a long-run earnings impact of a bachelor's degree of 22%.

[33] We have also examined impacts on labor force participation and found that college completion is associated with an approximately 7 percentage point increase in participation based on both the OLS and the IV specifications (with the exception of the linear specifications where the impacts appear negative and insignificant).

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. We anticipate that having a college degree is associated with a higher probability of having private health insurance. 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. We expect that both having a college degree and being a veteran are associated with a higher probability of having any health insurance. Note that in the ACS individuals can be classified as having both private and public health insurance.

The main results are reported in Table 7. The IV estimates suggest that increasing college completion rates from the level of the state with the lowest rates to that with the highest (a 42 percentage point change) leads to a 5.7 percentage point higher rate of having any health insurance coverage (which is about 6.1 percent of the mean of 92.4). Next, we consider private coverage. While the coefficient on college graduate is only significant in the OLS specification, the magnitude is similar in both OLS and IV and suggests that a cohort having a 100 percent college graduation rate versus 0 percent is associated with a 10-16 percentage point higher probability of individuals in that cohort having private health insurance coverage. Finally, we see that rates of college completion are unrelated to public coverage rates, although the OLS relationship is negative. Veterans have higher rates of public coverage in both the OLS and IV 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 National Health Interview Survey (NHIS) Health Insurance Supplement. The 1980 NHIS is chosen because this is the baseline year from our 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:

$$insurance_{irc} = a_1 C_{irc} + a_2 V_{irc} + \tau trend_{rc} + \delta_r + \varepsilon_{irc} \tag{14}$$

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 an indicator for whether the individual has a college degree. 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 theoretically 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.[34]

[34] 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.

The results from linear probability model regressions, with standard errors clustered by region of residence, are reported in the top of Table 8. 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 very similar to those from the more recent ACS data, with the exception of the probability of having public health insurance coverage. We find that being a college graduate is associated with a 9.4 percentage point higher probability of having any health insurance, which is approximately 10 percent of the mean of 89 percent. Similarly, college graduates are 11.1 percentage points more likely to have private health insurance coverage than non-college graduates.

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 reason, to which they may respond, among other things, that they receive services through Medicaid, welfare, or CHAMPUS. Only 2 percent of the sample report having any public insurance coverage in the NHIS. The restrictiveness of the 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. Indeed, when we look at individuals who are ages 55-64 in 1980 (born between 1916-1925), we see that 11.9 percent have public health insurance coverage, which is much closer to the ACS mean. Although the estimated coefficient on college graduate 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 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's status and public health insurance coverage, while the effect is positive and significant in the ACS. It is likely that the restrictive definition of public coverage in the NHIS, combined

with lower rates of public coverage take-up in our cohorts in 1980 (the year of the NHIS) as compared to the ACS years, account for the low rates of public coverage among veterans in the NHIS as compared to the ACS. For these reasons, we prefer the ACS estimates and conclude that veteran status increases access to health insurance by providing higher rates of public coverage.

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

Our final set of results considers how acute health conditions and health behaviors are associated with college attainment and veteran status for our cohort of men. Here we use the 1985 NHIS Health Promotion and Disease Prevention (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 (14) 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 college graduate and veteran status in the bottom of Table 8.

We first consider several acute health measures that could help to explain the association between education and mortality. First, we find that college graduates have significantly lower body mass index (BMI) and are significantly less likely to be obese. We find that college graduates are 3.9 percentage points more likely to report having been under a lot of stress in the past two weeks relative to non-college graduates. This result is in contrast to the finding in Table 5 that college graduates are significantly less likely to die from heart disease. However, it may be that higher stress levels among college graduates are due to reporting differences across education groups, rather than true underlying stress levels. We do not find any relationship between being a college graduate and the probability of missing work or having restricted activity or bed days in the past two weeks.

The bottom of Table 8 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 college graduates in our cohorts are less likely to smoke while veterans are more

likely to smoke. Similarly, we see that college graduates are more likely to exercise. Therefore, we observe that college graduates are exhibiting healthier behaviors than non-college graduates, while veterans are more likely to smoke. The evidence on heavy drinking is mixed. According to Table 8, college graduates report similar levels of heavy drinking to less educated respondents in two of the three categories, but 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 5, although competing risks and reporting differences across education groups may explain the higher mortality rates from liver disease among college graduates. Health behaviors also cannot explain the protective effect of veteran status on mortality.

VII. Conclusion

This paper examines the causal impact of a 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 completion. Using Census data from 1980 to 2000 and Vital Statistics data from 1981 to 2007, we show that the effect of higher college completion rates on mortality as estimated by 2SLS is quite similar in magnitude to the OLS gradient. In our cohorts, these estimates indicate that increasing college completion rates from the level of the birth-state with the lowest rate to that with the highest would decrease deaths by about 39 per 1,000 men. This is equivalent to 28 percent of the mean of total mortality over our period, or a full standard deviation in birth state-cohort mortality. Alternatively, given the fraction of our sample with a college degree, our IV coefficient indicates that mortality rates over the period are about 2.21 times greater for non-college graduates than for college graduates. We find large negative effects of college completion on deaths from cancer and heart disease, the leading causes of mortality among older adults. Increasing college completion rates in the lowest-education birth state-cohort to that of the highest

would reduce early cancer deaths by 25.9 per 1,000 (71% relative to the mean) and early heart disease deaths by 18.9 (42% 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).[35] We find that college graduates are over twice as likely to report being a smoker in 1980 relative to non-college graduates. Thus, 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 completion on auxiliary outcomes such as health insurance and earnings, as well as the cross-sectional relationship between college completion and various 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 completion leads to both higher earnings and higher rates of health insurance. Using data from the National Health Interview Study, we show that college completion is also associated with less smoking and more exercise for our cohorts of interest.[36] These results suggest that the impact of college completion 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 between college graduates and non-graduates.

Our conclusion is that the causal impact of college education on health is large. Our estimates suggest that the entire mortality differential between college graduates and non-graduate in our sample can be attributed to college completion. It is important to consider the plausibility of impacts of this magnitude. We think there are two reasons to be confident that the magnitude of the impact of college education on health is large. First, our analysis suggests that a large portion of the mortality differential

[35] 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).
[36] See Cutler and Lleras-Muney (2010) for similar results in the NHIS for a broader set of cohorts.

can be explained by differences in smoking rates. As long as the difference in smoking rates between college graduates and other is causal – which, admittedly remains a point for future research – then a large portion of our impacts can be attributed to this channel alone. Second, our estimates of the impact of college completion on earnings are in line with what others have found using a variety of alternative identification approaches.

On the other hand, 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 overall? In answering this question, we feel 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 small positive earnings impacts we find can also be explained by this, but there are no relevant impacts in the literature to which we can compare ours.[37]

[37] 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.

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. Given that higher education policy in recent years has been focused on increasing college completion, this represents a particularly important margin of analysis. 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 graduation. To the extent that this paper provides useful information about the benefits of college completion in reducing adult mortality, we hope that it might help address one possible source of inefficient investments in human capital.

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.

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

Bertrand, Marianne; Esther Duflo; and Sendhil Mullainathan. 2004. "How Much Should We Trust Difference-in-Differences Estimates?" *Quarterly Journal of Economics*. 119(1): 249-275.

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

Black, Dan A.; Mark C. Berger; and Frank A Scott. 2000. "Bounding Parameter Estimates with Nonclassical Measurement Error," *Journal of the American Statistical Association,* Sep 2000, 95(451): 739-738.

Black, Dan; Seth Sanders; and Lowell Taylor. 2003. "Measurement of Higher Education in the Census and Current Population Survey," *Journal of the American Statistical Association,* Sep 2003, 98 (463): 545-554.

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. 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. 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.

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

Kane, Thomas J., and Cecilia Elana Rouse. 1995. "Labor-Market Returns to Two- and Four-Year College," *American Economic Review*, 85(3): 600-614.

Kane, Thomas; Cecilia Rouse and Douglas Staiger. 1999. "Estimating Returns to Schooling When Schooling is Misreported." *Industrial Relations Section Working Paper #419*. Princeton University.

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.

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.

Malamud, Ofer and Wozniak, Abigail. Forthcoming. "The Impact of College on Migration: Evidence from the Vietnam Generation." *Journal of Human Resources*.

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.

Office of the Vice President. Press release for March 22, 2011. "Vice President Issues Call to Action to Boost College Graduation Rates Nationwide." http://www.whitehouse.gov/the-press-office/2011/03/22/vice-president-biden-issues-call-action-boost-college-graduation-rates-n. Accessed 29 May 2012.

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, James P. 1999. "Healthy bodies and thick wallets: The dual relationship between health and economic status," *Journal of Economic Perspectives*, 13(2): 145-166.

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.

Shapiro, Andrew O. and Striker, John M. *Mastering the Draft: A Comprehensive Guide to Solving Draft Problems*. Boston: Little, Brown and Company (1970).

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.

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.

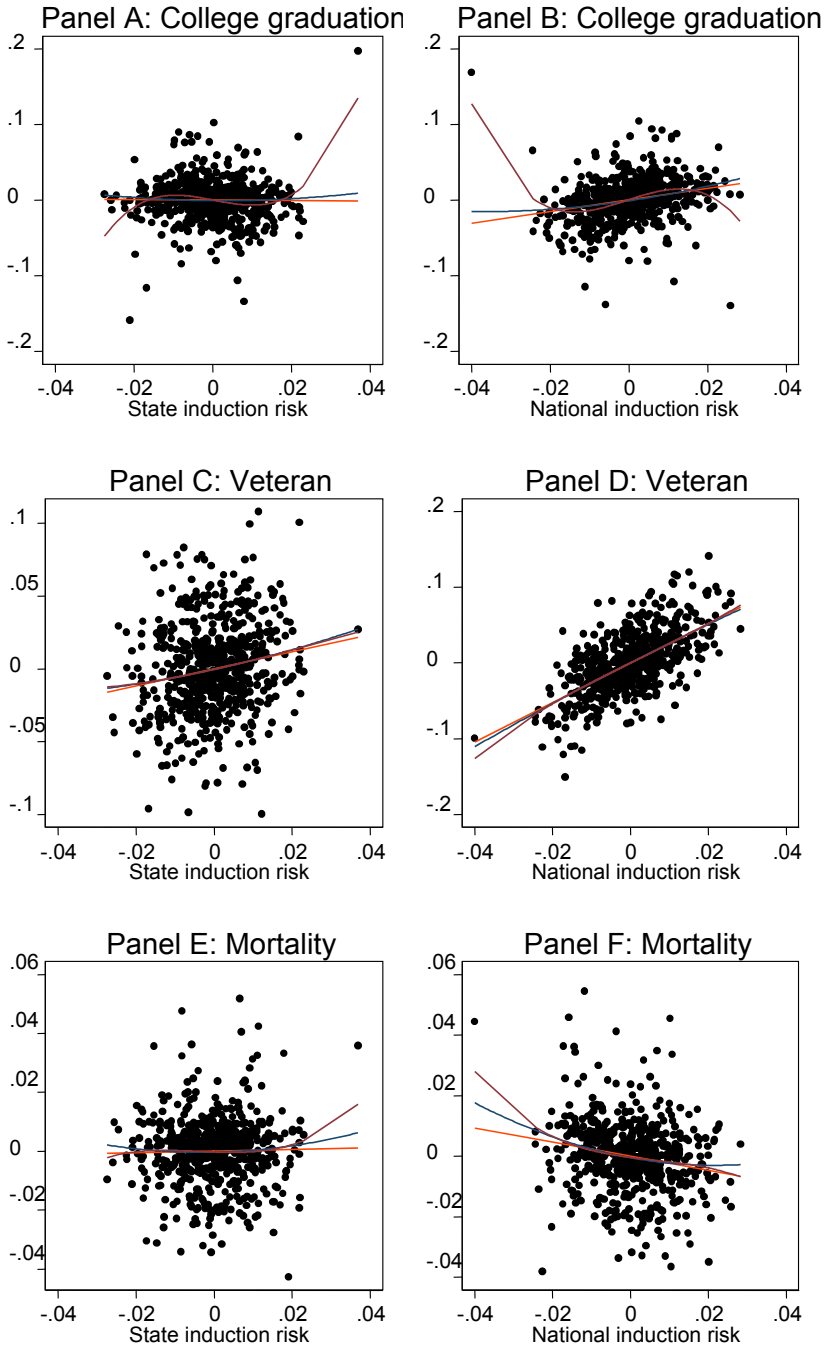
Wozniak, Abigail. (2010) "Are College Graduates More Responsive to Distant Labor Market Opportunities?" *Journal of Human Resources.*

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. Forthcoming. "The Returns to College Admissions for Academically Marginal Students." *Journal of Labor Economics,* forthcoming.

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



Notes: State induction risk and national induction risk are defined as in equations (10) and (11) 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). Linear, quadratic, and cubic fits are also plotted above.

Table 1: Summary Statistics

VARIABLES	Mean	SD
Census Data		
Years of Higher Education	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, 1990, and 2000 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.

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

	Dependent Variable:					
	College Graduate	Veteran	College Graduate	Veteran	College Graduate	Veteran
National Induction Risk	0.935**	2.349**	1.899**	2.865**	2.368**	1.237*
	(0.091)	(0.185)	(0.263)	(0.354)	(0.437)	(0.728)
National Risk ^2			-7.484**	-3.891*	-22.748**	25.194**
			(1.644)	(2.068)	(5.946)	(9.238)
National Risk ^3					101.276**	-154.080**
					(29.076)	(42.970)
State Induction Risk	-0.240**	0.768**	-0.380	0.653**	0.588	1.171*
	(0.089)	(0.162)	(0.233)	(0.311)	(0.375)	(0.702)
State Risk ^ 2			0.770	0.627	-12.254**	-5.907
			(1.260)	(1.731)	(4.646)	(8.736)
State Risk ^ 3					54.507**	25.502
					(19.888)	(36.988)
Cragg-Donald F-stat	14.23		11.64		13.14	
[p-value]	[0.00]		[0.00]		[0.00]	
Angrist-Pischke F-Stat	19.18	123.00	16.07	269.77	21.89	277.53
F-Stat	255.23	2154.74	169.47	1137.29	129.27	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. Specification refers to the functional form of the induction risk variables in the first stage regression. See the text for a discussion of the null hypotheses of the three F-statistics. See Table 1 notes for data sources, sample restrictions and variable definitions. 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 defined at the birth-year level), and birth-region trends.

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

		IV Specification:				
	OLS	Linear-2SLS	Quadratic-2SLS	Quadratic-LIML	Cubic-2SLS	Cubic-LIML
	[1]	[2]	[3]	[4]	[5]	[6]
College Graduate	-101.97***	-99.26	-170.76***	-172.46***	-93.93***	-93.53***
	(19.59)	(77.55)	(38.34)	(39.22)	(25.56)	(26.69)
Veteran Status	-46.36***	-42.93***	-27.21***	-26.82***	-44.45***	-44.50***
	(5.14)	(16.43)	(9.67)	(9.86)	(6.81)	(7.03)
Observations	600	600	600	600	600	600
R-squared	0.9516	0.9515	0.9505	0.9504	0.9515	0.9515

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. See Table 1 notes for data sources, sample restrictions and variable definitions. The IV specifications use the national and state-level induction risk to instrument for college graduate and veteran status. 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.

Table 4: OLS and IV Estimates for the Impact of College Completion Rates on Mortality by Decade

	1981-1990		1991-2000		2001-2007	
VARIABLES	OLS	IV	OLS	IV	OLS	IV
College Graduate	-24.51***	-24.38***	-20.00**	-79.46***	-38.58***	-100.37***
	(4.56)	(6.43)	(9.10)	(30.46)	(11.26)	(37.05)
Veteran Status	-9.93***	-9.33***	-35.54***	-27.32***	-36.99***	-28.80***
	(1.10)	(1.36)	(2.20)	(4.49)	(2.53)	(6.06)
Cragg-Donald F-Stat	13.14		5.46		3.60	
[p-value]	[0.00]		[0.00]		[0.00]	
A-P F-Stat for:						
College Grad	21.89		8.91		5.81	
Veteran	277.53		100.11		101.06	
F-Stat for:						
College Grad	129.27		38.51		50.81	
Veteran	851.75		1002.92		977.61	
Mean Mortality Rate	25.46		54.88		73.33	
Total Deaths	366,365		727,108		901,008	

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. See the text for a discussion of the null hypotheses of the three F-statistics. See Table 1 notes for data sources, sample restrictions and variable definitions. The IV specifications use a third-order polynomial in the national and state-level induction risks to instrument for college graduate and veteran status 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.

Table 5: OLS and IV Estimates for the Impact of College Completion Rates 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										
College Graduate	-38.34***	-32.29***	-5.21**	-0.68	-2.62	-1.19	-2.41	-2.98***	-5.33***	-1.62**
	(6.57)	(5.95)	(2.11)	(1.47)	(2.46)	(1.54)	(1.52)	(0.86)	(1.56)	(0.63)
Veteran Status	-15.88***	-10.16***	-3.20***	-1.71***	-0.980**	-1.250***	-0.730**	-1.79***	-4.00***	0.04
	(1.69)	(1.81)	(0.60)	(0.46)	(0.45)	(0.41)	(0.36)	(0.28)	(0.36)	(0.16)
Panel B: IV										
College Graduate	-61.90***	-45.07***	-0.71	4.24	14.75***	12.20***	-7.04***	-6.61***	-9.11***	-0.10
	(9.18)	(11.32)	(4.92)	(3.02)	(5.19)	(4.16)	(2.21)	(1.94)	(2.65)	(1.49)
Veteran Status	-9.72***	-6.64**	-3.81***	-2.74***	-4.56***	-4.04***	0.35	-0.88*	-3.05***	-0.22
	(2.45)	(3.20)	(1.15)	(0.80)	(1.18)	(0.94)	(0.48)	(0.45)	(0.56)	(0.35)

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. See Table 1 notes for data sources, sample restrictions and variable definitions. The IV specifications use a third-order polynomial in the national and state-level induction risks to instrument for college graduate and veteran status 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.

Table 6: OLS and IV Estimates for the Impact of College Completion Rates on Wages and Earnings, 1980

	Log Earnings		Log Wages	
VARIABLES	OLS	IV	OLS	IV
College Graduate	0.456***	0.720***	0.225***	0.293***
	(0.074)	(0.167)	(0.061)	(0.122)
Veteran Status	0.159***	0.0810*	0.115***	0.0890***
	(0.020)	(0.046)	(0.015)	(0.038)
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. First stage specifications and estimating sample are based on the cubic specifications in Table 2. Dependent variable is log real earnings or log real wages. 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.

Table 7: College Completion Rates and Health Insurance Access

	Any Insurance	Private Coverage	Public Coverage
Panel A: OLS			
College Graduate	0.093***	0.163***	-0.083**
	(0.025)	(0.028)	(0.035)
Veteran Status	0.049***	-0.038*	0.241***
	(0.011)	(0.021)	(0.028)
Panel B: IV			
College Graduate	0.135*	0.107	0.082
	(0.072)	(0.108)	(0.135)
Veteran Status	0.074***	-0.027	0.306***
	(0.017)	(0.020)	(0.023)
Mean of dep. variable	92.4%	83.7%	17.9%

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. Each "College Graduate" and "Veteran" pair are coefficients from a single estimation using the cubic OLS and IV specifications from Table 3. C-D F statistic is 0.72 (p-value of 0.49). First stage standard F statistics are 13.6 and 11.2 for college graduate and veteran status, respectively; A-P F statistics are 13.3 and 146.1 for college graduate and veteran status, respectively.

Table 8: Acute Health Outcomes and Health Behaviors

Dependent Variable	N	Mean / Percent	College Graduate		Veteran	
1980 Health Insurance Supplement:						
Any Health Insurance	6,292	89.0%	0.094***	(0.010)	0.036**	(0.009)
Private Coverage	6,292	86.8%	0.111***	(0.010)	0.035***	(0.005)
Public Coverage	6,292	2.0%	-0.019***	(0.002)	-0.001	(0.004)
1985 NHIS HPDP Supplement:						
Body Mass Index (BMI)	3,069	25.82	-1.081**	(0.294)	0.084	(0.093)
Obese (BMI ≥30)	3,069	10.9%	-0.066*	(0.022)	-0.011	(0.008)
Ever told had:						
Hypertension	3,031	18.7%	-0.017	(0.012)	0.006	(0.007)
High cholesterol	3,035	4.5%	0.013	(0.014)	-0.015**	(0.004)
Heart Condition	3,033	3.2%	-0.003	(0.006)	0.002	(0.011)
Stroke	3,035	0.5%	-0.004*	(0.001)	0.003	(0.002)
Has Diabetes	3,032	1.3%	0.002	(0.002)	-0.006	(0.003)
Any Restricted Days in Past 2 Wks	3,069	8.0%	-0.008	(0.009)	0.013	(0.013)
Any Work Loss Days in Past 2 Wks	3,069	4.7%	0.000	(0.006)	0.016	(0.015)
Any Bed Days in Past 2 Wks	3,069	4.2%	-0.001	(0.006)	0.010	(0.009)
A Lot of Stress in Past 2 Weeks	3,018	22.7%	0.039*	(0.012)	-0.010	(0.017)
Ever Smoked 100 Cigarettes	3,030	67.2%	-0.200***	(0.008)	0.074***	(0.012)
Smoke Now	3,006	37.6%	-0.213***	(0.016)	0.062**	(0.013)
Any days in Past Year 5+ Drinks	2,977	44.1%	-0.030	(0.011)	0.071**	(0.015)
25+ days in Past Year 5+ Drinks	2,977	14.0%	-0.082**	(0.023)	0.005	(0.009)
Any Drunk Driving in Past Year	2,996	20.9%	-0.008	(0.014)	0.014	(0.023)
Exercise or play sports regularly	3,036	45.0%	0.249***	(0.019)	0.011	(0.014)

Notes: The sample is restricted to white, male respondents born between 1942 and 1953. Each "College graduate" 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 30.1% are college graduates 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, 32.6% are college graduates and 38.1% are veterans.

Appendix Table 1: OLS and IV Estimates for the Impact of Years of College on Mortality

	Cumulative Mortality		Cancers		Heart Disease	
VARIABLES	OLS	IV	OLS	IV	OLS	IV
Years of College	-22.31***	-17.31***	-8.05***	-11.59***	-6.67***	-8.37***
	(3.80)	(5.42)	(1.24)	(1.98)	(1.21)	(2.36)
Veteran Status	-42.82***	-45.07***	-14.94***	-9.92***	-9.49***	-6.86**
	(5.38)	(7.34)	(1.63)	(2.70)	(1.84)	(3.38)

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 IV specifications use a third-order polynomial in the national and state-level induction risks to instrument for years of college and veteran status using a 2SLS model. The first stage for the regression of years of college on the draft risk variables has an A-P F-statistic of 23.81 and a standard F-statistic of 145.68. See Table 1 notes for data sources, sample restrictions and variable definitions. 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.

Appendix Table 2: OLS Estimates for the Impact of College Completion Rates on Cumulative Mortality, 1980-2007, for Early and Late Cohorts

	Baseline OLS		OLS - Early Cohorts		OLS - Late Cohorts	
	[1]	[2]	[3]	[4]	[5]	[6]
College Graduate	-101.97***	-115.40***	-101.88	-146.15**	1.42	-3.76
	(19.59)	(19.50)	(65.98)	(66.38)	(23.46)	(23.62)
Veteran Status	-46.36***	-49.62***	-73.29	-97.98	-44.85	-70.11**
	(5.14)	(5.25)	(63.50)	(66.05)	(32.35)	(31.93)
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	0.9516	0.9491	0.9212	0.9112	0.9568	0.9559

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. See Table 1 notes for data sources, sample restrictions and variable definitions. "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.

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

		IV Specification:				
	OLS	Linear-2SLS	Quadratic-2SLS	Quadratic-LIML	Cubic-2SLS	Cubic-LIML
	[1]	[2]	[3]	[4]	[5]	[6]
Panel A: Log earnings						
College Graduate	0.456***	-0.131	0.424**	0.418*	0.720***	0.769***
	(0.0737)	(0.295)	(0.205)	(0.238)	(0.167)	(0.197)
Veteran Status	0.159***	0.330***	0.169***	0.171**	0.0810*	0.0667
	-0.0203)	(0.0924)	-0.0633)	-0.0726)	-0.0463)	-0.0543)
Mean log earnings	9.543	9.543	9.543	9.543	9.543	9.543
Panel B: Log wages						
College Graduate	0.225***	-0.0395	0.0699	0.0661	0.293**	0.299**
	(0.0606)	(0.228)	(0.173)	(0.176)	(0.122)	(0.131)
Veteran Status	0.115***	0.187***	0.156***	0.157***	0.0890**	0.0872**
	(0.015)	(0.0679)	(0.0525)	(0.0535)	(0.0381)	(0.0405)
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. First stage specifications and estimating sample are identical to those in Table 2. Dependent variable is log real earnings or log real wages. 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.