The Complicated Links Between Health and the Genetic Endowments for Smoking*

Yeongmi Jeong
University of Georgia

Meghan M. Skira
University of Georgia

June 28, 2023

Abstract

Many health conditions have a strong genetic component. We examine whether genetic factors related to risky health behaviors explain variation in health. Using Health and Retirement Study data, we estimate associations between health and the genetic endowments linked to smoking initiation and intensity. Among those aged 50–65, higher genetic predisposition for smoking associates with worse health, even after flexibly controlling for individual smoking behavior and among never smokers, suggesting the genetic endowments correlate with health through non-smoking channels. The genetic endowments correlate with longevity expectations, planning horizons, and measures of conscientiousness, but these channels do not fully explain the estimated relationship between genetic risk for smoking and health. Furthermore, an increase in a spouse's genetic risk for smoking intensity has adverse spillovers for own health. Overall, our results suggest the genetic factors linked to smoking capture an array of traits that correlate with less engagement in health-promoting activities.

*Jeong: Department of Economics, Terry College of Business, University of Georgia, Athens, GA 30602 (email: yeongmi.jeong@uga.edu). Skira: Department of Economics, Terry College of Business, University of Georgia, Athens, GA 30602 (email: skira@uga.edu). The paper has benefited from useful discussions with Josh Kinsler, Emily Lawler, Laura Zimmermann as well as seminar participants at the University of Georgia.

1 Introduction

Many chronic conditions and health outcomes, such as diabetes, cardiovascular disease, and body weight, have a strong genetic basis. For example, studies that compare identical and fraternal twins find that genetic differences explain 50–90 percent of the variation in body mass index (Wehby 2016). Evidence on the heritability of health largely comes from these twin studies, which tell us how much genes collectively matter, but do not identify the particular genetic markers or pathways that influence health.

In this paper, we take advantage of recent advances in genetics that have led to the discovery of genetic markers that correlate with health behaviors. We explore the empirical relationship between health and genetic risk for smoking to understand whether and how genes linked to risky health behaviors explain variation in health outcomes. Understanding the factors underlying differences in health is important as the effects of public health policies will vary depending on the sources of health variation. Many policies directly restrict risky health behaviors, such as bans on smoking in public places or minimum legal purchase ages of cigarettes and alcohol. But, if the factors that predispose one toward risky behaviors also influence health even conditional on engagement in the risky behavior, then targeting these factors may be a more attractive policy option. For example, if genetic endowments linked to smoking are also tied to time preferences or returns on health investments, then commitment devices or monetary rewards for engaging in health-promoting behaviors may be more effective in modifying behavior and improving health.

We estimate associations between the genetic risk for smoking and health outcomes among those aged 50–65 using the Health and Retirement Study (HRS). We focus on two genetic measures—the polygenic score for smoking initiation (hereafter, the initiation score) and the polygenic score for smoking intensity as measured by the number of cigarettes smoked per day (hereafter, the CPD score). A polygenic score is a linear index of genetic markers that associate with the outcome or trait of interest. A higher initiation score means an individual is more likely to ever regularly smoke and a higher CPD score means they are more likely to smoke heavily.

Our estimates reveal a meaningful and statistically significant relationship between the smoking polygenic scores and several adverse health outcomes, and the associations are present even after flexibly controlling for individual smoking behavior. Higher initiation scores and CPD scores are associated with worse health on a variety of dimensions including self-reported health, diabetes, arthritis, body mass index, and obesity as well as an index that aggregates the individual health outcomes. Furthermore, a higher CPD score is robustly associated with worse health among both those who have ever smoked and never smokers, while never smokers drive the relationship between the initiation score and health. Taken together, our results suggest the genetic endowments for

smoking associate with health through channels beyond smoking.

As individuals inherit their genes, those with a higher genetic risk for smoking may grow up with parents who smoked or were in poor health. Variation in the smoking polygenic scores could reflect childhood environments, such as differential exposure to secondhand smoke. We add controls for parental smoking, parental mortality, and parental risky behaviors (problematic drug use or drinking), and find these controls associate significantly with health, but generate only modest changes in the estimated relationship between the smoking polygenic scores and health. Given evidence that health outcomes and behaviors are correlated between spouses and that smokers are more likely to marry smokers, we also add controls for spousal smoking behavior and spousal smoking polygenic scores. The addition of these controls, particularly spousal smoking behavior, diminishes the estimated relationship between one's own CPD score and health but by less than 14 percent. We find that the spousal CPD score has a statistically significant adverse relationship with own health (controlling for spousal smoking behavior), a result consistent with growing evidence of social-genetic effects (i.e., the influence of others' genotypes) (e.g., Domingue and Belsky 2017, Kong et al. 2018, Cawley et al. 2019, Sotoudeh et al. 2019, Brunello et al. 2020, Cawley et al. 2023, Jeong 2023, Otten and Mandemakers 2023). As couples likely engage in similar health behaviors, such as those related to diet, nutrition, and physical activity, these results support the notion that genetic endowments for smoking reflect factors that influence health and health behaviors more generally beyond smoking.[1]

Given the robust associations between the smoking polygenic scores and health that remain even after including parental and spousal controls, we then explore channels through which the genetic factors may operate, namely risk preferences, longevity expectations, time preferences, and impulsivity. Studies show smokers tend to be more risk-prone, impatient, and impulsive than non-smokers (e.g., Khwaja et al. 2006, 2007a). We find that increased genetic risk for smoking significantly associates with lower expectations of living to age 75, shorter financial planning horizons, and lower degrees of conscientiousness (more impulsiveness). Collectively, these channels account for at most a quarter of the relationship between the CPD score and health, and 25–45 percent of the relationship between the initiation score and health among never smokers.

Our focus on genetic endowments for smoking rather than other risky behaviors is motivated in part by data availability. We want to understand whether genetic markers associated with risky health behaviors explain variation in health and whether these markers influence health through channels other than that particular behavior. As such, we need detailed information about one's engagement in the risky behavior as well as that of their parents and spouse. The HRS includes rich

[1]These results are consistent with those of Otten and Mandemakers (2023), which finds significant associations between spousal genetic endowments and own BMI, smoking, and drinking.

data on smoking that meets these requirements. While polygenic scores for other health behaviors like alcohol use and physical activity exist, these behaviors are not measured as comprehensively in the HRS and are likely more difficult to recall. Furthermore, several of the genetic variants linked to smoking relate to specific biological systems, which provides a lens through which to interpret our results. For example, variants near dopamine receptors strongly correlate with smoking intensity, and variants associated with ever smoking relate to systems that affect reward-based learning and addiction (Liu et al. 2019). Thus, the smoking polygenic scores reflect, in part, biological factors linked to addiction, which may be important for health even conditional on smoking behavior. However, the polygenic scores aggregate various traits and biological pathways that associate with smoking, and should not be interpreted narrowly as a predisposition to addiction.

Our work is tied to several strands of literature. One strand focuses on the role of genes in explaining variation in health. Twin studies, which compare identical and fraternal twins to calculate the fraction of the variance in outcomes attributed to genes, find a number of chronic conditions, such as cardiovascular disease, obesity, diabetes, and arthritis, have a strong genetic component (see reviews in Polderman et al. 2015, Wehby 2016). A more nascent related literature studies the role of genes in the intergenerational transmission of health by comparing adoptive parent-child pairs and biological parent-child pairs (e.g., Thompson 2014, Classen and Thompson 2016) or exploiting information about adoptees and their biological and adopting parents (e.g., Björkegren et al. 2022). These studies find genetic transmission accounts for an important share of the intergenerational correlation in health (e.g., asthma, diabetes, obesity, hospitalizations, mortality). Twin studies and adoptee-based designs tell us how much genes matter collectively for a given health outcome or for health transmission, but not the specific genes or genetic pathways.[2] As a result, our knowledge of the channels through which genes influence health is limited. We demonstrate that genetic endowments for smoking, a policy-relevant and not uncommon risky behavior, predict health even after accounting for smoking behavior. Our results are consistent with individuals with a higher genetic propensity to smoke being less likely to engage in health-promoting behaviors.

This study also relates to recent literature that shows the polygenic score for educational attainment (hereafter, the EA score) predicts economic and health outcomes even after accounting for educational attainment and sheds light on the channels through which the EA score operates (Belsky et al. 2016, Barth et al. 2020, Papageorge and Thom 2020, Bolyard and Savelyev 2021).

[2]Linnér and Koellinger (2022) is a recent exception. They investigate how 27 polygenic scores for common medical conditions and mortality risk, including the CPD and smoking initiation scores, predict differences in longevity in the HRS. They find several of the polygenic scores, including the CPD score, have a non-negligible association with mortality. Consistent with our results, the CPD score significantly predicts mortality even after accounting for observed smoking.

Our paper is similar in that we show a polygenic score predicts outcomes it was not designed to predict and we explore the mechanisms underlying that relationship, but our focus is distinct in that we consider smoking polygenic scores and health outcomes.

We also contribute to the economic literature on the genetic risk for smoking, which typically focuses on the role of gene-environment interactions in smoking behavior (e.g., Boardman et al. 2011, Fletcher 2012, Meyers et al. 2013, Domingue et al. 2016, Schmitz and Conley 2016, Wedow et al. 2018, Biroli and Zwyssig 2021, Slob and Rietveld 2021, Bierut et al. 2023). Our focus is not on gene-environment interactions or smoking decisions, but rather understanding whether the smoking polygenic scores are predictive of health and the channels that mediate the association. By providing a comprehensive picture of how genetic endowments for smoking function, our analysis may provide an interpretative framework for the gene-environment interactions found in prior studies.

Our paper also contributes to a large strand of literature that documents differences in the characteristics of smokers and non-smokers with respect to preferences and subjective beliefs (e.g., Barsky et al. 1997, Khwaja et al. 2006, 2007a, Anderson and Mellor 2008, Chabris et al. 2008, Ida and Goto 2009). These studies generally find smokers are more impatient, risk tolerant, and impulsive than non-smokers. Khwaja et al. (2006) find no significant change in these characteristics among those who change their smoking status, which they interpret as evidence that these characteristics are innate and not caused by smoking. We contribute to this literature by exploring whether individuals who are genetically predisposed to smoking initiation and heavy smoking differ along these margins. Our results imply those with higher genetic risk for smoking expect earlier mortality, have shorter financial planning horizons, and exhibit more impulsiveness. We therefore provide a possible genetic microfoundation for some of the documented differences among smokers and non-smokers.

Finally, while our analysis is descriptive in nature, the results have implications for the literature that estimates causal effects of smoking on health, underscoring the importance of time-invariant unobserved heterogeneity. For example, Darden (2017) and Darden et al. (2018) uncover strong evidence of unobserved heterogeneity—individuals who are more likely to smoke are also at increased risk of poor chronic health and more likely to die independently of smoking. The sources of unobserved heterogeneity are not modeled, though the authors speculate genetics or differences in the likelihood of engaging in correlated risks may be at play. Our results shed light on the black box of unobserved heterogeneity, highlighting a non-trivial role for genetic endowments, and our findings are consistent with a genetic underpinning for engagement in risky and health-degrading behaviors. Our findings also have implications for interpreting instrumental variable estimates of smoking on health. In particular, if compliers systematically differ in their genetic endowments for smoking, substantial treatment effect heterogeneity is likely, limiting the generalizability of the local average treat-

ment effect of smoking on health. In addition, our results underscore that the smoking polygenic scores should not be used as instruments for smoking as the exclusion restriction will be violated.[3]

The rest of the paper is organized as follows. Section 2 provides a brief overview of polygenic scores, including the ones used in our analyses. Section 3 describes the Health and Retirement Study data and sample construction. Section 4 describes the empirical analysis. Section 5 presents the estimates of the relationship between health outcomes and smoking polygenic scores, and Section 6 explores channels that may mediate the observed relationship. Section 7 presents the results of robustness exercises. Section 8 provides some final thoughts.

2 Background on Polygenic Scores

Large-sample twin studies reveal that smoking is moderately to highly heritable (Li et al. 2003, Agrawal and Lynskey 2008), and recent advances in genetics have led to the discovery of particular genetic markers that predict smoking behavior. We rely on polygenic scores, which aggregate these genetic markers. We provide a brief overview of polygenic scores here and refer the reader to Benjamin et al. (2011) and Beauchamp et al. (2011) for a detailed discussion of the human genome and Barth et al. (2020) and Papageorge and Thom (2020) for a detailed discussion of polygenic scores.[4]

Broadly, a polygenic score is a linear index of genetic markers (or variants) associated with an observable trait or outcome, such as smoking behavior, educational attainment, body mass index, blood pressure, etc. A polygenic score is obtained by aggregating millions of individual genetic variants across the human genome and weighting those variants by the strength of their association with the outcome of interest. The estimated associations come from genome-wide association studies (GWAS), where thousands or millions of linear regressions are estimated linking specific genetic variants called single nucleotide polymorphisms (SNPs) to the outcome of interest. The polygenic score is a linear combination of the GWAS coefficients and is calculated as follows:

$$PGS_i = \sum_{j=1}^{J} \beta_j SNP_{ij} \tag{1}$$

where SNP_{ij} is the genotype (i.e., the number of copies of the reference allele–0, 1, or 2) for individual i at SNP j and β_j is the effect size for SNP j estimated in the GWAS. A higher polygenic score means an individual possesses more of the genetic variants that associate with the trait or outcome. That is, a higher score means the individual has a higher genetic risk for that trait or outcome.

We use the polygenic scores constructed and released by the Health and Retirement Study.

[3]We are not the first to caution against the use of polygenic scores as instruments for specific health conditions or behaviors. See for example Benjamin et al. (2011), Lehrer and Ding (2017), and Fletcher (2018).

[4]Benjamin et al. (2011), Beauchamp et al. (2011), Lehrer and Ding (2017), and Fletcher (2018) provide excellent overviews of the literature that incorporates genetic concepts and data into economic analysis.

The scores include all available SNPs that overlap between the HRS genetic database and the relevant GWAS, without accounting for linkage disequilibrium (i.e., correlations between nearby SNPs) or p-value thresholds, as this set of decisions led to the largest predictive power of the polygenic scores (Ware et al. 2017).[5]

In our analysis, we focus on two polygenic scores related to smoking behavior—one for smoking initiation (the initiation score) and one for cigarettes per day (the CPD score). The initiation score measures the genetic endowments linked to whether an individual ever smoked regularly while the CPD score measures genetic endowments linked to heaviness of smoking. The scores are based on GWAS meta-analyses results from Liu et al. (2019), which used a discovery sample of 1.2 million individuals of European ancestry for the initiation score and a subsample of 337,334 current and former smokers for the CPD score.[6] The discovery sample did not include the HRS, but the HRS was used to assess out-of-sample predictive power. The out-of-sample prediction results in Liu et al. (2019) imply that the initiation score explains about 4 percent of the variation in smoking initiation and the CPD score likewise explains about 4 percent of the variation in smoking intensity (among current and former smokers) in the HRS. We document the relationship between the polygenic scores and smoking behavior in our sample in the next section.

Several of the genetic variants implicated in the smoking polygenic scores are linked to biological systems and functions of note. For example, variants near dopamine receptors are associated with smoking heaviness but not smoking initiation. Some variants associated with smoking initiation relate to systems that affect reward-based processing and addiction. Nicotine receptor and nicotine metabolizing genes are also associated with smoking behaviors (Liu et al. 2019). Importantly, although the polygenic scores contain variants related to addiction, nicotine response, and reward-related learning, they should not be interpreted narrowly as a predisposition to nicotine addiction given they aggregate many biological systems, pathways, and traits.

3 Data

The Health and Retirement Study (HRS) is a longitudinal survey of individuals aged 50 and over in the United States and their spouses. The survey started in 1992 and followup surveys are conducted biennially. We use data from 1994–2016 in our analysis.[7] The data contain rich information on an array of topics, such as health, family structure, employment, assets, and

[5]There are various ways to construct a polygenic score. Sometimes only the variants with the strongest GWAS associations are included, namely SNPs that are considered genome-wide significant (e.g., p-value $< 5 \times 10^{-8}$). Other times, all SNPs are included regardless of genome-wide significance, like the scores released by the HRS.

[6]Although the genetic markers that relate to smoking intensity were discovered on a sample of current and former smokers, the CPD score can be constructed for all individuals regardless of their smoking history.

[7]We exclude the 1992 survey wave due to survey question differences.

expectations about future events.

Genetic data was collected during four waves (2006, 2008, 2010, and 2012) from respondents who consented to provide DNA samples. The HRS has made several polygenic scores publicly available for the approximately 15,000 individuals who have been genotyped. We use the fourth release of the HRS's polygenic score data. Most polygenic scores are for diseases and clinical outcomes, such as Alzheimer's disease, high blood pressure, and body mass index. More recently, due in part to the increased availability of genetic data, polygenic scores linked to behavioral traits, such as smoking, have been calculated from GWAS and released by the HRS. As discussed above, we focus on two polygenic scores linked to smoking—one for smoking initiation (initiation score) and one for cigarettes smoked per day (CPD score). The HRS normalizes the scores to have mean zero and a standard deviation equal to one within ancestry groups (i.e., European and African). A higher score implies more genetic risk for the corresponding trait (e.g., to ever smoke or smoke heavily).

3.1 Sample Construction

Our sample includes HRS respondents aged 50–65 who have been genotyped. We focus on this age group to minimize survival bias concerns and avoid capturing potential impacts of retirement on health.[8] The HRS provides polygenic scores for about 12,000 respondents with European ancestry and 3,000 respondents with African ancestry. The scores we use were discovered based on the GWAS of individuals with European ancestry, and scores based on such GWAS tend to lack predictive power when applied to other ethnic groups. We follow the convention in this literature (e.g., Barth et al. 2020, Papageorge and Thom 2020) and limit our sample to respondents classified as genetic Europeans and who self-identify as white.

3.2 Health Outcomes

The HRS provides rich information on various health outcomes. We consider both self-reported measures of physical health as well as doctor-diagnosed health conditions.

Respondents are asked to rate their general health on a 1–5 scale, where 1 is excellent, 2 is very good, 3 is good, 4 is fair, and 5 is poor. We create a "bad" self-reported health indicator that takes on a value of one if the individual reports their health as fair or poor, and zero otherwise. We create an indicator for whether an individual reports any overnight hospital stay in the last two years. We consider an individual's body mass index (BMI) based on self-reports of weight and height as well as an indicator for being obese (i.e., BMI greater than or equal to 30 kg/m^2).

[8] See for example Coe and Zamarro (2011), Behncke (2012), Insler (2014), Eibich (2015), Godard (2016), Mazzonna and Peracchi (2017), Gorry et al. (2018), Shai (2018). Later, we verify the robustness of our results to only including those aged 50–61, i.e., below the early Social Security retirement age.

We also create several indicators based on whether an individual reports that they have ever been diagnosed with certain conditions by a doctor, including high blood pressure; diabetes; cancer; chronic lung disease (e.g., chronic bronchitis, emphysema); heart problems (e.g., heart attacks coronary heart disease, angina, congestive heart failure); stroke; and arthritis. We construct an indicator for having difficulties with activities of daily living (ADLs) as well as an indicator for having difficulties with instrumental activities of daily living (IADLs). ADLs include walking across a room, dressing, bathing, eating, getting in and out of bed, and using the toilet. IADLs include using the phone, taking medications, managing money, shopping for groceries, preparing meals, using a map, using a calculator, using a microwave, and using a computer.

We construct a summary standardized index of these health measures in the spirit of Kling et al. (2007), Hoynes et al. (2016), and Bütikofer et al. (2021). To calculate this index, we create the z-score for each individual health measure and then take the average of the z-scores. The z-score is created by subtracting the mean and dividing by the standard deviation. Aggregating the measures in this way improves statistical power, limits the number of statistical tests, and allows us to draw general conclusions about health and genetic endowments for smoking. We create the index using all the health measures described above (except the obesity indicator as BMI is already included). As all the components of the index are associated with poor health, an increase in the index suggests a deterioration in health. In our analysis, we largely focus on this summary standardized index of health, but we also examine each health outcome separately to understand whether observed changes in the summary index are driven by particular components.[9]

3.3 Own Smoking Behavior

To examine whether the smoking polygenic scores are associated with health even after accounting for smoking behavior, we control for an individual's current and past smoking behavior as flexibly as possible. We make use of a wealth of smoking data available in the HRS described below.

We create an indicator for current smoking that equals one if the individual reports that they currently smoke, and zero otherwise. Similarly, we create an indicator variable for ever smoking that equals one for individuals who currently smoke or ever smoked in their lifetime. Among those who currently smoke, the HRS asks about the average number of cigarettes smoked per day. We set this variable to zero for those who do not currently smoke. For those who ever smoked but are not current smokers, we have information on the maximum number of cigarettes smoked per day when they were smokers. We use information on the age when ever smokers started smoking and quit smoking (if they have quit). We also create an indicator for teenage smoking that equals

[9]We also show our baseline results are robust when we consider a summary standardized index based only on the seven doctor-diagnosed conditions.

one if the individual reports they regularly smoked cigarettes when they were in grade school or high school, and zero otherwise. When data on smoking variables are missing, we set the value to zero and include an indicator for missing values of that variable.

3.4 Parental Smoking, Risky Behaviors, and Mortality

In some analyses, we control for parental behaviors during the respondent's childhood as well as parental mortality. The survey includes information about whether neither parent smoked, one of them smoked, or both smoked during the respondent's childhood. We create indicators for each potential response. We also construct an indicator for whether either parent drank or used drugs so often that it caused family problems during the respondent's childhood. We construct various measures of parental mortality. For a small subset of our sample, we can create an indicator for whether a parent died before the individual turned age 16.[10] The HRS also contains information on whether parents are alive as of the current survey wave, and if not, the parent's age when they passed away. Using this information, we create a categorical variable for the respondent's mother's mortality status: alive, died before age 65, and died after age 65 as well as analogous measures for the respondent's father. When data on any of these variables are missing, we set the value to zero and include an indicator for missing values of that variable.

3.5 Descriptives

Summary statistics for our sample are provided in Table 1. In our analyses, we estimate models on the full sample and separately for ever and never smokers. Thus, we present statistics for the full sample and by whether an individual has ever smoked. Average age in the sample, regardless of ever smoking status, is 58. About half the ever smoker sample is male compared to a third of the never smoker group.

Among current smokers, the average number of cigarettes smoked per day is almost 18, and among those who do not currently smoke but have smoked in the past, their smoking peaked, on average, at 22 cigarettes per day. On average, ever smokers began smoking between ages 17–18, and among those who have quit, the average age of quitting is 37. Almost 40 percent of ever smokers smoked during their teenage years.

As expected, the health of the never smokers is better than that of the ever smokers along several dimensions, and the differences in health are statistically significant in almost all cases. Consistent with smoking being one of the leading causes of coronary and pulmonary diseases in the US (US Department of Health and Human Services 2020), the proportion of individuals with

[10]Parental mortality before the respondent reached age 16 was only collected in recent years via life history mail surveys that are distributed to a subsample of respondents and have far less than 100 percent response rates.

diagnosed lung disease and heart problems is higher in the ever smoker sample. In addition, the ever smokers are more likely to report that their parents smoked and that one or both parents drank or used drugs so often it caused problems during their childhood.

Only HRS respondents who consented to provide genetic data are included in our sample. In Appendix Table A1, we provide summary statistics from 1994–2016 for non-genotyped HRS respondents aged 50–65 of European descent who identify as white. On average, the genotyped sample has more females and is healthier. If anything, the genotyped sample is positively selected on health, which will bias any negative association between health and the polygenic scores associated with smoking toward zero.

In Figure 1, we show the smoothed densities of the smoking polygenic scores for individuals in our sample. Figure 2 (a) shows the non-parametric (lowess) relationship between the initiation score and the probability of ever smoking, and Figure 2 (b) shows the relationship between the CPD score and average cigarettes smoked per day among current smokers.[11] In both cases, the relationship is positive and approximately linear. For former smokers, we have information on the maximum number of cigarettes smoked per day when they smoked. Figure 2 (c) plots this measure of peak smoking among former smokers against their CPD score. We observe a positive relationship, which is particularly strong among those with relatively high scores.

We also quantify the relationship between the smoking polygenic scores and smoking behavior in our sample. Appendix Table A2 presents estimates from regressions of the maximum number of cigarettes smoked per day among ever smokers (columns 1–3) or an indicator for whether an individual ever smoked (columns 4–6) on the CPD score only, the initiation score only, and both scores.[12] Unsurprisingly, the CPD score explains more of the variation in smoking intensity than the initiation score, with an incremental R^2 of 0.015 compared to 0.004 for the initiation score. A one standard deviation increase in the CPD score increases the number of cigarettes smoked by 2, while a one standard deviation increase in the initiation score is associated with a 1 cigarette increase. Similarly, the initiation score explains more of the variation in the probability of ever smoking compared to the CPD score, with an incremental R^2 of 0.031 compared to 0.003 for the CPD score. A one standard deviation increase in the CPD score is associated with a 1.6–2.7 percentage point increase in the probability of ever smoking, whereas a one standard deviation increase in the initiation score is associated with about a 9 percentage point increase. In sum, the polygenic scores predict their respective smoking outcomes in our sample.

[11] For each respondent, we calculate their average reported number of cigarettes smoked per day across periods when they report currently smoking and plot that average against their CPD score.

[12] In all specifications we include the standard controls and genetic principal components described in Section 4.

4 Empirical Analysis

To examine the link between health and genetic endowments for smoking, we estimate the following model via OLS:

$$Y_{it} = \beta_0 + \beta_1 CPDScore_i + \beta_2 InitiationScore_i + \beta_3 X_{it} + \varepsilon_{it}, \quad (2)$$

where Y_{it} is the health outcome of individual i at survey wave t, $CPDScore_i$ is the polygenic score for the number of cigarettes per day, $InitiationScore_i$ is the polygenic score for smoking initiation, and X_{it} is a vector of controls.[13] Given we observe the same individual potentially many times, we cluster standard errors at the individual level.

We categorize the variables in X_{it} into the following groups: standard controls, smoking controls, and genetic principal components. Following Barth et al. (2020) and Papageorge and Thom (2020), the standard controls include age dummies, a male dummy, birth year dummies, the interaction of age dummies and the male dummy, the interaction of birth year dummies and the male dummy, and survey year dummies. As we want to understand whether the smoking polygenic scores predict health outcomes even after accounting for smoking, we control for smoking behavior as flexibly as possible. We include a dummy for ever smoked status, dummies for each reported number of cigarettes smoked per day, and dummies for each reported number of maximum cigarettes smoked per day among former smokers. We additionally include a dummy for current smoking status, dummies for age started smoking, dummies for age quit smoking, and a dummy for teenage smoking status. For all of these smoking controls, we include dummies for missing values. Finally, to account for potential bias that may arise from population stratification and genetic differences across ethnic groups (even within our sample of individuals with European ancestry who identify as white), we follow standard practice (e.g., Price et al. 2006, Belsky et al. 2016, Barth et al. 2020) and include the 10 principal components of the full matrix of SNP data as well as their interactions with the male dummy.

We estimate equation 2 on the full sample of individuals, pooling across ever and never smokers. We also estimate the model separately by ever smoking status.[14] By estimating equation 2 on never smokers, we directly examine whether the smoking polygenic scores correlate with health through avenues other than own smoking. Thus, any observed relationship between the smoking scores and health for this sample must be explained by non-smoking channels. In Panels (a) and (b) of Figure 3,

[13] We have also allowed for an interaction between the initiation and CPD scores. Generally, the point estimate on the interaction is insignificant and very close to zero and its inclusion leaves the point estimates on the (non-interacted) initiation and CPD scores largely unchanged. We proceed without the interaction term for ease of exposition and interpretability of the estimates, but results with the interaction term are available by request.

[14] Only 4 individuals smoke for the first time during the sample period and hence change ever smoking status. Results are robust to excluding these individuals.

we show that distributions of the smoking initiation and CPD scores for ever and never smokers share common support, alleviating concerns about comparing estimates derived from these samples.[15]

5 Main Results

5.1 Health Outcomes and the Smoking Scores

In Table 2, we show results for the summary standardized index of the 12 health measures described in Section 3.2.[16] Estimates for the full sample are shown in columns (1)–(3). Column (1) contains results not including any controls. The estimates imply a one standard deviation increase in the CPD score increases the health index by 0.042 standard deviations and a one standard deviation increase in the initiation score increases the health index by 0.026 standard deviations; both estimates are significant at the 1 percent level. Recalling that an increase in the index means worse health, the results imply that higher genetic risk for smoking is associated with worse overall health. In Column (2), we add the standard controls and the 10 principal components of the genetic data. The coefficient on the initiation score grows in magnitude while the coefficient on the CPD score barely changes. In Column (3), we add the smoking controls, and unsurprisingly, the coefficients on the CPD and initiation scores attenuate. In this case, a one standard deviation increase in the CPD (initiation) score increases the health index by 0.035 (0.021) standard deviations. Thus, even when we flexibly control for current and past smoking behavior, higher genetic risk for smoking is significantly associated with worse health.

Columns (4)–(6) present results for the ever smoker sample. After including the standard controls, principal components, and smoking controls, we find a one standard deviation increase in the CPD score is correlated with a statistically significant 0.037 standard deviation increase in the summary index. A one standard deviation increase in the initiation score increases the health index by 0.011 standard deviations, but the estimate is not statistically significant. We present results for the never smoker sample in columns (7) and (8), a group for which there is no smoking behavior to control for. For this group, a one standard deviation increase in either the CPD score or initiation score is associated with a statistically significant 0.030 standard deviation increase in the health index.[17] We cannot reject equality of the CPD score coefficients across the ever and never smoker

[15] However, the p-values from Kolmogorov-Smirnov tests reject equality of the distributions by smoking status at the 5 percent level. This is unsurprising, particularly for the initiation score given it predicts ever smoking status.

[16] The outcomes include indicator variables for "bad" self-reported health status, having a hospital stay in the past two years, having doctor-diagnosed high blood pressure, diabetes, cancer, lung disease, heart problems, stroke, arthritis, having difficulties with ADLs, having difficulties with IADLs as well as BMI measured continuously.

[17] In Appendix Table A3, we present results for a summary standardized index that aggregates just the seven doctor-diagnosed conditions (high blood pressure, diabetes, cancer, lung disease, heart problems, stroke, and arthritis). The estimates are qualitatively similar to those using the summary index that includes the broader set of health measures.

samples, while we reject equality of the initiation score coefficients at the 5 percent level.

In Table 3, we present estimates for the individual health measures that appear in the summary standardized index for the full sample (Panel A), ever smokers (Panel B), and never smokers (Panel C). All models include the standard controls, genetic principal components, and the controls for smoking behavior (except in Panel C where there is no smoking behavior to account for). For most of the measures, a higher genetic risk for smoking, particularly smoking intensity, is associated with worse health in the full sample. We find some differences in these relationships among the ever and never smokers, though in most cases, we cannot reject equality of the CPD score coefficients or initiation score coefficients across the two groups.[18] For example, the smoking initiation score has a stronger relationship with health outcomes among the never smokers, and the CPD score correlates significantly with more health outcomes among the ever smokers. We explore these patterns in detail later.[19] Notably, we observe qualitatively similar and precisely estimated point estimates on the CPD score across the ever and never smoker samples for the following outcomes: self-reported "bad health," diabetes diagnosis, arthritis diagnosis, BMI, and obesity. Diabetes and obesity are influenced by health behaviors such as diet and exercise, and being overweight is a common cause of arthritis. One interpretation of these results is that genetic endowments for smoking correlate with behaviors that are health degrading besides smoking.

Overall, the results in Tables 2 and 3 imply higher genetic risk for smoking is associated with worse health even after flexibly controlling for smoking behaviors. This finding is robust across ever and never smokers. The fact that these associations are present and precisely estimated for never smokers is particularly compelling as there are no own smoking channels through which the scores can operate, suggesting non-smoking mechanisms are at play. Furthermore, if there is concern that the relationships between the scores and health in the full and ever smoker samples simply reflect smoking behavior that is not well-captured in our smoking controls, the fact that we find robust and meaningful associations in the ever and never smoker samples should alleviate that concern.

The most notable difference in the results across the ever and never smoker samples is the point estimate on the initiation score. In the specification with the most rigorous set of controls, we find a positive but not precisely estimated relationship between the initiation score and the summary health index among ever smokers but a larger and precisely estimated relationship among never smokers. Earlier, we showed the initiation score distribution is, unsurprisingly, shifted left for never smokers compared to ever smokers. If the association between the initiation score and health

[18]We reject equality of the CPD score coefficients for the following outcomes: recent hospital stay, diagnosed lung disease, and BMI. We reject equality of the initiation score coefficients for recent hospital stay, BMI, and obesity.

[19]These results are in line with those in Linnér and Koellinger (2022), which finds larger associations between the CPD score and mortality among ever smokers relative to never smokers.

is non-linear and the samples represent different parts of the distribution, this could explain the difference in the point estimates. Appendix Figure A1 shows lowess plots of the smoking polygenic scores and the summary health index. For most of the range of support, the relationship between the polygenic scores and the health index appears approximately linear, but in the tails of the polygenic score distributions, there is some evidence of non-linearities and interactions between the scores and smoking status. We therefore examine the relationship between terciles of the polygenic score distributions and health.

The results for the summary standardized index and the individual health measures are presented in Tables 4 and 5, respectively, for the full sample, ever smokers, and never smokers. For brevity, we focus on the summary health index. The indicators for being in the second tercile are normalized. The estimates reveal some interesting patterns. In the full sample, those in the first tercile of the CPD score distribution experience a 0.034 standard deviation decrease (i.e., improvement) in the summary health index relative to those in the middle of the distribution, and this improvement is amplified for never smokers. In the full sample, those in the third tercile of the CPD score distribution experience a 0.046 standard deviation increase (i.e., deterioration) in the health index relative to those in the second tercile. In this case, the relationship is larger in magnitude for ever smokers. A similar but even more striking pattern is apparent for the initiation score. Ever smokers in the first tercile of the initiation score see no significant difference in health relative to those in the second tercile, whereas the third tercile is associated with worse health, but the association is not significant once smoking controls are included. Never smokers with initiation scores in the first tercile experience significantly better health relative to those in the middle tercile, but the third tercile does not significantly correlate with health. We are only able to reject equality of coefficients across the ever and never smoker samples for the first tercile of the initiation score ($p < 0.05$).

Taken together, there is suggestive evidence of non-linearities in the health-smoking genes relationship.[20] Moreover, the difference in the baseline point estimates on the initiation score for ever and never smokers is driven by those with low initiation scores. In particular, the genetic factors associated with a low initiation score combined with never smoking seem to be protective of health.

5.2 Parental Smoking, Risky Behaviors, and Mortality

One potential explanation for the relationship between health outcomes and the smoking polygenic scores is parental smoking, risky behaviors, and health and mortality. As one inherits their genes, an individual's genetic risk for smoking is correlated with their parents' genetic risk. Parental

[20]However, we cannot reject symmetry of the associations along the distribution of each score within each sample. For example, we cannot reject that the coefficient on being in the first tercile of the CPD score equals minus the coefficient of being in the third tercile.

genetic predisposition toward smoking could impact their child's health in at least four ways. First, it could impact the child's health directly through smoking and secondhand smoke exposure. Second, parents with higher smoking polygenic scores may have been in worse health during the respondent's childhood due to smoking and/or the influence of the smoking genes that operates outside of smoking, which may have led to a disadvantaged childhood. Third, if the smoking polygenic scores are correlated with poor or risky health behaviors more broadly, children may have learned those behaviors from their parents. Fourth, if parents with a genetic predisposition for smoking experience worse health or die at an earlier age and individuals view their parents' health as foreshadowing their own health, they may adjust their own health-related behaviors in response.[21]

Table 6 presents the estimated relationship between the summary standardized index of health and the smoking polygenic scores after controlling for parental smoking, parental risky behaviors, and parental mortality during the respondent's childhood as well as parental mortality as of the current survey wave. We do not have parental genetic information, and hence cannot directly explore the role of parents' genetic endowments for smoking. We again present results for the full sample and separately by ever smoker status. Across the samples, there is a modest decline in the magnitude of the associations between the summary index of health and the smoking polygenic scores after including the parental variables, but they are still statistically significant (except the initiation score among ever smokers). We still find a non-trivial and robust association between the CPD score and health, and the correlation between the initiation score and health, like in the baseline results, is largely driven by the never smoker sample.

For clarity and ease of comparison, in Appendix Figure A2, we present the estimates for the individual health measures for the full sample, ever smokers, and never smokers, controlling for the parental measures described above. The inclusion of parental controls does not meaningfully change the estimated relationships between the smoking polygenic scores and the individual health measures. Taken together, the results thus far suggest the relationship between genetic endowments for smoking and health largely operates through channels other than own smoking and parental behaviors and mortality.

5.3 Spousal Genetic Smoking Endowments and Smoking Behavior

Another potential avenue through which the smoking polygenic scores may relate to own health is spousal smoking and spousal genetic endowments for smoking. It is well-documented that health outcomes and health behaviors are correlated between spouses, and smokers are more likely to marry smokers (Clark and Etilé 2006, Oreffice and Quintana-Domeque 2010, Chiappori et al.

[21] Darden and Gilleskie (2016) find limited evidence that offspring smoking is sensitive to parental health. One exception is women decrease their smoking following a parent's smoking-related cardiovascular event.

2012, 2018). Thus, an individual's own smoking behavior may be correlated with that of their spouse. Less clear is whether genetic risk for smoking is correlated among spouses.

Following Charles et al. (2013) and Barth et al. (2020), in Appendix Table A4 we consider married couples where both members provided genetic data and are aged 50–65 and categorize them into quartiles based on their CPD score (Panel A) and their initiation score (Panel B). We calculate the fraction of men in each quartile within a given female quartile. The diagonals of these matrices would be 100 percent and off-diagonal elements would be 0 percent with perfect assortative mating, while random assignment would imply 25 percent in each cell. For the CPD score, while we reject the random assignment null hypothesis (i.e., all cell entries equal to 25 percent) at the 10 percent level, the extent of assortative mating is small, with cell entries ranging from 21–30 percent. For the initiation score, we cannot reject the random assignment null hypothesis, and cell entries range from 22–28 percent. Furthermore, the within-couple correlation of the CPD scores is 0.03 as is the correlation of the initiation scores, while the correlation of ever smoking status is 0.24. Thus, we find little evidence of assortative mating on genetic endowments related to smoking, but smokers are more likely to marry smokers. These patterns are similar to those in Otten and Mandemakers (2023), which investigates social-genetic effects among married and cohabiting couples using the HRS.

We next estimate the relationship between the summary standardized index of health and own smoking polygenic scores after controlling for spousal smoking behaviors and spousal smoking polygenic scores.[22] We first add spousal smoking controls only, then spousal smoking polygenic scores only, and then both sets of spousal measures. The sample size decreases as we only include married individuals where both members were genotyped and aged 50–65. The results are presented in Table 7 for the full sample, ever smokers, and never smokers. The addition of spousal controls decreases the point estimate on the CPD score by at most 14 percent, and it is still statistically significant. For ever smokers, the point estimate on the initiation score is not statistically significant, and for never smokers, the association between the initiation score and health declines by about 17 percent and remains statistically significant. The decline in the size of the association between the polygenic scores and health is mostly driven by the inclusion of the spousal smoking variables, not the spousal polygenic scores. This is not surprising given we found little evidence of assortative mating on smoking-related genes. Interestingly, we find a robust and significant relationship between the spouse's CPD score and own health. A one standard deviation increase in the spouse's CPD score generates a 0.02 standard deviation increase in the health index. Thus, spousal genetic risk for smoking generates negative spillovers for own health.[23] To further explore this idea, we focus

[22]We control for spousal smoking identically to how we control for own smoking as described in Section 3.3.

[23]In Appendix Figure A3, we present results for the individual health measures controlling for spousal polygenic smoking scores and spousal smoking behavior. We again find modest declines in the associations between own

on couples where neither person has ever smoked, eliminating any own or spousal smoking influences. The results are presented in Appendix Table A5. Focusing on the summary standardized index of health in column (1), the association between own genetic risk for smoking and health attenuates somewhat, but we still find a significant association between a spouse's CPD score and own health. These results are consistent with smoking polygenic scores reflecting a propensity to engage in health-degrading behaviors more generally besides smoking, as a lack of health-promoting behaviors by a spouse could adversely affect one's own health. More broadly, these findings add to a growing body of evidence on social-genetic effects among genetically-unrelated groups (e.g., Domingue and Belsky 2017, Sotoudeh et al. 2019, Brunello et al. 2020, Jeong 2023, Otten and Mandemakers 2023).[24]

Overall, spousal smoking behavior plays a small mediating role in the relationship between own smoking genes and health. Spousal genetic endowments for smoking have a direct correlation with own health, but do not explain the association between own genetic endowments for smoking and own health. Even after including these spousal controls, we are left with a robust and meaningful association between the smoking polygenic scores and own health. This conclusion also holds when including *both* the parental and spousal variables. Appendix Table A6 shows that the association between the CPD score and the health index decreases by less than 20 percent across all samples, and in the case of never smokers, the association between the initiation score and the health index declines by 38 percent when parental and spousal controls are added.

6 Additional Mechanisms

In Section 5, we found higher genetic risk for smoking associates with worse health even after flexibly controlling for respondent's own smoking behaviors. Parental smoking, parental risky behaviors, parental mortality, and spousal smoking behavior play a role, but do not fully explain the estimated relationship. In this section, we investigate other potential non-smoking channels through which the smoking polygenic scores may operate including risk preferences, longevity expectations, planning horizons, and impulsivity. Prior studies suggest smokers and non-smokers differ along these margins; we explore whether those with a higher genetic predisposition for smoking differ along these characteristics.[25]

smoking polygenic scores and health, as well as adverse spillovers of the spouse's CPD score on some health measures.

[24]In particular, Otten and Mandemakers (2023) find that a partner's CPD score is positively associated with own smoking (controlling for own CPD score). More generally, they find individuals have higher BMI, drink more, and smoke more if their partners have higher polygenic scores for those particular behaviors and outcomes.

[25]For the results presented in this section, we use the estimation sample for the summary standardized index of health (i.e., those with non-missing variables for all of the individual health measures that compose the index). Results are robust if we instead use a larger sample consisting of anyone who has a non-missing value for any individual health measure we consider.

6.1 Risk Preferences

Risk aversion is negatively and significantly associated with smoking, heavy drinking, and being overweight or obese (e.g., Barsky et al. 1997, Khwaja et al. 2006, Anderson and Mellor 2008, Ida and Goto 2009, Sutter et al. 2013). If individuals with higher genetic risk for smoking tend to be less risk averse (i.e., more risk prone), they may engage in fewer health-promoting behaviors and more risky behaviors (besides smoking), which adversely impact health. We explore whether there is a relationship between risk preferences and the smoking polygenic scores.

From 1998–2008, the HRS elicited information on risk tolerance based on hypothetical income gambles. Individuals make a choice between a guaranteed stream of income and a 50-50 gamble that doubles that amount of income or cuts it by various amounts (10, 20, 33, 50, and 75 percent). The questions are asked such that they separate individuals into six distinct risk aversion (RA) categories ordered from least (category 1) to most (category 6) risk averse.[26] We create an indicator variable equal to one if the individual always makes the most risk averse decision in the gambles (i.e., RA category 6) as well as separate indicators for an RA category greater than or equal to 5, 4, 3, or 2. For example, those with an RA category greater than or equal to 4 compose the most, second most, and third most risk averse individuals.

We use other risk-related questions available in the 2014 and 2016 survey waves. The HRS asks "Are you generally a person who tries to avoid taking risks or one who is fully prepared to take risks?" Respondents answer on a 0–10 scale, where 0 means not willing to take any risks and 10 means fully willing to take risks. We create an indicator variable for whether respondents are generally *not* willing to take risk, corresponding to a response of 0 or 1 on the 0–10 scale, as well as an indicator variable for being very willing to take risk, corresponding to a response of 9 or 10. We also consider the 0–10 score itself as an outcome.[27]

Table 8 presents the estimated relationship between various measures of risk aversion and the smoking polygenic scores. Panels A, B, and C show the results for the full, ever smoker, and never smoker samples, respectively. Generally, we find little to no evidence that genetic risk for smoking correlates significantly with risk preferences. One exception is that the CPD score is associated

[26]The least risk averse category corresponds to those who would take a job with even chances of doubling their income or cutting it by 75 percent, while the most risk averse category corresponds to always choosing the guaranteed stream of income. Risk categories 2, 3, 4, and 5 consist of those who would take a job with even chances of doubling income or cutting it by 50 percent, a third, 20 percent, and 10 percent, respectively.

[27]The HRS also asks "How willing are you to take risks [while driving/in financial matters/during leisure and sport/in your occupation/with your health]?" These questions were asked only in the 2014 and 2016 waves via a leave-behind questionnaire that participants complete and return by mail; thus, the number of individuals with non-missing information is relatively small. We find no significant correlation between the smoking polygenic scores and these measures of risk aversion. Results are available by request.

with a higher likelihood of exhibiting the highest degree of risk aversion based on the income gambles, especially for never smokers. If anything, this result runs counter to the hypothesis that the relationship between genetic endowments for smoking and health is explained by less risk aversion among those with higher polygenic smoking scores. Overall, the results do not reveal a systematic relationship between risk preferences and the polygenic scores; thus, risk tolerance does not appear to explain the estimated relationship between the smoking polygenic scores and health.

6.2 Longevity Expectations

Longevity expectations are another channel that may explain the correlation between the smoking polygenic scores and health. An increase in expected longevity should increase the incentive to invest in health, a mechanism referred to as the "Mickey Mantle effect" (Fang et al. 2007). As explained in Fang et al. (2007), if one expects to live longer, there is more incentive to lead a healthier lifestyle to improve quality of life in old age.[28] A priori, it is unclear how genetic risk for smoking correlates with longevity expectations, though the literature has documented systematic differences in such beliefs by smoking status (e.g., Khwaja et al. 2007b, Wang 2014, Bissonnette et al. 2017). For example, Khwaja et al. (2007b) use the HRS and find smokers' longevity expectations are relatively optimistic (compared to objectively estimated probabilities) while never smokers are relatively pessimistic.[29] We explore whether the genetic risk for smoking correlates with longevity expectations (though not the accuracy of those beliefs). Given the "Mickey Mantle effect" above, those who expect to live longer may make larger investments in health and engage in more health-promoting activities, which would lead to better health outcomes.

In each survey wave, the HRS asks respondents younger than 65 about the probability of living to age 75, with answers on a 0–100 scale. We examine how these responses correlate with the genetic endowments for smoking. Results for the full sample, ever smokers, and never smokers are presented in Table 9. We find a one standard deviation increase in the CPD score decreases the self-reported probability of living to age 75 by about 1–1.7 percentage points and the estimates are significant at the 1 percent level. The magnitude of the decline is larger in the ever smoker sample. We do not find a statistically significant relationship between the initiation score and the self-reported probability of living to age 75 in any of the samples.

[28] In a similar vein, Picone et al. (2004) find individuals who expect to live longer are more likely to undergo screenings for early detection of breast and cervical cancer.

[29] Wang (2014) finds important differences when comparing adult smokers' own subjective longevity expectations to objective ones estimated using a rational expectations framework. In particular, they attach less weight to their health conditions and smoking choices and more weight to factors like age, race, and parents' longevity. Bissonnette et al. (2017) find at the median, smokers and non-smokers in the HRS correctly perceive their remaining life expectancy, but the distributions have heavy tails, such that on average, smokers are too optimistic.

Since we find evidence of a correlation between longevity expectations and genetic risk for smoking, we reestimate the health specifications including the respondent's self-reported probability of living to age 75. The results are presented in Table 10. An increase in the self-reported probability of living to age 75 is associated with an improvement in the summary health index, consistent with the "Mickey Mantle effect." Furthermore, the inclusion of longevity expectations decreases the magnitude of the point estimate on the CPD score by 13–21 percent and leaves the initiation score point estimate largely unchanged. Thus, the negative association between longevity expectations and the CPD score explains part of the estimated relationship between the CPD score and health. However, there still remains a non-trivial, positive, and significant link between both smoking scores and health after accounting for longevity expectations.

6.3 Planning Horizons

Another mechanism through which the smoking polygenic scores may operate is time preferences. Those with higher discount rates (i.e., more impatient, less future-oriented) derive less value from future health and are less likely to invest in their health and more likely to engage in risky behaviors.[30] Indeed, some studies find a significant association between higher rates of time preference and smoking (e.g., Khwaja et al. 2006, 2007b, Scharff and Viscusi 2011).[31] Using data from the Survey on Smoking, Khwaja et al. (2007b) find subjective rates of time discounting obtained via committed choice scenarios do not differ by smoking status, but more general measures of self-control, impulsivity, and financial planning do. In particular, those who are more impulsive and plan less for the future are more likely to smoke. They conclude that smoking may be a marker for problems of self-control.

We explore whether and how the smoking polygenic scores associate with financial planning horizons, noting that these horizons likely capture a combination of present (versus future) orientation, planning abilities, and/or more general aspects of self-control. We consider personality measures related to conscientiousness in the next subsection.

The HRS asks individuals "In planning your (family's) saving and spending, which of the following time periods is most important to you (and your husband/and your wife/and your partner), the next few months, the next year, the next few years, the next 5–10 years, or longer than 10 years?" Similar to Barth et al. (2020), we construct indicator variables that equal one if the respondent's financial planning horizon is longer than next year, next few years, next 5–10

[30]For example, Picone et al. (2004) find those with lower time preference, as measured by financial planning horizons, are more likely to undergo cancer screening.

[31]Scharff and Viscusi (2011) examine workers' fatality risk-wage decisions and find smokers have significantly higher rates of time preference than non-smokers with respect to years of life. Khwaja et al. (2006) find smokers in the HRS are more impatient than non-smokers, measured indirectly in terms of their financial planning horizons.

years, or longer than 10 years, respectively, and zero otherwise.[32]

Table 11 presents the estimated relationship between planning horizons and the smoking polygenic scores for the full sample (Panel A), ever smokers (Panel B), and never smokers (Panel C). In the full sample, a one standard deviation increase in either the CPD score or the initiation score is significantly associated with about a 0.8 percentage point decrease in the probability of reporting planning horizons greater than one year or the next few years. The point estimates in the ever and never smoker samples are similar to those in the full sample but generally not statistically significant. The exception is a higher CPD score is associated with a significant decline in the probability of ever smokers having a planning horizon longer than one year. Thus, there is evidence that a higher genetic risk for smoking correlates with shorter planning horizons. We reestimate our baseline health specifications including mutually exclusive indicator variables for reported financial planning horizon. Results are presented in Table 12. Longer financial planning horizons are associated with better health (i.e., a lower summary health index), but the inclusion of these horizons decreases the magnitude of the CPD and initiation score coefficients only slightly. Thus, planning horizons play a very small role in explaining the link between the smoking polygenic scores and health.

6.4 Impulsivity and Conscientiousness

The final channels we explore are impulsivity and conscientiousness. Khwaja et al. (2007a) find current and former smokers are significantly more impulsive than never smokers based on an index that measures ability to set goals and exercise self-control from the Survey on Smoking.[33] Studies find conscientiousness, one of the "big five" personality traits, is a good proxy for self-control (Ameriks et al. 2007) and is negatively correlated with smoking (e.g., Terracciano and Costa Jr 2004, Malouff et al. 2006). Conscientiousness is also positively correlated with a range of health-promoting behaviors like seeking preventative care and healthy eating (see reviews by Bogg and Roberts 2004, 2013). If the genetic endowments for smoking reflect impulsiveness and conscientiousness, these characteristics may be important mediators of the relationship between the smoking polygenic scores and health.

In various waves, the HRS ask questions related to personality traits to subsets of the sample via leave-behind questionnaires that respondents mail back to the HRS. We construct a variety of measures based on these questions. From 2010–2016, 31 items assessed the "big five" personality traits of neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness. Respondents indicated how well various adjectives described them on a scale of 1 (a lot) to 4 (not

[32]The financial planning horizon question is not asked in all waves nor to all respondents. Khwaja et al. (2006) describe the availability of this information across the survey waves.

[33]The index is based on the degree to which individuals agree with statements like "I make hasty decisions," "I do not control my temper," and "I act on impulse."

at all). We focus on conscientiousness, which is based on the following adjectives: organized, responsible, hardworking, careless, thorough, reckless, self-disciplined, impulsive, cautious, and thrifty.[34] Scales of the original adjectives are reversed so that 1 corresponds to not at all and 4 corresponds to a lot except for the careless, reckless, and impulsive adjectives. We then average the scores across the conscientiousness adjectives, and we refer to this measure as the big five conscientiousness score. The score ranges from 1–4, with a higher score indicating a higher degree of conscientiousness. We also create a score based on items the HRS included in 2010 and 2012 from the Multidimensional Personality Questionnaire (hereafter, the MPQ conscientiousness score) as well as a score based on an inventory developed in Roberts et al. (2005) (hereafter, RCSG conscientiousness score) that reflects six facets of conscientiousness (self-control, traditionalism, responsibility, industriousness, order, and virtue) included in the HRS in 2008 and 2010. Both the MPQ and RCSG conscientiousness scores range from 1–6 and a higher score indicates a higher degree of conscientiousness.[35]

In Table 13, we present estimated associations between the smoking polygenic scores and the various conscientiousness measures. For the full sample (columns 1–3), we find the CPD score is significantly and negatively correlated with the three measures we consider. That is, a higher genetic risk for heavy smoking is correlated with a lower degree of conscientiousness (i.e., more impulsiveness). We also find a negative relationship between the initiation score and the conscientiousness scores, though the point estimates are smaller and somewhat less precisely estimated. When we turn to the ever and never smoker samples, the picture is more complex. We always estimate a negative correlation between the polygenic scores and the conscientiousness measures, but the point estimates tend to be larger and more precisely estimated in the never smoker sample.[36]

Given the negative correlation between the smoking polygenic scores and the conscientiousness scores, we reestimate our health specifications including the conscientiousness measures. Results are shown in Table 14. Notably, conscientiousness has a large and statistically significant negative correlation with the summary standardized index of health. A one unit increase in the conscientiousness measures is associated with a 0.1–0.3 standard deviation decline (i.e., improvement) in the health index. Generally, the inclusion of the conscientiousness scores decreases

[34] In 2006 and 2008, the conscientiousness scale included only five adjectives (organized, responsible, hardworking, careless, thorough). The HRS added adjectives in 2010 to expand coverage of the sub-facets of conscientiousness. We prefer the scale based on the expanded set of items.

[35] The MPQ score is based on how strongly a respondent agrees with statements like "I often stop one thing before completing it and start another," "I often act without thinking," and "I am often not as cautious as I should be." The RCSG score is based on how strongly a respondent agrees with statements like "I rarely jump into something without first thinking about it," "I hardly ever lose or misplace things," and "I have high standards and work toward them." See Smith et al. (2017) for a detailed description of the various measures and their construction.

[36] These results are consistent with Hicks et al. (2021), which finds higher smoking initiation scores associate with behavioral disinhibition, including low conscientiousness, among adolescents.

the magnitude of the smoking polygenic score point estimates. Thus, the observed relationship between the smoking genetic endowments and health is mediated, in part, by impulsiveness and conscientiousness. However, there still remains a statistically significant relationship between the smoking polygenic scores and health, and this relationship is particularly robust for the CPD score.

6.5 Synthesis

We next explore how much of the association between the genetic endowments for smoking and health remains after including controls for risk aversion, longevity expectations, planning horizons, and conscientiousness simultaneously.[37] Results are shown in Table 15. The inclusion of these measures dampens the associations between the summary health index and the CPD score, but by less than 25 percent. Similarly, for never smokers, over 75 percent of the association between the initiation score and the health index remains. We then control for these mechanisms as well as parental and spousal variables, so the sample only includes married individuals where both members were genotyped and aged 50–65. We again find that about 75 percent of the association between the CPD score and health summary measure remains (see Table 16). For never smokers, the association between the initiation score shrinks by 45 percent and is no longer statistically significant. In sum, the potential mediating variables we study explain some of the relationship between the genetic endowments for smoking and health, but much of the relationship remains.

7 Robustness Exercises

We next explore the robustness of our results to different specifications and samples. For brevity, we present results for the summary standardized health index. Results for the individual health outcomes are available by request.

We first examine whether the associations between the genetic endowments for smoking and health are robust to alternate polygenic scores. We consider the CPD and initiation scores from a 2010 GWAS conducted by the Tobacco and Genetics Consortium (The Tobacco and Genetics Consortium 2010) that used a much smaller discovery sample compared to Liu et al. (2019). The results are presented in Appendix Table A7 Panel A. The estimated associations are smaller in magnitude but qualitatively similar to those in Table 2, which use the scores built from Liu et al. (2019). We still find the CPD score is associated with worse health for both ever and never smokers, and a higher initiation score is associated with worse health for never smokers.

[37] Information about longevity expectations is available in all waves, but risk aversion, planning horizons, and conscientiousness measures are not. For the above exercise only, we forward-fill as many missing values as possible and backward-fill the remaining. We also performed this exercise using data only from 1998–2006 as we only need to impute the conscientiousness measures in those years. The results are nearly identical to those reported here and are available by request.

In addition to smoking initiation and heaviness, Liu et al. (2019) identified genetic variants associated with smoking cessation and age of initiation of regular smoking. We do not consider these polygenic scores in our main analysis for several reasons. First, the older polygenic scores available in the HRS, which we used above to explore the robustness of our results, are only constructed for smoking initiation and smoking intensity. Second, in the Liu et al. (2019) GWAS, the scores related to smoking cessation and age of smoking initiation were substantially less predictive in- and out-of-sample than the smoking initiation and smoking intensity scores, explaining about 1 percent of the variation in their respective outcomes in the HRS (versus 4 percent for the initiation and CPD scores). For the sake of completeness, we present results including all four smoking-related polygenic scores based on the GWAS estimates in Liu et al. (2019). A higher cessation score implies a higher propensity for current smoking (rather than being a former smoker) and a higher initiation age score reflects a later age of initiation. The results are shown in Appendix Table A7 Panel B. Once we include the smoking controls, we find no significant association between the initiation age score and the summary health index. A higher cessation score (a lower propensity to quit) is significantly associated with worse health. The inclusion of the cessation and initiation age scores slightly dampens the point estimates on the CPD and initiation scores, but the main takeaways remain—the CPD score is correlated with worse health for both ever and never smokers, and the initiation score is correlated with worse health for the never smokers.

Our main estimation sample includes those aged 50–65. Individuals may claim Social Security retirement benefits as early as age 62, thus, our sample contains some individuals who are retired. Given the potential influence of retirement on health, we explore the robustness of our results to restricting the sample to those younger than the early retirement age, i.e., those aged 50–61. The results are presented in Appendix Table A7 Panel C and they are nearly identical to the full sample results.

We also found little to no change in the standard errors when clustering at the household level rather than the individual level. Those results are available by request.

8 Discussion and Conclusion

We study how genetic endowments linked to smoking associate with health using rich data from the Health and Retirement Study. Among individuals aged 50–65, we find a higher genetic predisposition for smoking associates with worse health, even after flexibly controlling for individual smoking behavior and even among those who have never smoked. The results suggest the smoking polygenic scores correlate with health through non-smoking channels. Several of the health outcomes where we find robust and meaningful relationships (e.g., diabetes, arthritis, body mass index, obesity) have

a strong link to health behaviors like diet and exercise. Our results are consistent with the idea that the smoking polygenic scores capture a complex array of traits that correlate with less engagement in health-promoting activities. This hypothesis is further supported by our finding that a spouse's genetic predisposition toward heavy smoking adversely and independently correlates with own health, consistent with a spouse's propensity toward health-degrading behaviors generating negative spillovers. While our results are descriptive in nature, they shed light on the black box of time-invariant unobserved heterogeneity emphasized in examinations of the effects of smoking on health (e.g., Darden 2017, Darden et al. 2018) as they highlight a non-trivial role for genetic endowments.

An advantage of the genetic data is we can learn who may be most affected by changes in behavior. While we find significant associations between the smoking polygenic scores and health for both ever and never smokers, the magnitudes vary across these groups and by one's place in the polygenic score distributions. In particular, those with relatively low initiation scores who never smoke exhibit better health, while ever smokers with low initiation scores do not experience such benefits. The results also suggest that both ever smokers and never smokers with relatively high CPD scores experience worse health outcomes, but the deterioration in health is smaller for never smokers. Thus, efforts to prevent those with a high risk for heavy smoking from ever starting could have important health benefits beyond those directly associated with not smoking.

One concern with interpretation of our results involves pleiotropy. That is, genetic variants can have multiple functions and affect multiple outcomes. One particular concern is that genetic markers that correlate with smoking also correlate with biological mechanisms that impact health outcomes (e.g., poor cholesterol metabolism or lipid transport). While we cannot dismiss this notion, we do not think it solely drives our results.[38] First, as noted above, several (though not all) of the genetic markers implicated in the Liu et al. (2019) GWAS are involved in systems related to neurotransmission, reward-related learning and memory, and stress response. Second, an increase in a spouse's CPD score is associated with declines in own health, including among couples where neither member ever smoked. If the smoking polygenic scores operated solely through biological mechanisms, we should not observe spousal spillovers. Furthermore, to the extent that genetic variants that predict smoking overlap with those that predict other health outcomes, our analysis sheds light on the mechanisms through which those variants link to health outcomes.

We note some important limitations of our work. First, we are quick to caution that our results should be interpreted as descriptive, not causal, relationships. Second, we shed some light on why those with higher smoking polygenic scores exhibit worse health, but there is

[38]Liu et al. (2019) document correlations between the smoking polygenic scores and polygenic scores for other outcomes and traits. Of note, the CPD and initiation scores are statistically significantly and positively correlated with polygenic scores for BMI and obesity, coronary artery disease, and proinsulin.

much left to understand. Parental smoking, parental risky behaviors, parental mortality, spousal smoking behavior, longevity expectations, planning horizons, and conscientiousness play a role, but do not fully explain the health-smoking genes relationship. A more complete understanding of the channels underlying the relationship between health and the smoking polygenic scores may allow policy-makers to better design policies aimed at improving health, particularly for those disproportionately endowed with a propensity to smoke. For example, to the extent the genetic endowments for smoking reflect differences in time preferences or impulsiveness, this has implications for the design of interventions like exercise incentives and workplace wellness programs.

References

Agrawal, A. and M. T. Lynskey (2008): "Are there genetic influences on addiction: Evidence from family, adoption and twin studies," *Addiction*, 103, 1069–1081.

Ameriks, J., A. Caplin, J. Leahy, and T. Tyler (2007): "Measuring self-control problems," *American Economic Review*, 97, 966–972.

Anderson, L. R. and J. M. Mellor (2008): "Predicting health behaviors with an experimental measure of risk preference," *Journal of Health Economics*, 27, 1260–1274.

Barsky, R. B., F. T. Juster, M. S. Kimball, and M. D. Shapiro (1997): "Preference parameters and behavioral heterogeneity: An experimental approach in the Health and Retirement Study," *The Quarterly Journal of Economics*, 112, 537–579.

Barth, D., N. W. Papageorge, and K. Thom (2020): "Genetic endowments and wealth inequality," *Journal of Political Economy*, 128, 1474–1522.

Beauchamp, J. P., D. Cesarini, M. Johannesson, M. J. van der Loos, P. D. Koellinger, P. J. Groenen, J. H. Fowler, J. N. Rosenquist, A. R. Thurik, and N. A. Christakis (2011): "Molecular genetics and economics," *Journal of Economic Perspectives*, 25, 57–82.

Behncke, S. (2012): "Does retirement trigger ill health?" *Health Economics*, 21, 282–300.

Belsky, D. W., T. E. Moffitt, D. L. Corcoran, B. Domingue, H. Harrington, S. Hogan, R. Houts, S. Ramrakha, K. Sugden, B. S. Williams, et al. (2016): "The genetics of success: How single-nucleotide polymorphisms associated with educational attainment relate to life-course development," *Psychological Science*, 27, 957–972.

Benjamin, D. J., D. Cesarini, C. F. Chabris, E. L. Glaeser, D. I. Laibson, V. Guðnason, T. B. Harris, L. J. Launer, S. Purcell, et al. (2011): "The promises and pitfalls of genoeconomics," *Annual Review of Economics*, 4, 627–662.

Bierut, L., P. Biroli, T. J. Galama, and K. Thom (2023): "Challenges in Studying the Interplay of Genes and Environment: A Study of Childhood Financial Distress Moderating Genetic Predisposition for Peak Smoking," *Journal of Economic Psychology*, 102636.

Biroli, P. and L. Zwyssig (2021): "Moral Hazard Heterogeneity: Genes and Health Insurance Influence Smoking after a Health Shock," IZA Discussion Paper No. 14176.

Bissonnette, L., M. D. Hurd, and P.-C. Michaud (2017): "Individual survival curves comparing subjective and observed mortality risks," *Health Economics*, 26, e285–e303.

Björkegren, E., M. Lindahl, M. Palme, and E. Simeonova (2022): "Pre-and post-birth components of intergenerational persistence in health and longevity: Lessons from a large sample of adoptees," *Journal of Human Resources*, 57, 112–142.

Boardman, J. D., C. L. Blalock, F. C. Pampel, P. K. Hatemi, A. C. Heath, and L. J. Eaves (2011): "Population composition, public policy, and the genetics of smoking," *Demography*, 48, 1517–1533.

Bogg, T. and B. W. Roberts (2004): "Conscientiousness and health-related behaviors: A meta-analysis of the leading behavioral contributors to mortality." *Psychological Bulletin*, 130, 887.

——— (2013): "The case for conscientiousness: Evidence and implications for a personality trait marker of health and longevity," *Annals of Behavioral Medicine*, 45, 278–288.

Bolyard, A. and P. A. Savelyev (2021): "Understanding the educational attainment polygenic score and its interactions with SES in determining health in young adulthood," SSRN Working Paper 3397735.

Brunello, G., A. Sanz-de Galdeano, and A. Terskaya (2020): "Not only in my genes: The effects of peers' genotype on obesity," *Journal of Health Economics*, 72, 102349.

Bütikofer, A., J. Riise, and M. Skira (2021): "The Impact of Paid Maternity Leave on Maternal Health." *American Economic Journal: Economic Policy*, 13.

Cawley, J., E. Han, J. Kim, and E. C. Norton (2019): "Testing for family influences on obesity: The role of genetic nurture," *Health Economics*, 28, 937–952.

——— (2023): "Genetic nurture in educational attainment," *Economics & Human Biology*, 49, 101239.

Chabris, C. F., D. Laibson, C. L. Morris, J. P. Schuldt, and D. Taubinsky (2008): "Individual laboratory-measured discount rates predict field behavior," *Journal of Risk and Uncertainty*, 37, 237–269.

Charles, K. K., E. Hurst, and A. Killewald (2013): "Marital sorting and parental wealth," *Demography*, 50, 51–70.

Chiappori, P.-A., S. Oreffice, and C. Quintana-Domeque (2012): "Fatter attraction: Anthropometric and socioeconomic matching on the marriage market," *Journal of Political Economy*, 120, 659–695.

——— (2018): "Bidimensional matching with heterogeneous preferences: Education and smoking in the marriage market," *Journal of the European Economic Association*, 16, 161–198.

Clark, A. E. and F. Etilé (2006): "Don't give up on me baby: Spousal correlation in smoking behaviour," *Journal of Health Economics*, 25, 958–978.

Classen, T. J. and O. Thompson (2016): "Genes and the intergenerational transmission of BMI and obesity," *Economics & Human Biology*, 23, 121–133.

Coe, N. B. and G. Zamarro (2011): "Retirement effects on health in Europe," *Journal of Health Economics*, 30, 77–86.

Darden, M. (2017): "Smoking, Expectations, and Health: A Dynamic Stochastic Model of Lifetime Smoking Behavior," *Journal of Political Economy*, 125, 1465–1522.

Darden, M. and D. Gilleskie (2016): "The effects of parental health shocks on adult offspring smoking behavior and self-assessed health," *Health Economics*, 25, 939–954.

Darden, M., D. B. Gilleskie, and K. Strumpf (2018): "Smoking and Mortality: New Evidence from a Long Panel," *International Economic Review*, 59, 1571–1619.

Domingue, B. W. and D. W. Belsky (2017): "The social genome: Current findings and implications for the study of human genetics," *PLoS Genetics*, 13, e1006615.

Domingue, B. W., D. Conley, J. Fletcher, and J. D. Boardman (2016): "Cohort effects in the genetic influence on smoking," *Behavior Genetics*, 46, 31–42.

Eibich, P. (2015): "Understanding the effect of retirement on health: Mechanisms and heterogeneity," *Journal of Health Economics*, 43, 1–12.

Fang, H., M. Keane, A. Khwaja, M. Salm, and D. Silverman (2007): "Testing the mechanisms of structural models: The case of the Mickey Mantle effect," *American Economic Review*, 97, 53–59.

Fletcher, J. M. (2012): "Why have tobacco control policies stalled? Using genetic moderation to examine policy impacts," *PloS One*, 7, e50576.

——— (2018): "Economics and Genetics," in *Oxford Research Encyclopedia of Economics and Finance*.

Godard, M. (2016): "Gaining weight through retirement? Results from the SHARE survey," *Journal of Health Economics*, 45, 27–46.

Gorry, A., D. Gorry, and S. N. Slavov (2018): "Does retirement improve health and life satisfaction?" *Health Economics*, 27, 2067–2086.

Hicks, B. M., D. A. Clark, J. D. Deak, M. Liu, C. E. Durbin, J. D. Schaefer, S. Wilson, W. G. Iacono, M. McGue, and S. I. Vrieze (2021): "Polygenic score for smoking is associated with externalizing psychopathology and disinhibited personality traits but not internalizing psychopathology in adolescence," *Clinical Psychological Science*, 9, 1205–1213.

Hoynes, H., D. W. Schanzenbach, and D. Almond (2016): "Long-run impacts of childhood access to the safety net," *American Economic Review*, 106, 903–34.

Ida, T. and R. Goto (2009): "Simultaneous measurement of time and risk preferences: Stated preference discrete choice modeling analysis depending on smoking behavior," *International Economic Review*, 50, 1169–1182.

Insler, M. (2014): "The health consequences of retirement," *Journal of Human Resources*, 49, 195–233.

Jeong, Y. (2023): "The Effect of Peers' Genetic Predisposition to Depression on Own Mental Health," Working Paper.

Khwaja, A., D. Silverman, and F. Sloan (2007a): "Time preference, time discounting, and smoking decisions," *Journal of Health Economics*, 26, 927–949.

Khwaja, A., F. Sloan, and S. Chung (2007b): "The relationship between individual expectations and behaviors: Mortality expectations and smoking decisions," *Journal of Risk and Uncertainty*, 35, 179–201.

Khwaja, A., F. Sloan, and M. Salm (2006): "Evidence on preferences and subjective beliefs of risk takers: The case of smokers," *International Journal of Industrial Organization*, 24, 667–682.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007): "Experimental analysis of neighborhood effects," *Econometrica*, 75, 83–119.

Kong, A., G. Thorleifsson, M. L. Frigge, B. J. Vilhjalmsson, A. I. Young, T. E. Thorgeirsson, S. Benonisdottir, A. Oddsson, B. V. Halldorsson, G. Masson, et al. (2018): "The nature of nurture: Effects of parental genotypes," *Science*, 359, 424–428.

Lehrer, S. F. and W. Ding (2017): "Are genetic markers of interest for economic research?" *IZA Journal of Labor Policy*, 6, 2.

Li, M. D., R. Cheng, J. Z. Ma, and G. E. Swan (2003): "A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins," *Addiction*, 98, 23–31.

Linnér, R. K. and P. D. Koellinger (2022): "Genetic risk scores in life insurance underwriting," *Journal of Health Economics*, 81, 102556.

Liu, M., Y. Jiang, R. Wedow, Y. Li, D. M. Brazel, F. Chen, G. Datta, J. Davila-Velderrain, D. McGuire, C. Tian, et al. (2019): "Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use," *Nature Genetics*, 51, 237–244.

Malouff, J. M., E. B. Thorsteinsson, and N. S. Schutte (2006): "The five-factor model of personality and smoking: A meta-analysis," *Journal of Drug Education*, 36, 47–58.

Mazzonna, F. and F. Peracchi (2017): "Unhealthy retirement?" *Journal of Human Resources*, 52, 128–151.

Meyers, J., M. Cerda, S. Galea, K. Keyes, A. Aiello, M. Uddin, D. Wildman, and K. Koenen (2013): "Interaction between polygenic risk for cigarette use and environmental exposures in the Detroit neighborhood health study," *Translational Psychiatry*, 3, e290–e290.

Oreffice, S. and C. Quintana-Domeque (2010): "Anthropometry and socioeconomics among couples: Evidence in the United States," *Economics & Human Biology*, 8, 373–384.

Otten, K. and J. J. Mandemakers (2023): "Partners in Health: Investigating Social Genetic Effects Among Married and Cohabiting Couples," *Behavior Genetics*, 1–11.

Papageorge, N. W. and K. Thom (2020): "Genes, education, and labor market outcomes: Evidence from the Health and Retirement Study," *Journal of the European Economic Association*, 18, 1351–1399.

Picone, G., F. Sloan, and D. Taylor (2004): "Effects of risk and time preference and expected longevity on demand for medical tests," *Journal of Risk and Uncertainty*, 28, 39–53.

Polderman, T. J., B. Benyamin, C. A. De Leeuw, P. F. Sullivan, A. Van Bochoven, P. M. Visscher, and D. Posthuma (2015): "Meta-analysis of the heritability of human traits based on fifty years of twin studies," *Nature Genetics*, 47, 702–709.

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich (2006): "Principal components analysis corrects for stratification in genome-wide association studies," *Nature Genetics*, 38, 904–909.

Roberts, B. W., O. S. Chernyshenko, S. Stark, and L. R. Goldberg (2005): "The structure of conscientiousness: An empirical investigation based on seven major personality questionnaires," *Personnel Psychology*, 58, 103–139.

Scharff, R. L. and W. K. Viscusi (2011): "Heterogeneous rates of time preference and the decision to smoke," *Economic Inquiry*, 49, 959–972.

Schmitz, L. and D. Conley (2016): "The long-term consequences of Vietnam-era conscription and genotype on smoking behavior and health," *Behavior Genetics*, 46, 43–58.

Shai, O. (2018): "Is retirement good for men's health? Evidence using a change in the retirement age in Israel," *Journal of Health Economics*, 57, 15–30.

Slob, E. A. W. and C. A. Rietveld (2021): "Genetic predispositions moderate the effectiveness of tobacco excise taxes," *PLoS One*, 16, e0259210.

Smith, J., G. Fisher, L. Ryan, A. Sonnega, and D. Weir (2017): "Psychosocial and lifestyle questionnaire," Survey Research Center, Institute for Social Research at University of Michigan.

Sotoudeh, R., K. M. Harris, and D. Conley (2019): "Effects of the peer metagenomic environment on smoking behavior," *Proceedings of the National Academy of Sciences*, 116, 16302–16307.

Sutter, M., M. G. Kocher, D. Glätzle-Rützler, and S. T. Trautmann (2013): "Impatience and uncertainty: Experimental decisions predict adolescents' field behavior," *American Economic Review*, 103, 510–31.

Terracciano, A. and P. T. Costa Jr (2004): "Smoking and the Five-Factor Model of Personality," *Addiction*, 99, 472–481.

The Tobacco and Genetics Consortium (2010): "Genome-wide meta-analyses identify multiple loci associated with smoking behavior," *Nature Genetics*, 42, 441–447.

Thompson, O. (2014): "Genetic mechanisms in the intergenerational transmission of health," *Journal of Health Economics*, 35, 132–146.

US Department of Health and Human Services (2020): "Smoking Cessation: A Report of the Surgeon General," Tech. rep., US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Atlanta, GA.

Wang, Y. (2014): "Dynamic implications of subjective expectations: Evidence from adult smokers," *American Economic Journal: Applied Economics*, 6, 1–37.

Ware, E. B., L. L. Schmitz, J. Faul, A. Gard, C. Mitchell, J. A. Smith, W. Zhao, D. Weir, and S. L. Kardia (2017): "Heterogeneity in polygenic scores for common human traits," *BioRxiv*, 106062.

Wedow, R., M. Zacher, B. M. Huibregtse, K. Mullan Harris, B. W. Domingue, and J. D. Boardman (2018): "Education, smoking, and cohort change: Forwarding a multidimensional theory of the environmental moderation of genetic effects," *American Sociological Review*, 83, 802–832.

Wehby, G. (2016): "How genetics can inform health economics," *The Oxford Handbook of Economics and Human Biology*, 366.

Tables and Figures

Table 1: Summary Statistics

	Full Sample			Ever Smoker Sample			Never Smoker Sample			
	Mean	SD	N	Mean	SD	N	Mean	SD	N	p-value
Demographics:										
Age	58.56	4.16	46,692	58.60	4.16	26,963	58.51	4.16	19,729	0.02
Birth Year	1945.24	7.64	46,692	1944.96	7.65	26,963	1945.63	7.61	19,729	0.00
Male	0.42	0.49	46,692	0.48	0.50	26,963	0.34	0.47	19,729	0.00
Own Smoking:										
# Cigs per Day (Among Current Smokers)	17.55	10.99	8,427	17.55	10.99	8,427				
Maximum # Cigs per Day (Among Former Smokers)	22.25	17.83	16,560	22.25	17.83	16,560				
Age Start Smoking (Among Ever Smokers)	17.63	5.01	25,788	17.63	5.01	25,788				
Age Quit Smoking (Among Quitters)	36.80	11.26	16,316	36.80	11.26	16,316				
Smoked as a Teenager	0.38	0.49	26,391	0.38	0.49	26,391				
Own Health:										
Poor Self-Reported Health	0.17	0.38	46,676	0.20	0.40	26,952	0.13	0.33	19,724	0.00
Any Hospital Stay	0.17	0.38	46,624	0.19	0.39	26,922	0.15	0.36	19,702	0.00
Ever Had High Blood Pressure	0.39	0.49	46,659	0.40	0.49	26,932	0.37	0.48	19,727	0.00
Ever Had Diabetes	0.11	0.32	46,663	0.12	0.32	26,949	0.11	0.31	19,714	0.18
Ever Had Cancer	0.08	0.28	46,637	0.08	0.28	26,919	0.09	0.28	19,718	0.47
Ever Had Lung Disease	0.06	0.25	46,671	0.09	0.29	26,956	0.03	0.17	19,715	0.00
Ever Had Heart Problem	0.13	0.34	46,667	0.15	0.36	26,951	0.11	0.31	19,716	0.00
Ever Had Stroke	0.03	0.17	46,669	0.04	0.19	26,947	0.02	0.14	19,722	0.00
Ever Had Arthritis	0.45	0.50	46,654	0.47	0.50	26,945	0.42	0.49	19,709	0.00
Any Limitations with ADLs	0.08	0.27	46,669	0.09	0.29	26,952	0.06	0.24	19,717	0.00
Any Limitations with IADLs	0.04	0.19	46,667	0.04	0.20	26,951	0.03	0.16	19,716	0.00
Body Mass Index (kg/m^2)	28.12	5.73	46,115	28.05	5.65	26,681	28.23	5.82	19,434	0.00
Obese	0.31	0.46	46,115	0.31	0.46	26,681	0.32	0.47	19,434	0.00
Parental Smoking:										
Neither Parent Smoked	0.18	0.38	46,692	0.15	0.36	26,963	0.21	0.41	19,729	0.00
One Parent Smoked	0.30	0.46	46,692	0.30	0.46	26,963	0.29	0.45	19,729	0.00
Both Parents Smoked	0.19	0.39	46,692	0.21	0.41	26,963	0.16	0.36	19,729	0.00
Missing	0.34	0.47	46,692	0.33	0.47	26,963	0.34	0.48	19,729	0.01
Parental Mortality:										
Neither Parent Died During Childhood	0.22	0.41	46,692	0.21	0.41	26,963	0.23	0.42	19,729	0.00
Parent(s) Died During Childhood	0.04	0.20	46,692	0.04	0.20	26,963	0.04	0.19	19,729	0.11
Parental Mortality During Childhood Missing	0.74	0.44	46,692	0.75	0.43	26,963	0.73	0.44	19,729	0.00
Mother Currently Alive	0.38	0.49	46,692	0.37	0.48	26,963	0.39	0.49	19,729	0.00
Mother Died Before Age 65	0.14	0.35	46,692	0.15	0.35	26,963	0.13	0.34	19,729	0.00
Mother Died Age 65+	0.47	0.50	46,692	0.47	0.50	26,963	0.46	0.50	19,729	0.02
Mother Mortality Information Missing	0.02	0.13	46,692	0.02	0.13	26,963	0.02	0.12	19,729	0.14
Father Currently Alive	0.17	0.37	46,692	0.15	0.36	26,963	0.19	0.39	19,729	0.00
Father Died Before Age 65	0.25	0.43	46,692	0.26	0.44	26,963	0.23	0.42	19,729	0.00
Father Died Age 65+	0.57	0.50	46,692	0.57	0.50	26,963	0.56	0.50	19,729	0.39
Father Mortality Information Missing	0.02	0.14	46,692	0.02	0.15	26,963	0.01	0.12	19,729	0.00
Parental Risky Behavior During Respondent's Childhood:										
Parent(s) Drank/Used Drugs Often	0.20	0.40	46,692	0.23	0.42	26,963	0.17	0.38	19,729	0.00
Neither Parent Drank/Used Drugs	0.73	0.44	46,692	0.70	0.46	26,963	0.77	0.42	19,729	0.00
Missing	0.07	0.25	46,692	0.07	0.26	26,963	0.06	0.23	19,729	0.00

Note: The table presents summary statistics for the full sample and separately for those who ever and never smoked in their life. The sample includes those aged 50–65 of European ancestry who self-identify as white and who have been genotyped. The last column presents p-values from a test of equality of means for ever and never smokers.

Table 2: Relationship between the Smoking Polygenic Scores and the Summary Standardized Index of Health

	Full Sample			Ever Smoker Sample			Never Smoker Sample	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CPD Score	0.042***	0.043***	0.035***	0.048***	0.047***	0.037***	0.032***	0.032***
	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Initiation Score	0.026***	0.035***	0.021***	0.019***	0.023***	0.011	0.021***	0.030***
	(0.004)	(0.005)	(0.005)	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)
Full Smoking Controls	No	No	Yes	No	No	Yes	No	No
Standard Controls	No	Yes	Yes	No	Yes	Yes	No	Yes
Principal Components	No	Yes	Yes	No	Yes	Yes	No	Yes
Sample Mean	-0.000	-0.000	-0.000	0.038	0.038	0.038	-0.052	-0.052
N	45,850	45,850	45,850	26,509	26,509	26,509	19,341	19,341
R^2	0.014	0.070	0.116	0.013	0.081	0.142	0.010	0.080

Note: Each column presents results from a separate regression where the outcome is the summary standardized health index. Results are presented for the full sample and then separately for ever smokers and never smokers. For each sample, we first present results without controls (columns 1, 4, 7). Then we add the standard controls and the genetic principal components described in Section 4 (columns 2, 5, 8). Then for the full sample and ever smoker sample, we add the full set of smoking controls described in Section 4 (columns 3, 6). * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Table 3: Relationship between the Smoking Polygenic Scores and Individual Health Measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
			Ever Diagnosed With:							Any Limitations With:			
	Poor Self-Reported Health	Any Hospitalization	High Blood Pressure	Diabetes	Cancer	Lung Disease	Heart Problem	Stroke	Arthritis	ADLs	IADLs	BMI	Obese
						Panel A: Full Sample							
CPD Score	0.019***	0.009***	0.017***	0.019***	0.000	0.011***	0.007**	-0.000	0.021***	0.010***	0.004***	0.501***	0.033***
	(0.003)	(0.002)	(0.005)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.005)	(0.002)	(0.001)	(0.061)	(0.005)
Initiation Score	0.018***	0.008***	0.005	0.003	-0.005	0.009***	0.005	0.004**	0.004	0.007***	0.005***	0.232***	0.017***
	(0.003)	(0.003)	(0.005)	(0.003)	(0.003)	(0.003)	(0.004)	(0.002)	(0.005)	(0.002)	(0.001)	(0.067)	(0.005)
Full Smoking Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Mean	0.170	0.174	0.386	0.114	0.084	0.064	0.132	0.030	0.451	0.080	0.036	28.122	0.312
N	46,676	46,624	46,659	46,663	46,637	46,671	46,667	46,669	46,654	46,669	46,667	46,115	46,115
R^2	0.068	0.026	0.058	0.058	0.032	0.088	0.051	0.033	0.076	0.037	0.031	0.084	0.060
						Panel B: Ever Smoker Sample							
CPD Score	0.018***	0.013***	0.021***	0.016***	-0.001	0.018***	0.010**	-0.000	0.022***	0.010***	0.003*	0.386***	0.027***
	(0.004)	(0.003)	(0.007)	(0.004)	(0.004)	(0.004)	(0.005)	(0.003)	(0.007)	(0.003)	(0.002)	(0.078)	(0.006)
Initiation Score	0.015***	0.004	-0.002	-0.001	-0.006	0.011**	0.001	0.003	-0.001	0.004	0.006**	0.111	0.008
	(0.005)	(0.004)	(0.007)	(0.004)	(0.004)	(0.004)	(0.005)	(0.003)	(0.007)	(0.003)	(0.002)	(0.086)	(0.006)
Full Smoking Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Mean	0.202	0.189	0.398	0.115	0.083	0.090	0.149	0.038	0.470	0.093	0.043	28.046	0.305
N	26,952	26,922	26,932	26,949	26,919	26,956	26,951	26,947	26,945	26,952	26,951	26,681	26,681
R^2	0.086	0.035	0.083	0.075	0.048	0.103	0.067	0.043	0.089	0.051	0.044	0.121	0.086
						Panel C: Never Smoker Sample							
CPD Score	0.021***	0.004	0.011	0.024***	0.002	0.004	0.002	-0.001	0.017**	0.009***	0.004**	0.641***	0.041***
	(0.004)	(0.004)	(0.008)	(0.005)	(0.004)	(0.003)	(0.005)	(0.002)	(0.008)	(0.003)	(0.002)	(0.097)	(0.007)
Initiation Score	0.019***	0.013***	0.014*	0.008	-0.004	0.007**	0.008	0.006***	0.012	0.011***	0.003*	0.373***	0.026***
	(0.005)	(0.004)	(0.008)	(0.005)	(0.005)	(0.003)	(0.005)	(0.002)	(0.008)	(0.003)	(0.002)	(0.104)	(0.008)
Sample Mean	0.126	0.154	0.370	0.111	0.085	0.029	0.109	0.020	0.424	0.063	0.026	28.227	0.321
N	19,724	19,702	19,727	19,714	19,718	19,715	19,716	19,722	19,709	19,717	19,716	19,434	19,434
R^2	0.032	0.017	0.050	0.053	0.032	0.031	0.039	0.026	0.070	0.024	0.014	0.062	0.049

Note: Each column within a panel presents results from a separate regression where the outcome is noted in the column heading. Results for the full sample are presented in Panel A, for the ever smoker sample in Panel B, and for the never smoker sample in Panel C. In all specifications, we include the standard controls and the genetic principal components described in Section 4. We also include the full set of smoking controls described in Section 4 for the full sample and ever smoker samples. * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Table 4: Relationship between the Terciles of the Smoking Polygenic Scores and the Summary Standardized Index of Health

	Full Sample			Ever Smoker Sample			Never Smoker Sample	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CPD Score Tercile 1	-0.044***	-0.045***	-0.034***	-0.040***	-0.041***	-0.026*	-0.042***	-0.048***
	(0.010)	(0.010)	(0.010)	(0.014)	(0.014)	(0.014)	(0.015)	(0.014)
CPD Score Tercile 3	0.052***	0.052***	0.046***	0.063***	0.060***	0.051***	0.036**	0.030*
	(0.011)	(0.011)	(0.011)	(0.015)	(0.015)	(0.015)	(0.016)	(0.016)
Initiation Score Tercile 1	-0.022**	-0.034***	-0.022**	0.001	-0.005	0.003	-0.036**	-0.046***
	(0.010)	(0.011)	(0.011)	(0.015)	(0.015)	(0.015)	(0.014)	(0.015)
Initiation Score Tercile 3	0.035***	0.039***	0.020*	0.041***	0.038***	0.023	0.010	0.015
	(0.011)	(0.011)	(0.011)	(0.015)	(0.015)	(0.014)	(0.017)	(0.017)
Full Smoking Controls	No	No	Yes	No	No	Yes	No	No
Standard Controls	No	Yes	Yes	No	Yes	Yes	No	Yes
Principal Components	No	Yes	Yes	No	Yes	Yes	No	Yes
Sample Mean	-0.000	-0.000	-0.000	0.038	0.038	0.038	-0.052	-0.052
N	45,850	45,850	45,850	26,509	26,509	26,509	19,341	19,341
R^2	0.011	0.066	0.115	0.010	0.079	0.141	0.010	0.079

Note: Each column presents results from a separate regression where the outcome is the summary standardized health index. Results are presented for the full sample and then separately for ever smokers and never smokers. For each sample, we first present results without controls (columns 1, 4, 7). Then we add the standard controls and the genetic principal components described in Section 4 (columns 2, 5, 8). Then for the full sample and ever smoker sample, we add the full set of smoking controls described in Section 4 (columns 3, 6). Indicators for being in tercile 2 of the CPD score and tercile 2 of the initiation score are omitted. * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Table 5: Relationship between the Terciles of the Smoking Polygenic Scores and Individual Health Measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
			Ever Diagnosed With:							Any Limitations With:			
	Poor Self-Reported Health	Any Hospitalization	High Blood Pressure	Diabetes	Cancer	Lung Disease	Heart Problem	Stroke	Arthritis	ADLs	IADLs	BMI	Obese
						Panel A: Full Sample							
CPD Score Tercile 1	-0.017**	-0.009	-0.009	-0.021***	-0.002	-0.010*	-0.001	0.001	-0.022*	-0.008*	-0.002	-0.630***	-0.044***
	(0.007)	(0.006)	(0.012)	(0.007)	(0.007)	(0.005)	(0.008)	(0.004)	(0.012)	(0.005)	(0.003)	(0.143)	(0.011)
CPD Score Tercile 3	0.027***	0.010*	0.028**	0.024***	-0.005	0.016***	0.018**	-0.001	0.030**	0.015***	0.004	0.543***	0.038***
	(0.008)	(0.006)	(0.012)	(0.008)	(0.007)	(0.006)	(0.008)	(0.004)	(0.012)	(0.006)	(0.003)	(0.148)	(0.011)
Initiation Score Tercile 1	-0.010	-0.008	-0.014	0.000	0.001	-0.002	-0.001	-0.006	-0.017	-0.006	-0.006**	-0.417***	-0.034***
	(0.008)	(0.006)	(0.012)	(0.008)	(0.007)	(0.005)	(0.008)	(0.004)	(0.012)	(0.005)	(0.003)	(0.147)	(0.011)
Initiation Score Tercile 3	0.028***	0.011*	0.000	0.001	-0.007	0.014**	0.002	0.002	0.002	0.011**	0.005	0.079	0.001
	(0.008)	(0.006)	(0.012)	(0.008)	(0.007)	(0.006)	(0.008)	(0.004)	(0.012)	(0.006)	(0.003)	(0.150)	(0.011)
Full Smoking Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Mean	0.170	0.174	0.386	0.114	0.084	0.064	0.132	0.030	0.451	0.080	0.036	28.122	0.312
N	46,676	46,624	46,659	46,663	46,637	46,671	46,667	46,669	46,654	46,669	46,667	46,115	46,115
R^2	0.067	0.025	0.058	0.058	0.033	0.087	0.051	0.033	0.076	0.036	0.030	0.083	0.060
						Panel B: Ever Smoker Sample							
CPD Score Tercile 1	-0.020*	-0.012	-0.001	-0.016	0.002	-0.016*	0.005	-0.002	-0.023	-0.005	-0.002	-0.257	-0.021
	(0.011)	(0.008)	(0.016)	(0.010)	(0.009)	(0.008)	(0.011)	(0.006)	(0.016)	(0.007)	(0.004)	(0.181)	(0.014)
CPD Score Tercile 3	0.018*	0.014*	0.042***	0.022**	-0.006	0.023**	0.034***	-0.004	0.035**	0.013*	0.003	0.509***	0.035**
	(0.011)	(0.008)	(0.015)	(0.010)	(0.009)	(0.009)	(0.011)	(0.006)	(0.016)	(0.008)	(0.005)	(0.182)	(0.014)
Initiation Score Tercile 1	-0.005	0.005	-0.001	0.018	0.001	-0.005	0.008	0.001	-0.006	0.002	-0.004	-0.267	-0.031**
	(0.011)	(0.008)	(0.017)	(0.011)	(0.009)	(0.009)	(0.012)	(0.006)	(0.017)	(0.007)	(0.005)	(0.194)	(0.015)
Initiation Score Tercile 3	0.028***	0.011	-0.004	0.005	-0.007	0.014	-0.004	0.005	0.005	0.015**	0.009*	0.055	-0.010
	(0.010)	(0.008)	(0.015)	(0.010)	(0.009)	(0.009)	(0.011)	(0.006)	(0.015)	(0.007)	(0.005)	(0.183)	(0.014)
Full Smoking Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Mean	0.202	0.189	0.398	0.115	0.083	0.090	0.149	0.038	0.470	0.093	0.043	28.046	0.305
N	26,952	26,922	26,932	26,949	26,919	26,956	26,951	26,947	26,945	26,952	26,951	26,681	26,681
R^2	0.086	0.035	0.083	0.075	0.048	0.102	0.068	0.043	0.090	0.051	0.044	0.120	0.086
						Panel C: Never Smoker Sample							
CPD Score Tercile 1	-0.015	-0.006	-0.021	-0.031***	-0.008	-0.006	-0.011	0.006	-0.024	-0.013*	-0.003	-1.154***	-0.077***
	(0.010)	(0.008)	(0.018)	(0.011)	(0.010)	(0.006)	(0.011)	(0.005)	(0.018)	(0.007)	(0.004)	(0.228)	(0.017)
CPD Score Tercile 3	0.036***	0.002	0.005	0.025**	-0.004	0.002	-0.007	0.002	0.013	0.015*	0.004	0.511**	0.036*
	(0.012)	(0.009)	(0.019)	(0.012)	(0.010)	(0.007)	(0.012)	(0.005)	(0.019)	(0.008)	(0.005)	(0.249)	(0.018)
Initiation Score Tercile 1	-0.012	-0.022***	-0.028	-0.017	0.003	0.003	-0.011	-0.015***	-0.025	-0.015**	-0.008**	-0.531**	-0.034**
	(0.010)	(0.008)	(0.018)	(0.011)	(0.011)	(0.006)	(0.012)	(0.005)	(0.018)	(0.007)	(0.004)	(0.224)	(0.017)
Initiation Score Tercile 3	0.024**	0.011	0.004	-0.002	-0.010	0.017**	0.011	-0.003	0.004	0.005	-0.001	0.085	0.015
	(0.012)	(0.009)	(0.020)	(0.013)	(0.011)	(0.008)	(0.013)	(0.006)	(0.020)	(0.009)	(0.005)	(0.257)	(0.019)
Sample Mean	0.126	0.154	0.370	0.111	0.085	0.029	0.109	0.020	0.424	0.063	0.026	28.227	0.321
N	19,724	19,702	19,727	19,714	19,718	19,715	19,716	19,722	19,709	19,717	19,716	19,434	19,434
R^2	0.031	0.017	0.050	0.053	0.032	0.031	0.040	0.027	0.070	0.024	0.014	0.063	0.051

Note: Each column within a panel presents results from a separate regression where the outcome is noted in the column heading. Results for the full sample are presented in Panel A, for the ever smoker sample in Panel B, and for the never smoker sample in Panel C. In all specifications, we include the standard controls and the genetic principal components described in Section 4. We also include the full set of smoking controls described in Section 4 for the full sample and ever smoker samples. Indicators for being in tercile 2 of the CPD score and tercile 2 of the initiation score are omitted. * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Table 6: Relationship between the Smoking Polygenic Scores and the Summary Standardized Index of Health Accounting for Parental Smoking, Risky Behaviors, and Mortality

	Full Sample		Ever Smoker Sample		Never Smoker Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
CPD Score	0.035***	0.033***	0.037***	0.034***	0.032***	0.029***
	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)
Initiation Score	0.021***	0.017***	0.011	0.010	0.030***	0.024***
	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)
Parental Controls	No	Yes	No	Yes	No	Yes
Full Smoking Controls	Yes	Yes	Yes	Yes	No	No
Standard Controls	Yes	Yes	Yes	Yes	Yes	Yes
Principal Components	Yes	Yes	Yes	Yes	Yes	Yes
Sample Mean	-0.000	-0.000	0.038	0.038	-0.052	-0.052
N	45,850	45,850	26,509	26,509	19,341	19,341
R^2	0.116	0.134	0.142	0.156	0.080	0.106

Note: Each column presents results from a separate regression where the outcome is the summary standardized health index. Results are presented for the full sample and then separately for ever smokers and never smokers. In all specifications, we include the standard controls and the genetic principal components described in Section 4. We also include the full set of smoking controls described in Section 4 for the full sample and ever smoker samples. In columns (2), (4), and (6), we add controls for parental smoking, parental risky behaviors, and parental mortality as described in Section 3.4. * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Table 7: Relationship between the Smoking Polygenic Scores and the Summary Standardized Index of Health Accounting for Spouse's Smoking Behavior and Smoking Polygenic Scores

	Full Sample				Ever Smoker Sample				Never Smoker Sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
CPD Score	0.030***	0.028***	0.030***	0.028***	0.030***	0.029***	0.031***	0.030***	0.029***	0.024***	0.029***	0.025***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Initiation Score	0.015**	0.011*	0.013**	0.009	0.005	0.001	0.003	-0.001	0.029***	0.025***	0.027***	0.024***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Spouse's CPD Score			0.025***	0.022***			0.022***	0.020***			0.026***	0.023***
			(0.006)	(0.006)			(0.008)	(0.008)			(0.008)	(0.008)
Spouse's Initiation Score			0.004	0.000			0.007	0.007			0.001	-0.005
			(0.006)	(0.006)			(0.008)	(0.008)			(0.009)	(0.009)
Spouse's Full Smoking Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Spouse's Principal Components	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
Full Smoking Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No
Standard Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Principal Components	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Mean	-0.035	-0.035	-0.035	-0.035	0.004	0.004	0.004	0.004	-0.085	-0.085	-0.085	-0.085
N	21,357	21,357	21,357	21,357	11,888	11,888	11,888	11,888	9,469	9,469	9,469	9,469
R^2	0.128	0.152	0.136	0.159	0.172	0.203	0.181	0.212	0.088	0.142	0.097	0.150

Note: Each column presents results from a separate regression where the outcome is the summary standardized health index. Results are presented for the full sample and then separately for ever smokers and never smokers, and we only include individuals in couples where both members are aged 50–65 and genotyped. In all specifications, we include the standard controls and the genetic principal components described in Section 4, and for the full sample and ever smoker sample, we always include the full set of own smoking controls described in Section 4. In columns (2), (6), and (10), we include the full set of the spouse's smoking controls. In columns (3), (7), and (11), we include the spouse's smoking polygenic scores and their genetic principal components. In columns (4), (8), and (12), we include both the full set of the spouse's smoking controls and the spouse's smoking polygenic scores and their genetic principal components. * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Table 8: Relationship between the Smoking Polygenic Scores and Risk Preferences

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Risk Aversion Category (Income Gamble)					General Willingness to Take Risk		
	$= 6$	≥ 5	≥ 4	≥ 3	≥ 2	Very Unwilling	Very Willing	0-10 Score
	Panel A: Full Sample							
CPD Score	0.013**	0.008	0.006	0.004	-0.001	0.002	0.002	-0.016
	(0.006)	(0.006)	(0.005)	(0.004)	(0.003)	(0.003)	(0.004)	(0.037)
Initiation Score	0.008	-0.002	0.003	0.005	0.001	0.002	-0.001	0.001
	(0.006)	(0.006)	(0.005)	(0.004)	(0.003)	(0.004)	(0.005)	(0.040)
Full Smoking Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Mean	0.379	0.602	0.772	0.872	0.948	0.045	0.067	5.878
N	9,211	9,211	9,211	9,211	9,211	5,518	5,518	5,518
R^2	0.046	0.046	0.043	0.048	0.045	0.061	0.057	0.073
	Panel B: Ever Smoker Sample							
CPD Score	0.008	0.003	0.008	0.006	-0.002	0.002	0.003	-0.045
	(0.008)	(0.008)	(0.007)	(0.005)	(0.003)	(0.005)	(0.006)	(0.051)
Initiation Score	0.008	0.002	0.002	0.006	0.004	-0.001	0.006	0.052
	(0.008)	(0.008)	(0.007)	(0.006)	(0.004)	(0.006)	(0.006)	(0.057)
Full Smoking Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Mean	0.385	0.601	0.765	0.862	0.944	0.049	0.073	5.863
N	5,346	5,346	5,346	5,346	5,346	3,011	3,011	3,011
R^2	0.068	0.065	0.059	0.065	0.065	0.107	0.102	0.118
	Panel C: Never Smoker Sample							
CPD Score	0.020**	0.013	0.002	0.002	-0.001	0.001	0.004	0.036
	(0.009)	(0.009)	(0.007)	(0.006)	(0.004)	(0.005)	(0.006)	(0.054)
Initiation Score	0.005	-0.004	0.005	0.004	-0.004	0.006	-0.006	-0.060
	(0.010)	(0.010)	(0.008)	(0.007)	(0.005)	(0.005)	(0.007)	(0.058)
Sample Mean	0.371	0.604	0.783	0.885	0.954	0.040	0.061	5.896
N	3,865	3,865	3,865	3,865	3,865	2,507	2,507	2,507
R^2	0.040	0.049	0.048	0.046	0.042	0.041	0.036	0.057

Note: Each column within a panel presents results from a separate regression. The outcomes in columns (1)–(5) are based on respondents' choice between a guaranteed stream of income and a 50-50 gamble that doubles that amount of income or cuts it by various amounts. Individuals are grouped into six categories from least (category 1) to most (category 6) risk averse. Category 6 corresponds to always choosing the guaranteed stream of income, and categories 1, 2, 3, 4, and 5 consist of those who would take a job with even chances of doubling income or cutting it by 75%, 50%, a third, 20%, and 10%, respectively. The outcomes in columns (6)–(8) are based on respondents' answers on a 0–10 scale to whether they avoid taking risks or are fully prepared to take risks. In column (6), the outcome is an indicator for not willing to take risk (response of 0 or 1). In column (7), the outcome is an indicator for very willing to take risk (response of 9 or 10). In column (8), the outcome is the score itself. Results for the full sample are presented in Panel A, for the ever smoker sample in Panel B, and for the never smoker sample in Panel C. In all specifications, we include the standard controls and the genetic principal components described in Section 4, and for the full sample and ever smoker sample, we always include the full set of smoking controls described in Section 4. * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Table 9: Relationship between the Smoking Polygenic Scores and Self-Reported Probability of Living to Age 75

	Full Sample	Ever Smoker Sample	Never Smoker Sample
	(1)	(2)	(3)
CPD Score	-1.420***	-1.749***	-1.018***
	(0.227)	(0.313)	(0.329)
Initiation Score	0.019	0.402	-0.429
	(0.248)	(0.339)	(0.361)
Full Smoking Controls	Yes	Yes	No
Sample Mean	66.672	64.362	69.806
N	42,001	24,182	17,819
R^2	0.082	0.109	0.028

Note: Each column presents results from a separate regression where the outcome is the self-reported probability of living to age 75. Results are presented for the full sample and then separately for ever smokers and never smokers. In all specifications, we include the standard controls and the genetic principal components described in Section 4, and for the full sample and ever smoker sample, we always include the full set of smoking controls described in Section 4. * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Table 10: Relationship between the Smoking Polygenic Scores and the Summary Standardized Index of Health Accounting for Longevity Expectations

	Full Sample		Ever Smoker Sample		Never Smoker Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
CPD Score	0.036***	0.029***	0.038***	0.030***	0.031***	0.027***
	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)
Initiation Score	0.020***	0.020***	0.011	0.013**	0.030***	0.028***
	(0.005)	(0.005)	(0.007)	(0.006)	(0.007)	(0.006)
Probability of Living to 75		-0.005***		-0.005***		-0.005***
		(0.000)		(0.000)		(0.000)
Full Smoking Controls	Yes	Yes	Yes	Yes	No	No
Sample Mean	-0.009	-0.009	0.029	0.029	-0.061	-0.061
N	42,001	42,001	24,182	24,182	17,819	17,819
R^2	0.115	0.191	0.141	0.214	0.075	0.151

Note: Each column presents results from a separate regression where the outcome is the summary standardized health index. Results are presented for the full sample and then separately for ever smokers and never smokers. In all specifications, we include the standard controls and the genetic principal components described in Section 4. We also include the full set of smoking controls described in Section 4 for the full sample and ever smoker samples. In columns (2), (4), and (6), we add controls for the individual's self-reported probability of living to age 75. * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Table 11: Relationship between the Smoking Polygenic Scores and Financial Planning Horizon

	(1) PH $\geq$ Next Year	(2) PH $\geq$ Few Years	(3) PH $\geq$ 5-10 Years	(4) PH $\geq$ 10 Years
		Panel A: Full Sample		
CPD Score	-0.008***	-0.008**	-0.007*	0.001
	(0.003)	(0.004)	(0.004)	(0.003)
Initiation Score	-0.007**	-0.009**	-0.005	-0.002
	(0.003)	(0.004)	(0.005)	(0.003)
Full Smoking Controls	Yes	Yes	Yes	Yes
Sample Mean	0.880	0.777	0.503	0.140
N	18,878	18,878	18,878	18,878
R^2	0.034	0.031	0.033	0.021
		Panel B: Ever Smoker Sample		
CPD Score	-0.008**	-0.007	-0.010*	-0.003
	(0.004)	(0.005)	(0.006)	(0.004)
Initiation Score	-0.005	-0.008	-0.005	-0.002
	(0.004)	(0.005)	(0.006)	(0.004)
Full Smoking Controls	Yes	Yes	Yes	Yes
Sample Mean	0.869	0.761	0.492	0.131
N	10,810	10,810	10,810	10,810
R^2	0.053	0.044	0.046	0.034
		Panel C: Never Smoker Sample		
CPD Score	-0.006	-0.007	-0.004	0.005
	(0.004)	(0.006)	(0.007)	(0.005)
Initiation Score	-0.007	-0.008	-0.004	-0.003
	(0.005)	(0.006)	(0.007)	(0.005)
Sample Mean	0.895	0.798	0.519	0.151
N	8,068	8,068	8,068	8,068
R^2	0.020	0.022	0.031	0.021

Note: Each column presents results from a separate regression where the outcome is an indicator for financial planning horizon (PH) length. Results are presented for the full sample and then separately for ever smokers and never smokers. In all specifications, we include the standard controls and the genetic principal components described in Section 4, and for the full sample and ever smoker sample, we always include the full set of smoking controls described in Section 4. * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Table 12: Relationship between the Smoking Polygenic Scores and the Summary Standardized Index of Health Accounting for Financial Planning Horizons

	Full Sample		Ever Smoker Sample		Never Smoker Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
CPD Score	0.040***	0.038***	0.042***	0.041***	0.034***	0.033***
	(0.005)	(0.005)	(0.008)	(0.008)	(0.007)	(0.007)
Initiation Score	0.024***	0.022***	0.015*	0.014*	0.032***	0.031***
	(0.006)	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)
PH = Next Year		-0.101***		-0.102***		-0.098***
		(0.017)		(0.023)		(0.025)
PH = Next Few Years		-0.157***		-0.162***		-0.144***
		(0.015)		(0.020)		(0.022)
PH = Next 5-10 Years		-0.188***		-0.192***		-0.175***
		(0.015)		(0.020)		(0.022)
PH > 10 years		-0.184***		-0.185***		-0.172***
		(0.017)		(0.023)		(0.025)
Full Smoking Controls	Yes	Yes	Yes	Yes	No	No
Sample Mean	0.017	0.017	0.064	0.064	-0.045	-0.045
N	18,878	18,878	10,810	10,810	8,068	8,068
R^2	0.123	0.139	0.150	0.165	0.077	0.094

Note: Each column presents results from a separate regression where the outcome is the summary standardized health index. Results are presented for the full sample and then separately for ever smokers and never smokers. In all specifications, we include the standard controls and the genetic principal components described in Section 4. We also include the full set of smoking controls described in Section 4 for the full sample and ever smoker samples. In columns (2), (4), and (6), we add dummy variables for the individual's financial planning horizon (PH). * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Table 13: Relationship between the Smoking Polygenic Scores and Impulsivity and Conscientiousness

	Full Sample			Ever Smoker Sample			Never Smoker Sample		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Big 5	MPQ Score	RCSG Score	Big 5	MPQ Score	RCSG Score	Big 5	MPQ Score	RCSG Score
	(1-4)	(1-6)	(1-6)	(1-4)	(1-6)	(1-6)	(1-4)	(1-6)	(1-6)
CPD Score	-0.019***	-0.046***	-0.036***	-0.021**	-0.016	-0.015	-0.017*	-0.080***	-0.062***
	(0.007)	(0.016)	(0.011)	(0.010)	(0.023)	(0.016)	(0.009)	(0.023)	(0.015)
Initiation Score	-0.015**	-0.040**	-0.020*	-0.001	-0.031	-0.030*	-0.030***	-0.048**	-0.001
	(0.007)	(0.017)	(0.011)	(0.011)	(0.025)	(0.016)	(0.010)	(0.023)	(0.016)
Full Smoking Controls	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Sample Mean	3.280	4.318	4.737	3.235	4.259	4.655	3.333	4.389	4.833
N	5,565	3,255	3,157	2,999	1,763	1,707	2,566	1,492	1,450
R^2	0.112	0.090	0.137	0.149	0.151	0.186	0.079	0.081	0.116

Note: Each column presents results from a separate regression. The outcome in columns (1), (4), and (7) is the Big 5 conscientiousness score, which ranges from 1 (less conscientious)–4 (more conscientious). The outcome in columns (2), (5), and (8) is the conscientiousness score from the Multidimensional Personality Questionnaire (MPQ), which ranges from 1 (less conscientious)–6 (more conscientious). The outcome in columns (3), (6), and (9) is the conscientiousness score from Roberts et al. (2005) (the RCSG score) which ranges from 1 (less conscientious)–6 (more conscientious). More details on these outcomes are provided in Section 6.4. Results are presented for the full sample and then separately for ever smokers and never smokers. In all specifications, we include the standard controls and the genetic principal components described in Section 4, and for the full sample and ever smoker sample, we always include the full set of smoking controls described in Section 4. * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Table 14: Relationship between the Smoking Polygenic Scores and the Summary Standardized Index of Health Accounting for Conscientiousness

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Full Sample			
CPD Score	0.033***	0.028***	0.036***	0.032***	0.039***	0.034***
	(0.008)	(0.008)	(0.009)	(0.009)	(0.008)	(0.008)
Initiation Score	0.023**	0.018**	0.027***	0.023**	0.032***	0.029***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Big 5 Conscientiousness Score		-0.274***				
		(0.021)				
MPQ Conscientiousness Score				-0.096***		
				(0.010)		
RCSG Conscientiousness Score						-0.152***
						(0.016)
Full Smoking Controls	Yes	Yes	Yes	Yes	Yes	Yes
Sample Mean	0.066	0.066	0.056	0.056	0.056	0.056
N	5,565	5,565	3,255	3,255	3,157	3,157
R^2	0.137	0.181	0.148	0.174	0.161	0.191
			Panel B: Ever Smoker Sample			
CPD Score	0.035***	0.029**	0.034**	0.032**	0.035***	0.033**
	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)	(0.013)
Initiation Score	0.016	0.016	0.026*	0.023	0.037***	0.033**
	(0.014)	(0.013)	(0.015)	(0.014)	(0.014)	(0.014)
Big 5 Conscientiousness Score		-0.300***				
		(0.030)				
MPQ Conscientiousness Score				-0.110***		
				(0.015)		
RCSG Conscientiousness Score						-0.134***
						(0.023)
Full Smoking Controls	Yes	Yes	Yes	Yes	Yes	Yes
Sample Mean	0.129	0.129	0.110	0.110	0.104	0.104
N	2,999	2,999	1,763	1,763	1,707	1,707
R^2	0.169	0.215	0.185	0.214	0.213	0.233
			Panel C: Never Smoker Sample			
CPD Score	0.031***	0.027**	0.038***	0.031***	0.043***	0.032***
	(0.011)	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)
Initiation Score	0.029**	0.022*	0.027**	0.023*	0.023*	0.023*
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Big 5 Conscientiousness Score		-0.239***				
		(0.029)				
MPQ Conscientiousness Score				-0.083***		
				(0.014)		
RCSG Conscientiousness Score						-0.172***
						(0.023)
Sample Mean	-0.008	-0.008	-0.009	-0.009	-0.001	-0.001
N	2,566	2,566	1,492	1,492	1,450	1,450
R^2	0.092	0.133	0.116	0.139	0.122	0.166

Note: Each column within a panel presents results from a separate regression where the outcome is the summary standardized health index. Results for the full sample are presented in Panel A, for the ever smoker sample in Panel B, and for the never smoker sample in Panel C. In all specifications, we include the standard controls and the genetic principal components described in Section 4. We also include the full set of smoking controls described in Section 4 for the full sample and ever smoker samples. In columns (1), (3), and (5), we do not include controls for conscientiousness. In columns (2), (4), and (6), we add controls for the Big 5 conscientiousness score (1–4), the MPQ conscientiousness score (1–6), and the RCSG conscientiousness score (1–6), respectively. * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Table 15: Relationship between the Smoking Polygenic Scores and the Summary Standardized Index of Health Accounting for Risk Aversion, Longevity Expectations, Planning Horizons, and Conscientiousness

	Full Sample		Ever Smoker Sample		Never Smoker Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
CPD Score	0.037***	0.029***	0.043***	0.033***	0.029***	0.024***
	(0.006)	(0.005)	(0.008)	(0.008)	(0.007)	(0.007)
Initiation Score	0.018***	0.013**	0.007	0.005	0.027***	0.021**
	(0.006)	(0.006)	(0.008)	(0.008)	(0.009)	(0.009)
Risk Aversion Controls	No	Yes	No	Yes	No	Yes
Expected Longevity Controls	No	Yes	No	Yes	No	Yes
Planning Horizon Controls	No	Yes	No	Yes	No	Yes
Conscientiousness Controls	No	Yes	No	Yes	No	Yes
Full Smoking Controls	Yes	Yes	Yes	Yes	No	No
Standard Controls	Yes	Yes	Yes	Yes	Yes	Yes
Principal Components	Yes	Yes	Yes	Yes	Yes	Yes
Sample Mean	-0.027	-0.027	0.003	0.003	-0.065	-0.065
N	25,818	25,818	14,565	14,565	11,253	11,253
R^2	0.130	0.219	0.166	0.248	0.095	0.192

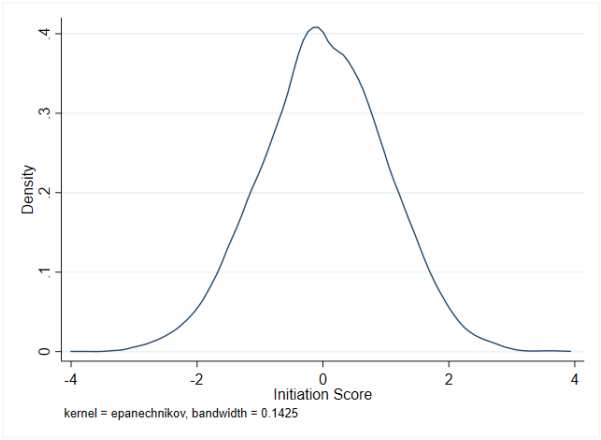
Note: Each column presents results from a separate regression where the outcome is the summary standardized health index. Results are presented for the full sample and then separately for ever smokers and never smokers. In all specifications, we include the standard controls and the genetic principal components described in Section 4. We also include the full set of smoking controls described in Section 4 for the full sample and ever smoker samples. In columns (2), (4), and (6), we add dummy variables for risk aversion categories, self-reported probability of living to age 75, dummy variables for financial planning horizon, and the Big 5 conscientiousness score. * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Table 16: Relationship between the Smoking Polygenic Scores and the Summary Standardized Index of Health Accounting for Risk Aversion, Longevity Expectations, Planning Horizons, Conscientiousness, and Parental and Spousal Variables

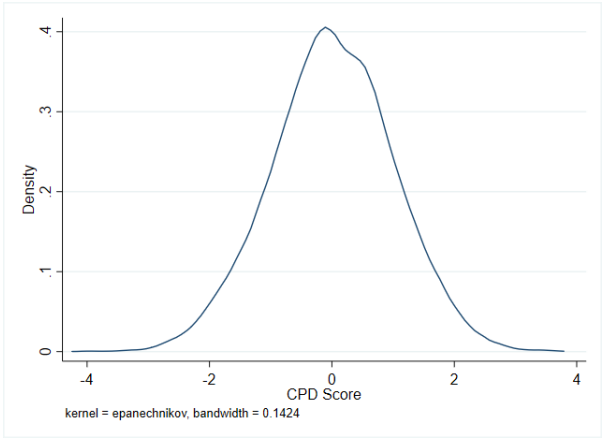
	Full Sample		Ever Smoker Sample		Never Smoker Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
CPD Score	0.035***	0.026***	0.039***	0.031***	0.032***	0.024***
	(0.007)	(0.007)	(0.010)	(0.010)	(0.010)	(0.009)
Initiation Score	0.015**	0.006	0.004	-0.001	0.029***	0.016
	(0.007)	(0.007)	(0.010)	(0.010)	(0.011)	(0.011)
Spouse's CPD Score		0.024***		0.016*		0.024**
		(0.007)		(0.009)		(0.010)
Spouse's Initiation Score		-0.008		0.000		-0.014
		(0.007)		(0.010)		(0.011)
Risk Aversion Controls	No	Yes	No	Yes	No	Yes
Expected Longevity Controls	No	Yes	No	Yes	No	Yes
Planning Horizon Controls	No	Yes	No	Yes	No	Yes
Conscientiousness Controls	No	Yes	No	Yes	No	Yes
Parental Controls	No	Yes	No	Yes	No	Yes
Spouse's Full Smoking Controls	No	Yes	No	Yes	No	Yes
Spouse's Principal Components	No	Yes	No	Yes	No	Yes
Full Smoking Controls	Yes	Yes	Yes	Yes	No	No
Standard Controls	Yes	Yes	Yes	Yes	Yes	Yes
Principal Components	Yes	Yes	Yes	Yes	Yes	Yes
Sample Mean	-0.058	-0.058	-0.029	-0.029	-0.093	-0.093
N	12,989	12,989	7,039	7,039	5,950	5,950
R^2	0.148	0.267	0.213	0.337	0.106	0.282

Note: Each column presents results from a separate regression where the outcome is the summary standardized health index. Results are presented for the full sample and then separately for ever smokers and never smokers. In all specifications, we include the standard controls and the genetic principal components described in Section 4. We also include the full set of smoking controls described in Section 4 for the full sample and ever smoker samples. In columns (2), (4), and (6), we add dummy variables for risk aversion categories, self-reported probability of living to age 75, dummy variables for financial planning horizon, the Big 5 conscientiousness score, controls for parental smoking, parental risky behaviors, and parental mortality as described in Section 3.4, the full set of the spouse's smoking controls, and the spouse's smoking polygenic scores and their genetic principal components. * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Figure 1: Distributions of the Smoking Polygenic Scores



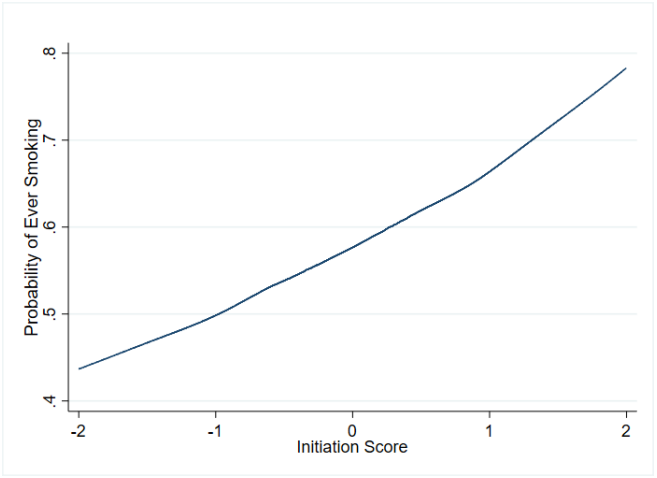
(a) Initiation Score



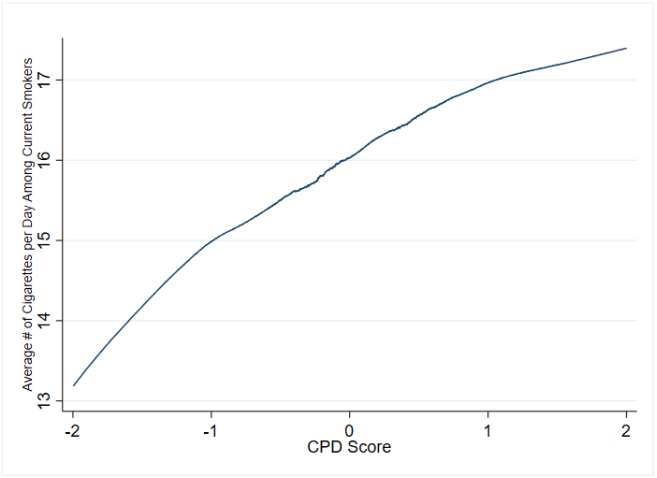
(b) CPD Score

Note: Panel (a) shows the smoothed density of the smoking initiation score and Panel (b) shows the smoothed density of the CPD (smoking intensity) score.

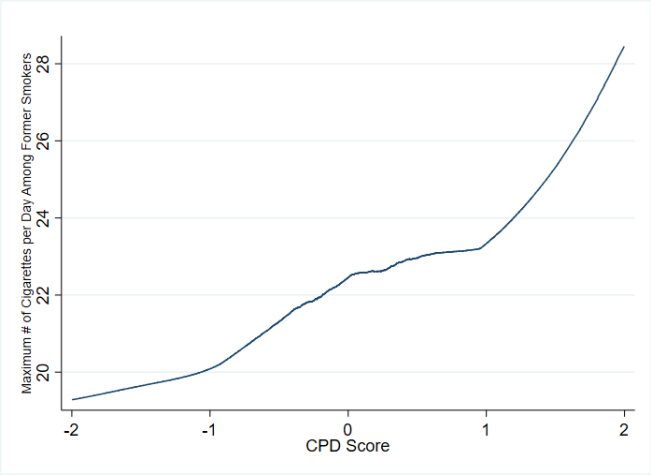
Figure 2: Relationship between Smoking Polygenic Scores and Various Smoking Measures



(a) Initiation Score and Probability of Ever Smoking



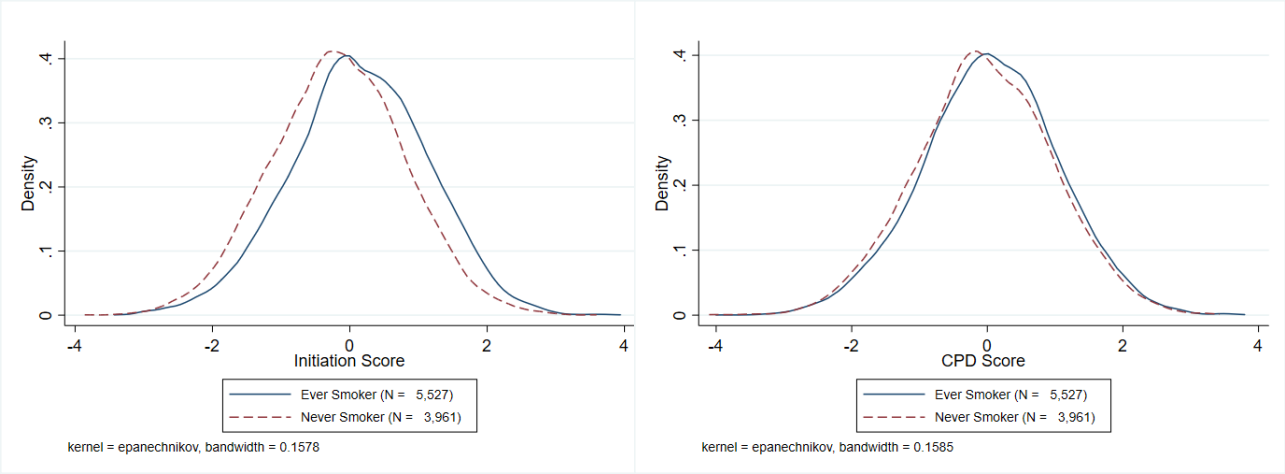
(b) CPD Score and Average # of Cigarettes Smoked per Day Among Current Smokers



(c) CPD Score and Maximum # of Cigarettes Smoked per Day Among Former Smokers

Note: Panel (a) shows the non-parametric (lowess) relationship between the smoking initiation score and the probability of ever smoking (as of the last time an individual is observed in our sample). Panel (b) shows the lowess relationship between the CPD score and average cigarettes smoked per day among current smokers. Panel (c) shows the lowess relationship between the CPD score and the maximum number of cigarettes smoked per day among former smokers.

Figure 3: Distribution of Smoking Polygenic Scores Among Ever and Never Smokers



(a) Initiation Score Among Ever and Never Smokers

(b) CPD Score Among Ever and Never Smokers

Note: Panel (a) shows the smoothed density of the smoking initiation score separately for ever and never smokers. Panel (b) shows the smoothed density of the CPD (smoking intensity) score separately for ever and never smokers.

Appendix

Table A1: Summary Statistics for Non-Genotyped Individuals

	Full Sample			Ever Smoker Sample			Never Smoker Sample			
	Mean	SD	N	Mean	SD	N	Mean	SD	N	p-value
Demographics:										
Age	58.20	4.18	31,015	58.38	4.15	18,877	57.93	4.21	12,138	0.00
Birth Year	1945.38	9.14	31,015	1944.59	9.01	18,877	1946.59	9.22	12,138	0.00
Male	0.46	0.50	31,015	0.53	0.50	18,877	0.35	0.48	12,138	0.00
Own Smoking:										
# Cigs per Day (Among Current Smokers)	17.76	12.30	7,076	17.76	12.30	7,076				
Maximum # Cigs per Day (Among Former Smokers)	19.43	17.76	8,183	19.43	17.76	8,183				
Age Start Smoking (Among Ever Smokers)	18.04	5.57	13,758	18.04	5.57	13,758				
Age Quit Smoking (Among Quitters)	38.97	11.58	9,800	38.97	11.58	9,800				
Smoked as a Teenager	0.36	0.48	11,722	0.36	0.48	11,722				
Own Health:										
Poor Self-Reported Healt	0.29	0.46	30,995	0.32	0.47	18,867	0.26	0.44	12,128	0.00
Any Hospital Stay	0.20	0.40	30,813	0.22	0.42	18,775	0.17	0.37	12,038	0.00
Ever Had High Blood Pressure	0.42	0.49	30,986	0.42	0.49	18,861	0.42	0.49	12,125	0.53
Ever Had Diabetes	0.16	0.37	30,967	0.16	0.37	18,843	0.16	0.37	12,124	0.76
Ever Had Cancer	0.08	0.28	30,951	0.09	0.29	18,837	0.07	0.26	12,114	0.00
Ever Had Lung Disease	0.08	0.26	30,997	0.11	0.31	18,867	0.03	0.17	12,130	0.00
Ever Had Heart Problem	0.14	0.35	30,986	0.17	0.37	18,857	0.10	0.30	12,129	0.00
Ever Had Stroke	0.04	0.20	30,998	0.05	0.23	18,869	0.02	0.16	12,129	0.00
Ever Had Arthritis	0.42	0.49	30,980	0.45	0.50	18,861	0.38	0.49	12,119	0.00
Any Limitations with ADLs	0.13	0.34	30,925	0.15	0.36	18,834	0.11	0.31	12,091	0.00
Any Limitations with IADLs	0.08	0.26	30,917	0.09	0.28	18,831	0.06	0.24	12,086	0.00
Body Mass Index (kg/m^2)	28.17	5.85	30,286	27.87	5.79	18,520	28.65	5.93	11,766	0.00
Obese	0.32	0.47	30,414	0.30	0.46	18,554	0.35	0.48	11,860	0.00
Parental Smoking:										
Neither Parent Smoked	0.18	0.39	31,015	0.14	0.35	18,877	0.25	0.43	12,138	0.00
One Parent Smoked	0.25	0.43	31,015	0.24	0.43	18,877	0.27	0.44	12,138	0.00
Both Parents Smoked	0.13	0.34	31,015	0.15	0.35	18,877	0.11	0.32	12,138	0.00
Missing	0.43	0.49	31,015	0.47	0.50	18,877	0.37	0.48	12,138	0.00
Parental Mortality:										
Neither Parent Died During Childhood	0.11	0.31	31,015	0.10	0.30	18,877	0.13	0.33	12,138	0.00
Parent(s) Died During Childhood	0.04	0.20	31,015	0.04	0.19	18,877	0.05	0.21	12,138	0.00
Parental Mortality During Childhood Missing	0.85	0.36	31,015	0.86	0.35	18,877	0.83	0.38	12,138	0.00
Mother Currently Alive	0.37	0.48	31,015	0.35	0.48	18,877	0.41	0.49	12,138	0.00
Mother Died Before Age 65	0.17	0.37	31,015	0.18	0.38	18,877	0.15	0.36	12,138	0.00
Mother Died Age 65+	0.44	0.50	31,015	0.45	0.50	18,877	0.41	0.49	12,138	0.00
Mother Mortality Information Missing	0.02	0.15	31,015	0.03	0.16	18,877	0.02	0.14	12,138	0.00
Father Currently Alive	0.16	0.37	31,015	0.15	0.35	18,877	0.19	0.40	12,138	0.00
Father Died Before Age 65	0.25	0.43	31,015	0.27	0.44	18,877	0.23	0.42	12,138	0.00
Father Died Age 65+	0.55	0.50	31,015	0.55	0.50	18,877	0.55	0.50	12,138	0.80
Father Mortality Information Missing	0.03	0.18	31,015	0.04	0.19	18,877	0.03	0.17	12,138	0.02
Parental Risky Behavior During Respondent's Childhood:										
Parent(s) Drank/Used Drugs Often	0.10	0.30	31,015	0.11	0.31	18,877	0.09	0.28	12,138	0.00
Neither Parent Drank/Used Drugs	0.40	0.49	31,015	0.37	0.48	18,877	0.45	0.50	12,138	0.00
Missing	0.50	0.50	31,015	0.53	0.50	18,877	0.46	0.50	12,138	0.00

Note: The table presents summary statistics for the full sample and separately for those who ever and never smoked. The sample includes those aged 50–65 of European ancestry who self-identify as white and who have *not* been genotyped. The last column presents p-values from a test of equality of means for ever and never smokers.

Table A2: Relationship between the Smoking Polygenic Scores and Smoking Behavior

	Max # of Cigarettes Smoked per Day			Ever Smoked		
	(1)	(2)	(3)	(4)	(5)	(6)
CPD Score	2.044***		1.949***	0.027***		0.016***
	(0.218)		(0.219)	(0.005)		(0.005)
Initiation Score		1.155***	0.924***		0.095***	0.093***
		(0.235)	(0.234)		(0.005)	(0.005)
Standard Controls	Yes	Yes	Yes	Yes	Yes	Yes
Principal Components	Yes	Yes	Yes	Yes	Yes	Yes
Sample Mean	22.468	22.468	22.468	0.583	0.583	0.583
N	5,392	5,392	5,392	9,488	9,488	9,488
R^2	0.103	0.092	0.105	0.049	0.078	0.079
Incremental R^2	0.015	0.004	0.018	0.003	0.031	0.032

Note: Each column presents results from a separate regression. In columns (1)–(3), the outcome is the maximum number of cigarettes smoked per day among current and former smokers. In columns (4)–(6), the outcome is an indicator for whether the individual has ever smoked as of the last time we observe them in our sample. In all specifications, we include the standard controls and the genetic principal components described in Section 4. In columns (1) and (4), we report the incremental R^2 of the CPD score; in columns (2) and (5), we report the incremental R^2 of the initiation score; and in columns (3) and (6), we report the incremental R^2 of both scores. * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Table A3: Relationship between the Smoking Polygenic Scores and Summary Standardized Index of Doctor-Diagnosed Conditions

	Full Sample			Ever Smoker Sample			Never Smoker Sample	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CPD Score	0.035***	0.036***	0.029***	0.044***	0.044***	0.034***	0.021***	0.022***
	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)
Initiation Score	0.020***	0.027***	0.013**	0.012*	0.016**	0.004	0.017**	0.023***
	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Full Smoking Controls	No	No	Yes	No	No	Yes	No	No
Standard Controls	No	Yes	Yes	No	Yes	Yes	No	Yes
Principal Components	No	Yes	Yes	No	Yes	Yes	No	Yes
Sample Mean	-0.000	-0.000	-0.000	0.037	0.037	0.037	-0.051	-0.051
N	45,850	45,850	45,850	26,509	26,509	26,509	19,341	19,341
R^2	0.008	0.083	0.123	0.009	0.093	0.145	0.004	0.092

Note: Each column presents results from a separate regression where the outcome is a summary standardized index of the doctor-diagnosed conditions. Results are presented for the full sample and then separately for ever smokers and never smokers. For each sample, we first present results without controls (columns 1, 4, 7). Then we add the standard controls and the genetic principal components described in Section 4 (columns 2, 5, 8). Then for the full sample and ever smoker sample, we add the full set of smoking controls described in Section 4 (columns 3, 6). * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Table A4: Smoking Polygenic Scores and Assortative Mating

Husband's CPD Score Quartile	**Panel A:** CPD Score (N=2,663)			
	Wife's CPD Score Quartile			
	Q1	Q2	Q3	Q4
Q1	30.4	25.4	25.5	24.7
Q2	22.3	25.7	25.0	26.2
Q3	22.6	23.2	25.0	28.2
Q4	24.7	25.7	24.4	20.9
Husband's Initiation Score Quartile	**Panel B:** Initiation Score (N=2,663)			
	Wife's Initiation Score Quartile			
	Q1	Q2	Q3	Q4
Q1	25.0	28.3	23.4	24.0
Q2	25.3	24.5	26.8	27.4
Q3	26.8	24.2	27.0	22.0
Q4	22.8	23.1	22.8	26.6

Note: Panel A (B) shows the distribution of the husband's CPD (initiation) score conditional on the quartile of the wife's CPD (initiation) score for all individuals in couples where both members were genotyped and aged 50–65. The column probabilities sum to 100.

Table A5: Relationship between the Smoking Polygenic Scores and Health Measures Accounting for Spouse's Smoking Polygenic Scores where Neither Member of Couple Ever Smoked

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
			Ever Diagnosed With:							Any Limitations With:				
	Summary Index	Poor Self-Reported Health	Any Hospitalization	High Blood Pressure	Diabetes	Cancer	Lung Disease	Heart Problem	Stroke	Arthritis	ADLs	IADLs	BMI	Obese
CPD Score	0.019*	0.009	0.002	-0.016	0.023**	-0.006	0.007*	0.001	-0.002	0.013	0.005	0.002	0.572***	0.030**
	(0.010)	(0.007)	(0.006)	(0.014)	(0.009)	(0.007)	(0.004)	(0.009)	(0.003)	(0.014)	(0.005)	(0.003)	(0.169)	(0.013)
Initiation Score	0.009	0.007	-0.002	0.005	-0.009	0.007	0.004	-0.013	0.000	0.029*	0.004	-0.002	0.253	0.017
	(0.011)	(0.008)	(0.007)	(0.015)	(0.010)	(0.008)	(0.005)	(0.010)	(0.003)	(0.015)	(0.005)	(0.002)	(0.178)	(0.015)
Spouse's CPD Score	0.026**	0.011	0.016**	0.025*	0.009	-0.001	0.008*	0.023***	0.006	0.014	0.002	0.000	-0.059	0.003
	(0.011)	(0.007)	(0.007)	(0.014)	(0.009)	(0.008)	(0.004)	(0.009)	(0.005)	(0.014)	(0.004)	(0.003)	(0.162)	(0.013)
Spouse's Initiation Score	-0.007	0.005	0.006	-0.015	-0.018*	-0.001	0.000	-0.006	-0.011***	0.013	0.000	0.001	0.130	-0.001
	(0.012)	(0.009)	(0.008)	(0.015)	(0.010)	(0.008)	(0.003)	(0.009)	(0.004)	(0.015)	(0.006)	(0.003)	(0.199)	(0.015)
Spouse's Principal Components	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Standard Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Principal Components	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Mean	-0.099	0.093	0.137	0.350	0.103	0.073	0.019	0.104	0.016	0.361	0.049	0.018	28.095	0.312
N	5,304	5,396	5,391	5,396	5,393	5,396	5,390	5,393	5,396	5,390	5,394	5,394	5,327	5,327
R^2	0.126	0.066	0.043	0.110	0.107	0.095	0.090	0.107	0.122	0.104	0.054	0.049	0.135	0.105

Note: Each column presents results from a separate regression where the outcome is noted in the column heading. The sample includes individuals in couples where both members were genotyped, aged 50–65, and neither member has ever smoked. In all specifications, we include the standard controls and the genetic principal components described in Section 4. We also include the spouse's smoking polygenic scores and their genetic principal components. * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Table A6: Relationship between the Smoking Polygenic Scores and the Summary Standardized Index of Health Accounting for Parental and Spousal Variables

	Full Sample				Ever Smoker Sample				Never Smoker Sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
CPD Score	0.030***	0.027***	0.028***	0.025***	0.030***	0.027***	0.030***	0.025***	0.029***	0.028***	0.025***	0.024***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Initiation Score	0.015**	0.012**	0.009	0.007	0.005	0.006	-0.001	0.001	0.029***	0.022***	0.024***	0.018**
	(0.006)	(0.006)	(0.006)	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Spouse's CPD Score			0.022***	0.020***			0.020***	0.018**			0.023***	0.022***
			(0.006)	(0.006)			(0.008)	(0.008)			(0.008)	(0.008)
Spouse's Initiation Score			0.001	-0.002			0.007	0.005			-0.005	-0.008
			(0.006)	(0.006)			(0.008)	(0.008)			(0.009)	(0.009)
Parental Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Spouse's Full Smoking Controls	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
Spouse's Principal Components	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
Full Smoking Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No
Standard Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Principal Components	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Mean	-0.035	-0.035	-0.035	-0.035	0.004	0.004	0.004	0.004	-0.085	-0.085	-0.085	-0.085
N	21,357	21,357	21,357	21,357	11,888	11,888	11,888	11,888	9,469	9,469	9,469	9,469
R^2	0.128	0.145	0.159	0.174	0.172	0.188	0.212	0.227	0.088	0.108	0.150	0.168

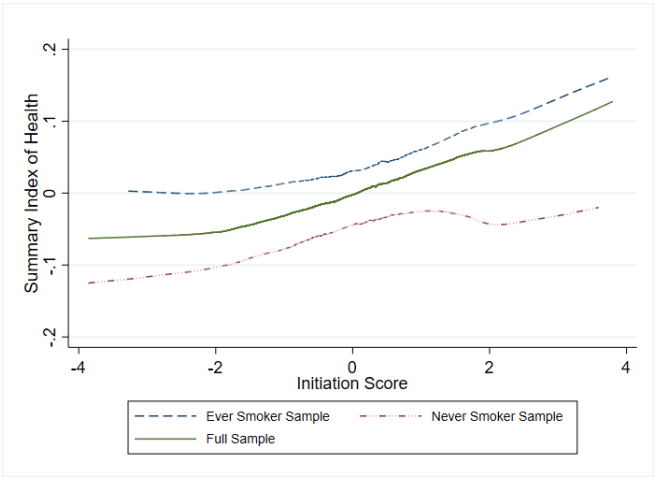
Note: Each column presents results from a separate regression where the outcome is the summary standardized health index. Results are presented for the full sample and then separately for ever smokers and never smokers, and we only include individuals in couples where both members are aged 50–65 and genotyped. In all specifications, we include the standard controls and the genetic principal components described in Section 4, and for the full sample and ever smoker sample, we always include the full set of smoking controls described in Section 4. In columns (2), (6), and (10), we add controls for parental smoking, parental risky behaviors, and parental mortality as described in Section 3.4. In columns (3), (7), and (11), we include the full set of the spouse's smoking controls and the spouse's smoking polygenic scores and their genetic principal components. In columns (4), (8), and (12) we include both sets of parental and spousal variables. * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Table A7: Robustness of the Relationship between the Smoking Polygenic Scores and the Summary Standardized Index of Health

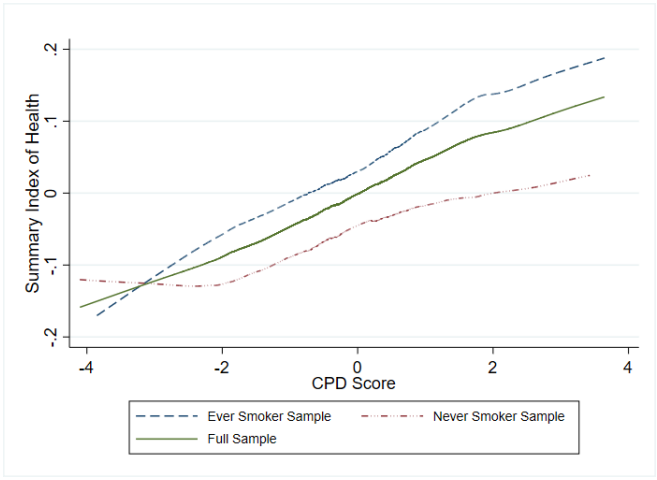
	Full Sample			Ever Smoker Sample			Never Smoker Sample	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Using Older Polygenic Scores								
CPD Score	0.034***	0.029***	0.024***	0.036***	0.030***	0.023***	0.031***	0.024***
	(0.005)	(0.005)	(0.005)	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)
Initiation Score	0.015***	0.019***	0.011**	0.003	0.007	0.002	0.020***	0.023***
	(0.004)	(0.005)	(0.005)	(0.006)	(0.007)	(0.006)	(0.006)	(0.007)
Full Smoking Controls	No	No	Yes	No	No	Yes	No	No
Standard Controls	No	Yes	Yes	No	Yes	Yes	No	Yes
Principal Components	No	Yes	Yes	No	Yes	Yes	No	Yes
Sample Mean	-0.000	-0.000	-0.000	0.038	0.038	0.038	-0.052	-0.052
N	45,850	45,850	45,850	26,509	26,509	26,509	19,341	19,341
R^2	0.007	0.059	0.111	0.006	0.073	0.138	0.008	0.074
Panel B: Including Additional Smoking Polygenic Scores								
CPD Score	0.037***	0.038***	0.032***	0.043***	0.043***	0.034***	0.026***	0.028***
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Initiation Score	0.018***	0.028***	0.016***	0.011*	0.016**	0.007	0.014**	0.025***
	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)
Cessation Score	0.017***	0.025***	0.022***	0.018***	0.022***	0.016**	0.019***	0.029***
	(0.005)	(0.005)	(0.005)	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)
Initiation Age Score	-0.025***	-0.011**	-0.007	-0.022***	-0.011	-0.007	-0.025***	-0.006
	(0.005)	(0.005)	(0.005)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
Full Smoking Controls	No	No	Yes	No	No	Yes	No	No
Standard Controls	No	Yes	Yes	No	Yes	Yes	No	Yes
Principal Components	No	Yes	Yes	No	Yes	Yes	No	Yes
Sample Mean	-0.000	-0.000	-0.000	0.038	0.038	0.038	-0.052	-0.052
N	45,850	45,850	45,850	26,509	26,509	26,509	19,341	19,341
R^2	0.017	0.072	0.118	0.015	0.083	0.143	0.015	0.083
Panel C: Only Those Aged 50–61								
CPD Score	0.043***	0.043***	0.036***	0.049***	0.048***	0.038***	0.031***	0.032***
	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Initiation Score	0.026***	0.034***	0.019***	0.020***	0.023***	0.011	0.020***	0.028***
	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)
Full Smoking Controls	No	No	Yes	No	No	Yes	No	No
Standard Controls	No	Yes	Yes	No	Yes	Yes	No	Yes
Principal Components	No	Yes	Yes	No	Yes	Yes	No	Yes
Sample Mean	-0.029	-0.029	-0.029	0.010	0.010	0.010	-0.081	-0.081
N	32,413	32,413	32,413	18,667	18,667	18,667	13,746	13,746
R^2	0.014	0.062	0.111	0.014	0.078	0.140	0.010	0.069

Note: Each column within a panel presents results from a separate regression where the outcome is the summary standardized health index. Results are presented for the full sample and then separately for ever smokers and never smokers. In Panel A, we use the smoking polygenic scores from The Tobacco and Genetics Consortium (2010). In Panel B, we include additional smoking polygenic scores from Liu et al. (2019). In Panel C, we present results for those aged 50–61 (i.e., below the early Social Security retirement age). * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Figure A1: Relationship between Smoking Polygenic Scores and the Summary Standardized Index of Health



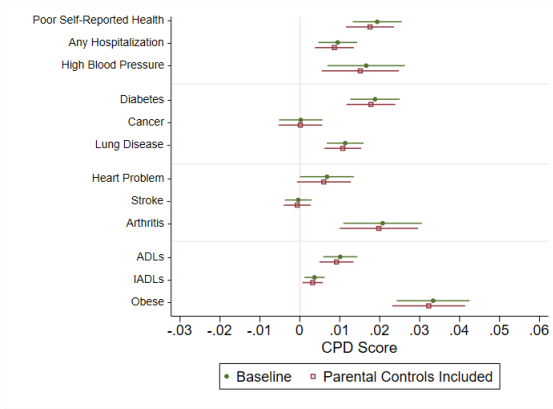
(a) Initiation Score and Summary Standardized Index of Health



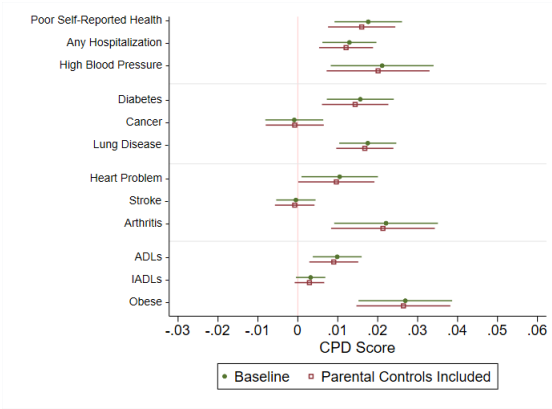
(b) CPD Score and Summary Standardized Index of Health

Note: Panel (a) shows the non-parametric (lowess) relationship between the smoking initiation score and the summary standardized index of health for the full sample, ever smokers, and never smokers. Panel (b) shows the non-parametric (lowess) relationship between the CPD score and the summary standardized index of health for the full sample, ever smokers, and never smokers.

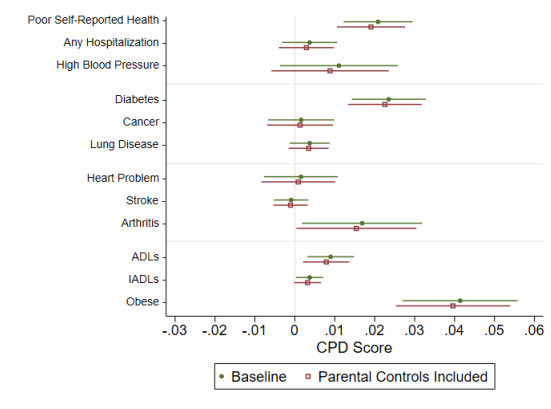
Figure A2: Relationship between the Smoking Polygenic Scores and Individual Health Measures Accounting for Parental Smoking, Risky Behaviors, and Mortality



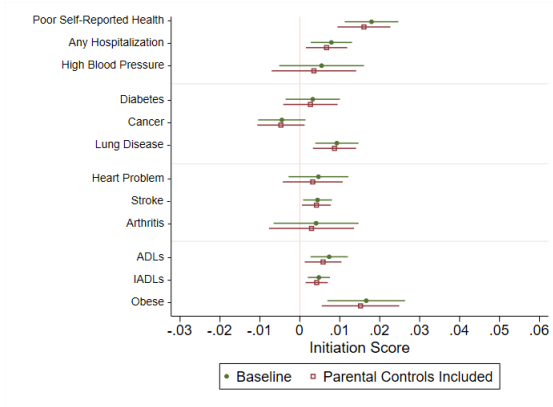
(a) CPD Score (Full Sample)



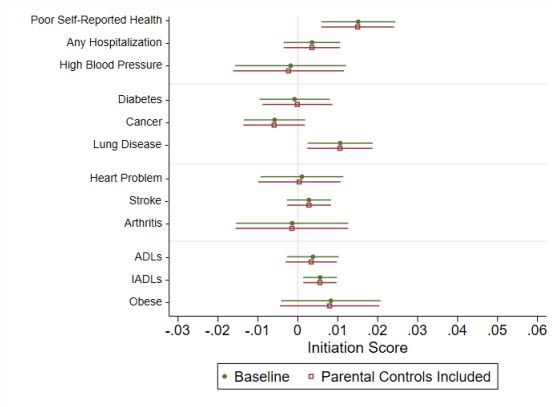
(b) CPD Score (Ever Smoker Sample)



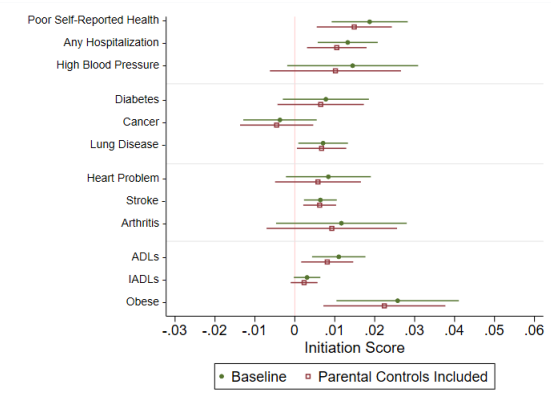
(c) CPD Score (Never Smoker Sample)



(d) Initiation Score (Full Sample)



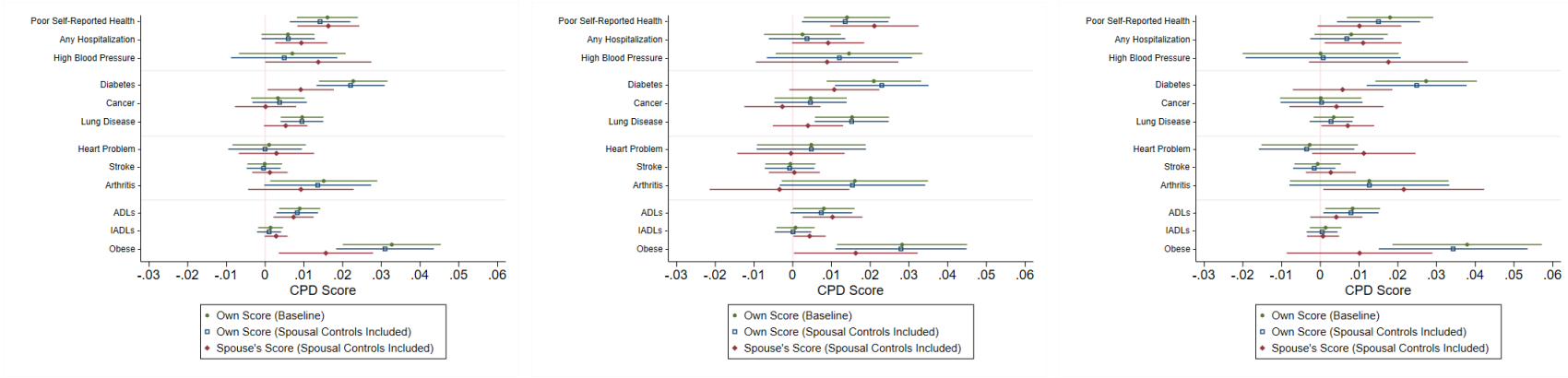
(e) Initiation Score (Ever Smoker Sample)



(f) Initiation Score (Never Smoker Sample)

Note: Panels (a), (b), and (c) present point estimates and 95% confidence intervals for the CPD score for the full sample, ever smokers, and never smokers, respectively. Panels (d), (e), and (f) present point estimates and 95% confidence intervals for the initiation score for the full sample, ever smokers, and never smokers, respectively. In all specifications, we include the standard controls and the genetic principal components described in Section 4, and for the full sample and ever smoker sample, we include the full set of smoking controls described in Section 4. We present results without (baseline) and with parental controls (i.e., parental smoking, parental risky behaviors, and parental mortality as described in Section 3.4).

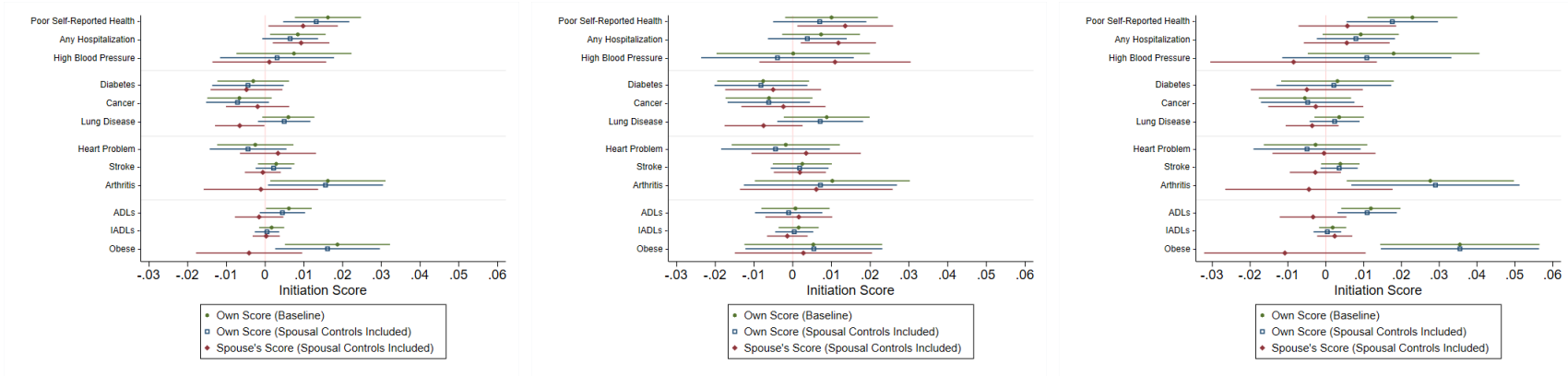
Figure A3: Relationship between the Smoking Polygenic Scores and Individual Health Measures Accounting for Spouse's Smoking Behavior and Smoking Polygenic Scores



(a) CPD Score (Full Sample)

(b) CPD Score (Ever Smoker Sample)

(c) CPD Score (Never Smoker Sample)



(d) Initiation Score (Full Sample)

(e) Initiation Score (Ever Smoker Sample)

(f) Initiation Score (Never Smoker Sample)

Note: Panels (a), (b), and (c) present point estimates and 95% confidence intervals for own CPD score and spousal CPD score for the full sample, ever smokers, and never smokers, respectively. Panels (d), (e), and (f) present point estimates and 95% confidence intervals for own initiation score and spousal initiation score for the full sample, ever smokers, and never smokers, respectively. In all specifications, we include the standard controls and the genetic principal components described in Section 4, and for the full sample and ever smoker sample, we include the full set of own smoking controls described in Section 4. We present results without (baseline) spousal controls and with spousal controls (i.e., the full set of the spouse's smoking controls and the spouse's smoking polygenic scores and their genetic principal components).