

NBER WORKING PAPER SERIES

THE LONG-TERM CONSEQUENCES OF VIETNAM-ERA CONSCRIPTION AND
GENOTYPE ON SMOKING BEHAVIOR AND HEALTH

Lauren Schmitz
Dalton Conley

Working Paper 21348
<http://www.nber.org/papers/w21348>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2015

Dalton Conley would like to thank the Russell Sage Foundation for financial support for this project. The Health and Retirement Study (HRS; accession number 0925-0670) is sponsored by the National Institute on Aging (grant numbers NIA U01AG009740, RC2AG036495, and RC4AG039029) and is conducted by the University of Michigan. Additional funding support for genotyping and analysis were provided by NIH/NICHHD R01 HD060726. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2015 by Lauren Schmitz and Dalton Conley. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Long-Term Consequences of Vietnam-Era Conscriptioin and Genotype on Smoking Behavior and Health

Lauren Schmitz and Dalton Conley

NBER Working Paper No. 21348

July 2015

JEL No. I1,I12,I14

ABSTRACT

Research is needed to understand the extent to which environmental factors mediate links between genetic risk and the development of smoking behaviors. The Vietnam-era draft lottery offers a unique opportunity to investigate whether genetic susceptibility to smoking is influenced by risky environments in young adulthood. Access to free or reduced-price cigarettes coupled with the stress of military life meant conscripts were exposed to a large, exogenous shock to smoking behavior at a young age. Using data from the Health and Retirement Study (HRS), we interact a genetic risk score for smoking initiation with instrumented veteran status in an instrumental variables (IV) framework to test for genetic moderation (i.e. heterogeneous treatment effects) of veteran status on smoking behavior and smoking-related morbidities. We find evidence that veterans with a high genetic predisposition for smoking were more likely to become regular smokers, smoke heavily, and are at a higher risk of being diagnosed with cancer or hypertension at older ages. Smoking behavior was significantly attenuated for high-risk veterans who attended college after the war, indicating post-service schooling gains from veterans' use of the GI Bill may have reduced tobacco consumption in adulthood.

Lauren Schmitz

New School for Social Research

New York, NY 10011

schml347@newschool.edu

Dalton Conley

New York University

249 West 29th Street #2E

New York, NY 10001-5230

and NBER

daltonclarkconley@gmail.com

1. Introduction

Cigarette smoking continues to be a significant and costly public health problem in the United States. Cigarette smoking and exposure to tobacco smoke resulted in at least 443,000 premature deaths, approximately 5.1 million years of potential life lost, and \$96.8 billion in productivity losses annually between 2000-2004 (CDC, 2008). Studies on the persistence of smoking across the life span have consistently found heavy smoking in young adulthood to be a strong predictor of smoking later in life (Merline, O'Malley, Schulenberg, Bachman, & Johnston, 2004). Moreover, few people begin smoking after adolescence (Kandel & Logan, 1984), and young adults who manage to quit smoking are generally able to avoid any significant lifetime health consequences (U.S. Department of Health & Human Services, 1990). These findings have motivated health policymakers and researchers to understand the complex influences that contribute to smoking initiation and cessation during this critical time period. However, the vast majority of research has telescoped its efforts around understanding either the social or the biological causes of early onset cigarette use, and has therefore neglected the extent to which environmental factors may moderate or mediate links between genetic risk and the development of smoking behaviors.

The Vietnam-era draft lottery offers a unique opportunity to investigate whether genetic predisposition to smoking is influenced by risky or stressful environments in young adulthood. Between August 1964 and May 1975 approximately 2.7 million Americans were inducted in the military—1,857,304 of which entered military service through the Selective Service System (U.S. Department of Veteran Affairs, 2015; U.S. Selective Service, 2015). Whether they were deployed to Vietnam or served outside the theater of war, conscripts were exposed to a large, exogenous shock to smoking behavior when they were approximately 19-22 years of age. All

men who served had access to tax-free cigarettes at military bases and commissaries, and men in combat received free cigarettes as part of their K- and C-rations (U.S. Department of Health and Human Services, 1989, pp. 425-426). Given the high concentration of extant smokers at military bases and facilities, peer effects may have further augmented the effect of cheap or reduced-price cigarettes (Eisenberg & Rowe, 2009). In addition, many servicemen were subjected to stressful, if not traumatic experiences, which could have led to elevated levels of depression and anxiety—both of which are associated with an increased risk of smoking (Glassman et al., 1990; Winefield, Winefield, Tiggemann, & Goldney, 1989). While a few studies have used the draft lotteries to evaluate the causal impact of military service on smoking behavior and health, this study is the first to examine whether genetic susceptibility to smoking is influenced by military service in young adulthood.

To properly test for gene-by-environment (GxE) interactions between cigarette consumption and military service, we incorporate a polygenic score or genetic risk score (GRS) for smoking behavior into an instrumental variables (IV) framework. Specifically, because selection into the military is far from random, and likely to be correlated with factors like socioeconomic background or underlying genotype, we follow prior studies and use respondent date of birth files for cohorts born between 1948 and 1952 in the Health and Retirement Study (HRS) as an instrumental variable to code for draft eligibility. Instrumented veteran status is then interacted with a GRS for smoking initiation to test for genetic moderation, or heterogeneous treatment effects, of veteran status on smoking behavior and smoking-related morbidities. Existing efforts to find associations between genetic variation and social behavior in large, multidisciplinary surveys are often unable to support causal inference because they use endogenous measures of the environment, genotype, or both (Conley, 2009; Fletcher & Conley,

2013). By exploiting natural variation in exposure to military service, this study lends insight into how physiological pathways might be moderated or mediated by environmental influences.

Our results show veterans with a high genetic predisposition for smoking—defined as having a GRS one or two standard deviations above the mean—were more likely to become smokers, or smoke more than 100 cigarettes in their lifetime, and smoke in greater quantities on a day-to-day basis than comparable non-veterans with the same GRS. Veterans with a GRS one to two standard deviations above the mean were approximately 57 to 71 percent more likely to become smokers, and smoked 18 to 27 cigarettes more per day than comparable non-veterans. Critically, we find these effects were significantly attenuated for high-risk veterans who attended college after the war, indicating post-service schooling gains from veterans’ use of the GI Bill may have reduced tobacco consumption in adulthood. We also find some evidence that high-risk veterans were more likely to be diagnosed with cancer and have higher mean arterial blood pressure (MAP) than comparable non-veterans.

The study is structured as follows. After reviewing the relevant literature on the genetics of smoking behavior and its relation to the draft lottery and military service in Section 2, we present an in-depth explanation of the data and methodology behind our causal GxE interaction model in Sections 3 and 4. Finally, Section 5 gives a detailed explanation of our results, and Section 6 concludes.

2. Background

2.1 *The Genetics of Smoking Behavior*

Twin studies on the genetics of smoking behavior have found genetic influences play an important role in smoking initiation, cessation, and response to antismoking interventions. Studies estimate heritability accounts for approximately 44 percent of the variation in the frequency of smoking (Slomkowski, Rende, Novak, Lloyd-Richardson, & Niaura, 2005) and 65 percent of the variation in current and lifetime tobacco use (Maes et al., 1999). Recently, genome-wide association studies (GWASs) of current and former smokers have identified genome-wide significant loci that are associated with heavy smoking in adulthood (Furberg et al., 2010; Liu et al., 2010; Thorgeirsson et al., 2010).

Importantly, factors that influence smoking initiation appear to differ from factors that contribute to the persistence and quantity of smoking behavior. In particular, shared environmental influences play a larger role in smoking initiation than persistence, and there is considerable evidence that peer effects fuel onset in adolescence. However, only a few studies have integrated environmental and biological explanations for cigarette use (Boardman, Saint Onge, Haberstick, Timberlake, & Hewitt, 2008; Daw et al., 2013; Kendler, Gardner, Jacobson, Neale, & Prescott, 2005; Timberlake et al., 2006). For example, using twin data from the National Longitudinal Study of Adolescent Health, Boardman et al. (2008) find the heritability of daily smoking is higher in schools where the most popular students are smokers and reduced in schools where the majority of students are non-Hispanic and white. On the other hand, using findings from GWASs to derive a GRS for smoking quantity, Belsky et al. (2013) find high genetic risk leads to persistent heavy smoking, nicotine dependence, and difficulty quitting for

individuals who progress rapidly from smoking initiation to heavy smoking during adolescence. The authors conclude interventions that disrupt the socially developmental progression of smoking behavior in young adulthood may mitigate the development of adult smoking problems for individuals with high genetic risk.

These findings signal the need for further research that can elucidate specific environmental mechanisms that moderate or mediate links between genetic risk and the development and persistence of smoking behaviors. This study is the first to use a polygenic score to measure genetic risk and a plausibly exogenous environmental exposure—the Vietnam-era draft lotteries—to identify whether military service shaped tobacco consumption across the life span.

2.2 *The Draft Lottery as a Quasi-Natural Experiment*

Between December 1969 and February 1972, the U.S. Selective Service held four Vietnam-era draft lotteries. Each of these draft lotteries randomly assigned men in eligible birth cohorts ordered induction numbers through a hand drawing of birthdates. Individual birthdates (including February 29th) were assigned a corresponding number between 1 and 366, placed in a blue capsule, and then individually drawn until each day of the year was paired with a random sequencing number (RSN). Depending on the U.S. Department of Defense's needs for manpower, a cut-off number for draft eligibility was chosen for the following calendar year. Men with RSNs below the cut-off were considered draft-eligible for that particular year and were therefore more likely to serve, whereas men with RSNs above the cutoff were not at risk of conscription. The first lottery was held on December 1, 1969 for men born between 1944 and 1950; subsequent lotteries drafted men from the 1951 and 1952 birth cohorts, respectively. A

final lottery was held in 1972 for men born in 1953, but since no new conscription orders were issued after 1972, these men were never called to service.

The random assignment mechanism of the draft lotteries has been used to identify the effects of war-time military service on a host of outcomes, including economic (Angrist, 1990; Angrist & Chen, 2011; Angrist, Chen, & Song, 2011), family (Conley & Heerwig, 2011; Heerwig & Conley, 2013) and health outcomes (Angrist, Chen, & Frandsen, 2010; Conley & Heerwig, 2012; Dobkin & Shabani, 2009). While earlier studies uncovered small but statistically significant effects of military service on mortality (Hearst, Newman, & Hulley, 1986) and earnings (Angrist, 1990), by and large studies have been unable to uncover a causal link between conscription and long-term health outcomes, lifetime earnings losses, employment, or labor force participation.

To our knowledge, only a few studies have estimated the causal effects of military service on smoking behavior and health. Bedard and Deschênes (2006) estimate the long run effects of military service in World War II and the Korean War on smoking, mortality, and other measures of health. Since detailed individual-level data is not available for these veterans over time, they construct panels of all-cause and cause-specific mortality rates for different birth cohorts and use instrumental variable approaches that exploit variation in the proportion of men who served across birth cohorts. They find military service during WWII and the Korean War era is associated with a 30 percentage point increase in the probability of lifetime smoking, and 36 to 79 percent of veteran deaths from heart disease and lung cancer are attributable to military-induced smoking.

Using the Vietnam-era draft lotteries, Dobkin and Shabani (2009) find no evidence of increased smoking behavior in the National Health Interview Survey (NHIS) over the years

1997-2004, indicating any increase in smoking behavior due to conscription in young adulthood likely dissipated by mid-life, or by the time veterans were aged 45-54. They also do not find evidence that Vietnam-era service resulted in poorer health outcomes as these populations age, or across three successive time periods in the NHIS (1974-1981, 1982-1996, and 1997-2004). Similarly, Eisenberg and Rowe (2009) find Vietnam-era military service increased the probability of smoking by 35 percentage points when veterans were aged 25 to 30, or as of 1978-1980 in the NHIS, but find any early effects were substantially attenuated later in life, and did not result in any long-term health consequences.

A major limitation of these studies is they are only able to estimate the *average* treatment effect of conscription on smoking behavior and health. In other words, studies have not been able to account for heterogeneous treatment effects by genotype, or the possibility that military-induced smoking varies across the spectrum of genetic risk. In addition, a significant link between military service and smoking behavior may be difficult to unearth due to the potentially offsetting effect of the GI Bill, which encouraged veterans to attend college after service. Prior research has shown the GI Bill increased schooling for WWII (Bound & Turner, 2002), Korea-era veterans (Stanley, 2003), and Vietnam veterans (Angrist & Chen, 2011). Among white Vietnam veterans, 50 percent used funds from the GI Bill for education—63 percent of which was put towards college courses, resulting in 0.26 more years of college on average compared to non-veterans (Angrist & Chen, 2011). Higher educational attainment is strongly associated with better health (e.g. Fuchs, 1982; Grossman, 2006; Lleras-Muney, 2005), and may have exerted an important influence on veterans by reinforcing positive health behaviors—either through improved decision-making and access to health-related information or because the boost in future earnings potential incentivized better health investments (Grossman, 1972). Along these

lines, numerous studies have found a causal link between education and a reduction in smoking behavior (e.g. de Walque, 2010; Farrell & Fuchs, 1982). In particular, studies have used the draft avoidance behavior documented by Card and Lemieux (2001) to infer causation from education to smoking. For example, De Walque (2007) and Grimard and Parent (2007) exploit the uptick in male college attendance in the early 1960s that occurred as a result of educational deferments and find convincing evidence that educational attainment reduced tobacco consumption.

Given the strong negative correlation between education and smoking in the literature, the GI bill may have acted as a second moderating environmental lever after the war, particularly for high-risk individuals. On the other hand, educational attainment may be endogenous in the GxE smoking model if it is driven by factors outside of the draft, such as socioeconomic background or unmeasured ability, that influence health behaviors. However, if draft status was orthogonal to standard socio-demographic variables at the time of the lottery and educational attainment after the war was directly related to draft eligibility, as past studies have shown, education may be another mechanism affecting the smoking behavior of conscripts. As a result, we also explore whether obtaining a college or advanced degree after the war moderated the smoking behavior of high-risk individuals in separate GxE interaction models.

3. Data and Descriptive Statistics

Our primary data source is the Health and Retirement Study (HRS). The HRS is a nationally representative, longitudinal dataset of individuals aged 51 plus and their spouses that was launched in 1992. It follows individuals from the time of their entry into the survey until their death. It introduces a new cohort of participants every six years, and interviews around

20,000 participants every two years. For the purposes of this study, the HRS genotyped 12,507 respondents in 2009 that provided DNA samples in 2006 and 2008. Genotype data is collected alongside detailed health, economic, and social data, including information about chronic illness, functional ability, depression, self-assessed health status, health related behaviors, and veteran status.

The majority of men who were draft eligible during the Vietnam era entered the HRS in 2004 as part of the “Early Baby Boomer” cohort—a nationally representative sample of men and women born between 1948-1953. However, to maximize sample size, we also include spouses of female respondents from former cohorts that were born between 1948-1952. After quality control analysis on the genotype data, 12,150 observations remained (see section 4.1 for QC specifics). We then exclude respondents who report being female (N=7,176), black (N=574), American Indian, Alaskan Native, Asian, or Pacific Islander (N=103), or who were born before 1948 (N=3,450) or after 1952 (N=216). Our final sample consists of 631 white men born between 1948-1952 who provided DNA samples in 2006 or 2008.

3.1 *Smoking Phenotypes*

The HRS asks several questions that pertain to current and former smoking behavior. All respondents were asked if they ever smoked more than 100 cigarettes in their lifetime (not including pipes or cigars). If the respondent answers yes to this question, they are asked if they currently smoke and if so, how many cigarettes or packs of cigarettes they smoke per day. If they report smoking in their lifetime but do not currently smoke, they are asked how many cigarettes or packs they smoked per day when they were smoking the most. The HRS also asks all current and former smokers when they started smoking, and if they no longer smoke, when

they stopped. Smoking questions were fielded to both current and former smokers in the 1992 and 1998-2010 waves.¹ Our primary phenotypes of interest for smoking behavior include a dichotomous variable for whether or not an individual reports smoking 100 cigarettes or more in their lifetime (“Ever Smoker”) and a continuous variable for either the amount of cigarettes the respondent reports smoking per day if they currently smoke or the maximum amount of cigarettes the respondent smoked per day if they no longer smoke (“Cigarettes Per Day”).

There are a few limitations with the measurement of both phenotypes that bear some acknowledgement. First, with respect to the “Ever Smoker” phenotype, the majority of current or former smokers in our sample (74.4%) report smoking before age 19, or before they were draft eligible—i.e. most men in our sample who served in the military did not initiate smoking during their tour of service or when they were draft eligible (only 16.1% report onset between the ages of 19 and 22). However, because we do not have any information on the frequency of smoking behavior at younger ages, it is impossible to discern at what point individuals crossed the 100-cigarette threshold. Tobacco use changes considerably during adolescence and young adults smoke fewer cigarettes daily and are less likely to smoke regularly than adults (CDC 2003; Johnston, O'Malley, Bachman, & Schulenberg, 2005; Rouse, Sanderson, & Feldmann, 2002). Thus, nondaily, experimental, or “social” smoking behavior early on may have evolved into established use as a result of military service. Likewise, men who started smoking after age 22 (9.4%) may have initiated because they experienced high levels of stress, difficulties assimilating back into civilian life, or other negative downstream effects from military service after the war. In both cases, because we cannot rule out the possibility that Vietnam-era service

¹ In 1994 and 1996, the smoking questions were fielded to current smokers only. However, since questions about former smoking behavior were asked in 1992, and a new cohort was not added until 1998, data on past smoking behavior is available for the majority of participants.

was a gateway to longer-term or more persistent tobacco use, we include the “Ever Smoker” phenotype in our results with the caveat that some men may have smoked 100 cigarettes or more before the war.

Second, since the HRS only asks former smokers about their lifetime maximum CPD, our “Cigarettes Per Day” phenotype measures current CPD for men who still smoke (approximately 45.2% of current and former smokers in our sample), and maximum CPD for former smokers. Current daily smoking may be an inaccurate measure for lifetime intensity. For example, it is unlikely that a person who currently smokes two packs a day has been consistently smoking this amount since they started. Smokers may have changed their lifetime smoking behavior for health reasons, to compensate for the changing nicotine yields in manufactured cigarettes, or to adjust for increases in the real price of cigarettes over time (CDC, 1994). The decreasing social acceptability of smoking may also cause current smokers—and heavy smokers in particular—to underreport the number of cigarettes they smoke.² A comparison of sample means finds there is no significant difference between self-reports of CPD for current and former smokers—i.e. current smokers are not smoking more on average than former smokers did at their peak. Thus, while the ideal measure would report maximum CPD for all respondents, if smoking behavior among current smokers tends to remain at or below peak levels, we would expect our results to be biased downward if anything as a result of this discrepancy.

² In general, self-reports of smoking behavior have been shown to be a reliable measure over time despite mounting social stigma (Hatziafreu, Pierce, Fiore, Grise, Novotny, & Davis, 1989) and findings from biochemical validation studies suggest self-reported usage is a valid estimate of smoking status in the population (Fortmann, Rogers, Vranizan, Haskell, Solomon, & Farquhar, 1984; Patrick, Cheadle, Thompson, Diehr, Koepsell, & Kinne, 1994).

3.2 *Smoking-Related Morbidities*

Genetically-at-risk smokers tend to not only smoke more, but may also experience increased toxin exposure at fixed quantities of smoking, and are at increased risk for heart disease and lung cancer (Le Marchand et al, 2008; Thorgeirsson et al., 2008; Spitz, Amos, Dong, Lin, & Wu, 2008). To see if smoking among higher-risk genotypes resulted in any long-term health consequences, we examine outcomes for two primary smoking-related health phenotypes: cancer and hypertension.³ We use a dichotomous variable for incidence of cancer, which is equal to “1” if the respondent reports ever being diagnosed with cancer or a malignant tumor of any kind except skin cancer and “0” otherwise. For hypertension, we calculate mean arterial blood pressure (MAP), or the average arterial pressure during a single cardiac cycle. MAP is often used as an alternate indicator of blood flow and is believed to be a better indicator of tissue perfusion than systolic blood pressure (SBP) since it accounts for the fact that two thirds of the cardiac cycle are spent in diastole.⁴ MAP has also been shown to be a significant independent predictor of all-cause, cardiovascular, and coronary mortality (e.g. Benetos, Safar, Rudnichi, Smulyan, Richard, Ducimetière, & Guize, 1997). To calculate MAP in the HRS sample, we use measures of systolic and diastolic blood pressure taken during enhanced face-to-face interviews with an automated sphygmomanometry (inflated blood pressure cuff).

³ We also explored outcomes related to lung disease and lung function but failed to find any significant effects. Results are available from the authors upon request.

⁴ *Mean arterial pressure (MAP)* $\cong \frac{2}{3} \times DBP + \frac{1}{3} \times SBP$

3.3 Draft Eligibility

To assess whether an individual was draft eligible, we follow prior studies that use the Vietnam-era draft lotteries as a quasi-natural experiment, and assign order of induction numbers based on restricted respondent date of birth files for cohorts born between 1948 and 1952 in the HRS. Table 1 shows the draft eligibility cutoffs and birth cohorts affected by each draft lottery. Based on the eligibility cutoffs reported, draft eligibility is equal to “1” if the lottery number corresponding to a respondent’s date of birth was called in the corresponding draft year and “0” otherwise.⁵

Table 1. Draft eligibility numbers by birth cohort and lottery year

Lottery Year	Birth Cohort(s)	Eligibility Ceiling
1969	1944-1950	195
1970	1951	125
1971	1952	95
1972	1953	--

Source: U.S. Selective Service.

Due to the high number of men who volunteered for the military or received educational deferments before the first draft lottery, the vast majority of studies exclude the older 1944-1949 birth cohorts, since this sample is far from representative (Angrist, 1990). For example, using the NHIS, Dobkin and Shabani (2009) regress veteran status on draft eligibility and find men born between 1950-1952 on draft eligible days were 14.8 percentage points more likely to serve, whereas the probability of service for draft eligible men from the 1948 or 1949 cohorts was only around 7 percent. We replicate this analysis with the HRS sample and do not find a significant

⁵ The results of the Vietnam draft lottery are available at: <http://www.sss.gov/lotter1.htm>.

difference in the likelihood of service for draft eligible men in the 1948-1952 and 1950-1952 birth cohorts.⁶ Thus, because our sample of genotyped veterans is significantly lower than past studies that use larger, nationally representative data sets like the NHIS, we include the 1948-1949 birth cohorts to maximize the power of this study.

3.4 *Descriptive Statistics*

Table 2 reports descriptive statistics for the sample by veteran and draft eligibility status. Independent t-tests comparing the means of variables between groups reveal veterans differ substantially from non-veterans along several dimensions. Veterans are more likely to become regular smokers (0.692) than non-veterans (0.577), smoke 4.327 more cigarettes per day on average, and are 9.6 percent more likely to continue smoking at older ages. Veterans have higher rates of high school completion (0.606) than non-veterans (0.418), but were less likely to obtain a college or advanced degree; only 23.2 percent of veterans completed college compared with 40.4 percent of non-veterans. On the other hand, descriptive statistics show no significant difference in marital status or smoking-related morbidities between veterans and non-veterans.

Importantly, since the HRS is a study of older adults and we are observing veterans nearly 40 years after the Vietnam war, the exogeneity of the GRS and the random nature of the draft eligible and ineligible populations may be called into question if there is substantial attrition due to poor health or selective mortality in the Vietnam cohort. For example, potential problems could arise if conscripts who survived to be genotyped in 2006 or 2008 have lower GRSs or are more likely to be educated and therefore have lower mortality rates than comparable populations.

⁶ Specifically, we regress veteran status on draft eligibility and a constant with controls for month of birth. Men born between 1948 and 1952 were 15.7 percentage points more likely to serve, while men born between 1950 and 1952 were 15.3 percentage points more likely to serve. Results are available from the authors upon request.

In both cases, if selective mortality is correlated with the GRS or with educational attainment, draft eligibility is no longer exogenous, and the exclusion restriction on which our 2SLS estimates rest will be violated.

To explore this further, we compare descriptive statistics for the draft eligible and ineligible populations. The results in Table 2 reveal no significant differences at the mean between draft eligible and ineligible men in the HRS by GRS, smoking status, health, demographics, or education. This suggests selection bias is modest if nonexistent in the aging Vietnam veteran sample—a finding that is supported in the literature. Conley and Heerwig (2012) find no effect of draft exposure on mortality, including for cause-specific death rates, in a larger sample of national death records, and Angrist et al. (2010) find little evidence of elevated mortality among draft-eligible men in the 2000 census. Importantly, Conley and Heerwig (2012) do find some evidence of elevated mortality among draft-eligible, college-educated men. However, they argue this effect works in the opposite direction of putative education-enhancing effects that could potentially violate the exclusion restriction in IV regression models—i.e. high SES men were not more likely to survive.

Finally, Table 3 compares descriptive statistics for white men born between 1948-1952 who provided genetic data (N=631) with same-age peers who were not genotyped (N=540). By and large there are few differences between the two groups. Genotyped men report smoking 2.4 more cigarettes per day, are 5.1 percent less likely to be Hispanic, and are 4.1 percent more likely to report being married or partnered than non-genotyped men.

4. Methodology

To properly test and identify GxE interaction effects, we incorporate the latest approaches from population genetics into an IV framework to ensure both G and E are independent of each other. Existing efforts to find associations between genetic variation and social behavior in large, multidisciplinary surveys are often unable to support causal inferences because they use endogenous measures of the environment, genotype, or both (Conley, 2009; Fletcher & Conley, 2013). In the case of military service, the descriptive statistics in Table 2 reveal selection into the military is far from random, and likely to be correlated with factors like socioeconomic background, prior health status, or genotype. Given these differences, it would be impossible to sort out the effects of military service from the effects of other GxE or gene-gene (GxG) interactions in a model that uses self-reported veteran status to estimate GxE effects.

In addition, because smoking behavior is highly polygenic, or reflects small contributions from many genetic loci, the use of single genetic variants in our GxE model may result in a form of omitted variable bias, whereby crucial genetic information is obscured and left sitting in the error term, confounding estimates. To yield a comprehensive measure of genetic risk, we calculate a genetic risk score (GRS) that sums risk alleles across independently segregating genotyped single nucleotide polymorphisms (SNPs). This condenses the amount of genetic information available for an individual into a single, quantitative measure of genetic risk, thus minimizing the possibility that “G” is acting as a proxy for other GxG, ExE, or GxE interactions while also maximizing statistical power. This approach also has the added advantage of using main effect analysis already extant in the literature, which benefits from large consortia of adequately powered data to efficiently estimate individual allelic effects.

4.1 *Genetic Risk Score*

We calculate linear polygenic scores for the HRS sample based on results from a recent GWAS meta-analysis conducted across 16 studies by the Tobacco and Genetics (TAG) Consortium among people of European ancestry (Furberg et al., 2010). Specifically, we construct a composite “risk score” or genetic risk score (GRS) composed of weighted effects of specific single nucleotide polymorphisms (SNPs) for both smoking initiation (ever versus current smoker) and smoking quantity (number of cigarettes per day or CPD) using the HRS Genotype Data Version 1 (2006-2008 samples).⁷ We analyzed original genotype data to construct the score—i.e. imputed data was not analyzed. SNPs in the HRS genetic database were matched to SNPs with reported results in the GWAS. In the HRS genetic data set, 802,191 SNPs were available to construct the smoking initiation score and 802,903 SNPs were available to construct the CPD score. The matched set of SNPs was “pruned” to account for linkage disequilibrium (LD) using the clumping procedure (which considers the level of association between the SNP and the phenotype, not simply LD) in the second-generation PLINK software (Chang, Chow, Tellier, Vattikuti, Purcell, & Lee, 2015; Purcell & Chang, 2014).⁸ For each of these SNPs, a loading was calculated as the number of smoking-associated alleles multiplied by the effect-size estimated in the original GWAS. We do not impose a p-value threshold; instead, SNPs with relatively large p-values or small effects are down weighted in the composite score. Loadings are then summed across the SNP set to calculate the polygenic score. After LD

⁷ Genotyping was performed on the HRS sample using the Illumina Human Omni-2.5 Quad beadchip (HumanOmni2.5-4v1 array). The median call-rate for the 2006-2008 samples is 99.7%.

⁸ Clumping takes place in two steps. The first pass is done in fairly narrow windows (250kb) for all SNPs (the p-value significance thresholds for both index and secondary SNPs is set to 1) with a liberal LD threshold ($r^2=0.5$). In a second pass, SNPs remaining after the first prune are again pruned in broader windows (5000kb) but with a more conservative LD threshold ($r^2=0.2$).

pruning/clumping was applied, 178,834 SNPs were used to generate the smoking initiation score and 179,245 SNPs were used to generate the CPD score.

The GRSs are standardized to have a mean of 0 and a standard deviation of 1 for the population of white males born between 1948 and 1952. Both the smoking initiation and smoking quantity scores are predictive of our two smoking phenotypes (cigarettes per day and ever smoker) in base, non-interactive main effects models for the entire sample of white, genotyped males in the HRS (see Table 4). However, in our sample of white males born between 1948 and 1952, the smoking quantity or CPD score is insignificant. Thus, we only report results from the GxE interaction analysis that use the GRS for smoking initiation.⁹

To control for confounding by population stratification—or the non-random distribution of genes across populations—we use principal components. The principal components measure the uncorrelated variation or dimensions in the data, accounting for ethnic or racial differences in genetic structures within populations that could bias estimates. We calculate the principal components from the entire sample of genotyped respondents in the HRS, and include the first 20 in our regression analysis—a dimensionality that has generally proven adequate in the literature (Price et al., 2006). Controlling for the first 20 principal components accounts for any systematic differences in ancestry that can cause spurious correlations while also maximizing the power that is needed to detect true associations.

To test the sensitivity of the genetic score to the environmental effect we regress smoking status on the smoking initiation GRS in the draft eligible and ineligible populations, and find preliminary evidence of a GxE effect (see Table 5); the smoking initiation score is positive and predictive of all four phenotypes in the draft eligible population, whereas the impact of the GRS is insignificant and considerably reduced in magnitude in the draft ineligible population. Yet,

⁹ GxE interaction models with the CPD GRS are available from the authors upon request.

commensurate with previous findings, we are unable to detect a main effect of our environmental exposure, or instrumented veteran status, on smoking behavior and health in the base 2SLS model (see Table 6). The significant gradient between the genotype score and smoking behavior in the draft eligible population may explain the absence of an average treatment effect—i.e. other studies may have been unable to detect a main effect precisely because outcomes vary considerably by genotype within the draft eligible population.

4.2 *Empirical Model*

To account for selectivity issues due to endogenous sorting into the military or unobserved heterogeneity that may bias results, we use an instrumental variables (IV) approach to properly identify GxE effects. Following prior studies, we use the Vietnam-era draft lottery as an instrument for veteran status to identify the causal effect of military service and genotype on veterans' smoking behavior and health after age 50. Since past research has shown draft eligibility is 1) highly correlated with actual veteran status, and 2) only affects outcomes through the first stage channel, or through its correlation with veteran status, it is considered a valid instrument for military service.¹⁰ Consider the following GxE interaction model:

¹⁰ The first condition is easy to verify, and standard first stage statistics (partial R^2 and F-statistic) for the significance of the instruments in the HRS sample show draft eligibility and its interaction with the GRS are robust predictors of veteran status and its interaction with the GRS (tables are available upon request). The exclusion restriction cannot directly be tested. In this study, a violation of the exclusion restriction could occur if the stress of having a low draft number triggered smoking behavior. Heckman (1997) shows the IV estimator is not consistent if heterogeneous behavioral responses to the treatment—or military service in this case—are correlated with the instrument (i.e. draft eligibility). However, past research has provided convincing counterfactuals that suggest the exclusion restriction holds. For example, Angrist (1990) found no significant relationship between earnings and draft eligibility status for men born in 1953 (where draft eligibility was defined using the 1952 lottery cutoff of 95). Since the 1953 cohort was assigned RSNs but never called to service, if the draft affected outcomes directly, we would expect outcomes to vary by draft eligibility for this cohort.

$$(1) \quad Y_i = \alpha_1 VET_i + \alpha_2 GRS_i \times VET_i + X_i' \alpha_3 + \varepsilon_i$$

Where, VET_i is coded as “1” if individual i reports serving in the military and “0” otherwise, GRS_i is the genetic risk score for smoking initiation, $GRS_i \times VET_i$ is their interaction, Y_i is the outcome of interest, X_i is a vector of exogenous controls, and ε_i is the disturbance term. The VET_i and $GRS_i \times VET_i$ terms are treated as endogenous and the $DRAFT_i$ and $GRS_i \times DRAFT_i$ terms are used as excluded instruments in a two-stage least squares (2SLS) IV framework, where $DRAFT_i$ is coded as “1” for draft eligible men and “0” for draft ineligible men according to the eligibility cutoffs reported in Table 1. For our GxE interaction model, the first-stage regressions are as follows:

$$(2) \quad VET_i = \pi_1 DRAFT_i + \pi_2 GRS_i \times DRAFT_i + X_i' \pi_3 + \eta_i$$

$$(3) \quad GRS_i \times VET_i = \gamma_1 DRAFT_i + \gamma_2 GRS_i \times DRAFT_i + X_i' \gamma_3 + \rho_i$$

where the model is exactly identified. The vector of exogenous controls (X_i) includes the GRS effect for draft ineligible non-veterans, the first 20 principal components to account for population stratification in the genotype data, and dummies for year and month of birth.¹¹ To increase precision in the first stage estimates, or to strengthen the correlation between the $GRS_i \times VET_i$ and $GRS_i \times DRAFT_i$ terms, we model the main effect of the GRS in X_i as $GRS_i \times NODRAFT_i$, where $NODRAFT_i$ is equal to “1” if individuals were not drafted and never served

¹¹ A mechanical failure in the implementation of the first round of the lottery (balls with the days of the year were not mixed sufficiently after having been dumped in a month at a time) resulted in a disproportionately high probability of being drafted for those born in the last few months of the year (Fienberg, 1971). This could bias estimates if those born later in the year differ in important ways from those born at other times during the year. For example, studies have shown health varies with season of birth.

in the military.¹² The first stage equations are then substituted into the structural equation to derive the reduced form in the second stage:

$$(4) \quad Y_i = \delta_1 DRAFT_i + \delta_2 GRS_i \times DRAFT_i + X_i' \delta_3 + \xi_i$$

For all limited and continuous dependent variables in this study, the second stage equation is estimated with a simple linear probability model because it is the ideal specification when faced with a set of simultaneous equations where the instrument, the endogenous regressor, and the dependent variable take on a limited set of values (Angrist & Pischke, 2008). As long as $DRAFT_i$ is exogenous and only affects the outcome through the first stage channel, or through its correlation with VET_i , we avoid any potential confounders, and the coefficient on δ_1 can be interpreted as the local average treatment effect (LATE) of military status, or the marginal effect of veteran status at the mean GRS.¹³ Similarly, because genetic inheritance is by and large the result of a random shuffling of paternal and maternal genotypes at conception, GRS_i is exogenous, and the coefficient on the GxE interaction term (δ_2) can be interpreted as the marginal effect of veteran status by genotype. Therefore, a statistically significant coefficient on

¹² In a typical linear regression model with an interaction term, the interaction term and each of the corresponding main effects are included as separate terms (e.g. “draft”, “GRS x draft” and “GRS”). Here, because we are using a 2SLS approach, and the “GRS” term is highly correlated with “GRS x draft”, we model the main effect of the GRS as “nodraft x GRS” to strengthen the correlation between the “GRS x veteran” and the “GRS x draft” terms in the first stage. Using the GxE interaction term for draft ineligible non-veterans instead of the main effect of the GRS does not change the meaning of this term, which can still be interpreted as the marginal effect for men who were not drafted and who did not serve. However, it does change the interpretation of δ_2 , which now represents the marginal effect for draft eligible veterans instead of the difference between the marginal effects for draft eligible veterans and draft ineligible non-veterans.

¹³ The IV estimates of effects of military service using the draft lottery capture the effect of military service on “compliers”, or men who served because they were draft eligible but who would not otherwise have served. It is not, therefore, an estimate of the effect of military service on men who volunteered. See Angrist and Pischke (2008) for a more detailed discussion of the interpretation of the LATE for the Vietnam-era draft.

δ_2 would indicate the effect of veteran status on smoking behavior varies by genotype, or the existence of a synergistic relationship between genotype and military service on the phenotypic outcome of interest at higher (or lower) values of the GRS. In this way, the model allows us to estimate the effects of military service across the entire distribution of genetic risk for smoking initiation.

5. Results

5.1 *GxE Estimates for Smoking Behavior and Smoking-Related Morbidities*

Our main results are presented in Table 7, which reports OLS and 2SLS estimates from the GxE interaction model. Results from the OLS models fail to find a significant GxE interaction effect among veterans. However, since all omitted genetic and environmental variables or pathways between self-reported veteran status and smoking behavior are not accounted for in this model, we cannot rule out the possibility that the naïve estimates are biased downward due to a host of other unobserved GxE, GxG, or ExE interactions.

Turning to the 2SLS models, we present the causal impact of veteran status on smoking behavior and smoking-related morbidities across the entire spectrum of genetic risk for smoking. 2SLS estimates of the “Veteran” coefficient capture the effect of conscription on our outcomes of interest at the mean GRS and the “GRS x Veteran” coefficient captures the effect of compulsory military service at all other values of the GRS. The “GRS x Non-Veteran” coefficient is the effect of the GRS on men who were not treated, or those who were not drafted and who did not serve in the military. The results indicate the presence of a statistically significant GxE interaction between Vietnam-era military service and the GRS for current or

former smoker status (“Ever Smoker”) and smoking quantity (“Cigarettes Per Day”). A one standard deviation increase in the GRS of draft-induced soldiers results in a 16 percent point increase in the likelihood of having ever been a smoker and increases quantity smoked per day by approximately 8 cigarettes. With respect to smoking-related morbidities, we find some evidence that veterans have a higher lifetime risk of being diagnosed with cancer and are more susceptible to hypertension at older ages. A one standard deviation increase in the GRS increases the risk of being diagnosed with cancer by 8.46 percent and elevates mean arterial pressure by 4.44 units.

Given our low sample size for detecting genetic associations, we conducted power analysis to evaluate whether the GxE coefficients (δ_2) reported in Table 7 are underpowered.¹⁴ Since our GRS is predictive of smoking behavior, effect sizes for δ_2 are slightly larger in the CPD (0.0161) and ever smoker (0.0088) models than they are in the more indirect, smoking-related morbidity models for cancer (0.0054), and MAP (0.0078).¹⁵ Using these effect sizes and the standard benchmark of 80% power ($\pi = 0.80$), the statistical power—or the probability of identifying a statistically significant GxE effect under the alternative hypothesis that a true association exists—is found to be adequate for the CPD model ($\pi = 0.89$). Not surprisingly, given their lower effect sizes, the ever smoker ($\pi = 0.656$), cancer ($\pi = 0.455$) and MAP ($\pi = 0.593$) models are underpowered by conventional standards, indicating further GxE

¹⁴ Post hoc power analysis was conducted for the GxE coefficients using the software package G*Power (Faul, Erdfelder, Buchner & Lang, 2009). The sample size of 631 was used for statistical power analysis on a multivariate linear regression equation with 83 predictors at the conventional, two-tailed 0.05 significance level ($\alpha = 0.05$).

¹⁵ The partial R^2 or Cohen’s effect size for the GxE coefficient was estimated from an OLS regression of Equation 4. To minimize any potential bias in the estimation of the effect sizes due to the endogeneity of self-reported veteran status, we model the GxE interaction terms in the OLS model using draft eligibility status instead of self-reported veteran status.

analysis on a larger sample of veterans would buttress the robustness of the findings we report here.

Across the board, we find no evidence that military service impacted the smoking behavior or health of conscripts with an average genetic predisposition for smoking—i.e. our findings are commensurate with other studies that are unable to identify a LATE of military service on smoking behavior and health (Dobkin & Shabani, 2009; Eisenberg & Rowe, 2009). These results indicate prior studies may in part have been unable to detect an overall LATE precisely because outcomes vary considerably by genotype. To illustrate this further, the first two rows in Table 8 estimate the total treatment effect of military service on smoking behavior for high-risk genotypes by taking a post-estimation linear combination of the marginal effects for veterans with GRSs one and two standard deviations above the mean.¹⁶ Veterans with a GRS one standard deviation above the mean are 56.9 percent more likely to have ever been smokers. Moreover, during their heaviest smoking years, these veterans smoked almost 18 more cigarettes per day than non-veterans with the same GRS—an effect that grows as we move up the genotype distribution. Veterans with a GRS two standard deviations above the mean were 70.5 percent more likely to become smokers and smoked more than a pack per day—or 27 cigarettes—at their peak compared to non-veterans with the same GRS.

The sizable variation in outcomes by genotype can be seen in Figure 1, which plots the total difference in predicted cigarettes smoked per day for veterans versus non-veterans at all values of the smoking initiation GRS with 95 percent confidence intervals. A strong treatment

¹⁶ Specifically, we estimate the total difference between veterans and non-veterans, or the total treatment effect, by adding the individual treatment effects for veterans (“Veteran” + “GRS x Veteran”) and subtracting them from the marginal treatment effect for draft ineligible non-veterans (“GRS x Non-Veteran”). To ensure the accuracy of the standard errors, this is done using a post-estimation linear combination.

effect persists for veterans with a GRS at approximately one standard deviation or higher, whereas the total treatment effect below one standard deviation is not statistically significant. Finally, total treatment effects are significant for the smoking behavior phenotypes only, or the GxE effect on smoking-related morbidities is only significant at the margin.

5.2 *The Influence of Post-Service Educational Attainment on Smoking*

To investigate the possibility that increases in smoking behavior due to military service were offset by post-service schooling gains from veterans' use of the GI Bill, we also estimate separate 2SLS models that include a control variable for education (coded as "1" if the respondent reports obtaining a college or advanced degree, and "0" otherwise) to calculate total effects by educational attainment (see Table 8). Education may be endogenous in the smoking model if it is correlated with omitted variables that influence smoking behavior, such as socioeconomic status or ability. However, if extra schooling is a downstream effect of military service (i.e. via the GI Bill) it may be a second mechanism through which draft eligibility influenced smoking behavior. Given the high correlation between educational attainment and health-enhancing behaviors in the literature, post-service educational attainment may have acted as a second environmental lever that attenuated any initial shocks to smoking behavior in young adulthood—particularly for high-risk genotypes.¹⁷

¹⁷ The exogenous downstream effect of educational attainment might be compromised if the high correlation between draft eligibility and schooling is a reflection of draft avoidance behavior rather than military service—i.e. if it is capturing the effect of men with low draft lottery numbers who "beat the draft" by obtaining educational deferments. Angrist and Chen (2011) find small but statistically significant positive effects of service on educational attainment for white men born between 1948-1952. Weighing this against the sharp decline in educational deferments during the draft lottery period, they argue there is little evidence to support the claim that increases in schooling among draft eligible men are due to draft avoidance behavior.

Consistent with our hypothesis, the statistically significant difference in smoking behavior between high-risk veterans and non-veterans completely disappears for college-educated men, while a strong, adverse effect persists for veterans who did not obtain a college degree. Veterans with a GRS one standard deviation above the mean who are not college educated were 66.8 percent more likely to become smokers and report smoking almost 22 cigarettes more per day than comparable non-veterans. Two standard deviations above the mean, high-risk veterans without a college degree were 80.8 percent more likely to become smokers and smoked more than a pack and a half per day (31 cigarettes) than high-risk non-veterans. The variation in outcomes by educational attainment can be seen in Figures 2 and 3, which plot the total effects from our CPD GxE model at all values of the GRS for veterans who did and did not obtain a college degree. GxE effects are significant at the 0.05 level for non-college educated men with GRSs at or above 0.4, whereas the results for college educated men are statistically significant for men at extreme risk only (i.e. GRSs greater than or equal to 3.4). These results indicate post-service education had a considerable moderating influence on the long-term smoking behaviors of genetically-at-risk men who were called to service.

6. Discussion

Tobacco is the leading cause of preventable death among civilian populations and continues to be a significant and pressing health problem within the US military. The military spends approximately \$1.6 billion a year on tobacco-related health care and lost productivity (Wedge & Bondurant, 2009), and smoking-attributable disease accounts for 16% of all deaths and approximately 10% of all hospital bed days (Helyer, Brehm, & Perino, 1998). Using a polygenic score to measure genetic risk and an instrumental variables approach to operationalize

the environmental exposure, this study finds evidence that military-induced smoking during the Vietnam era amplified polygenic risk for cigarette consumption. Vietnam veterans with a GRS one to two standard deviations above the mean were approximately 57 to 71 percent more likely to be smokers and smoked 18 to 27 more cigarettes per day more than non-veterans with the same GRS. On the other hand, we do not find a significant treatment effect for conscripts with GRSs at or below the mean. Heterogeneous response—by genotype—might explain why past studies that have used the draft lotteries to investigate smoking behavior have been unable to detect a LATE.

Limitations to these analyses should be mentioned. A significant drawback of our IV estimation strategy, and a common criticism of the natural experiment approach to IV estimation in general, is we cannot fully spell out the underlying theoretical relationships between military service and smoking behavior (Angrist & Krueger, 2001). For example, we cannot unravel the influence of specific mechanisms surrounding time in Vietnam—such as combat positions, year of arrival, or number of tours—on cigarette consumption, making it difficult to identify what aspects of the war experience intensified smoking behavior. In other words, there is nothing in our IV model that explains why Vietnam-era service affected smoking behavior.

However, due to endogeneity issues, current methods being used to uncover GxE interactions in observational studies are inadequate to support policy inference. Although we caution that estimates from this study apply specifically to veterans from the Vietnam era, and are therefore not externally valid or cannot be generalized to the population as a whole or even to present-era military personnel, their high degree of internal validity may direct practitioners to effective treatments for high-risk individuals. Thus, while inducing a military draft lottery, for example, would not be an intervention to promote public health (nor would wholesale

decommissioning of military personnel be politically feasible), to the extent that the Vietnam-era draft lottery serves as a proxy for stressful events in young adulthood, or exposure to combat, policymakers may want to design interventions to minimize similar traumatic events. In particular, the strong impact of college education on the attenuation of smoking behavior in adulthood signals educational attainment may be an important pathway for mitigating other potentially harmful GxE experiences at younger ages.

In addition, given the relatively small effect size between our GRS for smoking behavior and our phenotypes of interest, this study would benefit from the added power of a larger sample—particularly for the indirect smoking-related morbidities we report. For example, future research that exploits quasi-natural experiments that that can be operationalized in the UK Biobank or larger genotyped populations, such as variation in tobacco taxes, are needed to detect GxE associations between smoking behavior and longer-term health problems. Yet, despite our low sample size, we are able to detect adequately powered effects of military service on CPD or smoking quantity, underscoring the importance of using polygenic scores or genetic associations that benefit from prior analysis of large, well-powered consortia to estimate individual allelic effects.

More generally, we think this study provides an example for a way forward by which genotype can serve as the moderating prism to filter out discernable heterogeneity in treatment effects. It is primarily in this way, we believe, that genetics and social science will be fruitfully integrated. It is important for other scholars to not only be careful about estimating environmental factors using econometric techniques (or randomized controlled trials) that insure the exogeneity of the environmental regime in question but to also deploy adequate safeguards against population stratification. That is, where possible researchers should deploy within-

family models to guarantee that genotype is randomly assigned or—absent pedigreed data—
should control for principle components in cross-family models as we do here.

Works Cited

- Angrist, J. D. (1990). Lifetime earnings and the Vietnam-era draft lottery: Evidence from Social Security Administrative Records. *The American Economic Review*, 313-336.
- Angrist, J. D., & Chen, S. H. (2011). Schooling and the Vietnam-era GI Bill: Evidence from the draft lottery. *American Economic Journal: Applied Economics*, 3(2), 96-118.
- Angrist, J. D., Chen, S. H., & Frandsen, B. R. (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), 824-837.
- Angrist, J. D., Chen, S. H., & Song, J. (2011). Long-term consequences of Vietnam-era conscription: New estimates using Social Security data. *The American Economic Review*, 101(3), 334-338.
- Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *The Journal of Economic Perspectives*, 15(4), 69-85.
- Angrist, J. D., & Pischke, J. S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*: Princeton University Press.
- Bedard, K., & Deschênes, O. (2006). The long-term impact of military service on health: Evidence from World War II and Korean War veterans. *The American Economic Review*, 176-194.
- Belsky, D. W., Moffitt, T. E., Baker, T. B., Biddle, A. K., Evans, J. P., Harrington, H. L., . . . Williams, B. (2013). Polygenic risk and the developmental progression to heavy, persistent smoking and nicotine dependence: Evidence from a 4-decade longitudinal study. *JAMA Psychiatry*, 70(5), 534-542.
- Benetos, A., Safar, M., Rudnichi, A., Smulyan, H., Richard, J. L., Ducimetière, P., & Guize, L. (1997). Pulse pressure a predictor of long-term cardiovascular mortality in a French male population. *Hypertension*, 30(6), 1410-1415.
- Boardman, J. D., Saint O., Jarron M., Haberstick, B. C., Timberlake, D. S., & Hewitt, J. K. (2008). Do schools moderate the genetic determinants of smoking? *Behavior Genetics*, 38(3), 234-246.
- Bound, J., & Turner, S. (2002). Going to war and going to college: Did World War II and the GI Bill increase educational attainment for returning veterans? *Journal of Labor Economics*, 20(4), 784-815.
- Card, D., & Lemieux, T. (2001). Going to college to avoid the draft: The unintended legacy of the Vietnam War. *American Economic Review*, 97-102.
- Centers for Disease Control & Prevention. (1994). Surveillance for selected tobacco-use behaviors--United States, 1900-1994. *MMWR. Morbidity and Mortality Weekly Report*, 43(3), 1-43.
- Centers for Disease Control & Prevention. (2003). Prevalence of current cigarette smoking among adults and changes in prevalence of current and some day smoking--United States, 1996-2001. *MMWR. Morbidity and Mortality Weekly Report*, 52(14), 303.
- Centers for Disease Control & Prevention. (2008). Smoking-attributable mortality, years of potential life lost, and productivity losses--United States, 2000-2004. *MMWR. Morbidity and Mortality Weekly Report*, 57(45), 1226.

- Chang C.C., Chow C.C., Tellier LCAM, Vattikuti S., Purcell S.M., Lee J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4.
- Conley, D. (2009). The promise and challenges of incorporating genetic data into longitudinal social science surveys and research. *Biodemography and Social Biology*, 55(2), 238-251.
- Conley, D., & Heerwig, J. A. (2011). The war at home: Effects of Vietnam-era military service on postwar household stability. *American Economic Review*, 101(3), 350-354.
- Conley, D., & Heerwig, J. A. (2012). The long-term effects of military conscription on mortality: Estimates from the Vietnam-era draft lottery. *Demography*, 49(3), 841-855.
- Daw, J., Shanahan, M., Harris, K. M., Smolen, A., Haberstick, B., & Boardman, J. D. (2013). Genetic sensitivity to peer behaviors 5HTTLPR, smoking, and alcohol consumption. *Journal of Health & Social Behavior*, 54(1), 92-108.
- 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.
- De Walque, D. (2010). Education, information, and smoking decisions evidence from smoking histories in the United States, 1940–2000. *Journal of Human Resources*, 45(3), 682-717.
- Dobkin, C., & Shabani, R. (2009). The health effects of military service: Evidence from the Vietnam draft. *Economic Inquiry*, 47(1), 69-80.
- Eisenberg, D., & Rowe, B. (2009, February). The effect of smoking in young adulthood on smoking later in life: Evidence based on the Vietnam era draft lottery. In *Forum for Health Economics & Policy* (Vol. 12, No. 2).
- Farrell, P., & Fuchs, V. R. (1982). Schooling and health: The cigarette connection. *Journal of Health Economics*, 1(3), 217-230.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149-1160.
- Fienberg, S. E. (1971). Randomization and social affairs: The 1970 draft lottery. *Science*, 171(3968), 255-261.
- Fletcher, J. M., & Conley, D. (2013). The challenge of causal inference in gene–environment interaction research: Leveraging research designs from the social sciences. *American Journal of Public Health*, 103(S1), S42-S45.
- Fortmann, Stephen P., Todd Rogers, Karen Vranizan, William L. Haskell, Douglas S. Solomon, & John W. Farquhar. "Indirect measures of cigarette use: expired-air carbon monoxide versus plasma thiocyanate." *Preventive Medicine* 13, no. 1 (1984): 127-135.
- Fuchs, V. R. (1982). Time preference and health: An exploratory study. *Economic Aspects of Health*, 93.
- Furberg, H., Kim, Y., Dackor, J., Boerwinkle, E., Franceschini, N., Ardissino, D., . . . Merlini, P. A. (2010). Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature Genetics*, 42(5), 441-447.
- Glassman, A. H., Helzer, J. E., Covey, L. S., Cottler, L. B., Stetner, F., Tipp, J. E., & Johnson, J. (1990). Smoking, smoking cessation, and major depression. *JAMA*, 264(12), 1546-1549.
- Grimard, F., & Parent, D. (2007). Education and smoking: Were Vietnam war draft avoiders also more likely to avoid smoking? *Journal of Health Economics*, 26(5), 896-926.
- Grossman, M. (1972). On the concept of health capital and the demand for health. *The Journal of Political Economy*, 80(2), 223-255.

- Grossman, M. (2006). Education and nonmarket outcomes. *Handbook of the Economics of Education, 1*, 577-633.
- Hatziandreu, Evridiki J., John P. Pierce, Michael C. Fiore, Verner Grise, Thomas E. Novotny, & Ronald M. Davis. "The reliability of self-reported cigarette consumption in the United States." *American Journal of Public Health* 79, no. 8 (1989): 1020-1023.
- Hearst, N., Newman, T. B., & Hulley, S. B. (1986). Delayed effects of the military draft on mortality: A randomized natural experiment. *The New England Journal of Medicine, 314*(10), 620.
- Heckman, J. (1997). Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. *Journal of Human Resources*, 441-462.
- Heerwig, J. A., & Conley, D. (2013). The causal effects of Vietnam-era military service on post-war family dynamics. *Social Science Research, 42*(2), 299-310.
- Helyer, A. J., Brehm, W. T., & Perino, L. (1998). Economic consequences of tobacco use for the Department of Defense, 1995. *Military Medicine, 163*(4), 217-221.
- Institute of Medicine. (1994). Veterans and agent orange: Health effects of herbicides used in Vietnam. Washington, DC: National Academy Press.
- Johnston, L. D., O'Malley, P. M., Bachman, J. G., & Schulenberg, J. E. (2005). Monitoring the Future: National Survey Results on Drug Use, 1975-2004. Volume II: College Students & Adults Ages 19-45, 2004. *National Institutes of Health*.
- Kandel, D. B., & Logan, J. A. (1984). Patterns of drug use from adolescence to young adulthood: Periods of risk for initiation, continued use, and discontinuation. *American Journal of Public Health, 74*(7), 660-666.
- Kendler, K. S., Gardner, C., Jacobson, K. C., Neale, M. C., & Prescott, C. A. (2005). Genetic and environmental influences on illicit drug use and tobacco use across birth cohorts. *Psychological Medicine, 35*(09), 1349-1356.
- Le Marchand, L., Derby, K. S., Murphy, S. E., Hecht, S. S., Hatsukami, D., Carmella, S. G., ... & Wang, H. (2008). Smokers with the CHRNA lung cancer-associated variants are exposed to higher levels of nicotine equivalents and a carcinogenic tobacco-specific nitrosamine. *Cancer Research, 68*(22), 9137-9140.
- Liu, J. Z., Tozzi, F., Waterworth, D. M., Pillai, S. G., Muglia, P., Middleton, L., . . . Waeber, G. (2010). Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nature Genetics, 42*(5), 436-440.
- Lleras-Muney, A. (2005). The relationship between education and adult mortality in the United States. *The Review of Economic Studies, 72*(1), 189-221.
- Maes, H. H., Woodard, C. E., Murrelle, L., Meyer, J. M., Silberg, J. L., Hewitt, J. K., . . . Carbonneau, R. (1999). Tobacco, alcohol and drug use in eight-to sixteen-year-old twins: The Virginia Twin Study of Adolescent Behavioral Development. *Journal of Studies on Alcohol and Drugs, 60*(3), 293.
- Merline, A. C., O'Malley, P. M., Schulenberg, J. E., Bachman, J. G., & Johnston, L. D. (2004). Substance use among adults 35 years of age: Prevalence, adulthood predictors, and impact of adolescent substance use. *American Journal of Public Health, 94*(1), 96-102.
- Patrick, Donald L., Allen Cheadle, Diane C. Thompson, Paula Diehr, Thomas Koepsell, and Susan Kinne. (1994). The validity of self-reported smoking: a review and meta-analysis. *American Journal of Public Health, 84*(7), 1086-1093.
- Purcell, S., & Chang, C. (2014). PLINK [Computer software]. <<https://www.cog-genomics.org/plink2>>.

- Rouse, B., Sanderson, C., & Feldmann, J. (2002). Results from the 2001 National Household Survey on Drug Abuse: Volume I. Summary of National Findings. *Rockville: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.*
- Slomkowski, C., Rende, R., Novak, S., Lloyd-Richardson, E., & Niaura, R. (2005). Sibling effects on smoking in adolescence: Evidence for social influence from a genetically informative design. *Addiction, 100*(4), 430-438.
- Spitz, M. R., Amos, C. I., Dong, Q., Lin, J., & Wu, X. (2008). The CHRNA5-A3 region on chromosome 15q24-25.1 is a risk factor both for nicotine dependence and for lung cancer. *Journal of the National Cancer Institute, 100*(21), 1552-1556.
- Stanley, M. (2003). College education and the midcentury GI Bills. *The Quarterly Journal of Economics, 118*(2), 671-708.
- Thorgeirsson, T. E., Geller, F., Sulem, P., Rafnar, T., Wiste, A., Magnusson, K. P., ... & Oskarsson, H. (2008). A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature, 452*(7187), 638-642.
- Thorgeirsson, T. E., Gudbjartsson, D. F., Surakka, I., Vink, J. M., Amin, N., Geller, F., . . . Walter, S. (2010). Sequence variants at CHRNA3-CHRNA6 and CYP2A6 affect smoking behavior. *Nature Genetics, 42*(5), 448-453.
- Timberlake, D. S., Rhee, S. H., Haberstick, B. C., Hopfer, C., Ehringer, M., Lessem, J. M., . . . Hewitt, J. K. (2006). The moderating effects of religiosity on the genetic and environmental determinants of smoking initiation. *Nicotine & Tobacco Research, 8*(1), 123-133.
- U.S. Department of Health and Human Services. (1989). Reducing the health consequences of smoking: 25 Years of Progress. A Report of the Surgeon General. U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control, Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. DHHS Publication No. (CDC) 89-8411.
- U.S. Department of Health and Human Services. (1990). A report of the surgeon general: The health benefits of smoking cessation. Washington D.C.
- U.S. Department of Veteran Affairs. (2015). Vietnam Veterans. < <http://www.benefits.va.gov/persona/veteran-vietnam.asp>>.
- U.S. Selective Service System. (2015). Induction Statistics. < <https://www.sss.gov/induct.htm>>.
- Wedge, R., & Bondurant, S. (Eds.). (2009). *Combating Tobacco Use in Military and Veteran Populations*. Washington D.C.: National Academies Press.
- Winefield, H.R., Winefield, A. H., Tiggemann, M., & Goldney, R.D. (1989). Psychological concomitants of tobacco and alcohol use in young Australian adults. *British Journal of Addiction, 84*(9), 1067-1073.

Table 2. Descriptive statistics by veteran status, 1948-1952 birth cohorts, white males

	Non- veteran	Veteran	t Statistic	Not Draft Eligible	Draft Eligible	t Statistic
Ever smoked GRS	0.023 (0.988)	-0.051 (1.028)	0.871	-0.130 (0.966)	-0.057 (0.93)	-0.960
Cigarettes per day GRS	0.178 (0.623)	0.467 (0.536)	-5.646**	0.249 (0.64)	0.293 (0.576)	-0.905
Ever smoker	0.577 (0.495)	0.692 (0.463)	-2.754**	0.596 (0.491)	0.635 (0.482)	-0.993
Current smoker	0.247 (0.432)	0.343 (0.476)	-2.517**	0.269 (0.444)	0.287 (0.453)	-0.499
Cigarettes per day	11.783 (17.263)	16.116 (16.852)	-2.948**	12.593 (16.919)	13.823 (17.635)	-0.891
Cancer	0.088 (0.393)	0.167 (0.619)	-1.461	0.115 (0.478)	0.110 (0.476)	0.360
Mean arterial pressure	99.843 (12.640)	98.246 (11.629)	1.491	99.723 (12.729)	98.879 (11.861)	0.845
Hispanic	0.171 (0.377)	0.045 (0.209)	4.385**	0.140 (0.348)	0.121 (0.326)	0.732
Married	0.866 (0.341)	0.864 (0.344)	0.082	0.871 (0.336)	0.858 (0.350)	0.471
No degree	0.143 (0.351)	0.040 (0.197)	3.854**	0.120 (0.326)	0.099 (0.299)	0.836
High school diploma	0.418 (0.494)	0.606 (0.490)	-4.450**	0.450 (0.498)	0.511 (0.500)	-1.520
Associates degree	0.035 (0.183)	0.121 (0.327)	-4.243**	0.069 (0.253)	0.053 (0.225)	0.807
College or adv. degree	0.404 (0.491)	0.232 (0.423)	4.252**	0.361 (0.481)	0.337 (0.473)	0.632
N	433	198		349	282	

Notes: Estimated from the Health and Retirement Study (HRS). Standard deviations are reported in parentheses. * $p < 0.05$, ** $p < 0.01$.

Table 3. Descriptive statistics by genotyped status, 1948-1952 birth cohorts, white males

	All	Not genotyped	Genotyped	t Statistic
Ever smoker	0.633 (0.482)	0.657 (0.475)	0.613 (0.487)	1.538
Current smoker	0.275 (0.447)	0.273 (0.446)	0.277 (0.448)	-0.176
Cigarettes per day	12.032 (17.231)	10.733 (17.145)	13.143 (17.24)	-2.390*
Cancer	0.077 (0.266)	0.07 (0.256)	0.113 (0.476)	-0.771
Mean arterial pressure	99.696 (12.405)	101.144 (12.577)	99.348 (12.340)	1.588
Hispanic	0.155 (0.362)	0.183 (0.387)	0.132 (0.338)	2.443*
Married	0.846 (0.361)	0.824 (0.381)	0.865 (0.342)	-1.951*
No degree	0.121 (0.327)	0.133 (0.34)	0.111 (0.314)	1.170
High school diploma	0.481 (0.5)	0.485 (0.5)	0.477 (0.499)	0.279
Associates degree	0.063 (0.243)	0.065 (0.246)	0.062 (0.241)	0.211
College or adv. degree	0.335 (0.472)	0.317 (0.466)	0.350 (0.477)	-1.213
Ever smoked GRS			0 (1.00)	
Cigarettes per day GRS			0 (1.00)	
N	1171	540	631	

Notes: Estimated from the Health and Retirement Study (HRS). Genetic risk scores are standardized. Standard deviations are reported in parentheses.

* $p < 0.05$, ** $p < 0.01$.

Table 4. Main effect of smoking genetic risk scores on smoking behavior and smoking-related morbidity, white males

Smoking Initiation GRS	Cigarettes Per Day	Ever smoker	Cancer	Mean arterial pressure
<i>All white males</i>	0.984*** (0.335)	0.0341*** (0.00756)	0.00532 (0.00679)	0.0321 (0.217)
N	4297	4276	4297	4243
R ²	0.0144	0.00957	0.00806	0.00718
<i>Veteran sample</i>	1.756** (0.722)	0.0447** (0.0204)	0.00670 (0.0117)	0.826 (0.523)
N	631	631	631	620
R ²	0.0506	0.0548	0.0232	0.0385

Cigarettes Per Day GRS	Cigarettes Per Day	Ever smoker	Cancer	Mean arterial pressure
<i>All white males</i>	1.090*** (0.350)	0.0211*** (0.00792)	-0.0102 (0.00710)	-0.0276 (0.227)
N	4297	4276	4297	4243
R ²	0.0146	0.00649	0.00840	0.00717
<i>Veteran sample</i>	1.077 (0.773)	0.0163 (0.0218)	0.00519 (0.0125)	-0.258 (0.566)
N	631	631	631	620
R ²	0.0444	0.0482	0.0230	0.0349

Notes: Each column reports a separate ordinary least squares (OLS) regression of the dependent variable on a constant and the smoking GRS. All regressions include controls for population stratification in the genotype data and Hispanic ethnicity. Genetic risk scores are standardized. Robust standard errors are in parentheses.

* p<0.10, ** p<0.05, *** p<0.01.

Table 5. Main effect of smoking initiation genetic risk score on smoking behavior and smoking-related morbidities by draft eligibility, 1948-1952 birth cohorts, white males

	Cigarettes Per Day	Ever smoker	Cancer	Mean arterial pressure
<i>Draft Eligible</i>	3.339*** (1.157)	0.0688** (0.0300)	0.0308* (0.0175)	1.871** (0.811)
N	282	282	282	275
R ²	0.133	0.170	0.123	0.127
<i>Draft Ineligible</i>	0.283 (0.973)	0.0173 (0.0283)	-0.0106 (0.0169)	0.304 (0.825)
N	349	349	349	345
R ²	0.0392	0.0615	0.0466	0.0374

Notes: Each column reports a separate ordinary least squares (OLS) regression of the dependent variable on a constant and the smoking GRS. All regressions include controls for population stratification in the genotype data and Hispanic ethnicity. Genetic risk score is standardized. Robust standard errors are in parentheses.

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

Table 6. OLS and 2SLS estimates of veteran status on smoking behavior and smoking-related health outcomes, 1948-1952 birth cohorts, white males

Dependent variable	OLS	2SLS
Cigarettes per day	5.230*** (1.605)	8.339 (9.621)
Ever smoker	0.147*** (0.0459)	0.424 (0.283)
Cancer	0.0424 (0.0264)	-0.0180 (0.158)
Mean arterial pressure	-1.666 (1.210)	-4.858 (7.092)

Notes: All regressions control for respondent month and year of birth, population stratification in the genotype data, and Hispanic ethnicity. Genetic risk score is standardized. The sample size is 631, except for mean arterial pressure (N=620). Robust standard errors are in parentheses.

Table 7. OLS and 2SLS estimates of military service and smoking genotype on smoking behavior and smoking-related morbidities, 1948-1952 birth cohorts, white males

	Cigarettes Per Day		Ever Smoker		Cancer		Mean Arterial Pressure	
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Veteran	5.176*** (1.590)	9.132 (9.518)	0.142*** (0.0450)	0.433 (0.285)	0.0439 (0.0285)	-0.00812 (0.154)	-1.689 (1.159)	-4.531 (7.303)
GRS x Veteran	1.982* (1.187)	8.194*** (2.874)	0.0220 (0.0332)	0.161** (0.0817)	0.0279 (0.0192)	0.0846* (0.0432)	1.201 (0.950)	4.438** (2.049)
GRS x Non-Veteran	1.173 (0.947)	-0.660 (1.075)	0.0540** (0.0270)	0.0249 (0.0351)	-0.00602 (0.0159)	-0.0165 (0.0175)	0.719 (0.723)	0.288 (0.986)
N	631		631		631		620	

Notes: Estimated from the Health and Retirement Study. IV estimates of the "GRS x Veteran" coefficient capture the effect of military service on men who served because they were draft eligible, while the "GRS x Non-Veteran" coefficient captures the effect of the GRS on draft ineligible non-veterans. All regressions control for respondent month and year of birth, population stratification in the genotype data, and Hispanic ethnicity. Genetic risk score is standardized. Robust standard errors are in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table 8. Total treatment effects for high-risk genotypes, 1948-1952 birth cohorts, white males

	Cigarettes Per Day	Ever Smoke	Cancer	Mean Arterial Pressure
Total Veteran Effect (GRS=1)	17.99* (9.881)	0.569** (0.287)	0.0930 (0.169)	-0.380 (7.679)
Total Veteran Effect (GRS=2)	26.84** (11.06)	0.705** (0.313)	0.194 (0.194)	3.770 (8.605)
<i>By educational attainment:</i>				
Total Veteran Effect (GRS=1 & College=1)	10.29 (12.95)	0.378 (0.377)	0.0779 (0.222)	-3.905 (9.917)
Total Veteran Effect (GRS=1 & College=0)	21.96*** (8.440)	0.668*** (0.242)	0.101 (0.144)	1.482 (6.657)
Total Veteran Effect (GRS=2 & College=1)	19.32 (13.79)	0.519 (0.395)	0.179 (0.242)	0.279 (10.63)
Total Veteran Effect (GRS=2 & College=0)	30.99*** (9.785)	0.808*** (0.273)	0.202 (0.171)	5.666 (7.693)
N	631	631	631	620

Notes: The total treatment effect of military service is estimated by adding the marginal treatment effects for veterans ("Veteran" + "GRS x Veteran") from the IV regressions in Table 6 and subtracting them from the marginal treatment effect for non-veterans who were not drafted ("Non-Veteran x GRS"). To ensure the accuracy of the standard errors, this is done using a post-estimation linear combination. The same technique is used to estimate the total effect of military service for veterans who did and did not obtain a college or advanced degree using results from IV regressions that include a control for college education. Robust standard errors are in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Figure 1. Difference in Predicted Cigarettes Per Day
Veterans vs. Non-Veterans

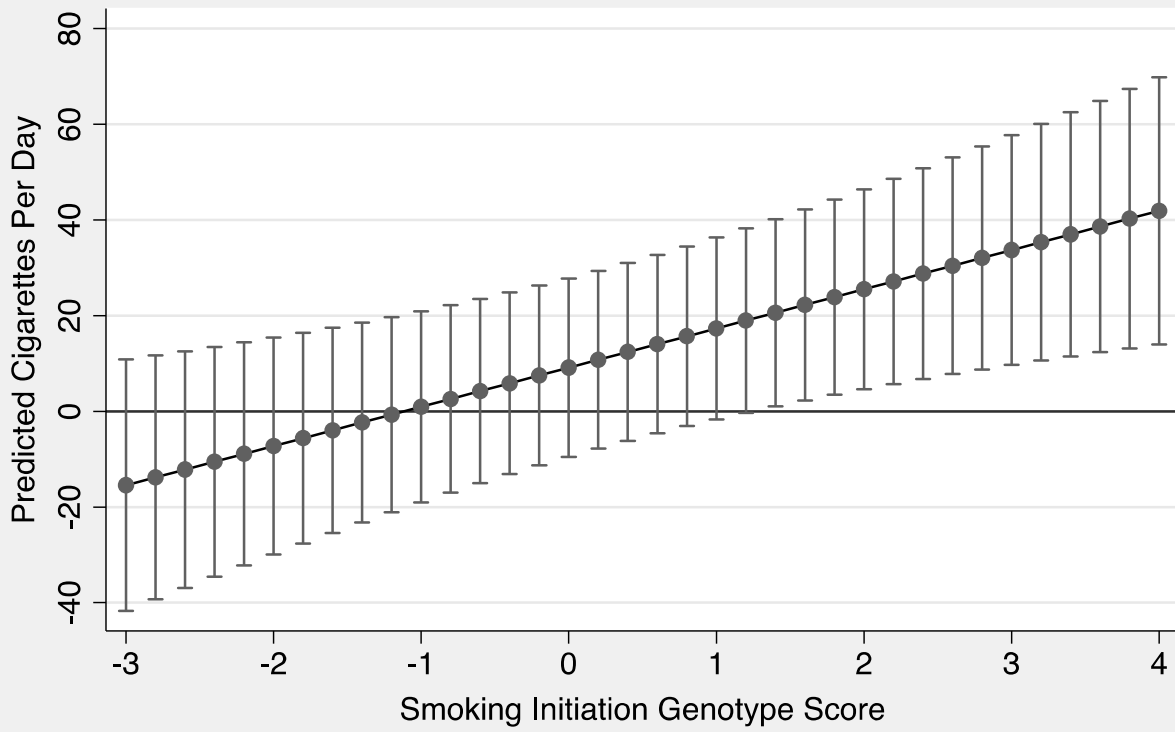


Figure 2. Difference in Predicted Cigarettes Per Day
Veterans vs. Non-Veterans (College Educated)

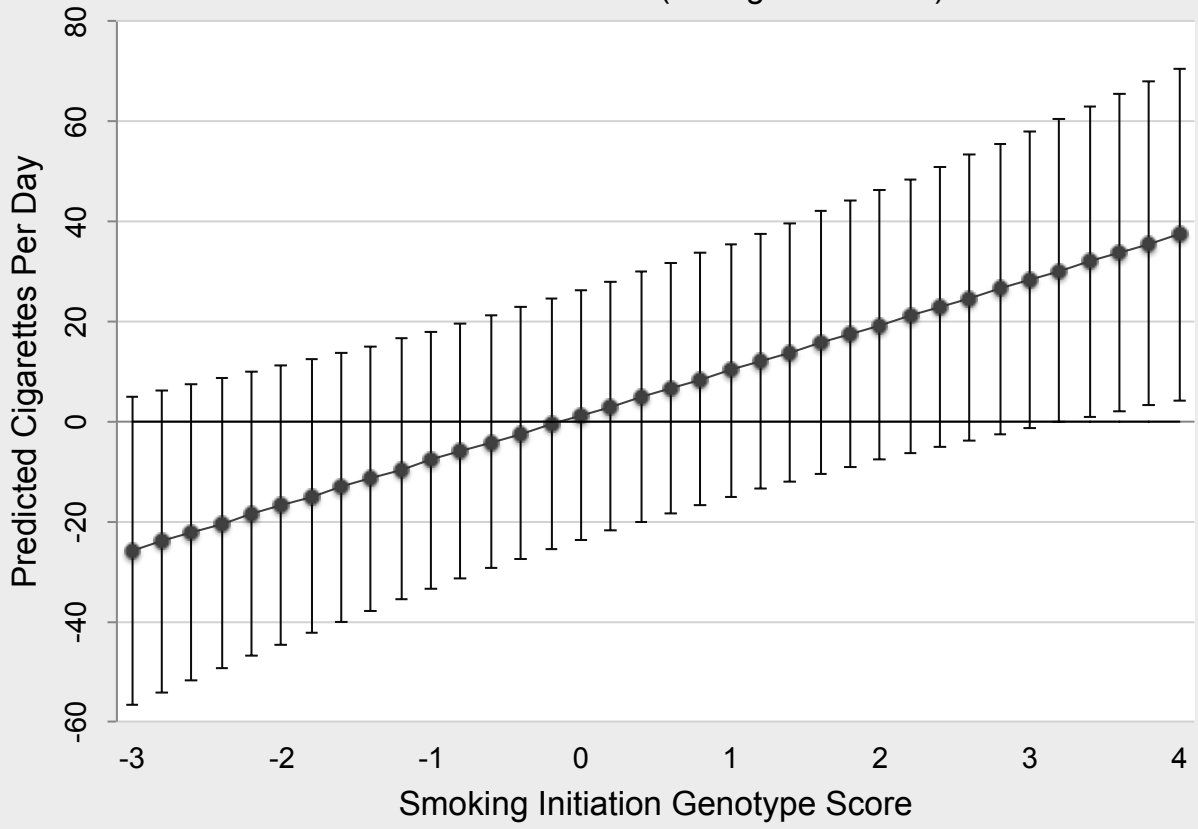


Figure 3. Difference in Predicted Cigarettes Per Day Veterans vs. Non-Veterans (Not College Educated)

