

Exposure to Higher Cigarette Taxes during Adolescence Reduces Lifetime Smoking in Individuals with Elevated Genetic Risk

Robel Alemu

University of California Los Angeles (UCLA)

Lauren L. Schmitz

llschmitz@wisc.edu

University of Wisconsin–Madison

Research Article

Keywords:

Posted Date: September 24th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9238462/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

**Exposure to Higher Cigarette Taxes during Adolescence Reduces Lifetime Smoking in
Individuals with Elevated Genetic Risk**

Robel Alemu^{1,2,3}, Lauren L. Schmitz^{*4}

¹University of California Los Angeles (UCLA), 110 Westwood Plaza, Los Angeles, CA 90095,
USA (rob.alemu02@gmail.com)

²Broad Institute of MIT and Harvard, Program in Medical and Population Genetics, Merkin
Building, 415 Main St, Cambridge, MA 02142, USA

³Adelaide Medical School, Faculty of Health and Medical Sciences, Adelaide University,
Adelaide, Australia

⁴University of Wisconsin-Madison, La Follette School of Public Affairs, 1225 Observatory Dr,
Madison, WI 53706 (llschmitz@wisc.edu)

***Correspondence to:** Lauren L. Schmitz (llschmitz@wisc.edu); 1225 Observatory Dr, Madison,
WI 53706; Phone #: 7189160830

Abstract

This study examines whether the effects of cigarette taxes on smoking behavior vary by an individual's genetic predisposition and the timing of policy exposure. Using data from the U.S. Health and Retirement Study, we link historical state-level cigarette excise taxes with polygenic indices (PGIs) for smoking behavior. We find that the timing of tax exposure has distinct effects on different margins of smoking behavior: taxes experienced during adolescence primarily influence long-term smoking participation, while taxes in older adulthood more strongly affect smoking intensity. Our key finding is a significant interaction between adolescent-era taxes and genetic risk. The smoking reduction caused by adolescent taxes is significantly greater for individuals with a higher genetic predisposition to smoke. Conversely, we find no evidence that the effects of taxes in older age are moderated by genetic risk. These results indicate that youth-targeted tax policies are particularly effective for genetically at-risk individuals, and that studies overlooking this heterogeneity may underestimate the long-run benefits of these interventions.

1. Introduction

Tobacco use remains a leading cause of preventable death and imposes substantial economic cost in the United States and globally(1–3). Cigarette taxes are a primary tool for reducing tobacco consumption, yet their effectiveness varies across individuals(4, 5) and their impact on aggregate consumption has recently plateaued(6, 7). This heterogeneity raises a crucial question for economic theory and policy design: what are the mechanisms driving these differential responses, and does the timing of a policy intervention fundamentally alter its long-run impact on population health? A recent literature has begun to explore these questions along two distinct lines of inquiry. One stream of research emphasizes the importance of the timing of policy exposure, demonstrating that higher cigarette taxes experienced during adolescence lead to a significant reduction in smoking and mortality decades later(8). Concurrently, an emerging literature uses genetic data to show that innate predispositions moderate behavioral response to contemporaneous environmental factors, including older-age cigarette taxes(9, 10). However, it remains unclear whether genetic risk moderates the long-run effects of cigarette taxes experienced during the critical window of adolescence.

This study bridges this gap by providing the first estimates of the interaction between adolescent cigarette taxes and genetic predisposition to smoking. We extend the literature in three important ways. First, we disentangle the effects of cigarette taxes experienced during adolescence from those faced in older age, analyzing their influence across multiple margins of smoking behavior: participation (current smoking), cessation, and intensity (cigarettes per day or CPD). Second, we introduce genetic heterogeneity as a key mechanism explaining differential policy responses, using polygenic indexes (PGIs, also called polygenic scores or genetic risk scores) to test whether the effectiveness of taxes is amplified for genetically predisposed individuals. Third, we use an instrumental variable (IV) approach, leveraging tax-induced variation in smoking to estimate its causal impact on long-run morbidity and mortality.

Researchers are increasingly using PGIs within natural or quasi-natural experiments to assess whether the effects of interventions vary by underlying genetic risk(10–18). Such advances were made possible in part due to the growing availability of genetic data in large population studies or biobanks(19, 20), and methodological advances in modeling gene-environment interactions (GxE)(12, 13, 15, 16, 21). PGIs are single-value, objective, time-invariant, continuous indices that summarize an individual's heritable liability to a phenotype or outcome of interest(22, 23). They are computed as weighted sum of common single nucleotide polymorphisms (SNPs), with each SNP's allele multiplied by an effect-size weight from genome-wide association study (GWAS) – large studies that tests millions of variant-phenotype associations and are often meta-analyzed across cohorts to improve precision(19, 20). By aggregating small SNP effects, PGIs explain more variation than single SNPs, hence are well suited for assessing the heterogeneous effects of policies or interventions(12, 22, 24). In the context of smoking,

GWAS of smoking behavior have identified hundreds of SNPs linked to smoking initiation, intensity, and cessation (25–27). These SNPs are often located in or near genes that regulate nicotine addiction, metabolism, and related neurotransmitter pathways in the brain(27–29). This genomic evidence aligns with twin and family studies which indicate that smoking behavior is highly heritable (31-60%)(29, 30).

In this study, we exploit plausibly exogenous changes in cigarette taxes across states and over time to assess whether PGIs for smoking behavior moderated the effect of cigarette taxes encountered during adolescence—a critical developmental window in which most individuals initiate smoking—on smoking behavior in adulthood. To conduct our analysis, we use data from the Health and Retirement Study (HRS), which combines rich, population-representative longitudinal data on smoking and health with genetic information for Americans over age 50. This allows us to construct individual histories of tax exposure from adolescence through adulthood and link them to respondents’ PGIs for smoking initiation, cessation, and intensity.

Our results highlight that the timing of tax exposure has notably different effects across the margins of smoking. Adolescent-era taxes are the primary driver of long-run smoking participation and cessation, while taxes experienced during older age are more influential in determining the number of cigarettes consumed by active smokers. Throughout, we use the terms “current smoker” and “lifetime smoker” interchangeably since individuals rarely initiate smoking after age 50, as nearly nine in ten adults in the United States who have ever been daily smokers tried their first cigarette before age 18(31). Our results show that a one-dollar increase in adolescent-era cigarette taxes is associated with a 5.9 percentage point reduction in the probability of becoming a current or lifetime smoker for an individual with average genetic risk. Notably, this protective effect is significantly larger for those with a higher genetic predisposition; the smoking-reducing effect of the tax increases by 2.7 percentage points for every one standard deviation increase in the smoking initiation PGI. In contrast, the interaction between genetic risk and older-age taxes is small and statistically insignificant for both smoking participation and cessation. The long-run adolescent tax elasticity of smoking participation is -0.41 , meaning that a 10% increase in cigarette taxes during adolescence reduces the likelihood of smoking in older adulthood by 4%. This effect is substantially larger in magnitude than the older-age tax elasticity of -0.29 .

Finally, our IV analysis reveals that ordinary least squares (OLS) models that do not account for the endogeneity of smoking behavior substantially underestimate the adverse health consequences of smoking. Using the sum of the average cigarette taxes that an individual was exposed to in adolescence and at older ages as an instrument, we find that being a current smoker causally increases the risk of lung disease, heart problems, and cancer. These results underscore the significant long-run welfare gains from tax policies that successfully curb smoking. Taken together, our findings demonstrate that cigarette taxes implemented during the critical window of adolescence have lasting effects that are amplified for

genetically at-risk individuals, highlighting the importance of early-life interventions in public health policy.

2. Results

2.1. Summary statistics

Our final estimation sample includes all genotyped European genetic ancestry HRS respondents with non-missing covariate data (N=11,558). A detailed description and schematic overview of the exclusion criteria are presented in Figure S1. Table S1 summarizes individual, household, and state-level time-invariant and time-varying characteristics for the full sample by smoking status. The average age of respondents was 66.96 (SD=11.01), and about 58 percent were female. A relatively larger share of the respondents (57.02 percent) were ever smokers, meaning they smoked at least 100 cigarettes in their lifetime, 43.29 percent of whom reported quitting smoking at the time of the survey. Due to increases in cigarette taxes over time (Figure S2), the average cigarette excise tax that HRS birth cohorts were exposed to during adolescence (\$0.96, SD=0.17) was much lower compared to excise taxes levied when they were surveyed in the HRS after age 50 (\$1.52, SD=0.63). Figure S3 depicts the considerable spatial variation in state tobacco excise taxes across states in 2016 (mean=\$1.63/ pack, SD=1.16). Current and former smokers have significantly higher PGIs for smoking initiation and intensity than never smokers, which means that, on average, they have more genetic variants that put them at higher risk for becoming smokers and smoking more CPD than never smokers. A standard two-sample t-test in column 6 of Table S1 shows that, compared to never smokers, current smokers had fewer years of education, earned less income, and were more likely to live in states with lower cigarette excise taxes and less stringent tobacco control policies ($p<0.01$). On average, relative to current or lifetime smokers, former smokers lived longer, were more educated, and resided in states with more stringent tobacco control policies ($p<0.01$) (column 6 of Table S1). Moreover, former smokers lived in counties with significantly higher travel distances to the nearest lower cigarette excise tax state border than current smokers ($p<0.01$).

2.2. Direct and interactive effects of cigarette taxes and PGIs on smoking behavior

Overall, the PGIs are strong and statistically significant predictors of their corresponding smoking phenotypes (Tables 1-3). As shown in Table S2, the PGI for smoking initiation explains 4% of the variance in current or lifetime smoking, while the PGIs for smoking cessation and CPD explain 2% and 0.56% of the variance in their respective outcomes. The higher predictive power of the smoking initiation PGIs is expected, as the GWAS for this trait can be conducted on much larger samples (including both smokers and non-smokers), whereas the GWAS for cessation and CPD are restricted to smaller samples of

individuals who have ever smoked. In comparison to other predictors, the smoking initiation PGI is 14, 3.1, and 1.4 times more predictive than respondents' age, years of education, and sex, respectively. For smoking cessation and CPD, however, sociodemographic factors such as age and education are more predictive than the corresponding PGIs (Table S2).

2.2.1. Current smoking

Table 1 presents results for current smoking status at older ages. The models are presented sequentially, demonstrating that our main findings are robust to the inclusion of a comprehensive set of individual and state-level controls. In our preferred specification (column 6), the main effect estimates show that both adolescent and older-age taxes are associated with a lower probability of smoking for an individual with average genetic risk. A one dollar increase in the adolescent-era tax is associated with a 5.9 percentage point reduction in the probability of smoking ($p < 0.01$), while a one-dollar increase in the older-age average tax is associated with a 2.6 percentage point reduction ($p < 0.01$). These effects correspond to a long-run adolescent tax elasticity of smoking participation of -0.41 and an older-age tax elasticity of -0.29. This shows that cigarette taxes experienced during formative years are especially powerful in shaping long-run smoking behavior, with a larger preventive effect than taxes during older adulthood. The magnitude of our main effect for adolescent taxes is larger than that found by Friedson et al.(8), who in a different sample (Tobacco Use Supplement to the Current Population Survey) and without incorporating genetic risk, report a 1.7 percentage point reduction.

Additionally, we find a significant interaction between adolescent taxes and genetic predisposition. The coefficient on the interaction term (β_4 in equation 4), reported in column 6 of Table 1, is -0.027 ($p < 0.10$), indicating that the protective effect of adolescent taxes is stronger for individuals with a higher genetic risk for smoking. In contrast, the interaction between older-age taxes and the PGI (β_5) is small and not statistically distinguishable from zero.

To illustrate the magnitude of this interaction, we can calculate the total effect of the adolescent tax for different genetic risk profiles. For an individual with a high genetic risk (1 standard deviation above the mean PGI), a one-dollar increase in the adolescent tax is associated with an 8.6 percentage point reduction in the probability of smoking ($-0.059 + (-0.027 \times 1)$). For an individual with low genetic risk (1 SD below the mean), the same tax increase is associated with a much smaller 3.2 percentage point reduction ($-0.059 + (-0.027 \times -1)$). This demonstrates that the long-term deterrent effect of taxes experienced during the critical window of adolescence is significantly amplified for those who are most genetically predisposed to smoking.

Our finding of a null interaction for average taxes at older ages differs from Slob and Rietveld(10), who, using the same HRS data, report a significant PGI-by-tax interaction of -0.012 for current smoking.

However, direct comparison is difficult as their model does not control for adolescent taxes and uses a different measure of older-age taxes. The null interaction in our model for adult taxes suggests that after accounting for adolescent taxes, the effect of taxes on current smoking after age 50 does not vary by genetic risk.

2.2.2. Smoking cessation

Table 2 presents the results for smoking cessation for former versus current smokers. In our preferred specification (column 6), we find that older-age taxes are associated with smoking cessation. A one dollar increase in adult taxes is associated with a 2.7 percentage point increase in the probability of being a former smoker ($p < 0.05$). In contrast, the main effect of adolescent taxes on cessation is positive but not statistically significant. The corresponding tax elasticities of cessation are 0.081 for adolescent taxes and 0.055 for older-age taxes.

Turning to the interaction effects, we find no evidence that the effect of adult taxes on quitting varies by genetic predisposition. Our finding of null interaction for older-age taxes is consistent with Slob and Rietveld(32), who also do not find a significant genetic gradient in the effect of older-age taxes on cessation, although their models do not account for adolescent taxes. However, we find a marginally significant interaction between adolescent taxes and the smoking cessation PGI ($p < 0.10$). While the main effect of adolescent taxes is not significant for the average-PGI individual, the significant interaction term indicates that the total effect of the tax does vary by genetic risk. The positive coefficient suggests that higher adolescent taxes were more effective in promoting cessation among individuals with a higher genetic predisposition to continue smoking. Specifically, for every one standard deviation increase in the PGI, the quitting-promoting effect of a one dollar adolescent tax increases by 4.2 percentage points.

2.2.3. Smoking Intensity

Table 3 presents the results for smoking intensity, measured as the natural log of cigarettes per day (CPD), for the subsample of current smokers. In our full specification (column 6), we find that smoking intensity is responsive to taxes experienced in older age, but not to those experienced during adolescence. The coefficient on the log of adult taxes is -0.286 ($p < 0.01$). In the log-log specification, this coefficient is interpreted as the tax elasticity of smoking intensity, indicating that a 10% increase in the older-age tax is associated with a 2.86 decrease in the number of cigarettes smoked per day. The magnitude of this elasticity is larger than the -0.03 to -0.04 tax elasticity range found by MacLean et al.(33) using the same HRS data; however, their study uses survey-year specific cigarette tax exposure while we use the average tax exposure during older adulthood, and they do not account for genetic risk.

We do not find evidence that the effects of either adolescent or adult taxes on CPD vary by an individual's genetic predisposition, as both interaction terms are small and statistically insignificant. Our null finding for older-age taxes differs from Slob and Rietveld(10), who find a marginally significant genetic gradient ($p < 0.10$); however, their model does not account for adolescent taxes, which likely explains the difference in results. This overall pattern suggests that while adolescent-era taxes are crucial in shaping the decision to smoke at all (the extensive margin), the quantity of cigarettes consumed by active smokers is more sensitive to contemporaneous tax rates. It is also important to note that these models may be underpowered, as they are run on a subsample of smokers who report smoking at least one cigarette per day, which comprises only 22% (2,533 out of 11,528) of the full analysis sample.

A more intuitive understanding of the genetic gradient in the effects of cigarette taxes on smoking can be gleaned from the two-way contour plots in Figures 1, S4, and S5. The plots depict the predicted probabilities of smoking across PGI values and cigarette excise tax exposure. The patterns in Figure 1 indicate that higher genetic risk individuals were more likely to become lifetime or current smokers when exposed to lower cigarette taxes than those experiencing higher cigarette taxes. Figure 2 further simplifies the interpretation of the observed genetic gradient by plotting coefficients from a model where the PGI is categorized into three groups: Low (bottom 20%), Medium (middle 60%), and High (top 20%). Relative to the Medium PGI reference group, the interaction for adolescent taxes (Panel A) is positive and statistically significant for the High PGI group, indicating a stronger smoking-reducing effect of the tax for this group. Conversely, the interaction for older-age taxes (Panel B) is not statistically significant for any group. This categorical analysis reinforces our primary finding from the continuous interaction model: the moderating effect of genetic risk is unique to taxes experienced during adolescence. Furthermore, a genetic risk gradient also emerges in the effects of cigarette taxes on the likelihood of quitting (Figure S4). Specifically, panel A of Figure S4 reveals that those with a higher genetic propensity to avoid quitting were more responsive to marginally higher adolescent cigarette taxes than those with a lower genetic predisposition to resist quitting.

2.3. Effects of cigarette-tax-induced changes in smoking on health

Our core results show that individuals with higher PGIs for smoking tend to benefit from higher cigarette excise taxes, as it increases the likelihood of smoking cessation, above and beyond what is expected for individuals with average PGIs. Having detected a robust PGI-by-tax interaction effect, we now assess how changes in smoking behavior induced by higher cigarette taxes impacted the likelihood of developing chronic illnesses and survival. To realize this, we employ an instrumental variable (IV) approach to estimate the causal link between smoking and various health indicators. We present coefficients from two-stage least squares (2SLS) models relative to naïve OLS models in Table 4 and Tables S4-S5.

Our instrument, the sum of average cigarette excise taxes experienced during adolescence and older age, meet the relevance assumption of 2SLS models as all of our 1st stage regressions had high F-statistics greater than 10, the rule of thumb threshold for 2SLS models.

Overall, we find that the endogenous OLS model consistently underestimates the effects of smoking (both in the extensive and intensive margins) on several chronic illnesses and survival. In Table 4, we show that being a lifetime smoker increases the risk of developing lung disease, heart problems, cancer (any), diabetes, and death at the time of the survey ($p < 0.01$). Our 2SLS estimates, for instance, show that being a current smoker increases the likelihood of reporting any form of lung disease by 68 percent, while the OLS indicates an increase of just 10.1 percent. Similarly, the 2SLS models show that being a current smoker increases the risk of reporting any heart problem (123%), any form of cancer (57.1%), and diabetes (66.1%), whereas the effects were null in the OLS models. Furthermore, being a lifetime smoker decreases survival at the time of the survey by 152.5 percent based on our 2SLS model, while the decline in survival was just 2 percent in the OLS model. In addition, we also show the impacts of other smoking phenotypes (quitting and CPD) in our supplementary information (Tables S4 and S5). Overall, we find qualitatively similar effects of smoking on the health outcomes in question, and that the OLS models are consistently downward-biased.

2.4. Robustness analyses

Our main analysis is restricted to HRS participants who survived until at least 2006 to provide genetic data. As shown by Domingue et al.(34), this can introduce mortality selection bias, as survivors tend to be healthier and less likely to smoke than those who died before 2006. To assess the sensitivity of our estimates to this potential bias, we re-estimate our main specification using an inverse probability weighting (IPW) approach to account for differential survival (see Supplementary Note A for details). We find that our results are robust to this correction. After applying the weights, the magnitude of the key interaction term for adolescent taxes and the smoking initiation PGI on current smoking is consistent, confirming that our main findings are not driven by mortality selection (Table 5).

Second, following Keller(35), we address potential bias that can arise when population structure correlates with environmental exposures (e.g., state cigarette taxes) by including interactions between each of the top ten genetic principal components (PCs) and both the adolescent and adult tax variables. This specification allows tax effects to vary along ancestry-related axes of genetic variation, reducing the risk that ancestry-correlated environmental structure is misattributed to the PGI. Estimates from this specification (Supplementary Table S3) are qualitatively unchanged relative to the main results: the adolescent tax main effect and its interaction with the smoking-initiation PGI remain negative and statistically significant, while the corresponding adult-tax interaction remains small and imprecise.

Together with the IPW results above, these findings indicate that our core conclusions are robust to potential biases related to selective survival and population structure(35).

3. Discussion

In this study, we examine how the long-run effects of cigarette taxes vary by genetic predisposition and the timing of policy exposure. Our results yield two primary findings. First, we show that the timing of tax exposure has distinct effects on different margins of smoking behavior. Taxes experienced during adolescence are the most powerful determinants of long-run smoking status, while taxes experienced in adulthood are associated with smoking cessation and smoking intensity among active smokers. Second, we provide evidence that the long-run deterrent effect of adolescent-era taxes on current smoking is significantly amplified for individuals with a higher genetic predisposition to smoke. In contrast, we find no evidence that the effects of older-age taxes on either smoking participation or intensity are moderated by genetic risk.

Our results have broader implications for both economic theory and public policy. Regarding theory, the canonical rational addiction model(36) in economics provides a framework for understanding addictive behaviors by assuming that individuals make forward-looking decisions, weighing current consumption against expected future costs. However, this model does not explicitly incorporate innate predisposition to addiction. Later extensions allow for stable individual differences – such as variation in how quickly the harmful effects of past smoking accumulate (the addictive stock), how heavily individuals value immediate gratification relative to long-term health (discounting behavior), or how sensitive they are to the reinforcing effects of consumption(37–41). By using a PGI as a direct proxy for such predispositions – rather than treating it as an unobserved parameter – we can examine heterogeneity in tax responsiveness along a dimension that the original model could not measure directly. The significant interaction between adolescent cigarette taxes and genetic risk suggests that early-life price shocks are particularly effective for those with a high genetic predisposition to smoke, which is in line with the predictions of the rational addiction model. This result would be obscured in analyses that ignore genetic heterogeneity(8, 33). From a policy perspective, this underscores that youth-targeted tobacco taxes are not only broadly effective but also disproportionately benefit those most at risk of initiating smoking. Moreover, our findings suggest that the recently observed stagnation in cigarette tax responsiveness may be temporary and could reverse as newer cohorts of high-risk individuals are exposed to progressively higher taxes during adolescence(8).

Beyond smoking behavior, our analysis demonstrates that the reductions in smoking attributable to higher cigarette taxes are associated with substantial long-term health benefits. Using an IV 2SLS approach to address endogeneity, we find that OLS models substantially understate the adverse effects of smoking

on chronic disease and survival. The 2SLS estimates indicate large causal impacts of smoking on lung disease, heart problems, cancer, diabetes, and mortality, highlighting that policies which effectively reduce smoking—whether by preventing initiation or promoting cessation—can yield significant gains in population health. These findings underscore the broader welfare implications of tobacco tax interventions.

Certain aspects of our study design point to important avenues for future research. First, our analysis is restricted to respondents of European genetic ancestry, meaning our findings may not generalize to other genetic ancestries. As genetic data from more diverse samples become available, future research will be crucial to evaluate the replicability of these findings across different genetic ancestries. Second, the PGIs used in this study are derived from standard (between-family) GWAS. Studies have shown that such PGIs are susceptible to biases from factors such as gene-environment correlation (including indirect genetic effects and population stratification) and assortative mating, which could, in turn, bias our estimates(42, 43). Future research could test the validity of the genetic gradient found here by using PGIs derived from within-family GWAS design, which mitigate such biases by more cleanly isolating direct genetic effects(42, 44, 45). Despite these considerations, this study provides novel evidence on the interplay between the timing of policy interventions and innate genetic risk, highlighting the long-run benefits of early-life policy interventions.

4. Data and Methods

4.1. Description of study cohort

The Health and Retirement Study (HRS) is a nationally representative longitudinal study of individuals aged 50 plus and their spouses in the U.S. We use the RAND Longitudinal File, which contains harmonized data from 13 waves of HRS interviews (1992-2018)(46, 47). This dataset provides comprehensive information on respondents' socioeconomic status, demographic characteristics, health behaviors, and physical/functional ability. In addition, we use restricted geographic data to identify respondents' state of residence during primary schooling years (adolescence) and while in the HRS (adulthood)(46, 47). The initial sample from these waves includes 42,719 individuals. From this sample, we apply several restrictions to arrive at our final estimation sample (see Figure S1 for a detailed description). First, we restrict the analysis to individuals of European genetic ancestry, as the PGIs used in this study were derived from SNP weights that were estimated in individuals of European ancestry and hence do not predict phenotypes as accurately in other genetic ancestries. The sample is further limited by excluding respondents who were deceased prior to the first wave of genotyping in 2006, as well as those who were alive in 2006 or in subsequent waves but were not genotyped. Finally, we exclude individuals with missing values for any covariates used in our regression models. These criteria yielded a final estimation sample of 11,558 individuals (Figure S1).

4.2. Assessment of smoking behavior

From the HRS survey waves, we construct three measures of smoking behavior: current smoking, smoking cessation, and smoking intensity. We create an indicator for whether a respondent currently smokes (Current Smoker) and an indicator for smoking cessation (Former Smoker). Given the low probability of initiating smoking after age 50, we also classify current smokers in the HRS as “lifetime smokers.” For smoking intensity, we construct a continuous variable for the number of cigarettes smoked per day (CPD) among current smokers. If responses were reported in packs per day they were converted to cigarettes (1 pack = 20 cigarettes). The use of maximum CPD for former smokers is supported by prior work that found a strong correlation between average and maximum CPD over time(48). While these self-reported measures do not capture all aspects of tobacco use, such as depth of inhalation or mode of consumption, they provide a reasonable proxy for consumption(49–51).

4.3. Measuring genetic risk for smoking

PGIs were constructed for each respondent by summing the number of risk-associated alleles they carry at each SNP, weighted by the strength of the SNP’s association with the smoking trait of interest estimated from a GWAS(52). Specifically, SNP weights are the effect-size estimates from a regression of the trait on each SNP, run one at a time:

$$Y_i = \alpha + \beta_m SNP_{im} + \delta X_i + \varepsilon \quad (1)$$

where Y_i is the smoking trait for individual i , and SNP_{im} is the genotype of individual i at locus m , coded as 0, 1, or 2 depending on the number of reference alleles present. The PGI for each individual is then constructed as the weighted sum of their genotypes across all M SNPs included in the index:

$$PGI_i = \sum_{m=1}^M \hat{\beta}_m SNP_{im} \quad (2)$$

PGIs constructed from the most recent GWAS explain approximately 5-10% of the variance in smoking behavior, net of the variance explained by standard covariates(26, 53, 54). The PGIs used in this study were constructed in the HRS using GWAS summary statistics from discovery samples that exclude HRS. For smoking initiation and cessation, we use PGIs developed by the Social Science Genetic Association Consortium (SSGAC)(53), which have superior predictive accuracy as they were generated by meta-analyzing summary statistics from several large-scale GWASs, including the UK Biobank and 23andMe. As a GWAS for current smoking is not yet available, we use the smoking initiation PGI in our analysis of current smoking status. For smoking intensity (CPD), we rely on the CPD PGI released by the HRS, as the SSGAC did not generate smoking intensity PGIs(55). For ease of interpretation, all PGIs are

standardized to have a mean of zero and a standard deviation of one. Higher PGIs for smoking initiation, cessation, and CPD are associated with worse outcomes (i.e., a higher probability of becoming a regular smoker, a higher probability of not quitting, and increases in CPD). A critical consideration when using PGIs is that GWAS are often performed within a specific genetic ancestry as differences in allele frequencies and linkage disequilibrium (the correlation structure between SNPs) can distort estimated relationship in pooled samples(56–58). To date, genotyped sample sizes for populations of non-European genetic ancestries have not been large enough to produce well-powered GWAS for smoking behavior(26). Consequently, we restrict our analysis to individuals of European genetic ancestry, as the available PGIs do not have the same predictive power for individuals from other genetic ancestries, and comparisons across groups could confound genetic effects with environmental differences(59). Even within our European genetic ancestry sample, we must account for population stratification – the possibility that subtle genetic structure arising from non-random mating could be spuriously correlated with our outcomes of interest(60, 61). To address this, we control for the first 10 principal components (PCs) of the genotype data in all regression models, a standard approach to account for fine-scale population structure(61–63).

4.4. Measuring exposure to cigarette taxes

Data on annual state-level excise tax per cigarette pack comes from the historical Tax Burden on Tobacco dataset(64). The data spans all years that HRS respondents were in adolescence and/or adulthood (1940-2018) (Figures S2 and S4). We adjust for inflation using the Consumer Price Index with 1991 as the base year. To measure cigarette tax exposure during adolescence, we calculate the average per-pack cigarette tax levied in the respondent’s state of residence when they were between 10 and 18 years of age. Specifically, the HRS asks respondents what state they lived in most of the time when they were in grade school, high school, or around age 10. Since this is the only information available on state of residence after birth until individuals are surveyed in the HRS after age 50, we assume the same state of residence for ages 10-18. Here, we closely following the setup employed by Friedson et al.(8):

$$Adolescent\ Cig.\ Tax_i = \frac{1}{9} \sum_{\eta=10}^{\eta=18} Tax_{s(i), y(i)+\eta} \quad (3)$$

where $y(i)$ is respondent i ’s year of birth and $s(i)$ is their state of residence during adolescence. Cigarette taxes at older ages (i.e., after age 50) were linked to respondents based on their current state of residence at each wave. Similarly, tax exposure at older ages was averaged across waves in which respondents were observed in the HRS as per equation (3).

4.5. Accounting for other tobacco control policies

It is plausible that certain time-varying factors across states, such as anti-smoking sentiment, could be correlated with both cigarette taxes and smoking behavior, biasing the estimated tax effects(65). To address this concern, we control for a comprehensive set of state-level policies and economic conditions. We compile data on smoke-free air (SFA) laws from the ImpacTeen Project and the Americans for Non-Smokers' Rights Foundation (ANRF)(66), from which we construct a variable averaging the restrictiveness score of bans across private and government worksites. These restrictiveness scores range from 0 (no ban) to either 3 or 5, with higher scores indicating more comprehensive and restrictive bans. For each respondent, we calculate the average score across all survey years to reflect cumulative exposure. We also source data on the minimum legal purchase age (MLPA) from the ImpacTeen Project (ImpacTeen.org, 2009). Finally, we include the annual state unemployment rate from the Bureau of Labor Statistics to account for economic conditions. All state-level datasets were merged with the HRS data by state and survey year.

4.6. Proxy measure for cigarette tax avoidance behavior

A potential source of bias in estimating the effects of cigarette taxes is tax avoidance behavior, as price-sensitive smokers may purchase cigarettes from neighboring states with lower taxes(65, 67). To address this, our models control for the opportunity to engage in cross-border cigarette purchases. We construct a proxy for this behavior by computing the travel distance in miles from the centroid of a respondent's county of residence in adulthood to the border of the closest state with a lower cigarette tax (level and squared). This calculation was performed for each respondent in each survey wave using geospatial analysis tools in ArcGIS Pro (Esri, v3.2, Redlands, CA, 2024) (details in Supplementary Note B).

4.7. Other covariates

Our regression models include a rich set of individual-level control variables to account for demographic and socioeconomic factors correlated with smoking behavior. Demographic controls include an indicator for sex (female, with male as the omitted category) and a quadratic for age. Socioeconomic status is controlled for using a quadratic for years of education and a series of indicator variables for marital status (divorced, separated, widowed, or never married, with married or partnered as the reference category). To control for household income, we convert annual household income to 2008 dollars using the CPI and construct a set of five indicators for income quintiles, with the first quintile as the omitted category.

4.8. Empirical Strategy

4.8.1. Testing for effects of cigarette taxes on smoking by genetic predisposition

Our empirical strategy is designed to test whether the demand response to adolescent-era cigarette taxes varies by an individual's genetic predisposition, while disentangling these effects from taxes faced in older adulthood. We analyze smoking behavior across three margins: current smoking and smoking cessation (extensive margins), and smoking intensity (as measured by CPD, intensive margin). Our primary analysis uses a pooled cross-section of all surveys, closely following the model specifications of prior literature(8, 33). The rationale for pooling observations is that there is little within-individual variation in smoking behavior for this older cohort. For our binary outcomes, we estimate a linear probability model (LPM), and for our continuous outcome, we run an ordinary least squares (OLS) model, both within a high-dimensional fixed-effects framework.

Our general specification is as follows:

$$Y_{ist} = \beta_0 + \beta_1 Tax_{is}^{Adol} + \beta_2 Tax_{is}^{Adult} + \beta_3 PGI_i + \beta_4 (Tax_{is}^{Adol} \times PGI_i) + \beta_5 (Tax_{is}^{Adult} \times PGI_i) + \delta X_{ist} + \lambda Z_{st} + \gamma_t + \gamma_b + \gamma_s^{Adol} + \gamma_s^{Adult} + \gamma_s^{Adult} \times t + \varepsilon_{ist} \quad (4)$$

This equation is adapted for each margin of smoking. For the extensive margin, Y_{ist} is a binary indicator for current or former smoker; Tax_{is}^{Adol} is the average annual cigarette excise tax that the HRS respondent experienced during adolescence, and Tax_{is}^{Adult} is the average annual cigarette tax that respondents were exposed to while in the HRS; PGI_i is the individual's PGI for the respective smoking outcome of interest (i.e., smoking initiation, cessation, or CPD). For current smoking and cessation, the coefficients β_1 and β_2 are interpreted as the percentage point change in the probability of the outcome for a one dollar increase in the adolescent or adult tax, respectively. For the intensive margin, Y_{ist} is the natural log of CPD and the tax variables are also log-transformed. In this log-log specification, the coefficients β_1 and β_2 are interpreted as tax elasticities. Current smoking models are estimated in the sample of current, former, and never smokers; cessation models are estimated in the subsample of current and former smokers; CPD models were estimated in the subsample of current smokers.

The primary coefficients of interest are β_4 and β_5 , which capture how the effects of adolescent and older-adulthood taxes are moderated by genetic risk. A negative value for β_4 in the current smoking model would indicate that the smoking-reducing effect of adolescent taxes is stronger for individuals with a higher genetic predisposition for smoking initiation.

In all specifications, we include a vector of individual-level controls X_{ist} (female, age, age squared, years of education and its square, marital status, household income, and the first 10 PCs of the genotype data), and a vector of time-varying state-level controls for current state of residence Z_{st} (unemployment rate, smoke-free air laws, and a proxy for cigarette tax avoidance). To control for unobserved heterogeneity,

the model includes a full set of fixed effects for survey year (γ_t), birth-year (γ_b), state of residence during adolescence (γ_s^{Adol}), and current state of residence (γ_s^{Adult}). The geographic fixed effects help absorb time-invariant differences at the state level, while the time fixed effects absorb factors that vary over time but are invariant across states. Furthermore, to account for smooth, unobserved changes within states over time, we also include state-specific linear time trends ($\gamma_s^{Adult} \times t$). ε_{ist} is the idiosyncratic error term. For our main pooled models, standard errors are two-way clustered at the individual level to correct for autocorrelation across survey waves for the same person, and at the adolescent state-of-residence level to account for correlated shocks among individuals who grew up in the same state.

Finally, we compute tax elasticities for all outcomes. In this context, elasticities express the percentage change in smoking behavior associated with a 1% change in cigarette taxes, making them a standard metric in health economics and public health for comparing the effectiveness of price-based interventions across settings and populations. For the log-log CPD model, the coefficients on the logged tax variables (adjusted for the interaction term) are the elasticities. For the LPM on current smoking and cessation, we calculate the elasticities at the sample mean using the standard formula ($Elasticity = [\beta_{Tax} + (\beta_{Interaction} \times PGI_{mean})] \times (\frac{Tax_{mean}}{Outcome_{mean}})$).

4.8.2. Estimating the causal effects of smoking on health

Our second objective is to estimate the causal impact of smoking on long-run health outcomes. We first estimate a naïve OLS model, which regresses health directly on smoking behavior, as a baseline for comparison:

$$Health_{ist} = \pi_0 + \pi_1 Smoking_{ist} + \pi_2 PGI_i + \Phi X_{ist} + \Lambda Z_{st} + \Gamma + v_{ist} \quad (5)$$

This model is likely biased due to the endogeneity of smoking behavior, which can arise from two primary sources. First, there may be unobserved confounding variables, such as an individual's risk preference or discount rate, that are correlated with both the decision to smoke and other health-related behaviors. For instance, individuals with a higher tolerance for risk may be more likely to smoke and less likely to engage in preventive healthcare, leading OLS to overstate the causal effect of smoking on poor health. Second, reverse causality is a significant concern; a negative health shock, such as a lung disease diagnosis, may cause an individual to quit or reduce their smoking. This would lead OLS to underestimate the true adverse effect of smoking, as the illness itself reduces the observed prevalence of smoking in the sample.

To address this, we use an instrumental variable (IV) approach with a two-stage least squares (2SLS) estimator. We instrument for an individual's smoking behavior – across all three margins of smoking – using their average lifetime exposure to state-level cigarette taxes. Specifically, our instrument

is the sum of the average cigarette excise tax a respondent was exposed to during adolescence and the average tax they were exposed to during older adulthood (i.e., while in the HRS). Cigarettes are a valid instrument as they strongly predict smoking behavior (relevance) but are plausibly uncorrelated with an individual's health outcomes except through their effect on smoking (exclusion restriction)(8, 68).

The first stage of our 2SLS model regresses the endogenous smoking variables on the tax instrument and the full set of exogenous controls:

$$Smoking_{ist} = \phi_0 + \phi_1(Tax_{is}^{Adol} + Tax_{is}^{Adult}) + \phi_2PGI_i + \Phi X_{ist} + \Lambda Z_{st} + \Gamma + \mu_{ist} \quad (6)$$

where $Smoking_{ist}$ is the measure of smoking behavior (current smoking, cessation and CPD); the model includes the same vector of individual controls (X_{ist}), state-level controls (Z_{st}), and the full set of fixed effects (Γ) as in equation (3).

The second stage uses the predicted value of smoking from the first stage to estimate the average causal effect of smoking on health:

$$Health_{ist} = \theta_0 + \theta_1 \widehat{Smoking}_{ist} + \phi_2PGI_i + \Phi X_{ist} + \Lambda Z_{st} + \Gamma + \omega_{ist} \quad (7)$$

where $Health_{ist}$ is the health outcome of interest, and $Smoking_{ist}$ is the predicted value from the first stage. The coefficient of interest, θ_1 , is interpreted as the causal effect of smoking on health. The model includes the same vector of individual controls (X_{ist}) and state-level controls (Z_{st}), as the first stage. The term Γ represents the full set of fixed effects for survey year, birth year, adolescent state, and current state. Consistent with our primary analysis, these IV models are estimated on the pooled cross-sectional sample as per prior literature(8, 10, 33).

Disclosure Statement

The authors declare no conflicts of interest.

Ethics

All analyses in this study were conducted using restricted Health and Retirement Study (HRS) data under an approved data-use agreement. Ethical approval for the research was granted by the University of Wisconsin–Madison Institutional Review Board (protocol 2019-0924-CP003). The HRS follows rigorous human subjects protections, including obtaining written informed consent from all participants prior to data collection. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

Data Availability

This study used restricted individual-level information from the HRS, and our contractual agreement does not permit public dissemination of the data. Details on how to access restricted data for the HRS can be found at <https://hrs.isr.umich.edu/data-products/restricted-data>. This study was approved by the Institutional Review Board (IRB) at the University of Wisconsin-Madison (2019-0924-CP003). HRS obtained informed consent from all participants.

Funding

Alemu received funding from the Neubauer Family Fellowship that covered the cost of attending his doctoral studies. Schmitz acknowledges funding from the National Institute on Aging (NIA): K99/R00 AG056599 and P30 AG017266. The content is solely the authors' responsibility and does not necessarily represent the official views of the National Institutes of Health.

Acknowledgments

We are grateful for the valuable feedback from Profs. William A. Masters, Steven A. Block, Cynthia Kinnan, Dan Benjamin, Patrick Turley, David Cesarini, Jason Fletcher, and Norma Coe. We thank seminar and conference participants at Tufts University, the “Frontiers in Genetics and Economics Conference” hosted by BRIDGE (BRInging Data on Genetics to Economics), and lab meeting at the Social Science Genetic Association Consortium (SSGAC). We are thankful for the research assistance by Lucinda Toyoma.

References

1. World Health Organization, *WHO report on the global tobacco epidemic, 2017: monitoring tobacco use and prevention policies* (2017).
2. W. Max, H.-Y. Sung, Y. Shi, Deaths From Secondhand Smoke Exposure in the United States: Economic Implications. *American journal of public health* (1971) **102**, 2173–2180 (2012).
3. S. S. Shrestha, *et al.*, Cost of Cigarette Smoking--Attributable Productivity Losses, U.S., 2018. *Am J Prev Med* **63**, 478–485 (2022).
4. D. B. Kandel, G. E. Kiros, C. Schaffran, M. C. Hu, Racial/Ethnic Differences in Cigarette Smoking Initiation and Progression to Daily Smoking: A Multilevel Analysis. *Am J Public Health* **94**, 128–135 (2004).
5. S. T. Yen, Zero observations and gender differences in cigarette consumption. *Appl Econ* **37**, 1839–1849 (2005).
6. American Lung Association, “State of Tobacco Control” (2025).
7. M. E. Cornelius, C. G. Loretan, T. W. Wang, A. Jamal, D. M. Homa, Tobacco Product Use Among Adults — United States, 2020. *MMWR Recommendations and Reports* **71**, 397–405 (2022).
8. A. Friedson, M. Li, K. Meckel, D. I. Rees, D. W. Sacks, Cigarette taxes, smoking, and health in the long run. *J Public Econ* **222** (2023).
9. J. M. Fletcher, Why Have Tobacco Control Policies Stalled? Using Genetic Moderation to Examine Policy Impacts. *PLoS One* **7**, e50576–e50576 (2012).
10. E. A. W. Slob, C. A. Rietveld, Genetic predispositions moderate the effectiveness of tobacco excise taxes. *PLoS One* **16**, e0259210 (2021).
11. L. L. Schmitz, D. Conley, The effect of Vietnam-era conscription and genetic potential for educational attainment on schooling outcomes. *Econ Educ Rev* **61**, 85–97 (2017).
12. P. Biroli, *et al.*, The Economics and Econometrics of Gene–Environment Interplay. *Review of Economic Studies* (2025). <https://doi.org/10.1093/restud/rdaf034>.
13. S. H. Barcellos, L. S. Carvalho, P. Turley, Education can reduce health differences related to genetic risk of obesity. *Proc Natl Acad Sci U S A* **115**, E9765–E9772 (2018).
14. J. M. Fletcher, Q. Lu, Health policy and genetic endowments: Understanding sources of response to Minimum Legal Drinking Age laws. *Health Economics (United Kingdom)* **30**, 194–203 (2021).
15. L. Schmitz, D. Conley, The Long-Term Consequences of Vietnam-Era Conscription and Genotype on Smoking Behavior and Health. *Behav Genet* **46**, 43–58 (2016).
16. J. Randolph Haber, *et al.*, Effect of Paternal Alcohol and Drug Dependence on Offspring Conduct Disorder: Gene—Environment Interplay. *J Stud Alcohol Drugs* **71**, 652–663 (2010).
17. E. Krapohl, *et al.*, Widespread covariation of early environmental exposures and trait-associated polygenic variation. *Proc Natl Acad Sci U S A* **114**, 11727–11732 (2017).
18. C. Mitchell, *et al.*, Role of mother’s genes and environment in postpartum depression. *Proc Natl Acad Sci U S A* **108**, 8189–8193 (2011).
19. A. Abdellaoui, L. Yengo, K. J. H. Verweij, P. M. Visscher, 15 years of GWAS discovery: Realizing the promise. *Am J Hum Genet* [Preprint] (2023).

- 635 20. M. C. Mills, C. Rahal, The GWAS Diversity Monitor tracks diversity by disease in real time.
636 *Nat Genet* (2020). <https://doi.org/10.5281/zenodo.3600471>.
- 637 21. T. Gracner, C. Boone, P. J. Gertler, Exposure to sugar rationing in the first 1000 days of
638 life protected against chronic disease. *Science* (1979) **386**, 1043–1048 (2024).
- 639 22. E. Uffelmann, *et al.*, Genome-wide association studies. *Nature Reviews Methods Primers*
640 [Preprint] (2021).
- 641 23. D. J. Benjamin, D. Cesarini, P. Turley, A. Strudwick Young, Social-Science Genomics:
642 Progress, Challenges, and Future Directions. *NBER Working Paper* (2024).
- 643 24. D. J. Benjamin, *et al.*, The promises and pitfalls of genoconomics. *Annu Rev Econom* **4**,
644 627–662 (2012).
- 645 25. A. M. Erzurumluoglu, *et al.*, Meta-analysis of up to 622,409 individuals identifies 40 novel
646 smoking behaviour associated genetic loci. *Mol Psychiatry* **25**, 2392–2409 (2020).
- 647 26. M. Liu, *et al.*, Association studies of up to 1.2 million individuals yield new insights into
648 the genetic etiology of tobacco and alcohol use. *Nat Genet* **51**, 237–244 (2019).
- 649 27. K. Xu, *et al.*, Genome-wide association study of smoking trajectory and meta-analysis of
650 smoking status in 842,000 individuals. *Nat Commun* **11** (2020).
- 651 28. D. J. K. Balfour, M. Munafò, *The neurobiology and genetics of nicotine and tobacco*
652 (Cham, Switzerland : Springer, 2015).
- 653 29. J. M. Vink, G. Willemsen, D. I. Boomsma, Heritability of Smoking Initiation and Nicotine
654 Dependence. *Behav Genet* **35**, 397–406 (2005).
- 655 30. L. C. Bidwell, R. H. C. Palmer, L. Brick, J. E. McGeary, V. S. Knopik, Genome-wide single
656 nucleotide polymorphism heritability of nicotine dependence as a multidimensional
657 phenotype. *Psychol. Med* **46**, 2059–2069 (2016).
- 658 31. S. L. Hedden, *et al.*, “Behavioral Health Trends in the United States: Results from the
659 2014 National Survey on Drug Use and Health” (2014).
- 660 32. E. Slob, PLOS ONE Genetic predispositions moderate the effectiveness of tobacco excise
661 taxes.
- 662 33. J. C. MacLean, A. S. Kessler, D. S. Kenkel, Cigarette Taxes and Older Adult Smoking:
663 Evidence from the Health and Retirement Study. *Health Economics (United Kingdom)* **25**,
664 424–438 (2016).
- 665 34. B. W. Domingue, *et al.*, Mortality selection in a genetic sample and implications for
666 association studies. *Int J Epidemiol* **46**, 1285–1294 (2017).
- 667 35. M. C. Keller, Gene × environment interaction studies have not properly controlled for
668 potential confounders: The problem and the (simple) solution. *Biol Psychiatry* [Preprint]
669 (2014).
- 670 36. Becker, K. Murphy, A Theory of Rational Addiction. *J Polit Econ* **96**, 675–700 (1988).
- 671 37. P. Arcidiacono, H. Sieg, F. Sloan, Living rationally under the volcano? An empirical
672 analysis of heavy drinking and smoking. *Int Econ Rev (Philadelphia)* **48**, 37–65 (2007).
- 673 38. B. H. Baltagi, J. M. Griffin, The econometrics of rational addiction: The case of cigarettes.
674 *Journal of Business and Economic Statistics* **19**, 449–454 (2001).
- 675 39. G. Becker, M. Grossman, K. Murphy, An empirical analysis of cigarette addiction.
676 *American Economic Review* **3** (1994).
- 677 40. F. Chaloupka, “Rational Addictive Behavior and Cigarette Smoking” (1991).

- 678 41. J. Gruber, B. Kořszegi, K. Kořszegi, "IS ADDICTION 'RATIONAL'? THEORY AND EVIDENCE*" (2001).
- 679
- 680 42. A. Kong, *et al.*, "The nature of nurture: Effects of parental genotypes" (2018).
- 681 43. A. I. Young, S. Benonisdottir, M. Przeworski, A. Kong, Deconstructing the sources of
- 682 genotype-phenotype associations in humans. *Science* (1979) [Preprint] (2019).
- 683 44. J. Guan, *et al.*, Family-based genome-wide association study designs for increased power
- 684 and robustness. *Nat Genet* **57**, 1044–1052 (2025).
- 685 45. A. I. Young, *et al.*, Mendelian imputation of parental genotypes improves estimates of
- 686 direct genetic effects. *Nat Genet* **54**, 897–905 (2022).
- 687 46. (RAND HRS Longitudinal File 2018 (V2)) public use dataset Health and Retirement Study,
- 688 "Produced and distributed by the University of Michigan with funding from the National
- 689 Institute on Aging (grant number NIA U01AG009740)" (2022).
- 690 47. H. L. F. 2018 RAND, "Produced by the RAND Center for the Study of Aging, with funding
- 691 from the National Institute on Aging and the Social Security Administration" (2018).
- 692 48. H. Furberg, *et al.*, Genome-wide meta-analyses identify multiple loci associated with
- 693 smoking behavior. *Nat Genet* **42**, 441–447 (2010).
- 694 49. W. H. Chan, *et al.*, Verifying the accuracy of self-reported smoking behavior in female
- 695 volunteer soldiers. *Sci Rep* **13** (2023).
- 696 50. D. L. Patrick, *et al.*, The Validity of Self-Reported Smoking: A Review and Meta-Analysis.
- 697 *Am J Public Health* **84** (1997).
- 698 51. R. J. Volk, *et al.*, Reliability of self-reported smoking history and its implications for lung
- 699 cancer screening. *Prev Med Rep* **17** (2020).
- 700 52. E. Uffelmann, *et al.*, Genome-wide association studies. *Nature Reviews Methods Primers*
- 701 [Preprint] (2021).
- 702 53. J. Becker, *et al.*, Resource profile and user guide of the Polygenic Index Repository. *Nat*
- 703 *Hum Behav* (2021). <https://doi.org/10.1038/s41562-021-01119-3>.
- 704 54. G. R. B. Saunders, *et al.*, Genetic diversity fuels gene discovery for tobacco and alcohol
- 705 use. *Nature* **612**, 720–724 (2022).
- 706 55. E. Ware, "HRS Documentation Report HRS Polygenic Scores-Release 5 2006-2012 Genetic
- 707 Data Sensitive Health Data" (2024).
- 708 56. B. Bulik-Sullivan, *et al.*, LD score regression distinguishes confounding from polygenicity
- 709 in genome-wide association studies. *Nat Genet* **47**, 291–295 (2015).
- 710 57. J. K. Pritchard, M. Przeworski, "Linkage Disequilibrium in Humans: Models and Data"
- 711 (2001).
- 712 58. J. Yang, *et al.*, Genomic inflation factors under polygenic inheritance. *European Journal of*
- 713 *Human Genetics* **19**, 807–812 (2011).
- 714 59. A. R. Martin, *et al.*, Human Demographic History Impacts Genetic Risk Prediction across
- 715 Diverse Populations. *Am J Hum Genet* **100**, 635–649 (2017).
- 716 60. S. Hu, *et al.*, Fine-scale population structure and widespread conservation of genetic
- 717 effect sizes between human groups across traits. *Nat Genet* **57**, 379–389 (2025).
- 718 61. J. H. Sul, L. S. Martin, E. Eskin, Population structure in genetic studies: Confounding
- 719 factors and mixed models. *PLoS Genet* [Preprint] (2018).
- 720 62. N. Patterson, A. L. Price, D. Reich, Population structure and eigenanalysis. *PLoS Genet* **2**,
- 721 2074–2093 (2006).

63. A. L. Price, *et al.*, Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–909 (2006).
64. Orzechowski and Walker, The Tax Burden on Tobacco, 1970-2014. **49** (2014).
65. P. DeCicca, D. S. Kenkel, M. F. Lovenheim, The Economics of Tobacco Regulation: A Comprehensive Review. *SSRN* (2020). <https://doi.org/10.3386/w26923>.
66. ANRF, “U.S. Tobacco Control Laws Database (1998 - 2020). Americans for Non-Smokers’ Rights Foundation (ANRF): Minimum Ages and Smoke Free Air Laws” (2021).
67. M. F. Lovenheim, How Far to the Border?: The Extent and Impact of Cross-Border Casual Cigarette Smuggling. *Natl Tax J* **61**, 7–33 (2008).
68. W. N. Evans, J. S. Ringel, Can higher cigarette taxes improve birth outcomes? *J Public Econ* **72**, 135–154 (1999).

Table 1. Effect of adolescent and adult cigarette taxes on current smoking at older ages by smoking initiation polygenic index (PGI)

	(1)	(2)	(3)	(4)	(5)	(6)
Adolescent Tax	-0.068*** (0.021)	-0.063*** (0.021)	-0.066*** (0.021)	-0.068*** (0.021)	-0.068*** (0.021)	-0.059*** (0.021)
Adult Tax		-0.024*** (0.008)	-0.025*** (0.007)	-0.025*** (0.008)	-0.025*** (0.008)	-0.026*** (0.008)
Smoking Initiation PGI			0.038*** (0.003)	0.067*** (0.014)	0.069*** (0.015)	0.066*** (0.016)
Adolescent Tax x PGI				-0.030** (0.013)	-0.029** (0.014)	-0.027* (0.015)
Adult Tax x PGI					-0.001 (0.004)	-0.001 (0.005)
Min. legal purchase age	No	No	No	No	No	Yes
State smoke free air laws	No	No	No	No	No	Yes
Travel distance to nearest lower tax state border (level, squared)	No	No	No	No	No	Yes
Mean Adolescent Tax						0.96
Mean Adult Tax						1.52
Mean of dependent variable						0.137
Adolescent Tax Elasticity						-0.413
Adult Tax Elasticity						-0.288
Individuals (N)	11,528	11,528	11,528	11,528	11,528	11,528
Person-wave observations	110,624	110,624	110,624	110,624	110,624	110,624
R-squared	0.095	0.095	0.107	0.108	0.108	0.109

Notes: This table presents coefficients from linear probability models that pool observations across all HRS survey waves, based on the specification in equation (4). The dependent variable is an indicator for whether a respondent is a current smoker. Adolescent Tax is the average state cigarette excise tax experienced during adolescence, and Adult Tax is the average state tax that respondents were exposed to after age 50 (i.e., while in the HRS). PGI is the standardized polygenic index for smoking initiation. All models control for sex (male as the omitted category), a quadratic for age, a quadratic for years of education, marital status indicators (with married or partnered as the omitted category), and indicators for annual household quintile (with the first quintile as the omitted category). All models include fixed effects for birth year, adolescent state of residence, current state of residence, survey year, a current-state-specific linear time trend, and the first 10 PCs of the genotype data. State-level controls include smoke-free air laws (average restrictiveness score for private and government worksites), minimum legal purchase age, and the state's unemployment rate. Robust standard errors, shown in parentheses, are two-way clustered at the individual and adolescent state-of-residence levels. The tax elasticities in Model 6 are calculated at the sample mean for an individual with average genetic risk, or a PGI=0, as follows: Adolescent Tax Elasticity = $(-0.059 + (-0.027 * 0)) * (0.96 / 0.137) \approx -0.413$; Adult Tax Elasticity = $(-0.026 + (-0.001 * 0)) * (1.52 / 0.137) \approx -0.288$. *** p<0.01, ** p<0.05, * p<0.1.

780 **Table 2. Effects of adolescent and adult cigarette taxes on smoking cessation by smoking cessation polygenic index (PGI)**
781

	(1)	(2)	(3)	(4)	(5)	(6)
Adolescent Tax	0.089** (0.043)	0.082* (0.043)	0.078* (0.043)	0.080* (0.043)	0.080* (0.043)	0.065 (0.041)
Adult Tax		0.025** (0.012)	0.025** (0.012)	0.025** (0.012)	0.025** (0.012)	0.027** (0.014)
Smoking Cessation PGI			-0.029*** (0.006)	-0.065*** (0.020)	-0.058*** (0.021)	-0.063*** (0.021)
Adolescent Tax x PGI				0.037 (0.023)	0.040 (0.024)	0.042* (0.024)
Adult Tax x PGI					-0.006 (0.007)	-0.006 (0.007)
Min. legal purchase age	No	No	No	No	No	Yes
State smoke free air laws	No	No	No	No	No	Yes
Travel distance to nearest lower tax state border (level, squared)	No	No	No	No	No	Yes
Mean Adolescent Tax						0.96
Mean Adult Tax						1.52
Mean of dependent variable						0.137
Adolescent Tax Elasticity						-0.413
Adult Tax Elasticity						-0.288
Individuals (N)	6,587	6,587	6,587	6,587	6,587	6,587
Person-wave observations	62,668	62,668	62,668	62,668	62,668	62,668
R-squared	0.141	0.141	0.144	0.144	0.144	0.146

782
783 *Notes:* This table presents coefficients from linear probability models that pool observations across all waves, based on equation (4). The dependent variable is an
784 indicator for whether a respondent is a former smoker. Adolescent Tax is the average state cigarette excise tax experienced during adolescence, and Adult Tax is
785 the average state tax experienced during older adulthood. PGI is the standardized polygenic index for smoking cessation. All models control for sex, a quadratic in
786 age, a quadratic in years of education, marital status indicators, and annual household income quintiles. All models also include fixed effects for birth year,
787 adolescent state of residence, current state of residence, survey year, a current-state-specific linear time trend, and the first 10 PCs of the genotype data. State-level
788 controls include smoke-free air laws, minimum legal purchase age, and the unemployment rate. Robust standard errors, shown in parentheses, are two-way clustered
789 at the individual and adolescent state-of-residence levels. The tax elasticities in Model 6 are calculated at the sample mean as follows: Adolescent Tax Elasticity =
790 $(0.065 + (0.042 * 0)) * (0.95 / 0.759) \approx 0.081$; Adult Tax Elasticity = $(0.027 + (-0.006 * 0)) * (1.51 / 0.759) \approx 0.055$. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
791
792
793
794

Table 3. Effects of adolescent and adult cigarette taxes on cigarettes smoked per day (CPD) by CPD polygenic index (PGI)

	(1)	(2)	(3)	(4)	(5)	(6)
ln(Adolescent Tax)	0.083 (0.134)	0.063 (0.134)	0.047 (0.126)	0.049 (0.127)	0.049 (0.127)	0.044 (0.114)
ln(Adult Tax)		-0.244*** (0.052)	-0.238*** (0.053)	-0.238*** (0.053)	-0.240*** (0.052)	-0.260*** (0.055)
CPD PGI			0.092*** (0.010)	0.065*** (0.010)	0.060*** (0.014)	0.060*** (0.015)
Adolescent Tax x PGI				-0.078 (0.077)	-0.085 (0.073)	-0.085 (0.074)
Adult Tax x PGI					0.016 (0.027)	0.009 (0.029)
Min. legal purchase age	No	No	No	No	No	Yes
State smoke free air laws	No	No	No	No	No	Yes
Travel distance to nearest lower tax state border (level, squared)	No	No	No	No	No	Yes
Mean Adolescent Tax						0.94
Mean Adult Tax						1.51
Mean of dependent variable						17.08
Adolescent Tax Elasticity						0.033
Adult Tax Elasticity						-0.259
Individuals (N)	2,495	2,495	2,495	2,495	2,495	2,495
Person-wave observations	13,323	13,323	13,323	13,323	13,323	13,323
R-squared	0.098	0.100	0.106	0.106	0.106	0.106

Notes: This table presents coefficients from Ordinary Least Squares (OLS) models that pool observations across all survey waves, based on the specification in equation (4). The dependent variable is the natural log of cigarettes smoked per day (CPD), and the analysis is conducted on the subsample of current-smokers who report smoking at least one cigarette per day. The key predictors, Adolescent Tax and Adult Tax, are also log-transformed. We use a log-log specification because CPD is a continuous outcome (unlike the binary participation/cessation margins) and is strongly right-skewed; logging CPD (and taxes) reduces the influence of high-intensity outliers, stabilizes variance (mitigating heteroskedasticity), and better captures proportional, multiplicative responses. It also simplifies interpretation, since coefficients correspond to elasticities. PGI is the standardized polygenic index for CPD. All models control for sex (male as the omitted category), a quadratic in age, a quadratic in years of education, marital status indicators, and annual household income quintiles (with the first quintile omitted). All models also include fixed effects for birth year, adolescent state of residence, current state of residence, survey year, a current-state-specific linear time trend, and the first 10 principal components of the genetic data. State-level controls include smoke-free air laws (average restrictiveness score for private and government worksites), minimum legal purchase age, and the unemployment rate. Robust standard errors, shown in parentheses, are two-way clustered at the individual and adolescent state-of-residence levels. In this log-log specification, the tax elasticities in Model 6 are calculated at the sample mean PGI as follows: Adolescent Tax Elasticity = $-0.023 + (0.066 * PGI_mean) \approx -0.023$; Adult Tax Elasticity = $-0.286 + (0.006 * PGI_mean) \approx -0.286$. (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table 4. Effects of being a current smoker on the likelihood of developing chronic diseases and on survival

	Lung Disease		Heart Problem		Any Cancer		Diabetic		Survival	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS	2SLS	OLS	2SLS	OLS	OLS	OLS	2SLS	OLS	2SLS
Current Smoker	0.101*** (0.006)	0.680*** (0.143)	0.005 (0.009)	1.226*** (0.255)	0.001 (0.008)	0.571*** (0.169)	-0.000 (0.007)	0.661*** (0.160)	-0.021*** (0.003)	-1.525*** (0.270)
F-stat (1 st Stage)	-	37.59	-	37.84	-	36.40	-	38.37	-	43.37
Per-wave observations	97,487	97,487	97,487	97,487	97,487	97,487	97,487	97,487	105,067	105,067
Individuals (N)	11,703	11,703	11,703	11,703	11,703	11,703	11,703	11,703	11,703	11,703

Notes: This table shows coefficients of naïve OLS (models 1, 3, 5, 7, and 9) and 2SLS (models 2, 4, 6, 8, and 10) regression models based on the specifications highlighted in equations (6-9). The dependent variables are indicators for whether respondents report ever being diagnosed with lung disease (models 1 and 2), heart problems (models 3 and 4), any cancer (models 5 and 6), diabetes (models 7 and 8), and survival at the time of survey (models 9 and 10). Models 1-8 were run on the HRS dataset that pools observations across all survey waves (1992-2016), conditional on survival at a given survey wave. The key predictor is an indicator variable for whether a respondent is a "Current smoker". In the 2SLS models, smoking behavior or "Current Smoker" is instrumented by the average state cigarette excise tax that respondents were exposed during adolescence and while in the HRS. All regression models control for gender, BMI (level, squared), education (level, squared), marital status, household income, state, birth year, and survey-year fixed effects. Robust standard errors clustered at the individual and state of residence levels are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

840 **Table 5. Effects of cigarette taxes on current smoking by smoking initiation PGI (adjusted for survival selection)**

Adolescent Tax	-0.048** (0.022)
Adult Tax	-0.046** (0.014)
Smoking Initiation PGI	0.076*** (0.018)
Adolescent Tax x PGI	-0.029* (0.017)
Adult Tax x PGI	-0.008 (0.006)
Per-wave observations	96,996
Individuals (N)	9,719
R-squared	0.118

841
842 *Notes:* This table presents coefficients from a linear probability model for current smoking status. The model is identical to the main specification (Table 1, Model
843 6) but is weighted using inverse probability weights (IPWs) to correct for mortality selection based on survival to 2006. The dependent variable is an indicator for
844 whether a respondent is a current smoker. Adolescent Tax is the average state cigarette excise tax experienced during adolescence, and Adult Tax is the average
845 state tax experienced during older adulthood. PGI is the standardized polygenic index for smoking initiation. The model includes the full set of individual, state-
846 level, and genetic controls, as well as all fixed effects detailed in the notes for Table 1. Robust standard errors, shown in parentheses, are two-way clustered at the
847 individual and adolescent state-of-residence levels. *** p<0.01, ** p<0.05, * p<0.1.
848

849

850

851

852

853

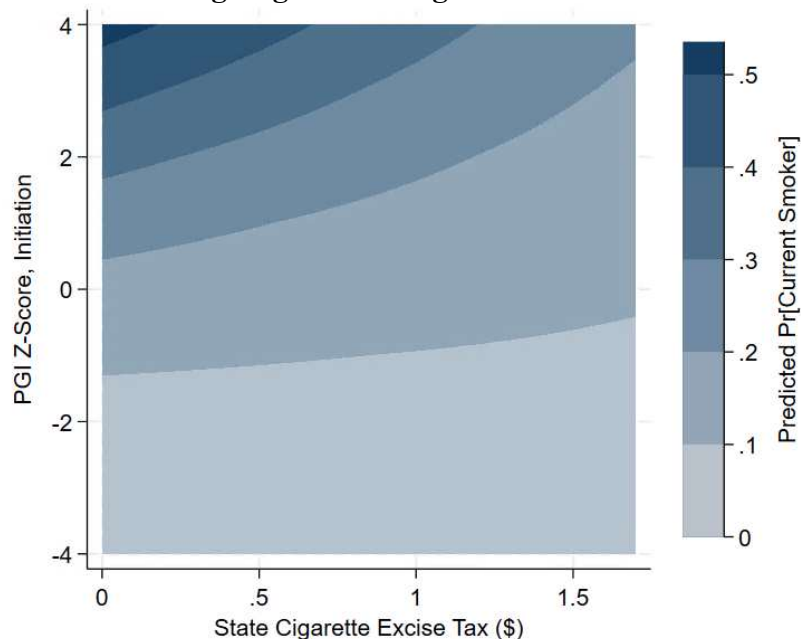
854

855

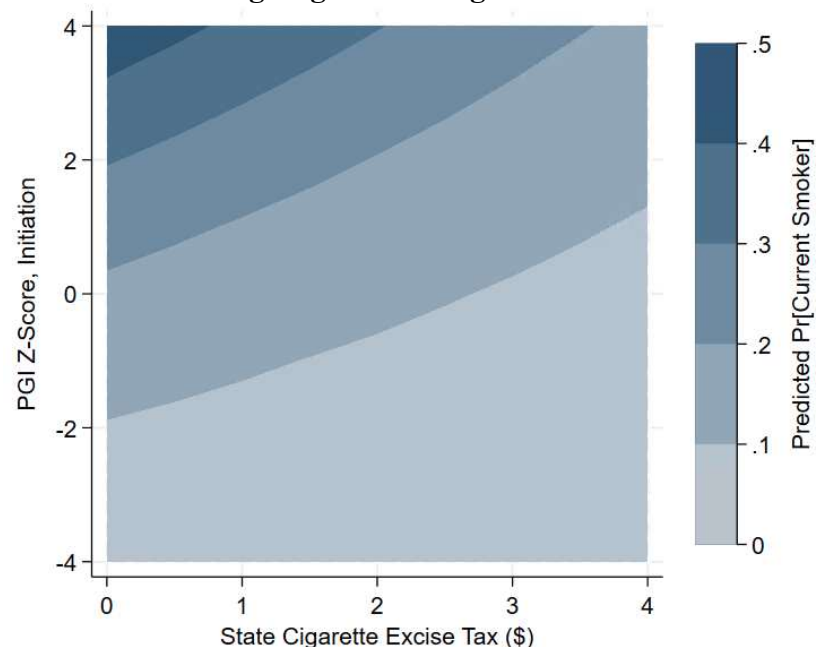
856

Figure 1. Predicted probability of becoming a current smoker by PGI and life course timing of cigarette excise taxes

Panel A: Average cig. tax during adolescence



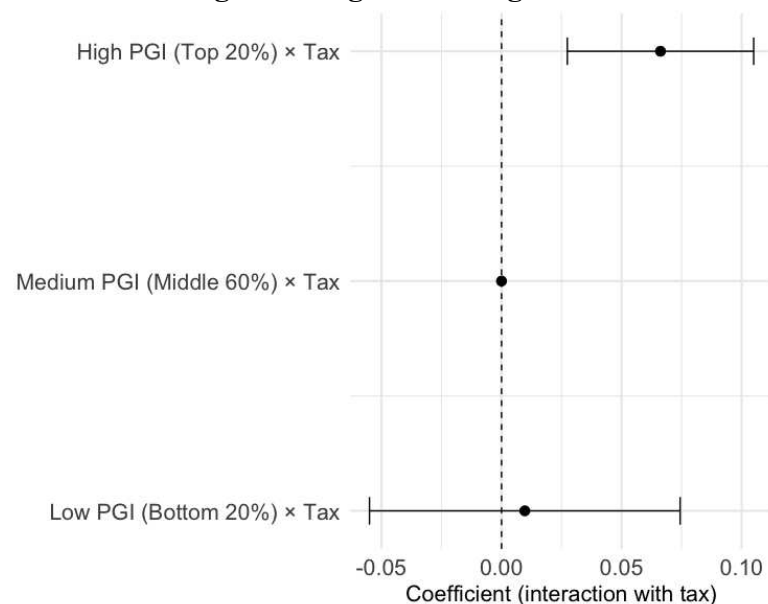
Panel B: Average cig. tax during older adulthood



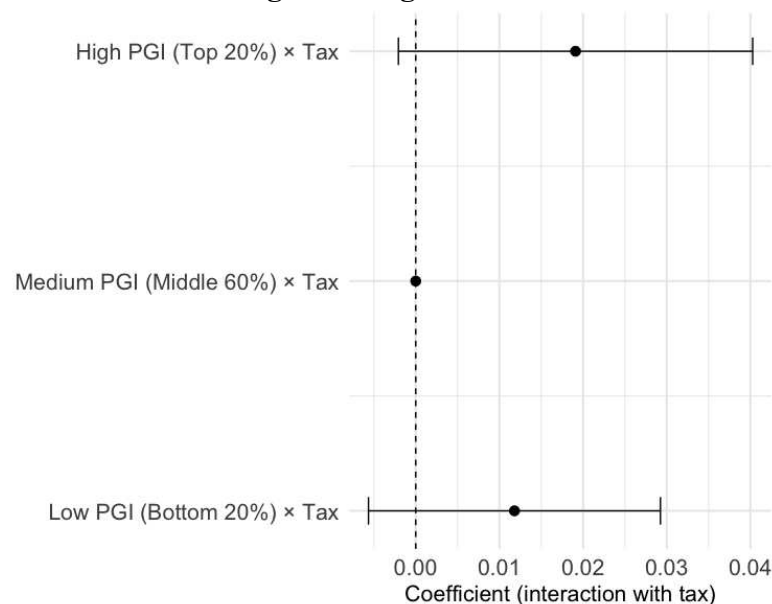
Notes: These contour plots show the predicted probability of becoming a current smoker for respondents who experienced different levels of average cigarette excise tax during their adolescence (panel A) and while in the HRS (panel B) by genetic predisposition to smoking initiation (PGI Z-score). The 3rd axis (z-axis) of the contour plot depicts the marginal predicted probabilities, which were estimated using a logistic regression model run on the HRS dataset that pools observations across all survey waves as per equation (4). Contours of similar color represent identical predicted probabilities of being a lifetime smoker.

Figure 2. Effects of cigarette taxes on the likelihood of being a lifetime smoker by genetic risk profile category

Panel A: Average state cig. tax during adolescence



Panel B: Average state cig. tax while in the HRS



Notes: Panels A and B present estimates from regression models similar to those in equation (4), except the continuous smoking initiation polygenic index (PGI) is replaced with a categorical version. The PGI is divided into three mutually exclusive groups: Low PGI (bottom 20% of the distribution), Medium PGI (middle 60%), and High PGI (top 20%). The Medium PGI group serves as the reference category and is shown as a point at zero without a confidence interval. Black dots represent the estimated interaction coefficients between cigarette excise taxes and the Low or High PGI categories relative to the Medium group, with horizontal lines indicating 95% confidence intervals. Panel A uses the average state cigarette excise tax respondents faced during their teenage years (“adolescent taxes”), and Panel B uses the annual state cigarette excise tax at the time respondents were observed at older ages. The dependent variable in both panels is an indicator for whether respondents are lifetime smokers.

Supplementary Information

A. Correcting for mortality selection

To address potential bias from mortality selection, we employ an inverse probability weighting (IPW) approach. This method re-weights the estimation sample to make it more representative of the original full cohort, including those who did not survive until the genotyping wave. The weights are constructed based on the predicted probability of survival until 2006, the first year of genotyping in the HRS. We model this survival probability using the following probit model:

$$P(Alive_{isb} = 1) = \Phi(\alpha + \delta X_i + \lambda H_i + \gamma_b) \quad (8)$$

where the dependent variable is an indicator for whether respondent i was alive in year 2006; $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, which ensures the predicted probability lies between 0 and 1. The model includes a vector of baseline individual-level covariates, X_i , and a vector of baseline health indicators, H_i (a quadratic in body mass index, and indicators for diabetes, heart problems, lung disease, cancer, stroke, self-reported health, and depression), along with birth year fixed effects, γ_b .

After estimating this model for the full initial sample, the inverse probability weight for each individual in our final estimation sample is calculated as the inverse of their predicted probability of survival. This procedure up-weights individuals who were observationally similar to those who did not survive, thereby correcting for potential selection bias.

B. Estimating travel distance as a proxy for tax avoidance behavior

To construct a proxy for tax avoidance behavior, we calculated the driving distance from each respondent's county of residence to the nearest state with a lower cigarette excise tax for each survey wave. This was accomplished using the Network Analyst extension in ArcGIS. We first obtained U.S. county and state boundary shapefiles and a national road network dataset from the U.S. Census Bureau. For each respondent in each survey wave, we identified their county of residence using restricted geographic data from the HRS and calculated its geographic centroid to serve as the origin point. For the destinations, we identified all neighboring states with a lower cigarette excise tax in that specific year and generated a series of points along their borders. Using the "Closest Facility" analysis tool, we then calculated the shortest driving route distance in miles along the road network from each origin (county centroid) to the nearest destination (point on a lower-tax state border). This wave-specific driving distance measure was then merged with our main dataset and included as a control variable in our regression models.

Table S1: Summary statistics of HRS study participants (1992-2016)

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Time invariant	Full Sample Mean (S.D.)	Current Smokers Mean (S.D.)	Never Smokers Mean (S.D.)	Former Smokers Mean (S.D.)	Difference (Current -Never) Mean (S.E.)	Difference (Former -Current) Mean (S.E.)
Female (%)	57.99	58.74	67.66	48.17	-0.090*** (0.005)	-0.105*** (0.005)
PGI Z-Score (Initiation)	-0.004(1.003)	0.363(0.985)	-0.246(0.979)	0.120(0.973)	0.140*** (0.010)	-0.243*** (0.009)
PGI Z-Score (CPD)	0.006 (1.00)	0.12 (1.00)	-0.02 (0.99)	-0.005 (1.01)	0.140*** (0.010)	-0.125*** (0.010)
PGI Z-Score (Cessation)	0.006 (1.00)	-0.12 (0.96)	-0.002 (0.98)	0.06 (1.03)	-0.119*** (0.009)	0.176*** (0.009)
Panel B: Time variant						
Age (years)	66.96 (11.01)	61.45 (9.19)	67.51 (11.39)	68.21 (10.62)	-6.063*** (0.094)	6.760*** (0.092)
Education (years)	13.19 (2.55)	12.41 (2.39)	13.43 (2.54)	13.20 (2.67)	-1.017*** (0.023)	0.785*** (0.023)
Income (log)	10.81(1.14)	10.57(1.34)	10.87(1.10)	10.84(1.09)	-0.307*** (0.013)	0.270*** (0.013)
Married (%)	64.49	59.94	64.44	66.21	-0.045*** (0.005)	0.064*** (0.005)
Ever smoked (%)	57.02	100	-	100	-	-
Currently smoking (%)	13.73	100	-	-	-	-
Former smoker (%)	43.29	-	-	100	-	-
Cigarettes per day	3.53 (7.95)	17.08 (8.40)	-	15.47 (9.97)	-	-
Minimum age for cig. sale (years)	18.05 (0.30)	18.05 (0.28)	18.05 (0.31)	18.05 (0.29)	-0.006* (0.003)	0.001(0.003)
Smoke-free air law (%)	49.75	42.15	50.63	51.35	-0.085*** (0.005)	0.092*** (0.005)
Cigarette tax during adolescence	0.96 (0.17)	0.94 (0.16)	0.97 (0.18)	0.96 (0.17)	-0.028*** (0.002)	0.021*** (0.002)
Cigarette tax while in HRS	1.52 (0.63)	0.86 (0.83)	0.97 (0.86)	0.99 (0.88)	-0.111*** (0.015)	0.128*** (0.008)
Travel distance to nearest lower tax state border (miles)	48.93 (79.43)	48.83 (77.18)	46.20 (75.57)	51.63 (83.59)	2.631*** (0.736)	2.804*** (0.755)
Individuals (N)	11,558	2,515	5,196	4,226	7,711	
Person-wave observations	104,651	14,347	44,910	42,654	59,257	

Notes: *** p<0.01, ** p<0.05, * p<0.1 with standard errors in parentheses. The coefficients in column 6 are from standard two-sample t-tests run assuming unequal variance of the covariates between current and never-smoker groups. The two-sample t-tests were run on the pooled sample, so the standard errors are adjusted for autocorrelation at the respondent level. All the coefficients in columns 2-5 represent the means and standard deviation of the variables. The PGI Z-scores are computed by subtracting the mean PGI from the raw PGI scores and dividing by the standard deviation.

Table S2. Incremental R-squared of each covariate predicting smoking behavior outcomes

	Smoking Behavior Outcomes			
	Ever smoker	Current smoker	Former smoker	CPD
Smoking initiation PGI	3.979	2.014	-	-
Smoking cessation PGI	-	-	0.556	-
CPD PGI	-	-	-	1.840
Sex	2.871	0.002	0.790	0.100
Age (level, squared)	0.280	3.934	6.710	3.290
BMI (kg/meters ²)	0.050	1.430	2.400	0.920
Education (years)	1.300	2.480	2.400	2.980
Marital status	0.460	0.960	1.000	0.660
Income (quintiles)	0.120	0.840	0.900	1.100
Cigarette excise taxes	0.070	0.000	0.000	0.500
Other predictors	0.050	0.000	0.100	0.000
Total R-Squared	9.130	11.660	14.756	11.390

Notes: This table shows the incremental R-squared of each covariate listed in the first column (expressed as a percent by multiplying by 100). The incremental R-squared of the PGIs is calculated by estimating the coefficient of determination (R^2) of the PGIs, after residualizing the effects of the first 20 principal components of the genetic relationship matrix. For the other covariates, the incremental R-squared is computed by taking the difference of R^2 between a base model that regresses the state-fixed effects on the smoking behavior outcome variables and subsequent models that sequentially add each predictor. Other predictors include a minimum age for sale, smoke-free air laws, and travel distance from the county of residence to the lower cigarette tax state border (level, squared, and cubed). The incremental R-squared values represent the share (%) of total variation each predictor explains.

Table S3. Effects of cigarette taxes on current smoking by genetic risk profile, controlling for cigarette tax by PCs interactions

Adolescent Tax	-0.005** (0.024)
Adult Tax	-0.024*** (0.008)
Smoking Initiation PGI	0.064*** (0.016)
Adolescent Tax x PGI	-0.026* (0.015)
Adult Tax x PGI	-0.001 (0.004)
Person-wave observations	96,996
Individuals (N)	9,719
R-squared	0.110

Notes: This table presents coefficients from a linear probability model for current smoking status. The model is identical to the main specification (Table 1, Model 6) but includes interaction terms between each of the top 10 genetic principal components (PCs) and the adolescent and adult cigarette tax variables. These interactions help adjust for potential confounding due to ancestry-related structure that may correlate with environmental exposures, following Keller (2014). The dependent variable is an indicator for whether a respondent is a current smoker. Adolescent Tax is the average state cigarette excise tax experienced during adolescence, and Adult Tax is the average state tax experienced during older adulthood. PGI is the standardized polygenic index for smoking initiation. The model includes the full set of individual, state-level, and genetic controls, as well as all fixed effects detailed in the notes for Table 1. Robust standard errors, shown in parentheses, are two-way clustered at the individual and adolescent state-of-residence levels. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table S4. Effects of smoking cessation on the likelihood of developing common chronic diseases and survival

	Lung Disease		Heart Problem		Any Cancer		Diabetic		Survival	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS	2SLS	OLS	2SLS	OLS	OLS	OLS	2SLS	OLS	2SLS
Former Smoker	-0.074*** (0.007)	-0.984*** (0.240)	0.011 (0.009)	-1.222*** (0.360)	0.003 (0.010)	-0.396* (0.206)	0.003 (0.007)	-0.686*** (0.181)	0.018*** (0.003)	1.623*** (0.350)
F-stat (1 st Stage)	-	21.91	-	22.56	-	21.85	-	22.53	-	25.96
Person-wave observations	55,348	55,348	55,550	55,550	55,503	55,503	55,449	55,449	60,207	60,20
Individuals (N)	6,659	6,659	6,659	6,659	6,659	6,659	6,659	6,659	6,659	6,659

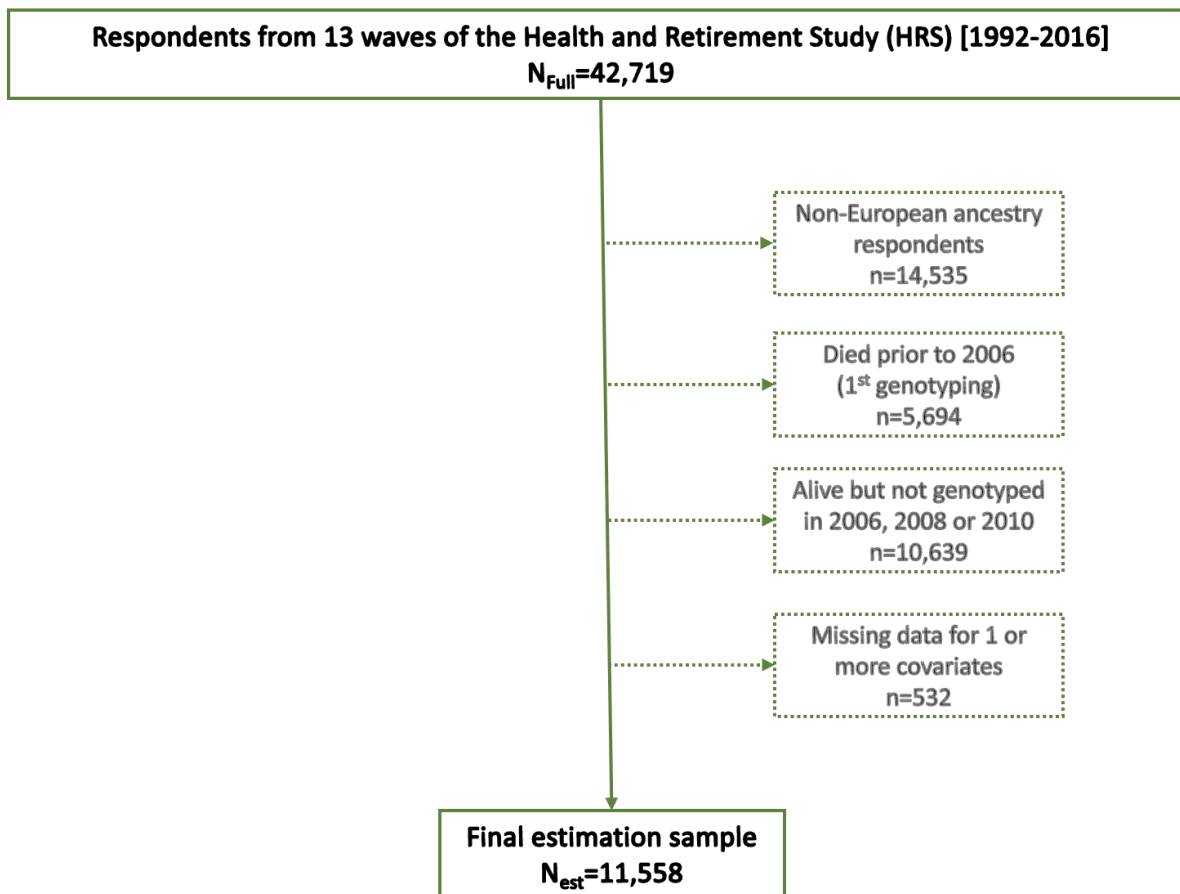
Notes: This table shows coefficients of naïve OLS (models 1, 3, 5, 7, and 9) and 2SLS (models 2, 4, 6, 8, and 10) regression models based on the specifications highlighted in equations (6-9). The dependent variables are indicators for whether respondents report ever being diagnosed with lung disease (models 1 and 2), heart problem (models 3 and 4), any cancer (models 5 and 6), diabetes (models 7 and 8), and survival at the time of survey (models 9 and 10). Models 1-8 were run on the HRS dataset that pools observations across all survey waves (1992-2016), conditional on survival at a given survey wave. The key predictor is an indicator variable for whether a respondent is a "Former Smoker". In the 2SLS models, smoking behavior or "Current Smoker" is instrumented by the average state cigarette excise tax that respondents were exposed during adolescence and while in the HRS. All regression models control for gender, BMI (level, squared), education (level, squared), marital status, household income, state, birth year, and survey-year fixed effects. Robust standard errors clustered at the individual and state of residence levels are shown in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Table S5. Effects of smoking intensity (CPD) on the likelihood of developing common chronic diseases and survival

	Lung Disease		Heart Problem		Any Cancer		Diabetic		Survival	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS	2SLS	OLS	2SLS	OLS	OLS	OLS	2SLS	OLS	2SLS
CPD	0.005*** 0.001	0.032*** (0.009)	0.001 (0.001)	0.022*** (0.006)	0.001 (0.001)	0.018*** (0.006)	0.001* (0.001)	0.014*** (0.005)	-0.001*** (0.000)	-0.035*** (0.008)
F-stat (1 st Stage)	-	16.10	-	15.93	-	16.73	-	16.14	-	21.43
Person-wave observations	20,475	20,475	20,615	20,615	20,569	20,569	20,559	20,559	20,564	20,564
Individuals (N)	2,563	2,634	2,634	2,634	2,634	2,634	2,634	2,634	2,634	2,634

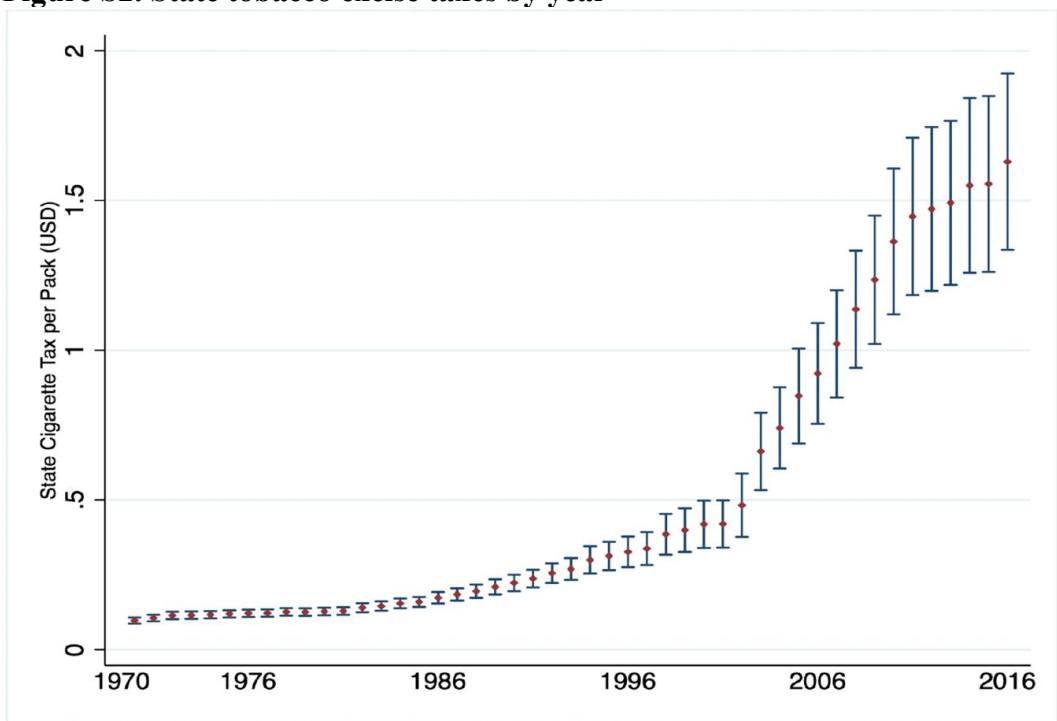
Notes: This table shows coefficients of naïve OLS (models 1, 3, 5, 7, and 9) and 2SLS (models 2, 4, 6, 8, and 10) regression models based on the specifications highlighted in equations (6-9). The dependent variables are indicators for whether respondents report ever being diagnosed with lung disease (models 1 and 2), heart problem (models 3 and 4), any cancer (models 5 and 6), diabetes (models 7 and 8), and survival at the time of survey (models 9 and 10). All models were run on the HRS dataset that pools observations across all survey waves (1992-2016), conditional on being a current smoker, while models 1-8 were further restricted to those who were alive in the specific wave under consideration. The key predictor is "CPD" representing the average number of cigarettes respondents (current smokers) reported smoking daily. In the 2SLS models, smoking behavior or "cigarettes smoked per day (CPD)" is instrumented by the average state cigarette excise tax that respondents were exposed to while in the HRS. All regression models control for gender, BMI (level, squared), education (level, squared), marital status, household income, state, birth year, and survey-year fixed effects. Robust standard errors clustered at the individual and state of residence levels are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

1 **Figure S1. Schematic showing derivation of the estimation sample**



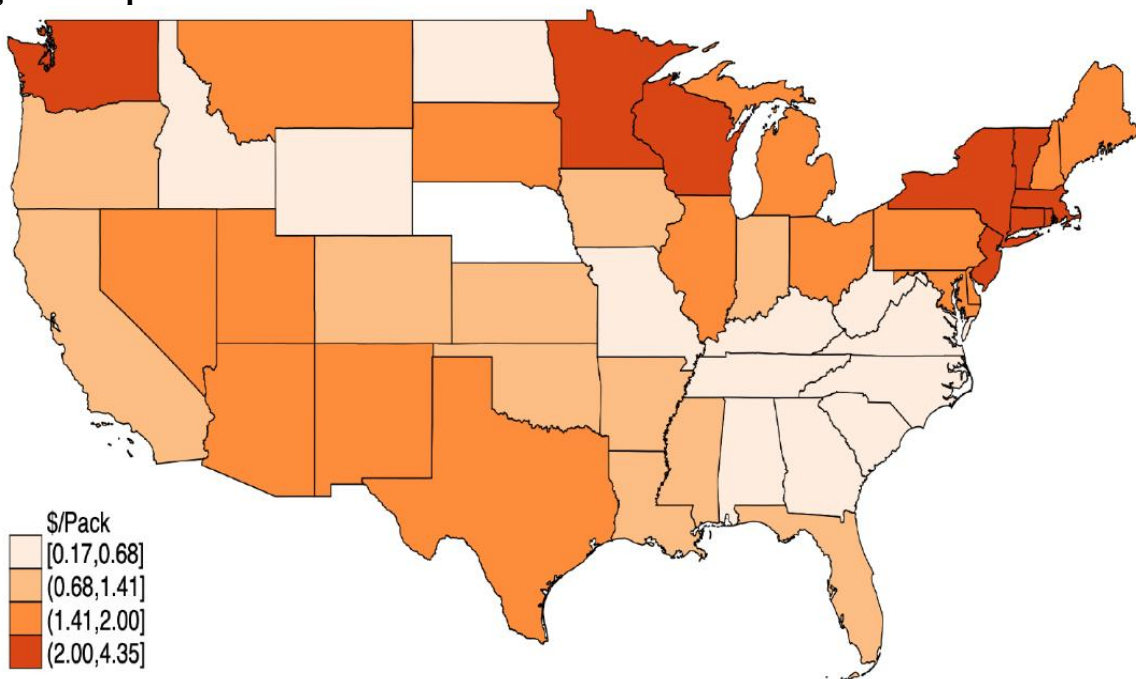
Notes: This schematic shows the steps followed to reach the estimation sample. The dashed rectangular boxes show the number of individuals excluded from our estimation sample. All non-European ancestry respondents were first excluded as PGIs developed using predominantly European-ancestry samples do not adequately predict phenotypes in non-European samples. The sample is further restricted by removing European-ancestry respondents who were deceased before the first genotyping (2006) and those who were alive in 2006 but were not genotyped. Finally, respondents with missing values for one or more regression model covariates were excluded, with most exclusions driven by missing adolescent state-of-residence information needed to construct cigarette tax exposure histories, yielding a final estimation sample of 11,558.

Figure S2. State tobacco excise taxes by year



Note: Red dots represent average tobacco excise taxes for each year and arrows show the respective 95 percent confidence interval. **Source:** Orzechowski & Walker (2016). Tax burden on tobacco: The historical compilation, vol. 51.

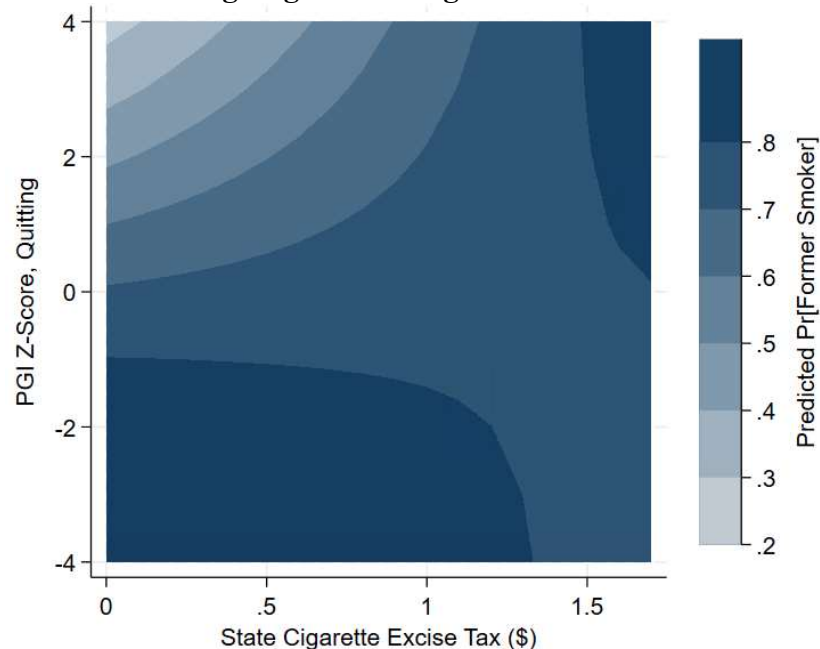
Figure S3. Spatial variation in state tobacco excise tax across states in 2016



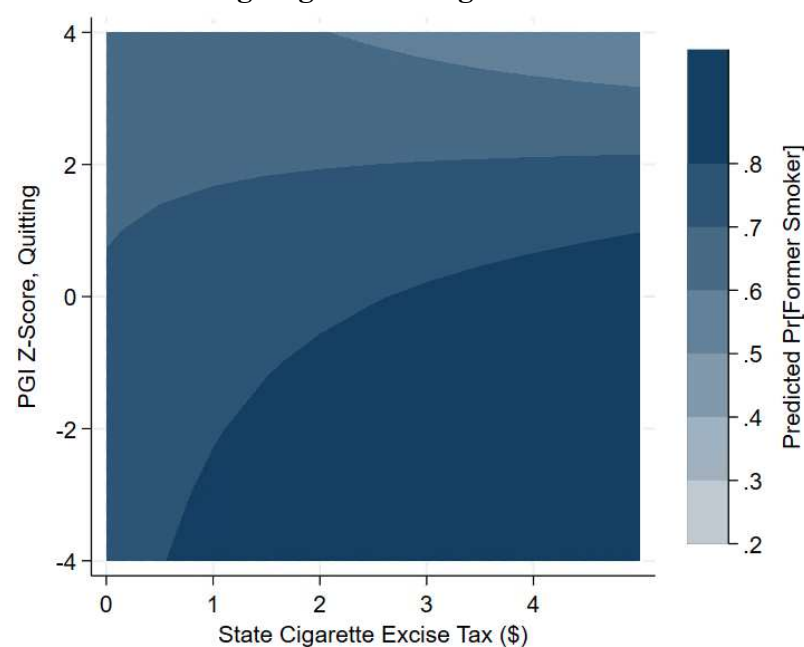
Source: Orzechowski & Walker (2016). Tax burden on tobacco: The historical compilation, vol. 51.

Figure S4. Predicted probability of becoming a former smoker by PGI and life course timing of cigarette excise taxes

Panel A: Average cig. tax during adolescence



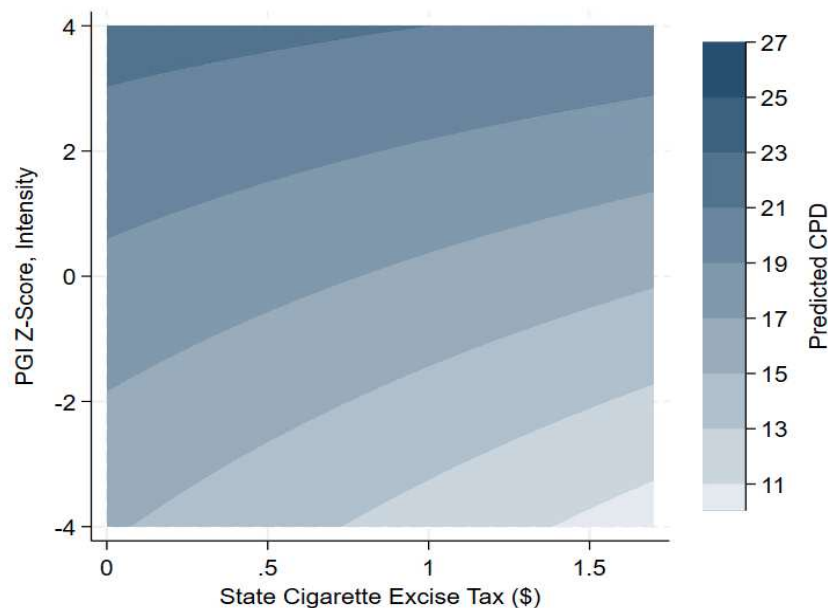
Panel B: Average cig. tax during older-adulthood



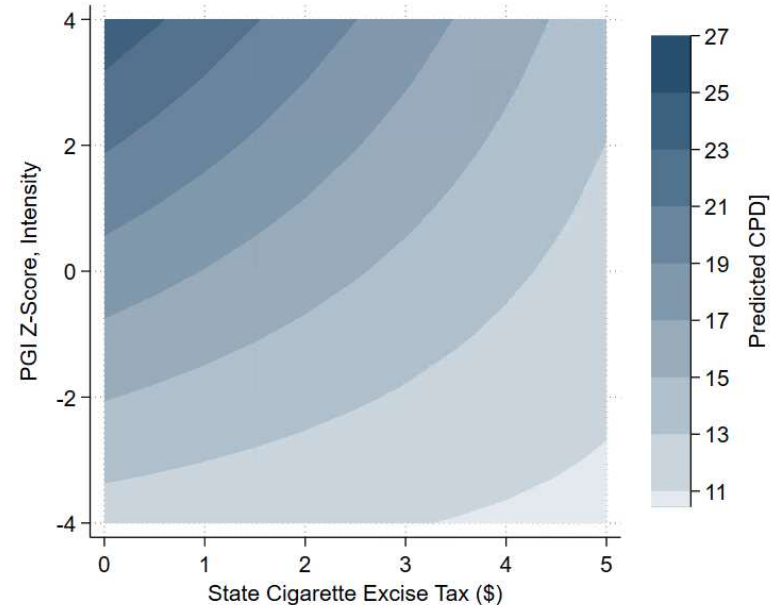
Notes: These contour plots show respondents' likelihood of quitting smoking by average cigarette excise tax during adolescence (panel A) and while in the HRS (panel B) by smoking intensity PGI for ever-smokers. The 3rd axis (z-axis) of the contour plot depicts the marginal predicted probabilities, which were estimated using a logit regression model run on the HRS dataset that pools observations across all survey waves as per equation (4). Contours of similar color represent identical predicted probabilities of being a lifetime smoker.

Figure S5. Predicted number of cigarettes smoked per day by PGI and life course timing of cigarette excise taxes

Panel A: Average cig. tax during adolescence



Panel B: Average cig. tax during older-adulthood



Notes: These contour plots show the predicted number of cigarettes smoked per day (CPD) for respondents who experienced different levels of average cigarette excise tax during their adolescence (panel A) and while in the HRS (panel B) by the genetic predisposition to smoking intensity (PGI Z-score), conditional on being a current smoker. The 3rd axis (z-axis) of the contour plot depicts the marginal predicted CPDs, and were estimated using an OLS regression model run on the HRS dataset that pools observations across all survey waves as per equation (4). Contours of similar color represent identical predicted probabilities of being a lifetime smoker.