Moral Hazard Heterogeneity: Genes and Health Insurance Influence Smoking after a Health Shock

Pietro Biroli * Laura Zwyssig †

December 22, 2021‡

Abstract

Decision-making is influenced both by biological and environmental factors. We show how the interplay between these biological and environmental constraints shapes smoking decisions after a cardio-vascular health shock. Individuals who suffer a health shock when uninsured are 25.6 percentage points more likely to reduce smoking, but this is true only for those who have a low index of genetic predisposition to smoking. This differential elasticity of response depending on genetic variants is evidence of individual-level heterogeneity in moral hazard. These results suggest that genetic heterogeneity should be considered when evaluating the effectiveness and fairness of health policies.

Keywords: Moral Hazard, Genetics, Smoking, Medicare
JEL CODES: I12, I13, D63, D91

*University of Zurich and UBS Center. Corresponding author. Email: pietro.biroli@uzh.ch
†University of Zurich. Email: laura.zwyssig@uzh.ch

‡We thank Vittorio Bassi, Gregory Crawford, Thomas Dohmen, Ernst Fehr, Titus Galama, Stephanie von Hinke, Alessandro Iaria, Hans van Kippersluis, Philipp Koellinger, Richard Karlsson Linner, Nicholas Papageorge, Kevin Thom, David Yanagizawa-Drott, Christian Zünd, all of the members of GEIGHEI team, and the various seminar participants for useful discussions, comments, and suggestions. Regina Seibel, Pia Arce, Jeremy Vollen, and Alex Thannhuber are the brave and fierce programmers whose research assistantship brought this project to a new level. All remaining errors are probably Pietro's fault. All the code necessary to replicate the results can be found on http://github.com/geighei/GxInsurancePublic We gratefully acknowledge financial support from NORFACE DIAL (Grant Reference 462-16-100) and the National Institute On Aging of the National Institutes of Health (R56AG058726).

1 Introduction

Moral hazard shapes behaviors in all domains characterized by asymmetric information (Arrow, 1963; Finkelstein, 2014). Propensity to moral hazard can differ across individuals, the same as preferences and beliefs. Unlike preferences and beliefs, heterogeneity in moral hazard has been proposed but never directly measured (Dubois and Vukina, 2009; Einav et al., 2013; Kowalski, 2018). Leveraging information on genetic variants associated with smoking behavior, we provide a novel approach to measuring heterogeneity in moral hazard.

To measure heterogeneity in moral hazard, we estimate how individuals carrying certain genetic variants differentially change their smoking behavior following a cardio-vascular health shock when they are not covered by health insurance. Combining the genetic and moral hazard factors shaping health behaviors, we contribute to the burgeoning literature studying gene-environment interactions (Barcellos, Carvalho and Turley, 2018; Belsky et al., 2018; Caspi et al., 2002; Fletcher, 2012; Papageorge and Thom, 2020; Schmitz and Conley, 2017*a*,*b*; Wedow et al., 2018). Specifically, this study is the first to show how genetic risk and insurance coverage jointly influence individual choices of health behavior. We show that moral hazard stemming from health insurance (financial risk) interacts with the genetic predisposition for smoking (genetic risk) in influencing the probability of smoking cessation following a health shock. Using data from the Health and Retirement Study (HRS), a longitudinal population-based study of elderly Americans, we estimated how experiencing a health shock between survey waves affected the smoking probability of individuals with different coverage levels and genetic predispositions for smoking.

To cleanly identify this interplay between genes and the environment, we leverage a key feature of the US health insurance system: Medicare, which provides public health insurance to all US citizens older than 65. Before the introduction of the Affordable Care Act (Obama, 2016), a significant fraction of the population younger than 65 was still uninsured (Barnett and Vornovitsky, 2016; Cohen et al., 2009). Exploiting the differential timing of health shocks before or after the Medicare-eligibility age of 65 for previously uninsured individuals, we estimate that Medicare eligibility reduced smoking cessation rates after a health shock by 25.6 percentage points, but *only* for those individuals with a low index of genetic propensity to smoking. Comparing this effect between individuals with a high versus a low index of genetic predisposition for smoking allows for an assessment of how the moral hazard problem found in previous research interacts with genetic risk. Our results suggests that genetic makeup can be a constraint and limit the response to financial incentives for behavior change.

We focus on smoking behavior because chronic diseases and health care costs caused by tobacco use are estimated to be one of the biggest health challenges in industrialized countries, and have rapidly increased in importance in the developing world (Goodchild, Nargis and Tursan d'Espaignet, 2018; United States Department of Health and Human Services, 2014). In the US, smoking is estimated to cause more than 400,000 premature deaths annually (Ma et al., 2018), and the economic costs of smoking-related illness amount to around $300 billion each year, including almost $170 billion for direct medical care and an additional $156 billion in lost productivity (United States Department of Health and Human Services, 2014; Xu et al., 2015). Both the health burden from smoking-related illness as well as the economic burden on an already strained health care system have made it a priority to understand what factors affect individuals' smoking decisions, and how health care can effectively encourage cessation.

Smoking cessation, like many other health behaviors, is influenced both by environmental and genetic factors. One environmental factor that has been associated with a reduction in tobacco consumption is health insurance coverage (Dave and Kaestner, 2009; Marti and Richards, 2017; Richards and Marti, 2014), especially after experiencing a severe smoking-related health shock, like the onset of a cardiovascular illness (Clark and Etilé, 2002; Falba, 2005; Keenan, 2009; Khwaja, Sloan and Chung, 2006*a*; Sundmacher, 2012; Wray et al., 1998). A driving mechanisms for this reduction in smoking might be financial costs associated with health shocks (Marti and Richards, 2017; Richards and Marti, 2014), and not only health considerations (Clark and Etilé, 2002; Khwaja, Sloan and Chung, 2006*a*,*b*; Smith et al., 2001). Alleviating the financial burden of health care costs, health insurance can have the unintended side effect of preventing beneficial behavior changes that would have taken place if the individuals were fully responsible for the financial consequences of poor health (Marti and Richards, 2017). This adverse incentive created by health insurance is a typical example of "moral hazard": the notion that individuals change their behavior in an undesired way because the consequences of their actions are not (fully) borne by themselves (Einav and Finkelstein, 2018; Zweifel and Manning, 2000).

Another factor that is tightly linked to smoking behavior is genetic makeup. Genetic factors can explain around 30% to 85% of the variance in regular smoking, according to several studies comparing identical and fraternal twins (Boardman, Blalock and Pampel, 2010; Hall, Madden and Lynskey, 2002; Heath et al., 1993; Li et al., 2003; Sullivan and Kendler, 1999). In recent years, significant progress has been made in identifying genetic variants associated with susceptibility to smoking (Liu et al., 2010, 2019; The Tobacco and Genetics Consortium et al., 2010; Thorgeirsson et al., 2010, 2008). In particular, most genetic variants from the *CHRNA5-CHRNA3-CHRNB4* gene cluster, which influences nicotine response and metabolism and which studies link to nicotine dependence, are strongly associated with smoking phenotypes (Stoker and Markou, 2013). Genetic variants near dopamine receptors are strongly associated with daily smoking and difficulty in cessation, but not with smoking initiation-related phenotypes, suggesting that dopamine-related variants become more relevant as an individual's nicotine use progresses (Liu et al., 2019). Indices of genetic predisposition to smoke have been shown to be related to smoking initiation, and individuals of higher genetic risk are more likely to develop dependence faster and more frequently and to fail in their cessation attempts (Belsky et al., 2013). Thus, genetic data may capture an individual's predisposition for smoking behaviors via multiple biological channels including addiction in addition to nicotine response.

These developments in mapping the genetic architecture of health behaviors, together with an increased availability of genetic measures in large representative surveys, allow for a better understanding of how genetics can interact with other environmental factors in determining smoking behavior and nicotine dependence. For example, adolescent environmental shocks have been shown to alter the influence of specific genetic variants on an individual's risk of developing nicotine dependence (Bierut, Johnson and Saccone, 2014; Chen et al., 2009; Johnson et al., 2010). More generally, the relationship between genetic variants and smoking has been shown to depend on neighborhood characteristics (Meyers et al., 2013) the cohort of birth (Domingue et al., 2016; Wedow et al., 2018), military service in the Vietnam era (Schmitz and Conley, 2016), and tobacco taxes (Fletcher, 2012; Slob and Rietveld, 2020).

Our results highlight the importance of considering genetic predisposition when evaluating behavioral responses to shocks and policies, such as health insurance coverage (Harper, 1993;

Morrison, 2005). Genetic predispositions can curb the negative behavioral consequences and the moral hazard associated with changes in health insurance status. Besides being relevant for the debate about equal and fair access to health insurance, our results highlight a new avenue of potential future research: leveraging recent advances in molecular and human genetics, we identify a new form of individual heterogeneity in the response to treatments. This heterogeneity used to be unobserved, which could lead to incorrect policy conclusions. In the era of genomics and personalized medicine, an individual genetic makeup can be a factor not only in the treatment they receive, but also in their response to insurance coverage (Khera et al., 2018; Ritz et al., 2017; Schork, Schork and Schork, 2018; Torkamani, Wineinger and Topol, 2018). A new form of individual heterogeneity in treatment effects (Papageorge and Thom, 2020). Since individuals are endowed at conception with their genetic makeup, and they have not done anything to deserve it or be held accountable for it (Barth, Papageorge and Thom, 2019; Harden, In Press; Kweon et al., 2020; Pereira, 2021), this differential response to insurance shocks raises questions of fairness and equality in the public debate over health insurance policies.

More generally, we show how leveraging recent developments in behavioral and molecular genetics can shed new light on fundamental economic concepts, such as elasticity of response to health shocks and heterogeneity in moral hazard. These concepts are formalized more precisely in the model in Appendix E. Indices of genetic predisposition to health behaviors provide readily available individual measures of heterogeneity that can enrich our economic models.

2 Data

2.1 Study Sample

The HRS is a nationally representative longitudinal household survey initiated in 1992 among US citizens aged 50 and older, followed for 12 biannual waves over 22 years, containing detailed medical, economic, social, psychological, and genetic information about the respondents (Sonnega et al., 2014).[1]

Genetic information comes from DNA samples collected in face-to-face interviews, for which a random subset of HRS households were selected to participate in 2006, 2008, and 2010. Saliva samples are collected and genotyped for more than 15,000 participants (Ware, Schmitz and Faul, 2017). More information on the genotyping process is provided in Appedix Section C.1.

The analysis is done on a subsample of the data selected based on criteria of age (between 60-70 at the time of interview), ever-smokers at baseline (their first observation), observed in at least 2 different time periods, and genetically of European descent. 5,854 HRS respondents satisfy these criteria and have non-missing values for the main variables of interest.[2] The age restriction imposed on the study sample increases comparability between those experiencing a health shock before or after age 65. The restriction imposed on genetic ancestry is best-practice in social-science genetics to avoid issues of population stratification(Hamer and Sirota, 2000). The restriction also increases the sample's similarity with the genetic profile of the discovery sample used to construct the index of genetic predisposition to smoking (Freese, 2018; Martin

[1] It is sponsored by the National Institute on Aging (grant number NIA U01AG009740).Response rates vary between 80% and 90% (HRS, 2017).

[2] See Section C.2 in the Appendix for more information on the construction of the study sample.

et al., 2017).[3] Also, we exclude respondents who reported the onset of a cardiovascular illness since the last survey wave when interviewed at ages 65 or 66. Since HRS surveys are only conducted every 2 years, it is not possible to determine Medicare eligibility status at the time of the health shock for these individuals.[4]

2.2 Variables of interest

Genetic propensity. As a measure of an individual's genetic propensity to smoke, we leverage recent developments in molecular and behavioral genetics to construct an index known as a polygenic score (PGS). A polygenic score is a weighted average of the number of risky genetic variants G_{ij} associated with the probability of smoking that are carried by a particular individual. For our measure of genetic variants G_{ij}, we follow the literature and use the most common form of genetic variation: Single Nucleotide Polymorphisms (SNP), a one base-pair substitution at a particular location (locus) on the human genome.

Specifically, the scores are calculated as follows:

$$PGS_i = \sum_{j=1}^{J} W_j G_{ij}, \tag{1}$$

where G_{ij} is the genotype for individual i at each of the more than 10 million SNP j; the weight W_j is the effect size for SNP j estimated via meta-analysis in a different sample than the HRS by the GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN) (Liu et al., 2019). The scores have been normalized to have a mean of zero and a standard deviation of one.[5]

To avoid functional form assumptions and maximize power in the statistical analysis, the sample is divided into high- and low-genetic propensity to smoking based on the PGS. A high-PGS indicator is defined to take the value 1 for individuals in the top 2/3 of the PGS distribution.[6] The distribution of the PGS is bell-shaped, it is shifted to the right for baseline smokers, and it is predictive of current smoking behavior: $\approx 30\%$ of individuals with a PGS higher than 2.5 are current smokers, while only $\approx 10\%$ of individuals with a PGS lower than -2.5 currently smoke (Appendix Figure 8).

[3] This reduces the external validity of any study that uses polygenic scores informed by GWAS of white participants. It might also exacerbate existing health disparities across ethnicity and hamper the potential for scientific knowledge and innovation (Martin et al., 2019). Thankfully, multi-ancestry GWAS are becoming more common, but not yet for smoking behavior (Peterson et al., 2019).

[4] The results do not change if these individuals are included.

[5] As a robustness check, shown in Appendix Section D.3.4, we also use the polygenic score provided by the HRS (Ware, Schmitz and Faul, 2017) for the smoking phenotype "regular smoking" (having smoked more than 100 cigarettes throughout one's life). This score is constructed as a weighted sum of the genotype over the 779,538 SNPs that overlap between the HRS genetic database and a 2010 GWAS meta-analysis conducted by the Tobacco and Genetics Consortium (The Tobacco and Genetics Consortium et al., 2010). See Section C.3.5 in the Appendix for more information on the construction of the PGS.

[6] We initially ran the analysis separately for each tertile of the PGS distribution, as shown in Appendix Section D.3.2, but the results are concentrated in the lower part of the distribution of the PGS, and therefore for simplicity we consider only high and low PGS. Results are robust to splitting the same above and below the median PGS score, as shown in Appendix Section D.3.3.

Health shocks. Following previous studies (Falba, 2005; Keenan, 2009; Khwaja, Sloan and Chung, 2006*a*,*b*; Smith et al., 2001), we focus on the first occurrence of an acute cardiovascular condition. Cardiovascular conditions include either a heart problem (heart attack, coronary heart disease, angina, congestive heart failure, or other heart problems) or a stroke (or transient ischemic attack). These conditions have strong links to tobacco use, have a relatively high incidence among older adults, and are likely to require immediate and intensive use of costly medical services (Lloyd-Jones et al., 2010; Teo et al., 2006; Thorpe, Florence and Joski, 2004). Additionally, they can occur repeatedly for the same individual and thereby incentivize improvements in health behaviors to prevent a recurrence.

The health shock is an indicator set to 1 if the respondent is diagnosed with a new cardiovascular condition during the time since the last HRS survey, but does not have a history of cardiovascular disease prior to this diagnosis. The exact timing of the health shock is not observed, and the diagnoses are all self-reported.

The rate of cardiovascular health problems increase with age from about 2% to 4% annually, but it is not strongly related to the PGS and does not jump around the age of 65 (Figure 4 and 9 in the Appendix).

Health Insurance. Before the age of 65, respondents are classified as uninsured in a given survey wave if they report not being covered by *any* health insurance plan. Respondents are classified as being *persistently uninsured* if they were uninsured in every wave between 60 and 65.

Smoking. The outcome of interest is a self-reported indicator for current smoking status. It is set to 1 if the respondent reports smoking at the time of the survey. The prevalence of smokers in the sample decreases with age and is about 5 percentage points higher for the individuals with a high PGS (Appendix Figure 3).

More information on the construction of all variables used in the analysis can be found in Appendix Section C.3.

2.3 Sample Characteristics

Characteristics of the study sample used in the analysis are summarized in the first column of Table 1. The study sample consists of 5,854 individuals (and 26,022 person-year observations). Individuals are observed for 4.5 waves on average, a good panel dimension even if restricted to the 60-70 age range. 5.9% of individuals in the sample are uninsured in all observations before the age of 65, and 12.3% experience a cardiovascular health shock during the observation period. Stratifying the sample by genetic propensity to smoking, we see that high-PGS respondents are 4 percentage points more likely to

smoke at baseline, 0.7 percentage points more likely to be uninsured before the age of 65, and included relatively more women compared to the low-PGS respondents.[7]

3 Empirical Analysis

In a within-person analysis with individual fixed effects, we compare the change in smoking rates after suffering a health shock across four types of persistently uninsured people: younger or older than 65 (as a proxy of exposure to health insurance via Medicare) and high or low genetic propensity to smoking. We focus on persistently uninsured people, i.e. people who reported being uninsured in every observation of the study sample before age 65, to zoom in on the potential for moral hazard.

The identification strategy builds on previous work (Marti and Richards, 2017) and leverages the differential timing of health shocks before or after Medicare eligibility at age 65. Conditional on having suffered a health shock, the exact *timing* of the shock is arguably both exogenous and unanticipated. While the probability of experiencing a health shock increases with age (Lloyd-Jones et al., 2010), no observable jump or change in trend can be seen in our data around the age of 65 (See Appendix C.4.1), which is consistent with the work of Card, Dobkin and Maestas (2009, 2008) who use much more granular data.[8]

The absence of a jump is reassuring for our identification strategy, which relies on a comparison before and after the age of 65: accounting for age, respondents experiencing the shock after 65 should not be systematically different from respondents experiencing the shock before 65. This setting thus provides a good framework for estimating the causal effect of health care eligibility on the smoking response to a health shock in individuals who are uninsured before the age of 65.

Methodologically, we use ordinary least squares (OLS) regression to estimate a linear probability model for smoking. Current smoking status (Y) is regressed on the full set of interactions between the indicators for the health shock ($shock$), being uninsured pre-65 ($uninsured$), Medicare eligibility ($post65$), and high polygenic score for smoking (g):

[7] This slight gender imbalance is a feature of our sampling strategy and not of the PGS, which is balanced across genders in the original paper (Liu et al., 2019). None of the results are driven by gender differences, which are always controlled for in the analysis. Appendix Section D.1 provides additional descriptive information about the study sample that experienced a health shock.

[8] Our dataset is not precise enough to estimate a possible discontinuity in the probability of suffering a cardiovascular health shock at age 65 since the exact age at the time of the cardiovascular health shock is not reported).

$$
\begin{aligned}
Y_{it} \;=\; & \beta\, shock_{it} + \gamma\, post65_{it} \\
& + \lambda_1 \left(shock_{it} \times post65_{it}\right) \\
& + \lambda_2 \left(shock_{it} \times uninsured_i\right) \\
& + \lambda_3 \left(post65_{it} \times uninsured_i\right) \\
& + \lambda_4 \left(shock_{it} \times g_i\right) \\
& + \lambda_5 \left(post65_{it} \times g_i\right) \\
& + \delta_1 \left(shock_{it} \times post65_{it} \times uninsured_i\right) \\
& + \delta_2 \left(shock_{it} \times uninsured_i \times g_i\right) \\
& + \delta_3 \left(shock_{it} \times post65_{it} \times g_i\right) \\
& + \delta_4 \left(post65_{it} \times uninsured_i \times g_i\right) \\
& + \zeta \left(shock_{it} \times post65_{it} \times uninsured_i \times g_i\right) \\
& + \sum_{a=1}^{3} \phi_a\, age_{it}^{a} + \eta_i + \tau_t + \varepsilon_{it}
\end{aligned}
\tag{2}
$$

Individual fixed effects (η_i) are included to control for unobserved time-invariant differences between respondents, and time fixed effects (τ_t) to control for time-specific confounders. In addition, a third-degree polynomial of the respondent's age in years at the time of the interview (age) is added to non-linearly control for age-reduction in smoking rates. Standard errors are clustered at the individual level.

The statistical method resembles a difference-in-differences approach: focusing on the group of previously uninsured individuals, we compared the effect of a health shock for those experiencing it before or after Medicare eligibility, and who have either a low or high genetic propensity to smoke.

We are aim to answer the following question: does the effect of health insurance vary according to individual genetic predisposition? Comparing the effect of Medicare eligibility on the post-shock smoking decision between the high- and the low-PGS groups we can assess the presence of genetic heterogeneity in moral hazard. The presence of genetic heterogeneity in moral hazard is estimated by the sum of δ_3 and ζ in Equation (2).[9]

[9] See Appendix Section C.5 for a derivation of these effects.

4 Results

4.1 Estimation Results

The effect of a health shock on the smoking probability of individuals who are uninsured before the age of 65—the subgroup of interest—for the 4 combinations of shock timing (before or after 65) and polygenic score (high or low) is shown in Figure 1 and Panel A of Table 2. Since we are controlling for individual fixed effects, these coefficients can be interpreted as the change in the probability of smoking after suffering a health shock. All of the regression coefficients are presented in the last column of Table 3.

Following a health shock, persistently uninsured individuals with a high polygenic score tend to reduce their smoking behavior by about 10 percentage points, irrespective of whether the health shock happens before or after the age of 65. Since the response to the health shock is the same before and after the age of Medicare eligibility, there is no evidence of moral hazard in this subgroup. With a 30.8% baseline probability of smoking, this 10 percentage point reduction is non-negligible, but it is nosily estimated and indistinguishable from zero.

On the contrary, the response of low polygenic score individuals differs markedly by the *timing* of the health shock: if the shock happens before the age of 65, while not covered by Medicare or another health insurance, they reduce smoking by 17 percentage points. With a baseline probability of smoking of 27%, this reduction is very sizable. On the other hand, if the health shock happens after the age of Medicare eligibility, low polygenic score individuals increase their smoking rate by 9 percentage points as compared to the rest of the population. Against the backdrop of a steady decline in smoking rates, this positive coefficient can be interpreted as a lower decrease in smoking for this particular sub-population.[10]

The difference in the smoking response depending on the timing of the health shock, before or after the age of 65, is summarized in Panel B of Table 2: individuals with a low polygenic score are 25.6 percentage points more likely to stop smoking if the health shock happens before the age of 65, when they are not covered by health insurance. High polygenic score individuals, on the other hand, are 2.3 percentage points more likely to do the opposite, a small and (statistically) insignificant difference.

Such marked difference in the response across the two genetic types is evidence of genetic heterogeneity in the effect of Medicare eligibility on the response to the shock. This genetic heterogeneity,[11] is summarized in Panel C of Table 2: low PGS individuals are 27.9 percentage points less likely than high PGS individuals to stop smoking if the

[10] Notice that the magnitude of this coefficient (0.09) is similar to the ones estimated for high polygenic score individuals (-0.108 and -0.13), but it is much more precisely estimated. The main reason behind smaller standard errors for the estimated effect of health shock after 65 for the low polygenic score individuals is the negative covariance $Cov(\hat{\delta}_1, \hat{\lambda}_2) = -0.0057$. See Appendix Table 7.

[11] A form of what behavioral geneticists call 'gene-environment interaction' (G×E) Haldane (1946); Plomin (1990).

health shock happens when they are uninsured.

The evidence in favor of genetic heterogeneity is remarkably stable to the inclusion of different controls and fixed effects, as shown in the different columns of Table 3. The first column does not include polynomial controls for age or any type of time or individual fixed effects; still, the sum of the estimated δ_3 ($shock_{it} \times post65_{it} \times g_i$) and ζ ($shock_{it} \times post65_{it} \times uninsured_i \times g_i$) coefficients is 32.6 percentage points, close to the 27.9 estimated with age controls and time and individual fixed effects.

How can we interpret this evidence of genetic heterogeneity? This result is consistent with the interpretation that high polygenic score individuals are less responsive and elastic in their response to shocks and changes in the environment. Regardless of whether the health shock happens when they are covered by health insurance, and therefore whether they are financially liable for the consequences of the health shock, they reduce smoking the same amount. Individuals with a low polygenic score are instead more flexible, strategic, and reactive to the environment: they reduce their smoking only when they bear the full cost of the consequences of negative health shocks. This result is therefore consistent with genetic heterogeneity in moral hazard.

4.2 Other Interpretations

Falsification Test: age 50 to 60 To dispel doubts of spurious results, we run a falsification or "placebo" test by focusing on the age range from 50 to 60, and estimating the effect of suffering a shock after the age of 55, where nothing happens in terms of health insurance.

We find no average or differential effect of suffering a health shock after the age of 55 (See Appendix Figure 10 and Appendix Table 9).

Is it really Medicare? Other factors besides eligibility to Medicare may be at play around the age of 65, the most prominent being retirement. Such factors might affect smoking behavior after health shocks and change the interpretation of our coefficient. Our empirical result about genetic heterogeneity would still stand, but the post-65 dummy could not be considered purely as a measure of health care eligibility, and the differential reduction in smoking before 65 would not be evidence of moral hazard heterogeneity.

These alternative interpretations are at odds with two findings: there is no effect for those already insured before the age of 65, and there is no differential effect on income or retirement.

First, the coefficient on the shock x post-65 interaction is close to 0 and not statistically significant, suggesting no differential effect of the health shock for who is already covered by health insurance before the age of 65 (See Table 3). If factors other than health insurance were to cause a differential effect of the health-shock pre- vs. post-65,

such effect should be noticeable also for those already insured before the age of 65.

Secondly, there is no big jump in income or retirement at the age of 65 in our sample (Appendix Figure 6), while there is a sharp change in enrollment in Medicare and drop in the share of uninsured (Appendix Figure 5). This suggests that the main event captured by the post-65 dummy for the sample of previously uninsured individuals is access to health coverage.

Finally, there is no differential effect of the health shock before or after the age of 65 on income or retirement for persistently uninsured individuals (Appendix Figure 15). The absence of any discernible effect suggests that neither income nor retirement are potential mechanisms behind the observed change in smoking responses.

Is it really genes? Other individual-level characteristics, besides genetic predispositions, might be driving this heterogeneity in moral hazard. Given the richness of the HRS data, we can run the same analysis outlined in equation 2 but replacing g_i with other individual characteristics.

None of the other individual measures that we have tried seems to be driving heterogeneity in moral hazard (Appendix Figure 2). Specifically, we tested for years of education (highly educated people might be more knowledgeable of the insurance system), a proxy of cognitive abilities(smart people might be more strategic), risk aversion (risk preferences might moderate smoking behaviors and the response to health shocks), the personality trait conscientiousness (conscientious people might be more likely to stop smoking after a health shock), gender (social scientist's favorite sample split to engage in ex-post rationalizations), and individual or household income (which might buffer negative shocks).[12] For completeness, results are shown split by tertiles for continuous measures, and in two for binary measures.

Is this driven by confounders? One might worry that the differential change in smoking behavior might be driven by other confounding factors that happen around the age of 65, or are triggered by the health shock. Suffering a heart attack might reduce people's income, induce them to retire, change their marital status, increase their out-of-pocket medical spending, or shorten their life expectancy. To dispel such concerns, we estimate equation 2 including each of these potential confounders as an outcome

[12]Notice that some of these variables are not quite predetermined, and might be mediators of the main effect, especially income. We still report the results for completeness, but caution the reader prone to causal interpretations. In this regard, genetic variants represent the ideal measure of heterogeneity of treatment effects: they are fixed since conception, immutable, and indexing plausible biological channels. In Appendix D.3.5, we report the results of using different polygenic scores as a proxy for g_i. Moral hazard heterogeneity is also detected in the group having a high PGS for educational attainment or cognitive abilities (Lee et al., 2018), and a low PGS for risk tolerance (Karlsson Linnér et al., 2019) and a low PGS for non-cognitive skills (Demange et al., 2020). These results suggest that potential mechanisms for the observed heterogeneity should include not only biological channels related to nicotine addiction, but also cognitive processes related to risky and strategic behavior, which are closely related to the concept of moral hazard.

variable Y_{it} (Pei, Pischke and Schwandt, 2018).

None of these potential confounders seems to be driving our results (Appendix Section D.4). If anything, the symmetric results for out-of-pocket medical expenditure shown in panel (e) suggest that the size of the health shock is comparable for both high and low PGS individuals.

4.3 Robustness checks

These results are robust to changes in the definition of the high-PGS indicator (above median PGS indicator or a linear PGS) and the definition of the pre-65 uninsured status indicator (uninsured in 33% or 66% of all pre-65 observations instead of 100%). Relaxing the definition of the uninsured indicator to include respondents uninsured in a minimum of 33% of pre-65 observations leaves the directions of the effects unchanged, but the magnitudes are smaller and not statistically significant. Furthermore, results are robust to the inclusion of HRS respondents who report a health shock at ages 65 or 66 (Medicare eligibility status at the time of the shock is unknown). Estimation results for all robustness checks are shown in Section D.3 in the Appendix.

4.4 Limitations

This study has a few limitations.

The identification hinges on a small number of respondents who suffer a health shock around the age of 65. Although the total sample size is 5,854 individuals and 26,022 person-year observations, the number of people who suffer a health shock, before or after the age of 65, and have a certain polygenic score is between 110 and 274.[13] We encourage and welcome a replication of our results, preferably in a within-family design which can better account for the thorny issue of ancestry and population structure, but we are unaware of an existing dataset that contains all of the necessary information to perform it ourselves.

Second, the analysis focused on the short-run smoking response to a health shock, and only considered changes in the extensive margin—i.e., changes between smoking and not smoking.

Third, regarding internal validity. We rely on self-reported smoking behavior, health diagnoses, and insurance status. Differential misreporting of smoking or health shocks by age or genotype might bias our analysis. We found no published evidence that misreporting might be associated with individual genotypes, or vary discontinously before or after 65. Finally, we may suffer from survivorship bias, attrition, and self-selection of HRS respondents into the DNA collection, which took place relatively late (starting in 2006). Participants with particularly high PGS and unhealthy smoking

[13] See Appendix Table 5.

habits may have passed away before this date or may have been less likely to reach the age of Medicare eligibility. Both problems would likely have lead to an underestimation of the difference between low- and high-PGS individuals.

Lastly, in terms of external validity. Medicare is very specific to the health care system in the US. While Medicare can generally be seen as just one example of universal health care coverage, it only applies to a specific age group, and it is not clear whether smoking behaviors in response to a health shock in this age group are representative for all ages in the population. Another concern is that Medicare eligibility does not necessarily translate into Medicare coverage. However, take-up rates in the study sample is high—approximately 90% at age 65, increasing to 98% at age 70. Additionally, take-up patterns do not differ much between the genetic groups. Table 16 in the Appendix shows that using actual Medicare enrollment status rather than Medicare eligibility status in the empirical analysis does not change the results.

5 Conclusions

Medicare eligibility significantly lowered the probability of smoking cessation after a health shock in individuals aged between 60 and 70 years who are uninsured before the age of 65 and have a low genetic predisposition for smoking. Health insurance can plausibly affect the smoking response to a health shock by lowering the financial risk associated with the shock, and thereby eroding additional incentives for behavior change. This change in behavior following Medicare eligibility is indicative of moral hazard, and we observe it only among individuals with a low genetic predisposition for smoking. The differential effect of Medicare eligibility for the two genetic groups suggests that biological constraints can overpower both health-related and financial incentives for smoking cessation, and provides a readily available measure of heterogeneity in moral hazard. Heterogeneity in moral hazard can be used to enrich economic models of health behavior[14] and our understanding of how individual biological characteristics can influence decision-making.

Building on previous work analyzing the interplay between genes and exogenous environmental changes (Barcellos, Carvalho and Turley, 2018; Fletcher, 2012; Schmitz and Conley, 2017*a*), this study provides a contribution to the centennial debate about nature *and* nurture (Galton, 1874; Haldane, 1946; Kong et al., 2018; Lundborg and Stenberg, 2010; Mulcaster, 1582), casting further doubts about genetic determinism. The influence of genetic variants on our choices and outcomes is modulated by the environment around us, just as the response to environmental shocks is filtered through the prism of our genetic predispositions.

Our results show that genetic factors can influence health decisions and strategic be-

[14] For example, see the model in Appendix E.

haviors, and therefore should be taken into consideration when evaluating the effectiveness and fairness of different policies, such as health insurance. Fairness considerations should take into account the fact that genetic endowments are passed down from one generation to the next, are fixed at conception, and cannot be changed by an individual's choices or effort, raising questions of deservedness, merit, and luck (Harden, In Press; Kweon et al., 2020; Pereira, 2021). Efficiency considerations should consider that genetic endowments are usually unobserved to the individual, the insurance companies, and the government. This unobservability raises important regulatory questions at the intersection between health and information economics: who, if anyone, should have this information? How should the information be provided? How will this information affect demand and supply? Should private contracts or public policies take this information into account? Under current US legislation, the 2008 Genetic Information Nondiscrimination Act prohibits health insurers from discriminating based on explicit genetic information. However, the recent rise of direct-to-consumer genetic testing services might force further discussion into the public debate.

More generally, future studies should build off of the idea of leveraging existing indices of genetic predispositions to provide biological measures of heterogeneity in human behaviors. This can enrich economic models and empirical studies, shedding new light on fundamental economic parameters.

References

Arrow, Kenneth J. 1963. "Uncertainty and the Welfare Economics of Medical Care." *The American Economic Review*, 53(5): 941–973.

Barcellos, Silvia H, Leandro S Carvalho, and Patrick Turley. 2018. "Education can reduce health differences related to genetic risk of obesity." *Proceedings of the National Academy of Sciences*, 201802909.

Barnett, Jessica C, and Marina Vornovitsky. 2016. "Health Insurance Coverage in the United States: 2015." US Census Bureau, Washington DC.

Barth, Daniel, Nicholas W Papageorge, and Kevin Thom. 2019. "Genetic Endowments and Wealth Inequality." *Journal of Political Economy*, 705415.

Belsky, Daniel W, Benjamin W Domingue, Robbee Wedow, Louise Arseneault, and Jason D Boardman. 2018. "Genetic analysis of social-class mobility in five longitudinal studies." *Proceedings of the National Academy of Sciences*, 1–10.

Belsky, Daniel W, Terrie E Moffitt, Timothy B Baker, Andrea K Biddle, James P Evans, HonaLee Harrington, Renate Houts, Madeline Meier, Karen Sugden, Benjamin Williams, et al. 2013. "Polygenic risk and the developmental progression to heavy, persistent smoking and nicotine dependence: evidence from a 4-decade longitudinal study." *JAMA psychiatry*, 70(5): 534–542.

Bierut, Laura Jean, Eric O Johnson, and Nancy L Saccone. 2014. "A glimpse into the future–Personalized medicine for smoking cessation." *Neuropharmacology*, 76: 592–599.

Boardman, Jason D, Casey L Blalock, and Fred C Pampel. 2010. "Trends in the Genetic Influences on Smoking." *Journal of Health and Social Behavior*, 51(1): 108–123.

Card, David, Carlos Dobkin, and Nicole Maestas. 2009. "Does Medicare Save Lives?" *Quarterly Journal of Economics*, 124(2): 597–636.

Card, D, C Dobkin, and N Maestas. 2008. "The impact of nearly universal insurance coverage on health care utilization: evidence from Medicare." *American Economic Review*, 98(5): 2242–58.

Caspi, Avshalom, Joseph McClay, Terrie E Moffitt, Jonathan Mill, Judy Martin, Ian W Craig, Alan Taylor, and Richie Poulton. 2002. "Role of genotype in the cycle of violence in maltreated children." *Science*, 297(5582): 851–4.

Chatterjee, Chirantan, Radhika Joshi, Neeraj Sood, and P. Boregowda. 2018. "Government health insurance and spatial peer effects: New evidence from India." *Social Science and Medicine*, 196: 131–141.

Chen, Li-Shiun, Eric O Johnson, Naomi Breslau, Dorothy Hatsukami, Nancy L Saccone, Richard A Grucza, Jen C Wang, Anthony L Hinrichs, Louis Fox, Alison M Goate, et al. 2009. "Interplay of genetic risk factors and parent monitoring in risk for nicotine dependence." *Addiction*, 104(10): 1731–1740.

Clark, Andrew, and Fabrice Etilé. 2002. "Do health changes affect smoking? Evidence from British panel data." *Journal of health economics*, 21(4): 533–62.

Cohen, Robin A, Diane M Makuc, Amy B Bernstein, Linda T Bilheimer, and Eve Powell-Griner. 2009. "Health insurance coverage trends, 1959-2007: estimates from the National Health Interview Survey." *National health statistics reports*, 17: 1–25.

Dave, Dhaval M, and Robert Kaestner. 2009. "Health insurance and ex ante moral hazard: evidence from Medicare." *International Journal of Health Care Finance and Economics*, 9(4): 367–390.

Demange, Perline A, Margherita Malanchini, Travis T Mallard, Pietro Biroli, Simon R Cox, Andrew D Grotzinger, Elliot M Tucker-drob, Abdel Abdellaoui, Louise Arseneault, Avshalom Caspi, David Corcoran, Benjamin W Domingue, Elsje Van Bergen, Dorret I Boomsma, Kathleen M Harris, Hill F Ip, Terrie E Moffitt, Richie Poulton, Joseph Prinz, Karen Sugden, Jasmin Wertz, Eveline L De Zeeuw, Daniel W Belsky, K Paige Harden, Michel G Nivard, Colter Mitchell, Elsje van Bergen, Dorret I Boomsma, Kathleen M Harris, Hill F Ip, Terrie Moffitt, Richie Poulton, Joseph Prinz, Karen Sugden, Jasmin Wertz, Benjamin Williams, Eveline de Zeeuw, Daniel W Belsky, K Paige Harden, and Michel G Nivard. 2020. "Investigating the Genetic Architecture of Non-Cognitive Skills Using GWAS-by-Subtraction." *bioRxiv*.

Domingue, Benjamin W, Dalton Conley, Jason Fletcher, and Jason D Boardman. 2016. "Cohort effects in the genetic influence on smoking." *Behavior genetics*, 46(1): 31–42.

Dubois, Pierre, and Tomislav Vukina. 2009. "Optimal incentives under moral hazard and heterogeneous agents: Evidence from production contracts data." *International Journal of Industrial Organization*, 27(4): 489–500.

Einav, Liran, Amy Finkelstein, Stephen P Ryan, Paul Schrimpf, and Mark R Cullen. 2013. "Selection on Moral Hazard in Health Insurance." *American Economic Review*, 103(1): 178–219.

Einav, Liran, and Amy Finkelstein. 2018. "Moral Hazard in Health Insurance: What We Know and How We Know It." *Journal of the European Economic Association*, 16(4): 957–982.

Falba, Tracy. 2005. "Health events and the smoking cessation of middle aged Americans." *Journal of behavioral medicine*, 28(1): 21–33.

Finkelstein, Amy. 2014. *Moral Hazard in Health Insurance.* New York Chichester, West Sussex:Columbia University Press.

Fletcher, Jason M. 2012. "Why have tobacco control policies stalled? Using genetic moderation to examine policy impacts." *PloS one*, 7(12): e50576.

Freese, Jeremy. 2018. "The Arrival of Social Science Genomics." *Contemporary Sociology: A Journal of Reviews*, 47(5): 524–536.

Galton, Francis. 1874. *English men of science: Their nature and nurture.* London:McMillan & Co.

Goodchild, Mark, Nigar Nargis, and Edouard Tursan d'Espaignet. 2018. "Global economic cost of smoking-attributable diseases." *Tobacco Control*, 27(1): 58–64.

Haldane, John Burdon Sanderson. 1946. "The interaction of nature and nurture." *Annals of Eugenics*, 13(3): 197–205.

Hall, W, P Madden, and M Lynskey. 2002. "The genetics of tobacco use: methods, findings and policy implications." *Tobacco control*, 11(2): 119–24.

Hamer, D, and L. Sirota. 2000. "Beware the chopsticks gene." *Molecular Psychiatry*, 5(1): 11–13.

Harden, Kathryn Paige. In Press. *The Genetic Lottery.* Princeton University Press.

Harper, Peter S. 1993. "Insurance and genetic testing." *The Lancet*, 341(8843): 495.

Health and Retirement Study. HRS polygenic scores 2006-2010 genetic data, public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI. 2018. n.d.*b*.

Health and Retirement Study. RAND HRS longitudinal file 2014, version P, public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI. 2018. n.d.*a*.

Heath, Andrew C., Randall Cates, Nicholas G Martin, Joanne Meyer, John K. Hewitt, Michael C. Neale, and Lindon J. Eaves. 1993. "Genetic contribution to risk of smoking initiation: Comparisons across birth cohorts and across cultures." *Journal of Substance Abuse*, 5(3): 221–246.

Hoffmann, Roman. 2017. "Following the peers: The role of social networks for health care utilization in the Philippines." *Working Paper.*

HRS. 2017. "Health and Retirement Study. Sample sizes and response rates." *https://hrs.isr.umich.edu/sites/default/files/biblio/ResponseRates_2017.pdf*., Accessed on April 20, 2018.

Jamal, Ahmed, Brian A. King, Linda J. Neff, Jennifer Whitmill, Stephen D. Babb, and Corinne M. Graffunder. 2016. "Current Cigarette Smoking Among Adults – United States, 2005–2015." *MMWR. Morbidity and Mortality Weekly Report*, 65(44): 1205–1211.

Johnson, Eric O, Li-Shiun Chen, Naomi Breslau, Dorothy Hatsukami, Tania Robbins, Nancy L Saccone, Richard A Grucza, and Laura J Bierut. 2010. "Peer smoking and the nicotinic receptor genes: an examination of genetic and environmental risks for nicotine dependence." *Addiction*, 105(11): 2014–2022.

Karlsson Linnér, Richard, Pietro Biroli, Edward Kong, S Fleur W Meddens, Robbee Wedow, Mark Alan Fontana, Maël Lebreton, Stephen P Tino, Abdel Abdellaoui, Anke R Hammerschlag, Michel G Nivard, Aysu Okbay, Cornelius A Rietveld, Pascal N Timshel, Maciej Trzaskowski, Ronald de Vlaming, Christian L Zünd, Yanchun Bao, Laura Buzdugan, Ann H Caplin, Chia-Yen Chen, Peter Eibich, Pierre Fontanillas, Juan R Gonzalez, Peter K Joshi, Ville Karhunen, Aaron Kleinman, Remy Z Levin, Christina M Lill, Gerardus A Meddens, Gerard Muntané, Sandra Sanchez-Roige, Frank JA van Rooij, Erdogan Taskesen, Yang Wu, Futao Zhang, Adam Auton, Jason D Boardman, David W Clark, Andrew Conlin, Conor C Dolan, Urs Fischbacher, Patrick J F Groenen, Kathleen Mullan Harris, Gregor Hasler, Albert Hofman, Mohammad A Ikram, Sonia Jain, Robert Karlsson, Ronald C Kessler, Maarten Kooyman, James MacKillop, Minna Männikkö, Carlos Morcillo-Suarez, Matthew B McQueen, Klaus M Schmidt, Melissa C Smart, Matthias Sutter, A Roy Thurik, André G Uitterlinden, Jon White, Harriet de Wit, Jian Yang, Lars Bertram, Dorret I Boomsma, Tõnu Esko, Ernst Fehr, David A Hinds, Magnus Johannesson, Meena Kumari, David Laibson, Patrik K E Magnusson, Michelle N Meyer, Arcadi Navarro, Abraham A Palmer, Tune H Pers, Danielle Posthuma, Daniel Schunk, Murray B Stein, Rauli Svento, Henning Tiemeier, Paul R H J Timmers, Patrick Turley, Robert J Ursano, Gert G Wagner, James F Wilson, Jacob Gratten, James J Lee, David Cesarini, Daniel J Benjamin, Philipp D Koellinger, and Jonathan P Beauchamp. 2019. "Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences." *Nature Genetics*, 51(2): 245–257.

Keenan, Patricia S. 2009. "Smoking and Weight Change After New Health Diagnoses in Older Adults." *Archives of Internal Medicine*, 169(3): 237.

Khera, Amit V, Mark Chaffin, Krishna G. Aragam, Mary E. Haas, Carolina Roselli, Seung Hoan Choi, Pradeep Natarajan, Eric S. Lander, Steven A. Lubitz, Patrick T. Ellinor, and Sekar Kathiresan. 2018. "Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations." *Nature Genetics*, 1.

Khwaja, Ahmed, Frank Sloan, and Sukyung Chung. 2006*a*. "Learning about individual risk and the decision to smoke." *International Journal of Industrial Organization*, 24(4): 683–699.

Khwaja, Ahmed, Frank Sloan, and Sukyung Chung. 2006*b*. "The Effects of Spousal Health on the Decision to Smoke: Evidence on Consumption Externalities, Altruism and Learning Within the Household." *Journal of Risk and Uncertainty*, 32(1): 17–35.

Kong, Augustine, Gudmar Thorleifsson, Michael L Frigge, Bjarni J Vilhjalmsson, Alexander I Young, Thorgeir E Thorgeirsson, Stefania Benonisdottir, Asmundur Oddsson, Bjarni V Halldorsson, Gisli Masson, Daniel F Gudbjartsson, Agnar Helgason, Gyda Bjornsdottir, Unnur Thorsteinsdot-

tir, and Kari Stefansson. 2018. "The nature of nurture: Effects of parental genotypes." *Science*, 359(6374): 424–428.

Kowalski, Amanda E. 2018. "Extrapolation using Selection and Moral Hazard Heterogeneity from within the Oregon Health Insurance Experiment." *NBER Working Paper Series*, 24647: 1–38.

Kweon, Hyeokmoon, Casper A P Burik, Richard Karlsson Linnér, Ronald De Vlaming, Aysu Okbay, Daphne Martschenko, Kathryn Paige Harden, Thomas A Di Prete, and Philipp D Koellinger. 2020. "Genetic Fortune: Winning or Losing Education, Income, and Health." *Working Paper*.

Lee, James J, Robbee Wedow, Aysu Okbay, Edward Kong, Omeed Maghzian, Meghan Zacher, Tuan Anh Nguyen-Viet, Peter Bowers, Julia Sidorenko, Richard Karlsson Linnér, Mark Alan Fontana, Tushar Kundu, Chanwook Lee, Hui Li, Ruoxi Li, Rebecca Royer, Pascal N Timshel, Raymond K Walters, Emily A Willoughby, Loïc Yengo, Maris Alver, Yanchun Bao, David W Clark, Felix R Day, Nicholas A Furlotte, Peter K Joshi, Kathryn E Kemper, Aaron Kleinman, Claudia Langenberg, Reedik Mägi, Joey W Trampush, Shefali Setia Verma, Yang Wu, Max Lam, Jing Hua Zhao, Zhili Zheng, Jason D Boardman, Harry Campbell, Jeremy Freese, Kathleen Mullan Harris, Caroline Hayward, Pamela Herd, Meena Kumari, Todd Lencz, Jian'an Luan, Anil K Malhotra, Andres Metspalu, Lili Milani, Ken K Ong, John R B Perry, David J Porteous, Marylyn D Ritchie, Melissa C Smart, Blair H Smith, Joyce Y Tung, Nicholas J Wareham, James F Wilson, Jonathan P Beauchamp, Dalton C Conley, Tõnu Esko, Steven F Lehrer, Patrik K E Magnusson, Sven Oskarsson, Tune H Pers, Matthew R Robinson, Kevin Thom, Chelsea Watson, Christopher F Chabris, Michelle N Meyer, David I Laibson, Jian Yang, Magnus Johannesson, Philipp D Koellinger, Patrick Turley, Peter M Visscher, Daniel J Benjamin, and David Cesarini. 2018. "Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals." *Nature Genetics*, 1.

Li, Ming D, Rong Cheng, Jennie Z Ma, and Gary E Swan. 2003. "A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins." *Addiction*, 98(1): 23–31.

Liu, Jason Z, Federica Tozzi, Dawn M Waterworth, Sreekumar G Pillai, Pierandrea Muglia, Lefkos Middleton, Wade Berrettini, Christopher W Knouff, Xin Yuan, Gérard Waeber, Peter Vollenweider, Martin Preisig, Nicholas J Wareham, Jing Hua Zhao, Ruth J F Loos, Inês Barroso, Kay-Tee Khaw, Scott Grundy, Philip Barter, Robert Mahley, Antero Kesaniemi, Ruth McPherson, John B Vincent, John Strauss, James L Kennedy, Anne Farmer, Peter McGuffin, Richard Day, Keith Matthews, Per Bakke, Amund Gulsvik, Susanne Lucae, Marcus Ising, Tanja Brueckl, Sonja Horstmann, H-Erich Wichmann, Rajesh Rawal, Norbert Dahmen, Claudia Lamina, Ozren Polasek, Lina Zgaga, Jennifer Huffman, Susan Campbell, Jaspal Kooner, John C Chambers, Mary Susan Burnett, Joseph M Devaney, Augusto D Pichard, Kenneth M Kent, Lowell Satler,

Joseph M Lindsay, Ron Waksman, Stephen Epstein, James F Wilson, Sarah H Wild, Harry Campbell, Veronique Vitart, Muredach P Reilly, Mingyao Li, Liming Qu, Robert Wilensky, William Matthai, Hakon H Hakonarson, Daniel J Rader, Andre Franke, Michael Wittig, Arne Schäfer, Manuela Uda, Antonio Terracciano, Xiangjun Xiao, Fabio Busonero, Paul Scheet, David Schlessinger, David St Clair, Dan Rujescu, Gonçalo R Abecasis, Hans Jörgen Grabe, Alexander Teumer, Henry Völzke, Astrid Petersmann, Ulrich John, Igor Rudan, Caroline Hayward, Alan F Wright, Ivana Kolcic, Benjamin J Wright, John R Thompson, Anthony J Balmforth, Alistair S Hall, Nilesh J Samani, Carl A Anderson, Tariq Ahmad, Christopher G Mathew, Miles Parkes, Jack Satsangi, Mark Caulfield, Patricia B Munroe, Martin Farrall, Anna Dominiczak, Jane Worthington, Wendy Thomson, Steve Eyre, Anne Barton, Vincent Mooser, Clyde Francks, Jonathan Marchini, and Jonathan Marchini. 2010. "Meta-analysis and imputation refines the association of 15q25 with smoking quantity." *Nature Genetics*, 42(5): 436–440.

Liu, Mengzhen, Yu Jiang, Robbee Wedow, Yue Li, David M Brazel, Fang Chen, Gargi Datta, Jose Davila-Velderrain, Daniel McGuire, Chao Tian, Xiaowei Zhan, Hélène Choquet, Anna R Docherty, Jessica D Faul, Johanna R Foerster, Lars G Fritsche, Maiken Elvestad Gabrielsen, Scott D Gordon, Jeffrey Haessler, Jouke-Jan Hottenga, Hongyan Huang, Seon-Kyeong Jang, Philip R Jansen, Yueh Ling, Reedik Mägi, Nana Matoba, George McMahon, Antonella Mulas, Valeria Orrù, Teemu Palviainen, Anita Pandit, Gunnar W Reginsson, Anne Heidi Skogholt, Jennifer A Smith, Amy E Taylor, Constance Turman, Gonneke Willemsen, Hannah Young, Kendra A Young, Gregory J M Zajac, Wei Zhao, Wei Zhou, Gyda Bjornsdottir, Jason D Boardman, Michael Boehnke, Dorret I Boomsma, Chu Chen, Francesco Cucca, Gareth E Davies, Charles B Eaton, Marissa A Ehringer, Tõnu Esko, Edoardo Fiorillo, Nathan A Gillespie, Daniel F Gudbjartsson, Toomas Haller, Kathleen Mullan Harris, Andrew C Heath, John K Hewitt, Ian B Hickie, John E Hokanson, Christian J Hopfer, David J Hunter, William G Iacono, Eric O Johnson, Yoichiro Kamatani, Sharon L R Kardia, Matthew C Keller, Manolis Kellis, Charles Kooperberg, Peter Kraft, Kenneth S Krauter, Markku Laakso, Penelope A Lind, Anu Loukola, Sharon M Lutz, Pamela A F Madden, Nicholas G Martin, Matt McGue, Matthew B McQueen, Sarah E Medland, Andres Metspalu, Karen L Mohlke, Jonas B Nielsen, Yukinori Okada, Ulrike Peters, Tinca J C Polderman, Danielle Posthuma, Alexander P Reiner, John P Rice, Eric Rimm, Richard J Rose, Valgerdur Runarsdottir, Michael C Stallings, Alena Stančáková, Hreinn Stefansson, Khanh K Thai, Hilary A Tindle, Thorarinn Tyrfingsson, Tamara L Wall, David R Weir, Constance Weisner, John B Whitfield, Bendik Slagsvold Winsvold, Jie Yin, Luisa Zuccolo, Laura J Bierut, Kristian Hveem, James J Lee, Marcus R Munafò, Nancy L Saccone, Cristen J Willer, Marilyn C Cornelis, Sean P David, David A Hinds, Eric Jorgenson, Jaakko Kaprio, Jerry A Stitzel, Kari Stefansson, Thorgeir E Thorgeirsson, Gonçalo Abecasis, Dajiang J Liu, and Scott Vrieze. 2019. "Association studies of up to 1.2 million individuals yield new insights

into the genetic etiology of tobacco and alcohol use." *Nature Genetics*.

Lloyd-Jones, Donald, Robert J. Adams, Todd M. Brown, Mercedes Carnethon, Shifan Dai, Giovanni De Simone, T. Bruce Ferguson, Earl Ford, Karen Furie, Cathleen Gillespie, Alan Go, Kurt Greenlund, Nancy Haase, Susan Hailpern, P. Michael Ho, Virginia Howard, Brett Kissela, Steven Kittner, Daniel Lackland, Lynda Lisabeth, Ariane Marelli, Mary M. McDermott, James Meigs, Dariush Mozaffarian, Michael Mussolino, Graham Nichol, Véronique L. Roger, Wayne Rosamond, Ralph Sacco, Paul Sorlie, Randall Stafford, Thomas Thom, Sylvia Wasserthiel-Smoller, Nathan D. Wong, and Judith Wylie-Rosett. 2010. "Heart Disease and Stroke Statistics–2010 Update." *Circulation*, 121(7).

Lundborg, Petter, and Anders Stenberg. 2010. "Nature, nurture and socioeconomic policy-what can we learn from molecular genetics?" *Economics and Human Biology*, 8(3): 320–30.

Ma, Jiemin, Rebecca L. Siegel, Eric J. Jacobs, and Ahmedin Jemal. 2018. "Smoking-attributable Mortality by State in 2014, U.S." *American Journal of Preventive Medicine*, 54(5): 661–670.

Marti, Joachim, and Michael R Richards. 2017. "Smoking Response to Health and Medical Spending Changes and the Role of Insurance." *Health Economics*, 26: 305–320.

Martin, Alicia R, Christopher R Gignoux, Raymond K Walters, Genevieve L Wojcik, Benjamin M Neale, Simon Gravel, Mark J Daly, Carlos D Bustamante, and Eimear E Kenny. 2017. "Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations." *The American Journal of Human Genetics*, 100(4): 635–649.

Martin, Alicia R, Masahiro Kanai, Yoichiro Kamatani, Yukinori Okada, Benjamin M Neale, and Mark J Daly. 2019. "Clinical use of current polygenic risk scores may exacerbate health disparities." *Nature Genetics 2019 51:4*, 51(4): 584.

Meyers, JL, Magdalena Cerda, S Galea, KM Keyes, AE Aiello, M Uddin, DE Wildman, and KC Koenen. 2013. "Interaction between polygenic risk for cigarette use and environmental exposures in the Detroit neighborhood health study." *Translational psychiatry*, 3(8): e290–e290.

Morrison, Patrick J. 2005. "Insurance, unfair discrimination, and genetic testing." *The Lancet*, 366(9489): 877–880.

Mulcaster, Richard. 1582. *Mulcaster's Elementaire.* London:Clarendon Press.

Obama, Barack. 2016. "United States Health Care Reform." *JAMA*, 27(6): 1718–1727.

Papageorge, Nicholas W, and Kevin Thom. 2020. "Genes, Education, and Labor Market Outcomes: Evidence from the Health and Retirement Study." *Journal of the European Economic Association*, 18(3): 1351–1399.

Pei, Zhuan, Jörn Steffen Pischke, and Hannes Schwandt. 2018. "Poorly Measured Confounders are More Useful on the Left than on the Right." *Journal of Business and Economic Statistics*, 1–12.

Pereira, Rita Dias. 2021. "Inborn Ability and Equality of Opportunity in Education." *Working Paper*.

Peterson, Roseann E, Karoline Kuchenbaecker, Raymond K Walters, Chia-Yen Chen, Alice B Popejoy, Sathish Periyasamy, Max Lam, Conrad Iyegbe, Rona J Strawbridge, Leslie Brick, Caitlin E Carey, Alicia R Martin, Jacquelyn L Meyers, Jinni Su, Junfang Chen, and Alexis C Edwards. 2019. "Leading Edge Primer Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations." *Cell*.

Plomin, Robert. 1990. *Nature and nurture: An introduction to human behavioral genetics.* Wadsworth Publishing Company.

Richards, Michael R, and Joachim Marti. 2014. "Heterogeneity in the smoking response to health shocks by out-of-pocket spending risk." *Health Economics, Policy and Law*, 9(04): 343–357.

Ritz, Beate R, Nilanjan Chatterjee, Montserrat Garcia-Closas, W James Gauderman, Brandon L Pierce, Peter Kraft, Caroline M Tanner, Leah E Mechanic, and Kimberly McAllister. 2017. "Lessons Learned From Past Gene-Environment Interaction Successes." *American Journal of Epidemiology*, 186(7): 778–786.

Schmitz, Lauren, and Dalton Conley. 2016. "The long-term consequences of Vietnam-era conscription and genotype on smoking behavior and health." *Behavior genetics*, 46(1): 43–58.

Schmitz, Lauren L, and Dalton C Conley. 2017*a*. "Modeling Gene-Environment Interactions With Quasi-Natural Experiments." *Journal of Personality*, 85(1): 10–21.

Schmitz, Lauren L, and Dalton C Conley. 2017*b*. "The effect of Vietnam-era conscription and genetic potential for educational attainment on schooling outcomes." *Economics of Education Review*, 61: 85–97.

Schork, Andrew J, M. Anthony Schork, and Nicholas J Schork. 2018. "Genetic risks and clinical rewards." *Nature Genetics*, 50(9): 1210–1211.

Slob, Eric A W, and Cornelius A Rietveld. 2020. "The moderating impact of the genetic predisposition to smoking behaviour on the response to tobacco excise taxes." *medRxiv*, 2020.12.02.20242388.

Smith, V. Kerry, Donald H. Taylor, Frank A. Sloan, F. Reed Johnson, and William H. Desvousges. 2001. "Do Smokers Respond to Health Shocks?" *Review of Economics and Statistics*, 83(4): 675–687.

Sonnega, Amanda, Jessica D Faul, M B Ofstedal, K M Langa, John W Phillips, and David R Weir. 2014. "Cohort Profile: the Health and Retirement Study (HRS)." *International Journal of Epidemiology*, 43(2): 576–585.

Stoker, Astrid K, and Athina Markou. 2013. "Unraveling the neurobiology of nicotine dependence using genetically engineered mice." *Current opinion in neurobiology*, 23(4): 493–499.

Sullivan, P F, and K S Kendler. 1999. "The genetic epidemiology of smoking." *Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco*, 1 Suppl 2: S51–7; discussion S69–70.

Sundmacher, Leonie. 2012. "The effect of health shocks on smoking and obesity." *The European Journal of Health Economics*, 13(4): 451–460.

Teo, Koon K, Stephanie Ounpuu, Steven Hawken, MR Pandey, Vicent Valentin, David Hunt, Rafael Diaz, Wafa Rashed, Rosario Freeman, Lixin Jiang, Xiaofei Zhang, Salim Yusuf, and INTERHEART Study Investigators. 2006. "Tobacco use and risk of myocardial infarction in 52 countries in the INTERHEART study: a case-control study." *The Lancet*, 368(9536): 647–658.

The Tobacco and Genetics Consortium, Helena Furberg, YunJung Kim, Jennifer Dackor, Eric Boerwinkle, Nora Franceschini, Diego Ardissino, Luisa Bernardinelli, Pier M Mannucci, Francesco Mauri, Piera A Merlini, Devin Absher, Themistocles L Assimes, Stephen P Fortmann, Carlos Iribarren, Joshua W Knowles, Thomas Quertermous, Luigi Ferrucci, Toshiko Tanaka, Joshua C Bis, Curt D Furberg, Talin Haritunians, Barbara McKnight, Bruce M Psaty, Kent D Taylor, Evan L Thacker, Peter Almgren, Leif C Groop, Claes Ladenvall, Michael Boehnke, Anne U Jackson, Karen L Mohlke, Heather M Stringham, Jaakko Tuomilehto, Emelia J Benjamin, Shih-Jen Hwang, Daniel Levy, Sarah Rosner Preis, Ramachandran S Vasan, Jubao Duan, Pablo V Gejman, Douglas F Levinson, Alan R Sanders, Jianxin Shi, Esther H Lips, James D McKay, Antonio Agudo, Luigi Barzan, Vladimir Bencko, Simone Benhamou, Xavier Castellsagué, Cristina Canova, David I Conway, Eleonora Fabianova, Lenka Foretova, Vladimir Janout, Claire M Healy, Ivana Holcátová, Kristina Kjaerheim, Pagona Lagiou, Jolanta Lissowska, Ray Lowry, Tatiana V Macfarlane, Dana Mates, Lorenzo Richiardi, Peter Rudnai, Neonilia Szeszenia-Dabrowska, David Zaridze, Ariana Znaor, Mark Lathrop, Paul Brennan, Stefania Bandinelli, Timothy M Frayling, Jack M Guralnik, Yusplitri Milaneschi, John R B Perry, David Altshuler, Roberto Elosua, Sekar Kathiresan, Gavin Lucas, Olle Melander, Christopher J O'Donnell, Veikko Salomaa, Stephen M Schwartz, Benjamin F Voight, Brenda W J H Penninx, Johannes H Smit, Nicole Vogelzangs, Dorret I Boomsma, Eco J C de Geus, Jacqueline M Vink, Gonneke Willemsen, Stephen J Chanock, Fangyi Gu, Susan E Hankinson, David J Hunter, Albert Hofman, Henning Tiemeier, Andre G Uitterlinden, Cornelia M van Duijn, Stefan Walter, Daniel I Chasman, Brendan M Everett, Guillaume Paré, Paul M Ridker, Ming D Li, Hermine H Maes, Janet Audrain-McGovern, Danielle Posthuma, Laura M Thornton, Caryn Lerman, Jaakko Kaprio, Jed E Rose, John P A Ioannidis, Peter Kraft, Dan-Yu Lin, and Patrick F Sullivan. 2010. "Genome-wide meta-analyses identify multiple loci associated with smoking behavior." *Nature Genetics*, 42(5): 441–447.

Thorgeirsson, Thorgeir E, Daniel F Gudbjartsson, Ida Surakka, Jacqueline M Vink, Najaf Amin, Frank Geller, Patrick Sulem, Thorunn Rafnar, Tonu Esko, Stefan Walter, Christian Gieger, Rajesh Rawal, Massimo Mangino, Inga Prokopenko, Reedik Mägi, Kaisu Keskitalo, Iris H Gudjonsdottir, Solveig Gretarsdottir, Hreinn Stefansson, John R Thompson, Yurii S Aulchenko, Mari Nelis, Katja K H Aben, Martin den Heijer, Asger Dirksen, Haseem Ashraf, Nicole Soranzo, Ana M Valdes, Claire Steves, Andre G Uitterlinden, Albert Hofman, Anke Tönjes, Peter Kovacs, Jouke Jan Hottenga, Gonneke Willemsen, Nicole Vogelzangs, Angela Döring, Norbert Dahmen, Barbara Nitz, Michele L Pergadia, Berta Saez, Veronica De Diego, Victoria Lezcano, Maria D Garcia-Prats, Samuli Ripatti, Markus Perola, Johannes Kettunen, Anna-Liisa Hartikainen, Anneli Pouta, Jaana Laitinen, Matti Isohanni, Shen Huei-Yi, Maxine Allen, Maria Krestyaninova, Alistair S Hall, Gregory T Jones, Andre M van Rij, Thomas Mueller, Benjamin Dieplinger, Meinhard Haltmayer, Steinn Jonsson, Stefan E Matthiasson, Hogni Oskarsson, Thorarinn Tyrfingsson, Lambertus A Kiemeney, Jose I Mayordomo, Jes S Lindholt, Jesper Holst Pedersen, Wilbur A Franklin, Holly Wolf, Grant W Montgomery, Andrew C Heath, Nicholas G Martin, Pamela A F Madden, Ina Giegling, Dan Rujescu, Marjo-Riitta Järvelin, Veikko Salomaa, Michael Stumvoll, Timothy D Spector, H-Erich Wichmann, Andres Metspalu, Nilesh J Samani, Brenda W J H Penninx, Ben A Oostra, Dorret I Boomsma, Henning Tiemeier, Cornelia M van Duijn, Jaakko Kaprio, Jeffrey R Gulcher, Mark I McCarthy, Leena Peltonen, Unnur Thorsteinsdottir, and Kari Stefansson. 2010. "Sequence variants at CHRNB3–CHRNA6 and CYP2A6 affect smoking behavior." *Nature Genetics*, 42(5): 448–453.

Thorgeirsson, Thorgeir E, Frank Geller, Patrick Sulem, Thorunn Rafnar, Anna Wiste, Kristinn P Magnusson, Andrei Manolescu, Gudmar Thorleifsson, Hreinn Stefansson, Andres Ingason, Simon N Stacey, Jon T Bergthorsson, Steinunn Thorlacius, Julius Gudmundsson, Thorlakur Jonsson, Margret Jakobsdottir, Jona Saemundsdottir, Olof Olafsdottir, Larus J Gudmundsson, Gyda Bjornsdottir, Kristleifur Kristjansson, Halla Skuladottir, Helgi J Isaksson, Tomas Gudbjartsson, Gregory T Jones, Thomas Mueller, Anders Gottsäter, Andrea Flex, Katja K H Aben, Femmie de Vegt, Peter F A Mulders, Dolores Isla, Maria J Vidal, Laura Asin, Berta Saez, Laura Murillo, Thorsteinn Blondal, Halldor Kolbeinsson, Jon G Stefansson, Ingunn Hansdottir, Valgerdur Runarsdottir, Roberto Pola, Bengt Lindblad, Andre M van Rij, Benjamin Dieplinger, Meinhard Haltmayer, Jose I Mayordomo, Lambertus A Kiemeney, Stefan E Matthiasson, Hogni Oskarsson, Thorarinn Tyrfingsson, Daniel F Gudbjartsson, Jeffrey R Gulcher, Steinn Jonsson, Unnur Thorsteinsdottir, Augustine Kong, and Kari Stefansson. 2008. "A variant associated with nicotine dependence, lung cancer and peripheral arterial disease." *Nature*, 452(7187): 638–642.

Thorpe, K. E., Curtis S Florence, and Peter Joski. 2004. "Which Medical Conditions Account For The Rise In Health Care Spending?" *Health Affairs*, Suppl Web: W4–437–45.

Torkamani, Ali, Nathan E Wineinger, and Eric J Topol. 2018. "The personal and clinical utility of polygenic risk scores." *Nature Reviews Genetics*, 1.

United States Department of Health and Human Services. 2014. "The Health Consequences of Smoking–50 Years of Progress A Report of the Surgeon General." *The Health Consequences of Smoking-50 Years of Progress A Report of The Surgeon General*.

Ware, E, L Schmitz, and J Faul. 2017. "HRS documentation report – HRS polygenic scores, 2006-2010 genetic data."

Wedow, Robbee, Meghan Zacher, Brooke M Huibregtse, Kathleen Mullan Harris, Benjamin W Domingue, and Jason D Boardman. 2018. "Education, Smoking, and Cohort Change: Forwarding a Multidimensional Theory of the Environmental Moderation of Genetic Effects." *American Sociological Review*, 83(4): 802–832.

Wray, L A, A R Herzog, R J Willis, and R B Wallace. 1998. "The impact of education and heart attack on smoking cessation among middle-aged adults." *Journal of health and social behavior*, 39(4): 271–94.

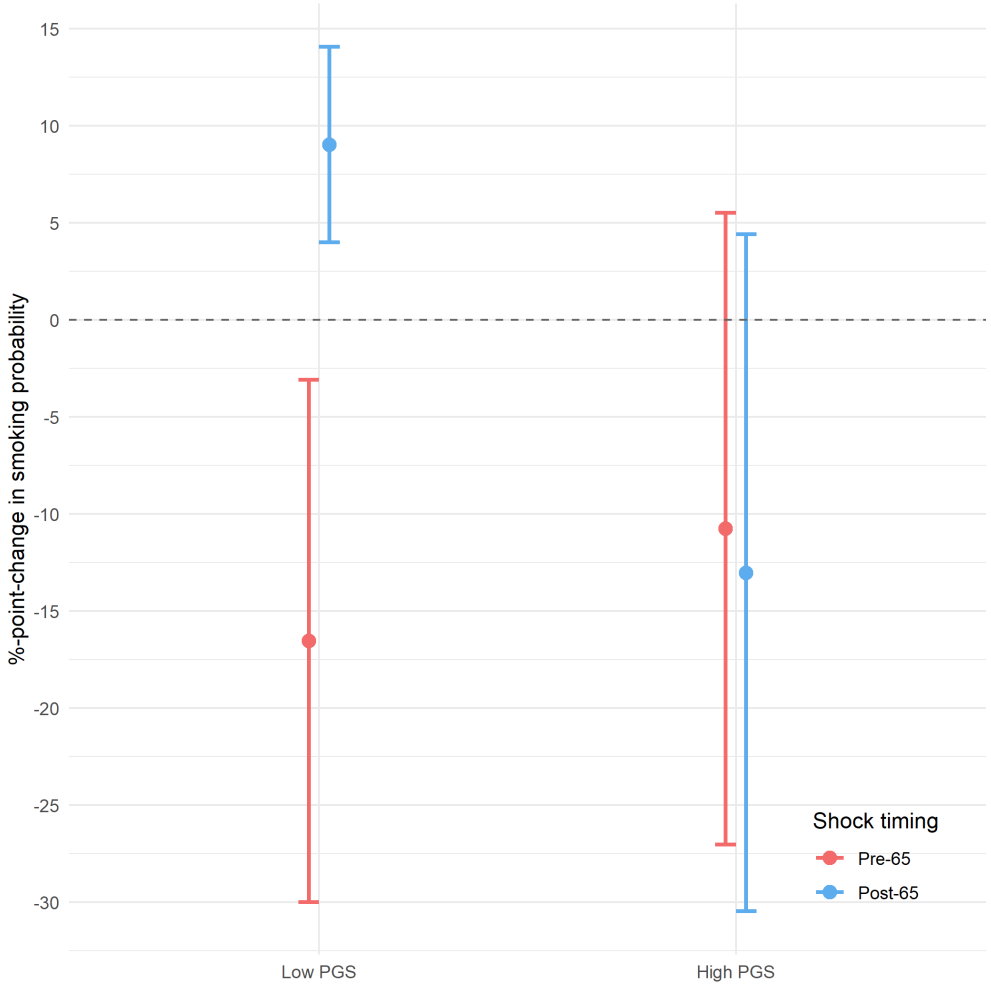
Xu, Xin, Ellen E. Bishop, Sara M. Kennedy, Sean A. Simpson, and Terry F. Pechacek. 2015. "Annual Healthcare Spending Attributable to Cigarette Smoking." *American Journal of Preventive Medicine*, 48(3): 326–333.

Yengo, Loic, Julia Sidorenko, Kathryn E Kemper, Zhili Zheng, Andrew R Wood, Michael N Weedon, Timothy M Frayling, Joel Hirschhorn, Jian Yang, Peter M Visscher, and GIANT Consortium. 2018. "Meta-analysis of genome-wide association studies for height and body mass index in ˜700,000 individuals of European ancestry." *bioRxiv*, 274654.

Zweifel, Peter, and Willard G. Manning. 2000. "Moral Hazard and Consumer Incentives in Health Care." *Handbook of Health Economics*, 1: 409–459.

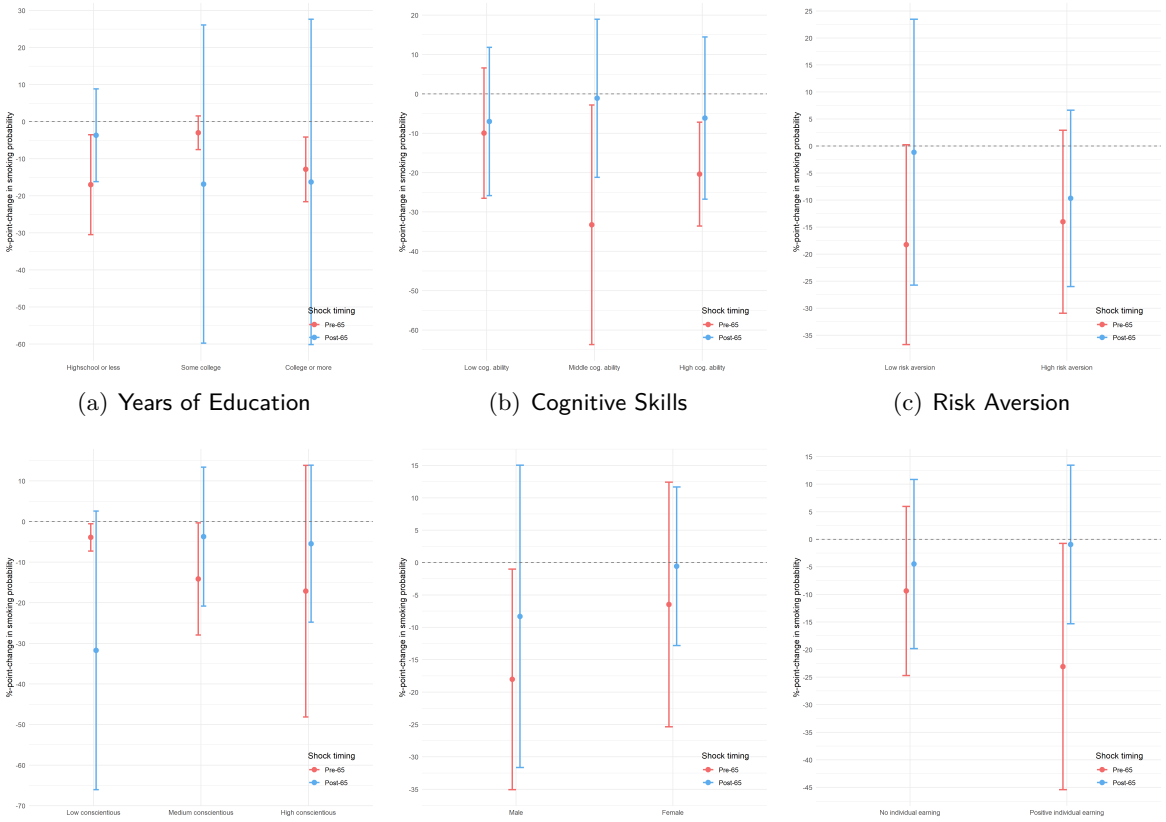
6 Figures

Figure 1: Effect of a Health Shock on the Smoking Probability in the Pre-65 Uninsured Subgroup, Stratified by Timing of the Shock and Genetic Type



Notes: Low PGS = lowest tertile of the polygenic score distribution; high PGS = upper two tertiles of the polygenic score distribution. Pre-65: Health shock since the last survey reported at ages 60-64. Post-65: Health shock since the last survey reported at ages 67-70. Estimates and standard errors are shown in Panel A of Table 2. Effects are estimated using the coefficients in the last column of Table 3 and following the derivation described in C.5. Bars show 95% confidence intervals, standard error clustered at the individual level.
Data used: HRS study sample, n = 5,854.

Figure 2: Other Individual Characteristics as Proxy for Moral Hazard Heterogeneity



(a) Years of Education (b) Cognitive Skills (c) Risk Aversion

(d) Conscientiousness (Big-5) (e) Gender (f) Earnings

Notes: The figures report the effect of suffering a health shock on the probability of smoking for the pre-65 uninsured subgroup, stratified by timing of the shock (before and after the age of 65) and different measures of individual characteristics. Effects are estimated using a combination of the coefficients from equation 2 where g_i is replaced by the different individual characteristics reported in the sub-figure captions, following the derivation described in C.5. Pre-65: Health shock since the last survey reported at ages 60-64. Post-65: Health shock since the last survey reported at ages 67-70. Bars show 95% confidence intervals, standard error clustered at the individual level.
Data used: HRS study sample, n = 5,854.

7 Tables

Table 1: Descriptive Statistics for the Study Sample, Stratified by Genetic Group

	All	**Low PGS**	**High PGS**
	Mean (SD)	Mean (SD)	Mean (SD)
Age (baseline)	61.17 (1.93)	61.18 (1.96)	61.16 (1.92)
Smoking PGS	0.11 (0.99)	-0.96 (0.51)	0.64 (0.71)
No. waves present	4.44 (1.38)	4.46 (1.37)	4.43 (1.39)
	%	%	%
Female	50.42	46.97	52.14
Smoking (baseline)	29.55	27.06	30.79
Persistently uninsured	5.85	5.39	6.08
CV health shock	12.44	11.82	12.74
No. of individuals	5813	1929	3884
No. Person-year observations	25800	8602	17198

Notes: Low PGS = lowest tertile of the polygenic score distribution; high PGS = upper two tertiles of the polygenic score distribution. Baseline = first wave used in the study sample for the empirical analysis. Smoking PGS = polygenic score for smoking constructed according to equation 1 used as a proxy for genetic propensity to smoke. Persistently uninsured = binary indicator for being uninsured in every observation of the study sample before the age of 65. CV health shock = binary indicator for having suffered a cardiovascular health shock for the first time since the previous wave. See section 2.2 for a thorough definition of the variables used and Section 2.1 for a definition of the study sample.
Data used: HRS study sample, n = 5,854.

Table 2: Summary of Statistical Results for the Pre-65 Uninsured Subgroup, Stratified by Timing of the Shock and Genetic Group

Effect of health shock on smoking probability		
	Low PGS	High PGS
Pre 65	-0.165	-0.108
	(0.069)	(0.083)
Post 65	0.09	-0.13
	(0.026)	(0.089)
Effect of health insurance on effect of health shock		
	Low PGS	High PGS
Post 65 - Pre 65	0.256	-0.023
	(0.079)	(0.121)
Differential effect of health insurance by genetic group		
	High PGS	- low PGS
Post 65 - Pre 65	-0.279	
	(0.144)	

Low PGS = lowest tertile of the polygenic score distribution; high PGS = upper two tertiles of the polygenic score distribution. Pre-65: Health shock since the last survey reported at ages 60-64. Post-65: Health shock since the last survey reported at ages 67-70. *P $<$ 0.1; **P $<$ 0.05; ***P $<$ 0.01. Robust standard errors in parentheses are clustered at the individual level. The covariance matrix used for calculating standard errors is shown in Appendix Table 7. Effects are estimated using the coefficients in the last column of Table 3 and following the derivation described in C.5.
Data used: HRS study sample, n = 5,854.

Table 3: Coefficients from estimating the linear probability model in equation (2) using OLS

	Dependent variable:				
	Smoking status				
	(1)	(2)	(3)	(4)	(5)
Health Shock	−0.027	−0.028	−0.027	−0.046	−0.045
	(0.040)	(0.040)	(0.040)	(0.024)	(0.024)
Post-65	−0.064	−0.020	−0.020	−0.010	−0.009
	(0.008)	(0.012)	(0.012)	(0.008)	(0.008)
Uninsured	0.169	0.169	0.169		
	(0.027)	(0.027)	(0.027)		
High PGS	0.033	0.033	0.033		
	(0.012)	(0.012)	(0.012)		
Shock × Post-65	0.008	0.018	0.016	0.026	0.025
	(0.056)	(0.056)	(0.056)	(0.032)	(0.032)
Shock × Uninsured	−0.193	−0.187	−0.188	−0.118	−0.120
	(0.184)	(0.185)	(0.184)	(0.073)	(0.073)
Post-65 × Uninsured	0.068	0.067	0.067	−0.052	−0.052
	(0.048)	(0.048)	(0.048)	(0.031)	(0.031)
Shock × High PGS	0.028	0.029	0.026	0.015	0.015
	(0.049)	(0.049)	(0.049)	(0.029)	(0.029)
Post-65 × High PGS	−0.0003	−0.0005	−0.0005	−0.005	−0.005
	(0.010)	(0.010)	(0.010)	(0.008)	(0.008)
Shock × Post-65 x Uninsured	0.343	0.336	0.336	0.230	0.230
	(0.256)	(0.256)	(0.255)	(0.086)	(0.085)
Shock × Uninsured × High PGS	0.184	0.182	0.183	0.041	0.042
	(0.219)	(0.220)	(0.219)	(0.112)	(0.112)
Shock × Post-65 × High PGS	−0.047	−0.047	−0.045	−0.078	−0.078
	(0.068)	(0.068)	(0.068)	(0.042)	(0.042)
Post-65 × Uninsured × High PGS	−0.116	−0.117	−0.117	0.042	0.042
	(0.061)	(0.062)	(0.062)	(0.036)	(0.036)
Shock × Post-65 × Uninsured × High PGS	−0.279	−0.278	−0.278	−0.199	−0.200
	(0.311)	(0.312)	(0.311)	(0.151)	(0.150)
Age		*Yes*	*Yes*	*Yes*	*Yes*
Year FE			*Yes*		*Yes*
Individual FE				*Yes*	*Yes*
Observations	26,022	26,022	26,022	26,022	26,022
R^2	0.015	0.017	0.017	0.823	0.823

Notes: Health shock = binary indicator for having suffered a cardiovascular health shock for the first time since the previous wave. Post-65 = indicator for age > 65 at the time of the interview. Uninsured = binary indicator for being persistently uninsured in every observation of the study sample before the age of 65. High PGS = indicator for being in the upper two tertiles of the polygenic score distribution. Age = controls for 3^{rd} degree polynomial in age. FE = adding fixed effects. Robust standard errors in parentheses are clustered at the individual level. *p<0.1; **p<0.05; ***p<0.01.
Data used: HRS study sample, n = 5,854.

A Appendix for Online Publication

Contents

B Genetics for Economists

The human genome consists of over 3 billion base pairs (6 billion bases) in each cell nucleus, with four possible bases: adenine (A), thymine (T), guanine (G), and cytosine (C).[15] Comparing any two unrelated human beings, over 99% of their genome will be identical. The remaining <1% differs between individuals, with a Single Nucleotide Polymorphism (or SNP, pronounced *snip*) being the most common form of genetic variation. A SNP is a single base-pair substitution at a particular location (locus) on the human genome.

To identify genetic variants that are associated with a particular trait of interest, such as coronary heart disease, so-called Genome-Wide Association Studies (GWAS) relate each SNP to the trait in a hypothesis free-approach. Stringent p-values are then used to identify SNPs that are robustly associated with the trait of interest, and replication is performed in other, independent samples. Only SNPs that have consistent associations across the different samples are interpreted as robust. However, this does not guarantee that individual SNPs have large effect sizes. Most human complex traits are polygenic, meaning that they are affected by many SNPs, each with a very small effect size. To increase the predictive power of SNPs, they can then be aggregated into so-called polygenic scores (PGS), defined as:

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

where G_{ij} is SNP j of individual i, and β_j is the effect size of that SNP, obtained from an independent Genome-Wide Association Study. GWAS sample sizes have grown substantially in recent years, meaning that (a) SNPs with very small effects are more likely to be identified, and (b) that the effect sizes are estimated with increased precision. Indeed, we have seen large improvements in genetic prediction, with initial PGS being able to explain less than 1% of the variation in the trait of interest, to more recent ones explaining up to 11-13% of the variation in educational attainment just by increasing the sample size of the discovery sample (see e.g., Lee et al., 2018). PGS have been shown to be powerful tools to identify patients with increased risk for coronary artery disease, atrial fibrillation, type 2 diabetes, inflammatory bowel disease, and breast cancer. Indeed, Khera et al. (2018) propose the use of polygenic prediction in clinical care.

C Methods

C.1 DNA Extraction and Genotyping

In 2006, the Health and Retirement Study (HRS) introduced enhanced face-to-face interviews (EFTFs), which expanded the core interview with measures of physical function, blood-based biomarkers, and DNA samples. Sample selection for the EFTFs was conducted as follows: A random 50% of the 2006 sample was preselected for an EFTF, and the other half was selected in 2008. A new cohort of households was added to the HRS in 2010. Of these new households, a random 50% was selected for EFTF data collection in 2010, while the other half was selected in 2012. The households selected

[15] A base *pair* is set of two bases, with A always pairing with T, and C always pairing with G.

for EFTFs in 2012 are not yet included in the polygenic scores data used in this study. In 2006, saliva collection was conducted using a mouthwash collection method. From 2008 onwards, the Oragene DNA Collection Kit (OG-250) was used.

Genotype data was obtained for over 15,000 HRS participants. Genotyping was conducted by the Center for Inherited Disease Research (CIDR) in 2011, 2012, and 2015, using the Illumina HumanOmni 2.5 BeadChips (HumanOmni2.4-4v1, HumanOmni2.5-8v1). Approximately 2.4 million single nucleotide polymorphisms (SNPs) were measured. Of the roughly 1.9 million genotyped SNPs that passed quality control, 21 million SNPs were imputed using the 1000 Genomes Reference Panels (phase 3, version 5). More details on genotyping and imputation can be found in the official HRS Documentation Report.

C.2 Study Sample

To improve replicability of the results, we mostly use data from the publicly available RAND HRS file (version P)(, n.d.*a*)—an easy-to-use longitudinal data set based on the HRS data—as well as the publicly available initial release of the HRS polygenic scores data (, n.d.*b*).

C.2.1 Reshaping, Merging, and Sample Restrictions

This section describes how the study sample was constructed from the RAND HRS version P data file. First, the data file was reshaped from wide to long format, with each observation corresponding to a respondent-wave entry. Second, polygenic risk scores (PGSs) for the HRS phenotype "smoking initiation" (referred to as "regular smoking" in this study, for clarity purposes) from the initial HRS PGS data release, using genetic data from 2006 to 2010, were merged for the 9,991 genotyped individuals of European ancestry. The following shows the list of restrictions that were then imposed to arrive at the study sample used in the main analysis. Regarding notation, note that `VARIABLE` refers to the long-format version of the variables that were called `R1VARIABLE` to `R12VARIABLE` in the RAND HRS data file. From the reshaped data file, the study sample was reached by carrying out the following steps (in this exact order):

1. Drop observations with an age (`AGEY_E`; see Section C.3.3) below 60 or above 70 years
2. Drop individuals with only 1 observation
3. Drop observations with missing values for the PGS for smoking (`PGS_EvrSmk_TAG10`; see Section C.3.5), the self-constructed health shock indicator (Section C.3.2), or the current smoking status (`SMOKEN`; see Section C.3.1)
4. Drop individuals who in their first observation (the baseline) reported never having smoked (`SMOKEV` equal to 0; see Section C.2.2)
5. Drop individuals with missing values for the self-constructed pre-65 uninsured status indicator (Section C.3.4)
6. Drop individuals who reported a health shock (self-constructed health shock indicator equal to 1; see Section C.3.2) when interviewed at ages 65 or 66

C.2.2 Ever-Smoker Status

"Ever-smoker" refers to the RAND variable `SMOKEV`, which indicates whether the respondent has ever smoked cigarettes. Ever smoking means having smoked more than 100 cigarettes throughout one's life, not including pipes or cigars. This is consistent with the Centers for Disease Control classification of the term "ever-smoker."Jamal et al. (2016) The ever-smoked question was usually only asked at the respondent's first interview and then carried forward for subsequent waves. For details on the survey questions and recodings for missings into yes/no answers, see the publicly available official RAND HRS documentation.

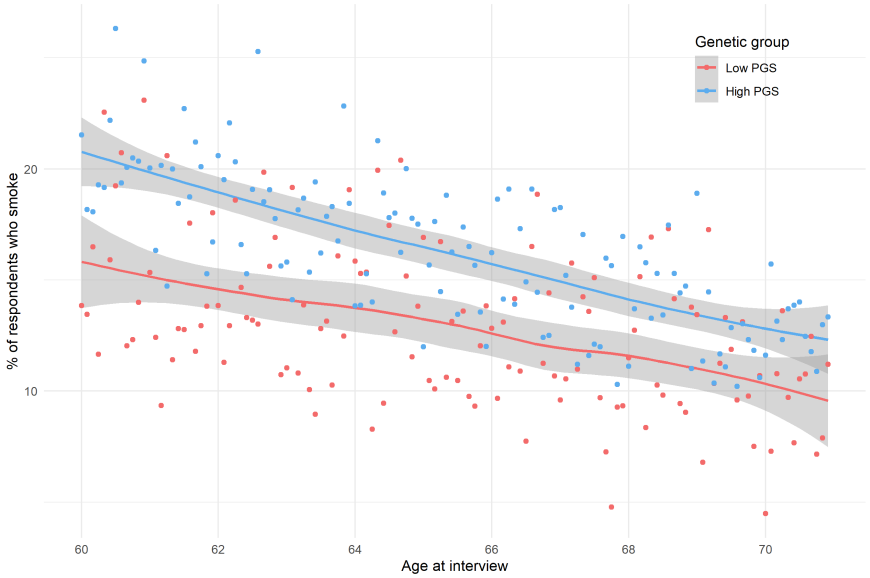
C.3 Outcome and Exposure Variables

C.3.1 Smoking Status

Current smoking status refers to the RAND variable `SMOKEN`, which indicates whether the respondent smokes at the time of the interview. The survey question about current smoking status was only asked for respondents who answered yes to being ever-smokers (having smoked more than 100 cigarettes in their lifetime). For details on the survey questions and recodings for missings into yes/no answers, see the RAND HRS documentation.

As shown in Figure 3, the prevalence of smokers in the sample decreases with age and is about 5 percentage points higher for the individuals with a high PGS.

Figure 3: Prevalence of smoking.



Notes: Self-reported smoking rate by age and high or low PGS. Bin-scattered plot and generalized linear smoothed correlation between age and smoking shown in red (low PGS = lowest tertile of polygenic score) and blue (high PGS = upper two tertiles of polygenic score).
Data used: HRS waves 1-13, restricted to observations with age between 60 and 70 years.

C.3.2 Health Shocks

The health shock indicator was defined using the RAND variables HEARTE and STROKE. The variable HEARTE indicates whether or not a doctor has ever told the respondent that he/she had 1 of the following conditions:

1. Heart attack
2. Coronary heart disease
3. Angina
4. Congestive heart failure
5. Other heart problems

The variable STROKE indicates whether or not a doctor has ever told the respondent that he/she had 1 of the following conditions:

1. Stroke
2. Transient ischemic attack

For details on the survey questions and the construction of these variables, see the RAND HRS documentation.

The health shock indicator used in this analysis was then defined as follows: It was set to 1 in a given wave if the lagged values of both HEARTE and STROKE were equal to 0, and if either one of the current values (or both) was equal to 1.

Figure 4 plots the share of people reporting a health shock since the last HRS survey over the different ages.

C.3.3 Medicare Eligibility Status

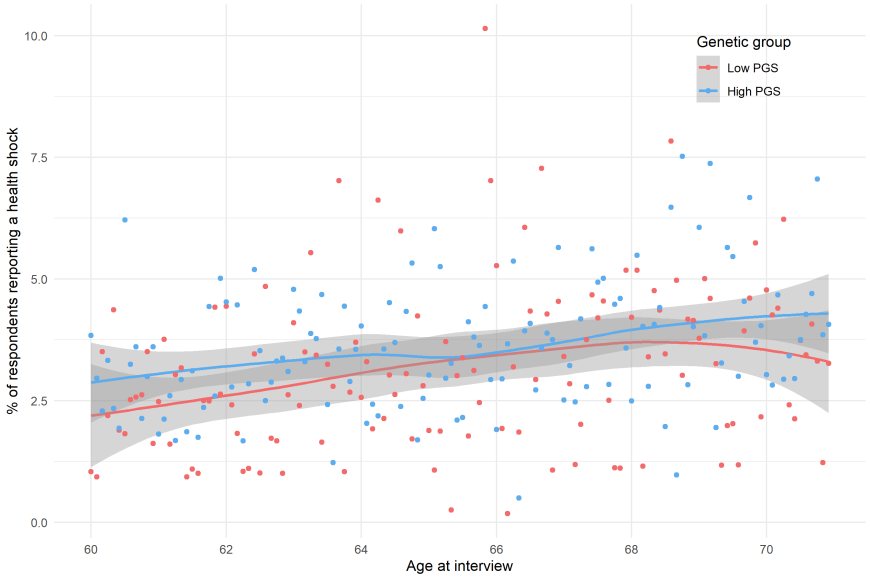
The Medicare eligibility status indicator was defined using the RAND variable AGEY_E. This variable indicates the respondent's age in years at the end of the HRS interview in a given wave. For details on the construction of this variable, see the RAND HRS documentation.

C.3.4 Pre-65 Uninsured Status

The pre-65 uninsured status indicator was defined using the RAND variables HIGOV, COVR, COVS, and HIOTHP. The variable HIGOV indicates whether the respondent was covered by any government health insurance program. COVR indicates whether the respondent was covered by health insurance from his/her current or previous employer. COVS indicates whether the respondent was covered by his/her spouse's employer. HIOTHP indicates whether the respondent was covered by any health insurance other than the government, employer-provided, or long-term care insurance. For details on the survey questions and the construction of these variables, see the RAND HRS documentation.

The pre-65 uninsured status indicator used in this analysis was then defined as follows: First, we defined a wave-specific uninsured indicator, which was set to 1 in a given wave if the values of HIGOV, COVR, COVS, and HIOTHP were all equal to 0. Then, the pre-65 uninsured status indicator was set to 1 for respondents whose wave-specific

Figure 4: Prevalence of cardiovascular health shock.



Notes: Self-reported indicator of having been diagnosed for the first time with a cardiovascular condition since the last HRS survey. Age refers to the time of the survey, not the time of the health shock, which is unknown up to a 2-year windows, since HRS surveys are bi-annual. Bin-scattered plot and generalized linear smoothed correlation between age at interview and cardiovascular health shock shown in red (low PGS = lowest tertile of polygenic score) and blue (high PGS = upper two tertiles of polygenic score).
Data used: HRS waves 1-13, restricted to observations with age between 60 and 70 years.

uninsured indicator was 1 in 100% of all pre-65 observations (i.e., where `AGEY_E` < 65).

Panel A of Figure 5 shows the share of persistently uninsured individuals. As shown in Panel B of Figure 5, the vast majority of individuals are aware of Medicare eligibility: more than 90% of individuals report being covered by Medicare after the age of 65.

Panels (a), (b), and (c) of Figure 6 show that there is no apparent jump in income or the probability of retirement exactly at the age of 65 in our sample.

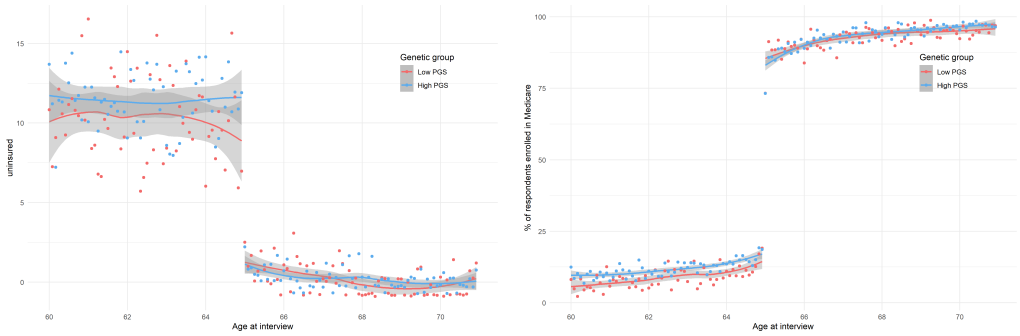
C.3.5 Polygenic Score for Regular Smoking

The high-PGS indicator was defined using the PGS for the phenotype "smoking initiation" (referred to as "regular smoking" in this study, for clarity purposes) using **all of the genetic variants** (SNPs).

The PGS was calculated using the effect sizes estimated in a genome-wide association study (GWAS) meta-analysis Liu et al. (2019) conducted by the GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN). The phenotype "smoking initiation" studied in this GWAS was defined as ever versus never having been a regular smoker, where regular smokers were individuals who reported having smoked ≥ 100 cigarettes throughout their life.

The PGSs used these GWAS-estimated effect sizes for all SNPs that overlapped between the HRS genetic database and the GWAS meta-analysis, without accounting for linkage disequilibrium between SNPs or considering P-value thresholds. Scores were calculated according to Equation (1) in the main text using the software packages PRSice and PLINK.

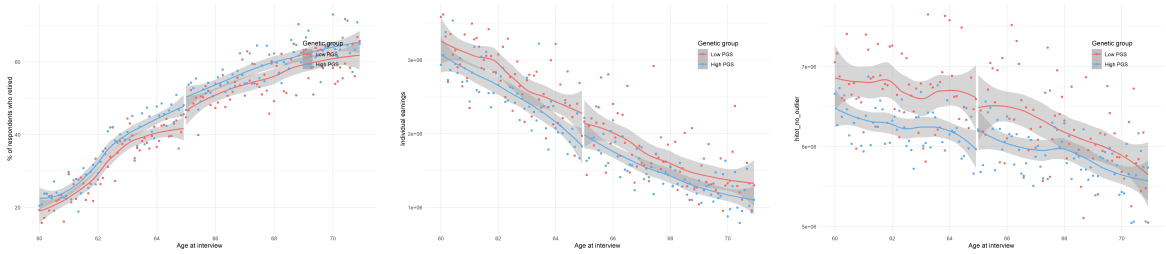
Figure 5: Share uninsured and enrolled in Medicare across the age 65 threshold by PGS.



(a) Share uninsured (b) Share Enrolled in Medicare

Notes: Self-reported indicator for lack of any insurance coverage and being enrolled in Medicare. Bin-scattered plot and generalized linear smoothed correlation between age and outcome variables shown in red (low PGS = lowest tertile of polygenic score) and blue (high PGS = upper two tertiles of polygenic score).
Data used: HRS waves 1-13, restricted to observations with age between 60 and 70 years.

Figure 6: Retirement and income across the age 65 threshold by PGS.



(a) Share retired (b) Earned Income (c) Total Household Income

Notes: Self-reported retirement status, earned income, and total-household income over time. Bin-scattered plot and generalized linear smoothed correlation between age and outcome variables shown in red (low PGS = lowest tertile of polygenic score) and blue (high PGS = upper two tertiles of polygenic score).
Data used: HRS waves 1-13, restricted to observations with age between 60 and 70 years.

The high-PGS indicator used in this analysis was then defined as follows: It was set to 1 for individuals with a PGS above the lowest tertile, and to 0 for individuals with a PGS below or equal to the lowest tertile. The two upper-tertiles of the PGS distribution were combined to improve statistical power and simplify the exposition. Initial results using an indicator for each tertile of the distribution, displayed in Appendix Section C.5.2, show that the results for the two upper-tertiles of the PGS distribution are very similar to each other.

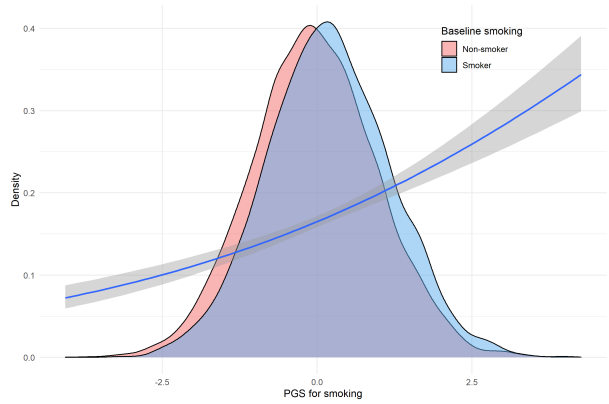
The polygenic score is predictive of smoking behavior, as expected and displayed in Figure 8, but not only. As shown in Appendix Figure 7—which displays the coefficients of simple OLS regressions of several outcomes on the linear PGS controlling for age, gender, and the 10 principal components of the genomewide matrix—the PGS is also predictive of other unfavorable outcomes: younger age at first birth as well as age started smoking, lower cognitive skills, perseverance, years of education, wealth, income, health rating, higher depressive symptoms, anxiety, non-cancer illnesses, drinking behavior. Reassuringly, the correlations with retirement, medications taken, and mortality are positive but extremely small and not distinguishable from zero.

Figure 7: PGS distribution and correlation with smoking behavior.

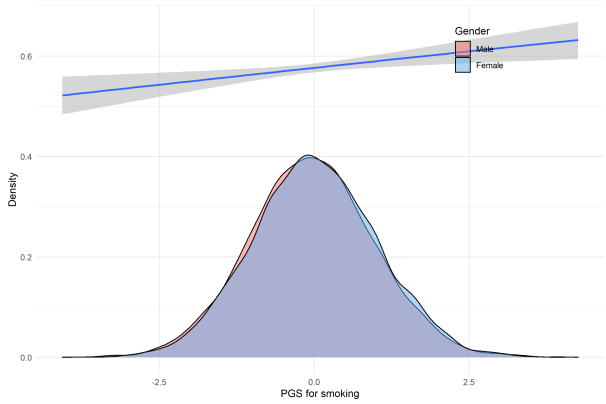
Notes: Plot of estimated coefficients associated with the PGS (entered linearly) from several OLS regressions of the different outcomes displayed on the y-axis on the PGS, age, age squared, age cubed, sex, and the first 10 principal components of the genomewide matrix.
Data used: HRS waves 1-13, restricted to observations with age between 60 and 70 years.

As shown in Figure 8, the PGS is mildly correlated with gender and almost uncorrelated with the probability of suffering from a health shock.

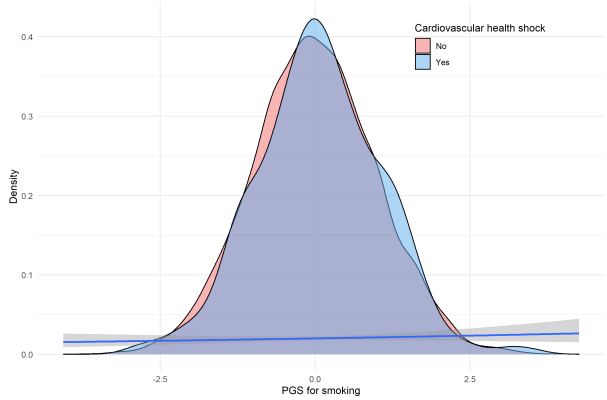
Figure 8: Distribution of PGS for Regular Smoking and its relation to the main analysis varaibles in the HRS Data



(a) PGS distribution and correlation with smoking behavior.



(b) PGS distribution and correlation with female gender.



(c) PGS distribution and correlation with the health shock.

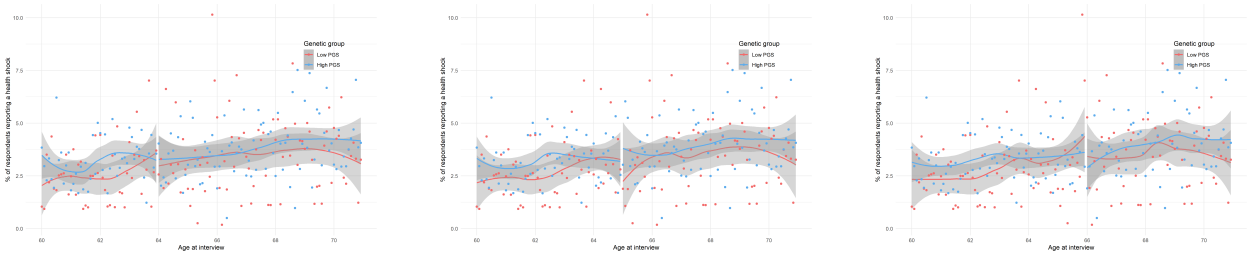
Notes: Distribution of Polygenic Score (PGS) for baseline smokers (blue) and non-smokers (red). Generalized linear smoothed correlation between current smoking (panel a), gender (panel b), and health shock (panel b) and the PGS shown in the blue line (with 95% confidence intervals in shaded grey area).
Data used: HRS waves 1-13, restricted to observations with age between 60 and 70 years.

C.4 Statistical Analysis

C.4.1 Age Pattern of Health Shock Incidence

With the data used in this study, it was not possible to narrow down the exact timing of a health shock to more than the between-survey 2-year window. Therefore, the probability of having a health shock at a specific age could not be determined. What could be said from this data about the age at the health shock is that for all shocks reported at ages 64 or below, the shocks must have occurred before the age of 65. Similarly, for all health shocks reported at ages 67 or above, the shocks must have occurred after the age of 65. For shocks reported at interview ages 65 or 66, it could not be determined whether the shock occurred before or after age 65 (as interviews were conducted biennially).

Figure 9: Percentage of Reported Health Shocks by Age



(a) Cutoff at age 64 (b) Cutoff at age 65 (c) Cutoff at age 66

Self-reported indicator of having been diagnosed for the first time with a cardiovascular condition since the last HRS survey. Age refers to the time of the survey, not the time of the health shock, which is unknown up to a 2-year windows, since HRS surveys are bi annual. Bin-scattered plot and generalized linear smoothed correlation between age at interview and cardiovascular health shock shown in red (low PGS) and blue (high PGS). Linear smoothed correlation estimated separately to the left and to the right of age cutoff: (a) age 64; (b) age 65; (c) age 66.
Data used: HRS waves 1-13, restricted to observations with age between 60 and 70 years.

Figure 9 visualizes the fraction of HRS respondents who reported experiencing a health shock since the last survey wave at a given interview age, stratified both by genetic group and by gender. By visual inspection, there seems to be a positive trend in the fraction of respondents reporting health events with age, with frequent deviations but no obvious jump between 64 and 67.

To formally test for a jump or change in trend in the incidence of health shocks around the age of 65, a segmented regression approach was used. Specifically, we tested if the change between the percentage of respondents who reported a health shock at age 64 and the percentage of respondents who reported a health shock at age 67 was larger than what could be explained by a linear age trend. For this test, all observations where respondents were aged 65 or 66 were excluded. For all remaining observations, the binary health shock indicator ($shock$) was regressed on the age variable (age), a post-age-67 indicator variable ($post67$), and a post-age-67 trend ($post67slope$):

$$shock_{it} = \beta_0 + \beta_1\, age_{it} + \beta_2\, post67_{it} + \beta_3\, post67slope_{it} + \varepsilon_{it} \tag{3}$$

The post-age-67 indicator variable was defined to take the value 1 if a respondent was aged 67 or older at the time of the HRS interview. Therefore, it guaranteed that any potential health shocks were experienced after the age of 65. The post-age-67 slope variable was a continuous variable coded 0 up to and including age 67, and increased sequentially from 1 thereafter. β_1 captured the general age trend in the probability of reporting a health shock; β_2 estimated the jump in the report of health shocks at age 67; β_3 reflected changes in the age trend of reported health shocks after age 67.

The linear probability model in Equation (3) was estimated using ordinary least squares (OLS) regression for (i) the study sample (HRS waves 1-12, restricted to observations with age between 60 and 70 years, and additionally excluding all observations with ages 65 or 66), (ii) both genetic groups separately, and (iii) both men and women separately. Estimation results are shown in Table 4. Across all groups, there were no

statistically significant jumps for health shocks reported at age 67 compared to age 64 (accounting for a linear age trend). Similarly, the age trend was not significantly different after age 67 than before. In the study sample and in the low-PGS group, the increase with age in the probability of reporting a health shock was statistically significant.

Table 4: Coefficients from Estimating the Linear Probability Model in Equation (3) Using OLS

	Dependent variable:				
	Probability of shock				
	All	Low PGS	High PGS	Male	Female
	(1)	(2)	(3)	(4)	(5)
Age	0.002*	0.003*	0.001	0.001	0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Post-67 dummy	−0.001	−0.007	0.002	−0.001	−0.001
	(0.005)	(0.008)	(0.006)	(0.008)	(0.006)
Post-67 slope	−0.0005	−0.002	0.0003	0.001	−0.002
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Constant	−0.068	−0.134	−0.036	−0.046	−0.082
	(0.049)	(0.082)	(0.060)	(0.079)	(0.061)
Observations	39,185	13,076	26,109	16,648	22,537
Note:					*p<0.1; **p<0.05; ***p<0.01

Low PGS = lowest tertile of the polygenic score distribution; high PGS = upper two tertiles of the polygenic score distribution.
Data used: HRS waves 1-13, restricted to observations with age between 60 and 70 years and non-missing smoking status, and additionally excluding all observations with ages 65 or 66.

C.5 Derivation of the effects from the OLS coefficients

C.5.1 OLS estimation: high and low PGS

Current smoking status (Y) is regressed on the full set of interactions between the indicators for the health shock (*shock*), being uninsured pre-65 (*uninsured*), Medicare eligibility (*post*65), and high polygenic risk for smoking (g):

$$\begin{aligned} Y_{it} \;=\; & \beta\, shock_{it} + \gamma\, post65_{it} \\ & + \lambda_1 \left(shock_{it} \times post65_{it}\right) \\ & + \lambda_2 \left(shock_{it} \times uninsured_i\right) \\ & + \lambda_3 \left(post65_{it} \times uninsured_i\right) \\ & + \lambda_4 \left(shock_{it} \times g_i\right) \\ & + \lambda_5 \left(post65_{it} \times g_i\right) \\ & + \delta_1 \left(shock_{it} \times post65_{it} \times uninsured_i\right) \\ & + \delta_2 \left(shock_{it} \times uninsured_i \times g_i\right) \\ & + \delta_3 \left(shock_{it} \times post65_{it} \times g_i\right) \\ & + \delta_4 \left(post65_{it} \times uninsured_i \times g_i\right) \\ & + \zeta \left(shock_{it} \times post65_{it} \times uninsured_i \times g_i\right) \\ & + \sum_{a=1}^{3} \phi_a\, age_{it}^a + \eta_i + \tau_t + \varepsilon_{it} \end{aligned}$$

Effect of the health shock on the outcome for the previously uninsured To calculate the effect of the shock on the outcome, we evaluate the derivative of the outcome with respect to shock:

$$\begin{aligned} \frac{\partial Y_{it}}{\partial shock_{it}} = & \beta + \lambda_1 post65_{it} + \lambda_2 uninsured_i + \lambda_4 g_i \\ & + \delta_1(post65 \times uninsured_i) + \delta_2(uninsured_i \times g_i) + \delta_3(post65_{it} \times g_i) \\ & + \zeta(post65_{it} \times uninsured_i \times g_i) \end{aligned} \tag{4}$$

We can look at the decomposition for the different genotypes (high and low PGS) and shock timing (before and after 65):

$$E\left[\frac{\partial Y_{it}}{\partial shock_{it}} | post65_{it} = 0, g_i = 0, uninsured_i = 1\right] = \beta + \lambda_2 \tag{5}$$

$$E\left[\frac{\partial Y_{it}}{\partial shock_{it}} | post65_{it} = 1, g_i = 0, uninsured_i = 1\right] = \beta + \lambda_1 + \lambda_2 + \delta_1 \tag{6}$$

$$E\left[\frac{\partial Y_{it}}{\partial shock_{it}} | post65_{it} = 0, g_i = 1, uninsured_i = 1\right] = \beta + \lambda_2 + \lambda_4 + \delta_2 \tag{7}$$

$$\begin{aligned} E\left[\frac{\partial Y_{it}}{\partial shock_{it}} | post65_{it} = 1, g_i = 1, uninsured_i = 1\right] = & \beta + \lambda_1 + \lambda_2 + \lambda_4 \\ & + \delta_1 + \delta_2 + \delta_3 + \zeta \end{aligned} \tag{8}$$

Calculating the first two differences as above::

$$(6) - (5) = \lambda_1 + \delta_1 \tag{9}$$
$$(8) - (7) = \lambda_1 + \delta_1 + \delta_3 + \zeta \tag{10}$$

And finally the genetic heterogeneity in this difference:

$$(10) - (9) = \delta_3 + \zeta \tag{11}$$

C.5.2 OLS estimation: low, middle and high PGS

Current smoking status (Y) is regressed on the full set of interactions between the indicators for the health shock ($shock$), being uninsured pre-65 ($uninsured$), Medicare eligibility ($post65$), medium polygenic risk for smoking (g^m) and high polygenic risk for smoking (g^h):

$$\begin{aligned}
Y_{it} \;=\; & \beta\, shock_{it} + \gamma\, post65_{it} \\
& + \lambda_1 \left(shock_{it} \times post65_{it}\right) \\
& + \lambda_2 \left(shock_{it} \times uninsured_i\right) \\
& + \lambda_3 \left(post65_{it} \times uninsured_i\right) \\
& + \lambda_4 \left(shock_{it} \times g_i^m\right) \\
& + \lambda_5 \left(post65_{it} \times g_i^m\right) \\
& + \lambda_6 \left(shock_{it} \times g_i^h\right) \\
& + \lambda_7 \left(post65_{it} \times g_i^h\right) \\
& + \delta_1 \left(shock_{it} \times post65_{it} \times uninsured_i\right) \\
& + \delta_2 \left(shock_{it} \times uninsured_i \times g_i^m\right) \\
& + \delta_3 \left(shock_{it} \times post65_{it} \times g_i^m\right) \\
& + \delta_4 \left(post65_{it} \times uninsured_i \times g_i^m\right) \\
& + \delta_5 \left(shock_{it} \times uninsured_i \times g_i^h\right) \\
& + \delta_6 \left(shock_{it} \times post65_{it} \times g_i^h\right) \\
& + \delta_7 \left(post65_{it} \times uninsured_i \times g_i^h\right) \\
& + \zeta_1 \left(shock_{it} \times post65_{it} \times uninsured_i \times g_i^m\right) \\
& + \zeta_2 \left(shock_{it} \times post65_{it} \times uninsured_i \times g_i^h\right) \\
& + \sum_{a=1}^{3} \phi_a\, age_{it}^a + \eta_i + \tau_t + \varepsilon_{it}
\end{aligned}$$

Effect of the shock on the outcome The derivative of the outcome with respect to shock is:

$$\begin{aligned}\frac{\partial Y_{it}}{\partial shock_{it}} &= \beta + \lambda_1 post65_{it} + \lambda_2 uninsured_i + \lambda_4 g_i^m + \lambda_6 g_i^h \\ &+ \delta_1(post65 \times uninsured_i) + \delta_2(uninsured_i \times g_i^m) + \delta_3(post65_{it} \times g_i^m) \\ &+ \delta_5(uninsured_i \times g_i^h) + \delta_6(post65_{it} \times g_i^h) \\ &+ \zeta_1(post65_{it} \times uninsured_i \times g_i^m) + \zeta_2(post65_{it} \times uninsured_i \times g_i^h)\end{aligned} \tag{12}$$

Again, we can look at the decomposition:

$$E\left[\frac{\partial Y_{it}}{\partial shock_{it}} | post65_{it} = 0, g_i^m = 0, g_i^h = 0, uninsured_i = 1\right] = \beta + \lambda_2 \tag{13}$$

$$E\left[\frac{\partial Y_{it}}{\partial shock_{it}} | post65_{it} = 1, g_i^m = 0, g_i^h = 0, uninsured_i = 1\right] = \beta + \lambda_1 + \lambda_2 + \delta_1 \tag{14}$$

$$E\left[\frac{\partial Y_{it}}{\partial shock_{it}} | post65_{it} = 0, g_i^m = 1, g_i^h = 0, uninsured_i = 1\right] = \beta + \lambda_2 + \lambda_4 + \delta_2 \tag{15}$$

$$\begin{aligned}E\left[\frac{\partial Y_{it}}{\partial shock_{it}} | post65_{it} = 1, g_i^m = 1, g_i^h = 0, , uninsured_i = 1\right] &= \beta + \lambda_1 + \lambda_2 + \lambda_4 \\ &+ \delta_1 + \delta_2 + \delta_3 + \zeta_1\end{aligned} \tag{16}$$

$$E\left[\frac{\partial Y_{it}}{\partial shock_{it}} | post65_{it} = 0, g_i^m = 0, g_i^h = 1, uninsured_i = 1\right] = \beta + \lambda_2 + \lambda_6 + \delta_5 \tag{17}$$

$$\begin{aligned}E\left[\frac{\partial Y_{it}}{\partial shock_{it}} | post65_{it} = 1, g_i^m = 0, g_i^h = 1, , uninsured_i = 1\right] &= \beta + \lambda_1 + \lambda_2 + \lambda_6 \\ &+ \delta_1 + \delta_5 + \delta_6 + \zeta_2\end{aligned} \tag{18}$$

Calculating the first two differences as above::

$$(29) - (28) = \lambda_1 + \delta_1 \tag{19}$$

$$(31) - (30) = \lambda_1 + \delta_1 + \delta_3 + \zeta_1 \tag{20}$$

$$(33) - (32) = \lambda_1 + \delta_1 + \delta_6 + \zeta_2 \tag{21}$$

And again the genetic heterogeneity in this difference:

$$(35) - (34) = \delta_3 + \zeta_1 \tag{22}$$

$$(36) - (34) = \delta_6 + \zeta_2 \tag{23}$$

D Results

D.1 Sample Characteristics

Table 5 displays summary statistics for the subset of the study sample that experienced a cardiovascular health shock over the course of the observation period, stratified by timing of the shock (pre-65 versus post-65) and genetic group. Within a genetic group, demographics are mostly similar across the timing strata. However, in both groups, those experiencing the shock after age 65 are on average older at baseline. In the high-

PGS group, there are also relatively more women experiencing the shock after the age of 65 than before 65. This is consistent with the general pattern that women experience cardiovascular disease later in life than men (Lloyd-Jones et al., 2010).

Table 5: Descriptive Statistics for the Subset of the study sample with a Health Shock, Stratified by Timing of the Shock and Genetic Group

	Low PGS	**High PGS**	**P value**
Shock at ages 60-64			
	Mean (SD)	Mean (SD)	
Age (baseline)	60.49 (0.57)	60.47 (0.64)	0.78
Smoking PGS	-0.97 (0.54)	0.63 (0.68)	0.00
Years of education	12.2 (3.41)	12.11 (3.17)	0.81
Income (nominal $ 1000)	19.79 (27.82)	18.95 (30.14)	0.79
No. waves present	4.65 (1.32)	4.59 (1.32)	0.65
	%	%	
Female	48.7	45.05	0.51
Smoking (baseline)	31.3	37	0.28
Persistently uninsured	4.35	6.59	0.36
Avg. cessation rate (baseline smokers)	12.05	12.17	0.97
No. of individuals	115	273	
No. of Person-year individuals	535	1252	
Shock at ages 67-70			
	Mean (SD)	Mean (SD)	
Age (baseline)	61.33 (2.15)	61.04 (1.4)	0.19
Smoking PGS	-0.94 (0.48)	0.75 (0.79)	0.00
Years of education	12.68 (3.13)	12.41 (3.09)	0.46
Income (nominal $ 1000)	17.81 (27.44)	15.74 (20.15)	0.48
No. waves present	5.06 (1.15)	5.12 (0.84)	0.65
	%	%	
Female	40.71	49.1	0.14
Smoking (baseline)	30.09	35.59	0.31
Persistently uninsured	8.85	6.31	0.42
Avg. cessation rate (baseline smokers)	10.92	11.03	0.97
No. of individuals	113	222	
No. of Person-year individuals	572	1136	

Pre-65: Health shock since the last survey reported at ages 60-64.
Post-65: Health shock since the last survey reported at ages 67-70.
P-values report significance tests for the difference in means between the shock timing strata.
Low PGS = lowest tertile of the polygenic score distribution; high PGS = upper two tertiles of the polygenic score distribution.
Cessation rates are defined as smoking in the previous but not in the current wave.
Data used: study sample restricted to individuals experiencing a health shock during the observation period.

The subset of respondents affected by cardiovascular illness during the observed years differed in some characteristics from those unaffected. Table 6 shows a comparison.

Table 6: Descriptive Statistics for the Study Sample, Stratified by Future Health Shock Status

	All	**New Shock**	**No new Shock**	**P value**
	Mean (SD)	Mean (SD)	Mean (SD)	
Age (baseline)	61.15 (1.87)	60.74 (1.14)	61.21 (1.94)	0.00
Smoking PGS	0.1 (0.99)	0.18 (1.02)	0.09 (0.99)	0.03
Years of education	12.48 (3.09)	12.33 (3.16)	12.5 (3.08)	0.18
Income (nominal $ 1000)	20.41 (34.97)	17.97 (26.62)	20.75 (35.98)	0.01
No. waves present	4.45 (1.36)	4.85 (1.15)	4.39 (1.38)	0.00
	%	%	%	
Female	49.86	45.9	50.42	0.02
Smoking (baseline)	29.5	34.08	28.86	0.01
Persistently uninsured	5.88	6.4	5.8	0.54
CV health shock	12.28	100	0	-
Avg. cessation rate (baseline smokers)	10.3	11.35	10.11	0.21
No. of individuals	5854	719	5135	

Health shock: diagnosed with a new cardiovascular condition during the time since the last HRS survey, but no history of cardiovascular disease prior to this diagnosis. P-values report significance tests for the difference in means between the health shock strata. Cessation rates were defined as smoking in the previous but not in the current wave.
Data used: HRS study sample, n = 5,854.

D.2 Main Results

Table 7 reports the covariance matrix of these estimated coefficients. These coefficients and standard errors are used to calculate the effect of health shocks on the smoking probability of individuals who are uninsured before the age of 65—the subgroup of interest—for the 4 combinations of shock timing (before or after 65) and polygenic score (high or low). The derivation of these effects is described in Appendix Section C.5.

D.3 Robustness Checks

D.3.1 Falsification Test

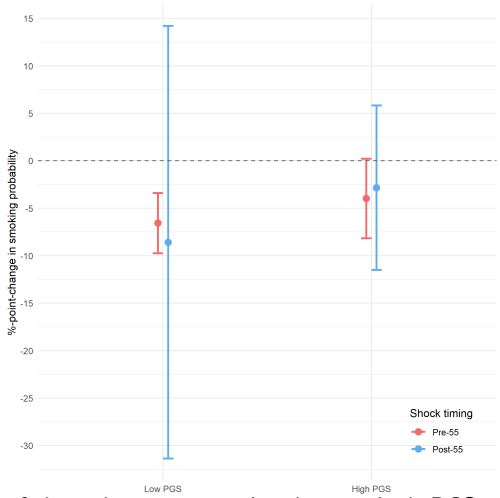
Figure 10 and Table 8 reports the summary results of the falsification test. We run the same analysis but focusing on the age range from 50 to 60, and estimating the effect of suffering a shock after the age of 55. Since there should be nothing special about age 55, at least in terms of health insurance, we do not expect any result and consider this to be a falsification or "placebo" test of our approach.

Table 9 reports the all of the coefficients estimated in the main regression of the falsification test.

D.3.2 Using all tertiles

In this section, we derive the results using all of the three tertiles of the PGS distribution, as derived in Appendix Section D.3.2. Results shown in Tables 10 and Figure 11, show that the effects for the upper two tertiles of the polygenic score are virtually the same. To improve statistical power and simplicity of exposition, the main results are always

Figure 10: Falsification test, age 50 to 60



Notes: Low PGS = lowest tertile of the polygenic score distribution; high PGS = upper two tertiles of the polygenic score distribution. Pre-55: Health shock since the last survey reported at ages 50-54. Post-55: Health shock since the last survey reported at ages 57-60. Effects are estimated using the coefficients in the last column of Table 9 and following the derivation described in C.5. Bars show 95% confidence intervals, standard error clustered at the individual level.
Data used: HRS study sample.

presented by pooling these two tertiles toghether into a single category labeled high-PGS.

D.3.3 Median Split of the polygenic score

In this section, we derive the results by using a median-split for the PGS, instead of tertiles. Results shown in Tables 11 12, as well as Figure 12, show that the effects are virtually the same, albeit a bit smaller in magnitude and less precisely estimated, as when splitting the PGS according to tertiles.

D.3.4 Older GWAS Summary Statistics

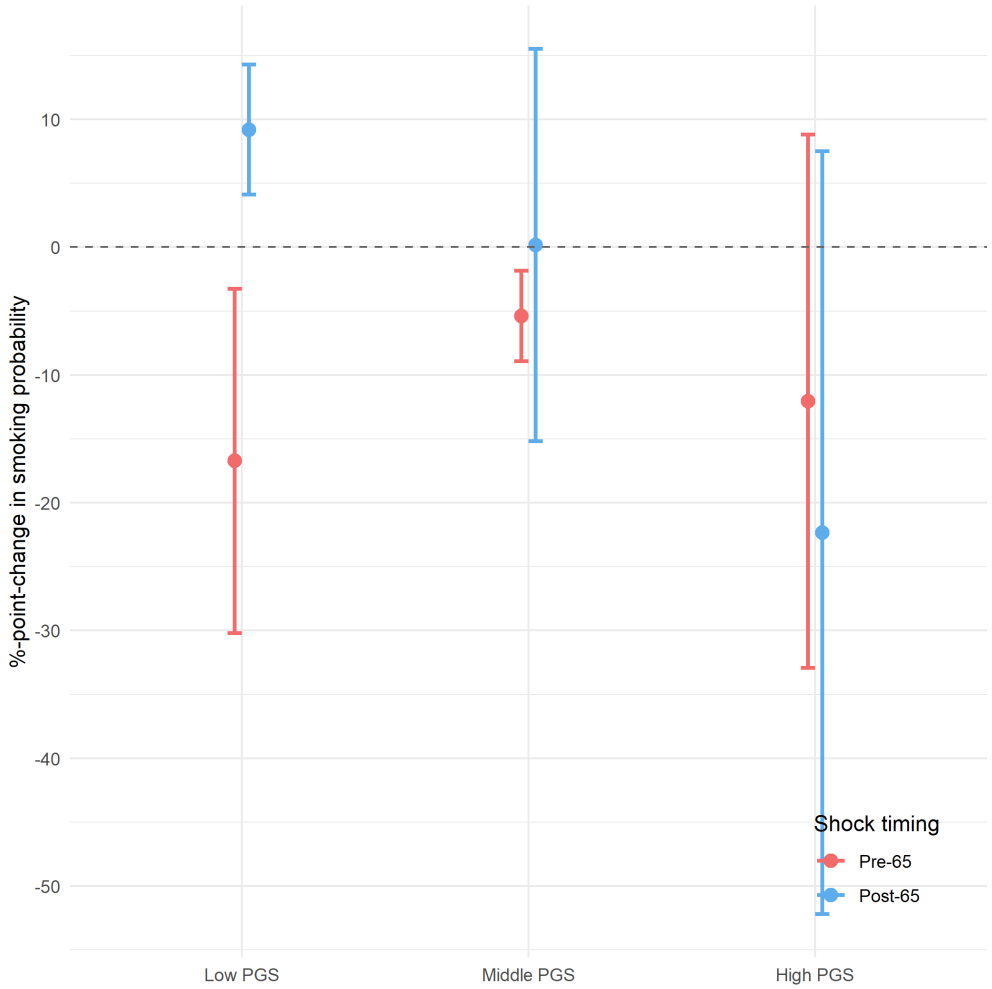
In this section, we also use the polygenic score publicly provided by the HRS (Ware, Schmitz and Faul, 2017) for the smoking phenotype "regular smoking" (having smoked more than 100 cigarettes throughout one's life). This score is constructed as a weighted sum of the genotype over the 779,538 SNPs that overlap between the HRS genetic database and a 2010 GWAS meta-analysis conducted by the Tobacco and Genetics Consortium (The Tobacco and Genetics Consortium et al., 2010).

Results shown in Tables 13 and Figure 13, show that the effects are virtually the same, possibly even a bit stronger.

D.3.5 Using different polygenic scores

Other polygenic scores (PGS), besides the one for being a smoker, might be driving this heterogeneity in moral hazard. We estimated the same analysis outlined in equation 2 but replacing g_i with the following PGS: the PGS for cigarettes per day (CPD) (Liu et al., 2019), the PGS for educational attainment and the one for cognitive abilities (Lee et al., 2018), the PGS for risk tolerance (Karlsson Linnér et al., 2019), the PGS for

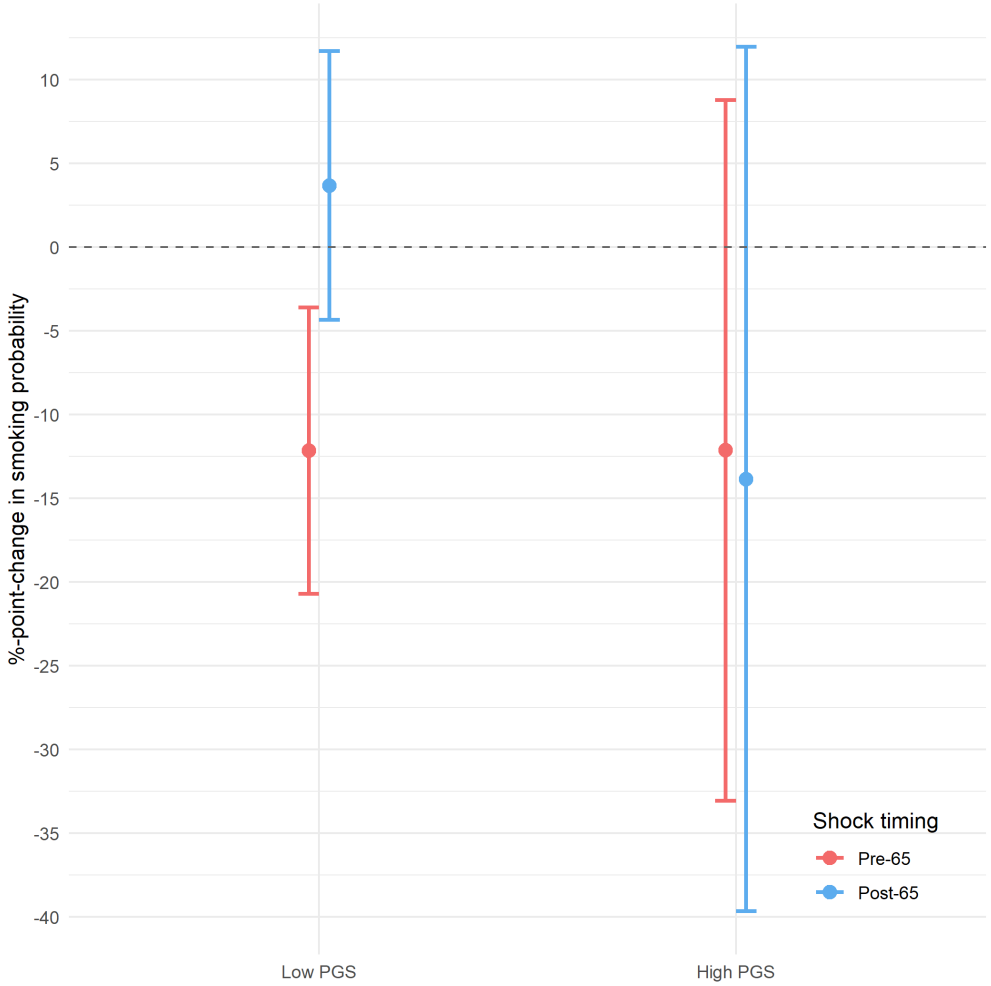
Figure 11: Effect of a Health Shock on the Smoking Probability in the Pre-65 Uninsured Subgroup, Stratified by Timing of the Shock and Genetic Type (median split)



Notes: Low PGS = lowest tertile of the polygenic score distribution; Middle PGS = high PGS = upper two tertiles of the polygenic score distribution. Pre-65: Health shock since the last survey reported at ages 60-64. Post-65: Health shock since the last survey reported at ages 67-70. Estimates and standard errors are shown in Panel A of Table 2. Effects are estimated using the coefficients in Table 3 and following the derivation described in C.5. Bars show 95% confidence intervals, standard error clustered at the individual level.
Data used: HRS study sample, n = 5,854. Supplement.

non-cognitive skills (Demange et al., 2020), and the PGS for Body-Mass-Index (Yengo et al., 2018). We chose polygenic scores for traits that are genetically correlated with smoking, such as CPD and risk tolerance, or plausibly related to strategic behaviors and moral hazard, such as education, cognition, and non-cognitive skills. BMI is meant more as a placebo.

Figure 12: Effect of a Health Shock on the Smoking Probability in the Pre-65 Uninsured Subgroup, Stratified by Timing of the Shock and Genetic Type (median split)



Notes: Low PGS = polygenic score below the median; high PGS = polygenic score above the median. Pre-65: Health shock since the last survey reported at ages 60-64. Post-65: Health shock since the last survey reported at ages 67-70. Estimates and standard errors are shown in Panel A of Table 2. Effects are estimated using the coefficients in Table 3 and following the derivation described in C.5. Bars show 95% confidence intervals, standard error clustered at the individual level.
Data used: HRS study sample, n = 5,854.

Table 7: Covariance Matrix for Regression Coefficients in the last column of Table 3

	Shock	Shock x post65	Shock x uninsured	Shock x high PGS	Shock x post65 x uninsured	Shock x uninsured x high PGS	Shock x post65 x high PGS	Shock x post65 x uninsured x high PGS
Shock	0.0006	-0.0006	-0.0006	-0.0006	0.0006	0.0006	0.0006	-0.0006
Shock x post65	-0.0006	0.0010	0.0006	0.0006	-0.0010	-0.0006	-0.0010	0.0010
Shock x uninsured	-0.0006	0.0006	0.0053	0.0006	-0.0057	-0.0053	-0.0006	0.0057
Shock x high PGS	-0.0006	0.0006	0.0006	0.0008	-0.0006	-0.0008	-0.0008	0.0008
Shock x post65 x uninsured	0.0006	-0.0010	-0.0057	-0.0006	0.0072	0.0057	0.0010	-0.0072
Shock x uninsured x high PGS	0.0006	-0.0006	-0.0053	-0.0008	0.0057	0.0124	0.0008	-0.0127
Shock x post65 x high PGS	0.0006	-0.0010	-0.0006	-0.0008	0.0010	0.0008	0.0017	-0.0017
Shock x post65 x uninsured x high PGS	-0.0006	0.0010	0.0057	0.0008	-0.0072	-0.0127	-0.0017	0.0225

Variances and covariances were used for calculating the standard errors of the effects of interest displayed in Table 2 and Figure 1.

Table 8: Summary of Statistical Results for the Pre-65 Uninsured Subgroup, Stratified by Timing of the Shock and Genetic Group

	Dependent variable:
	Smoking status
Health Shock	-0.084^{*}
	(0.047)
Post-55	-0.002
	(0.010)
Shock × Post-55	0.038
	(0.055)
Shock × Uninsured	-0.009
	(0.255)
Post-55 × Uninsured	-0.033
	(0.022)
Shock × High PGS	0.018
	(0.057)
Post-55 × High PGS	0.016^{**}
	(0.008)
Shock × Post-55 x Uninsured	-0.025
	(0.269)
Shock × Uninsured × High PGS	0.021
	(0.289)
Shock × Post-55 × High PGS	-0.094
	(0.068)
Post-55 × Uninsured × High PGS	-0.023
	(0.026)
Shock × Post-55 × Uninsured × High PGS	0.117
	(0.310)
Observations	22,356
R^2	0.004

Low PGS = lowest tertile of the polygenic score distribution; high PGS = upper two tertiles of the polygenic score distribution. Pre-55: Health shock since the last survey reported at ages 50-54. Post-55: Health shock since the last survey reported at ages 57-60. *P $<$ 0.1; **P $<$ 0.05; ***P $<$ 0.01. Robust standard errors in parentheses are clustered at the individual level. Effects are estimated using the coefficients in the last column of Table 9 and following the derivation described in C.5. *Data used*: HRS study sample.

Table 9: Coefficients from estimating the linear probability model in equation (2) using OLS

	Dependent variable:				
	Smoking status				
	(1)	(2)	(3)	(4)	(5)
Health Shock	0.086	0.091	0.098	−0.092**	−0.084**
	(0.089)	(0.089)	(0.088)	(0.042)	(0.043)
Post-55	−0.054***	0.009	0.007	−0.001	−0.002
	(0.012)	(0.016)	(0.016)	(0.010)	(0.010)
Uninsured	0.213***	0.216***	0.214***		
	(0.025)	(0.025)	(0.025)		
High PGS	0.021	0.021	0.021		
	(0.017)	(0.017)	(0.017)		
Shock × Post-55	−0.069	−0.060	−0.065	0.045	0.038
	(0.105)	(0.104)	(0.104)	(0.056)	(0.056)
Shock × Uninsured	0.335***	0.323***	0.319***	−0.003	−0.009
	(0.093)	(0.092)	(0.092)	(0.046)	(0.046)
Post-55 × Uninsured	−0.061	−0.066*	−0.064	−0.035	−0.033
	(0.040)	(0.040)	(0.040)	(0.024)	(0.024)
Shock × High PGS	−0.090	−0.092	−0.095	0.021	0.018
	(0.109)	(0.109)	(0.108)	(0.062)	(0.062)
Post-55 × High PGS	0.011	0.011	0.011	0.016	0.016
	(0.015)	(0.015)	(0.015)	(0.010)	(0.010)
Shock × Post-55 x Uninsured	−0.180	−0.167	−0.167	−0.030	−0.025
	(0.181)	(0.180)	(0.179)	(0.132)	(0.131)
Shock × Uninsured × High PGS	0.068	0.083	0.093	0.022	0.021
	(0.109)	(0.109)	(0.109)	(0.067)	(0.068)
Shock × Post-55 × High PGS	−0.050	−0.046	−0.045	−0.098	−0.094
	(0.127)	(0.126)	(0.126)	(0.077)	(0.077)
Post-55 × Uninsured × High PGS	0.028	0.029	0.029	−0.023	−0.023
	(0.050)	(0.050)	(0.050)	(0.030)	(0.030)
Shock × Post-55 × Uninsured × High PGS	−0.340	−0.356	−0.365	0.115	0.117
	(0.228)	(0.227)	(0.227)	(0.148)	(0.148)
Age		Yes	Yes	Yes	Yes
Year FE			Yes		Yes
Individual FE				Yes	Yes
Observations	22,356	22,356	22,356	22,356	22,356
R^2	0.017	0.019	0.019	0.841	0.841

Notes: Health shock = binary indicator for having suffered a cardiovascular health shock for the first time since the previous wave. Post-55 = indicator for age > 65 at the time of the interview. Uninsured = binary indicator for being persistently uninsured in every observation of the study sample before the age of 55. High PGS = indicator for being in the upper two tertiles of the polygenic score distribution. Age = controls for 3^{rd} degree polynomial in age. FE = adding fixed effects. Robust standard errors in parentheses are clustered at the individual level. *p<0.1; **p<0.05; ***p<0.01. *Data used*: HRS study sample.

Table 10: Summary of Statistical Results for the Pre-65 Uninsured Subgroup, Stratified by Timing of the Shock and Genetic Group (Median)

Effect of health shock on smoking probability			
	Low PGS	Middle PGS	High PGS
Pre 65	-0.165**	-0.051***	-0.123
	(0.069)	(0.018)	(0.107)
Post 65	0.09***	-0.039	-0.221
Effect of health insurance on effect of health shock			
	(0.026)	(0.078)	(0.152)
	Low PGS	Middle PGS	High PGS
Post 65 - Pre 65	0.256***	0.011	-0.097
Differential effect of health insurance by genetic group			
	(0.079)	(0.076)	(0.185)
	High PGS - low PGS	Middle PGS - low PGS	High PGS - middle PGS
Post 65 - Pre 65	-0.353*	-0.245**	-0.109
	(0.201)	(0.109)	(0.2)

Low PGS = polygenic score below the median; high PGS = polygenic score above the median. Pre-65: Health shock since the last survey reported at ages 60-64. Post-65: Health shock since the last survey reported at ages 67-70. *P $<$ 0.1; **P $<$ 0.05; ***P $<$ 0.01. Robust standard errors in parentheses are clustered at the individual level. The covariance matrix used for calculating standard errors is shown in Appendix Table 7. Effects are estimated using the coefficients in Table 3 and following the derivation described in C.5.
Data used: HRS study sample, n = 5,854.

Table 11: Coefficients from estimating the linear probability model in equation (2) using OLS (PGS median split)

	Dependent variable:				
	Smoking status				
	(1)	(2)	(3)	(4)	(5)
Health Shock	-0.034	-0.034	-0.033	-0.048^{**}	-0.047^{**}
	(0.031)	(0.032)	(0.031)	(0.019)	(0.019)
Post-65	-0.066^{***}	-0.022^{*}	-0.021^{*}	-0.013^{*}	-0.012^{*}
	(0.007)	(0.012)	(0.012)	(0.007)	(0.007)
Uninsured	0.168^{***}	0.169^{***}	0.169^{***}		
	(0.027)	(0.027)	(0.027)		
High PGS	0.039^{***}	0.040^{***}	0.040^{***}		
	(0.012)	(0.012)	(0.012)		
Shock × Post-65	0.017	0.024	0.024	0.019	0.019
	(0.045)	(0.045)	(0.046)	(0.028)	(0.028)
Shock × Uninsured	-0.055	-0.055	-0.056	-0.071	-0.073
	(0.160)	(0.160)	(0.159)	(0.048)	(0.048)
Post-65 × Uninsured	0.026	0.025	0.025	-0.045^{*}	-0.045^{*}
	(0.036)	(0.037)	(0.037)	(0.025)	(0.025)
Shock × High PGS	0.052	0.050	0.048	0.024	0.024
	(0.046)	(0.046)	(0.046)	(0.027)	(0.027)
Post-65 × High PGS	0.004	0.004	0.003	0.0001	-0.0003
	(0.010)	(0.010)	(0.010)	(0.007)	(0.007)
Shock × Post-65 x Uninsured	0.090	0.090	0.090	0.136^{**}	0.137^{**}
	(0.212)	(0.213)	(0.212)	(0.067)	(0.067)
Shock × Uninsured × High PGS	0.004	0.014	0.015	-0.028	-0.026
	(0.211)	(0.211)	(0.210)	(0.119)	(0.119)
Shock × Post-65 × High PGS	-0.082	-0.078	-0.077	-0.086^{**}	-0.088^{**}
	(0.064)	(0.064)	(0.064)	(0.042)	(0.042)
Post-65 × Uninsured × High PGS	-0.072	-0.072	-0.072	0.041	0.041
	(0.057)	(0.057)	(0.057)	(0.033)	(0.033)
Shock × Post-65 × Uninsured × High PGS	0.154	0.145	0.142	-0.131	-0.136
	(0.310)	(0.312)	(0.311)	(0.189)	(0.188)
Age		*Yes*	*Yes*	*Yes*	*Yes*
Year FE			*Yes*		*Yes*
Individual FE				*Yes*	*Yes*
Observations	26,022	26,022	26,022	26,022	26,022
R^2	0.016	0.017	0.018	0.823	0.823

Notes: Health shock = binary indicator for having suffered a cardiovascular health shock for the first time since the previous wave. Post-65 = indicator for age > 65 at the time of the interview. Uninsured = binary indicator for being persistently uninsured in every observation of the study sample before the age of 65. High PGS = indicator for being above the median polygenic score. Age = controls for 3^{rd} degree polynomial in age. FE = adding fixed effects. Robust standard errors in parentheses are clustered at the individual level. *p<0.1; **p<0.05; ***p<0.01.
Data used: HRS study sample, n = 5,854.

Table 12: Summary of Statistical Results for the Pre-65 Uninsured Subgroup, Stratified by Timing of the Shock and Genetic Group (Median)

Effect of health shock on smoking probability		
	Low PGS	High PGS
Pre 65	-0.12***	-0.122
	(0.044)	(0.107)
Post 65	0.036	-0.188
	(0.041)	(0.137)
Effect of health insurance on effect of health shock		
	Low PGS	High PGS
Post 65 - Pre 65	0.157**	-0.066
	(0.061)	(0.173)
Differential effect of health insurance by genetic group		
	High PGS	- low PGS
Post 65 - Pre 65	-0.223	
	(0.184)	

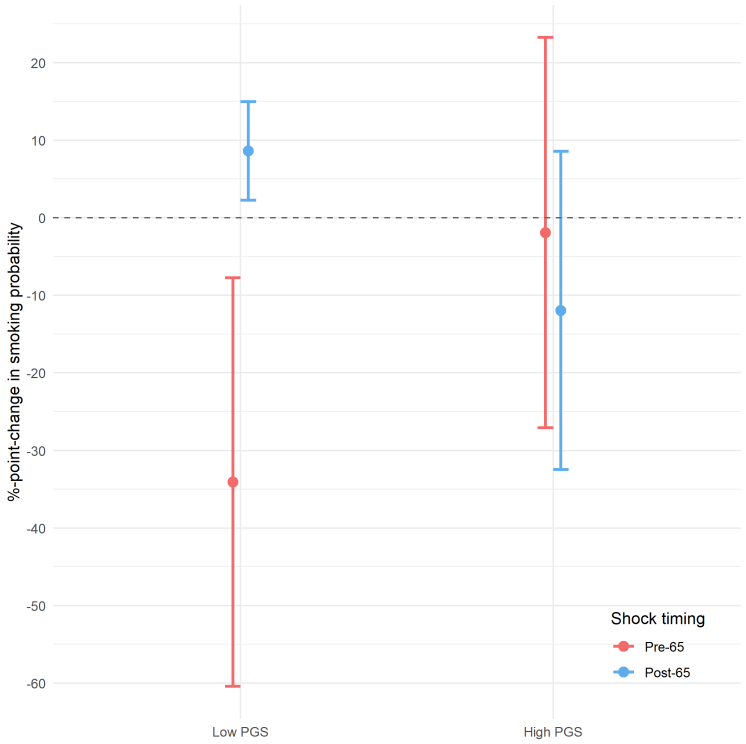
Low PGS = polygenic score below the median; high PGS = polygenic score above the median. Pre-65: Health shock since the last survey reported at ages 60-64. Post-65: Health shock since the last survey reported at ages 67-70. *P $<$ 0.1; **P $<$ 0.05; ***P $<$ 0.01. Robust standard errors in parentheses are clustered at the individual level. The covariance matrix used for calculating standard errors is shown in Appendix Table 7. Effects are estimated using the coefficients in Table 3 and following the derivation described in C.5.
Data used: HRS study sample, n = 5,854.

Table 13: Summary of Statistical Results for the Pre-65 Uninsured Subgroup, Stratified by Timing of the Shock and Genetic Group (older PGS)

Effect of health shock on smoking probability		
	Low PGS	High PGS
Pre 65	-0.338**	-0.018
	(0.135)	(0.128)
Post 65	0.06***	-0.12
	(0.02)	(0.105)
Effect of health insurance on effect of health shock		
	Low PGS	High PGS
Post 65 - Pre 65	0.399***	-0.103
	(0.139)	(0.163)
Differential effect of health insurance by genetic group		
	High PGS	- low PGS
Post 65 - Pre 65	-0.501**	
	(0.214)	

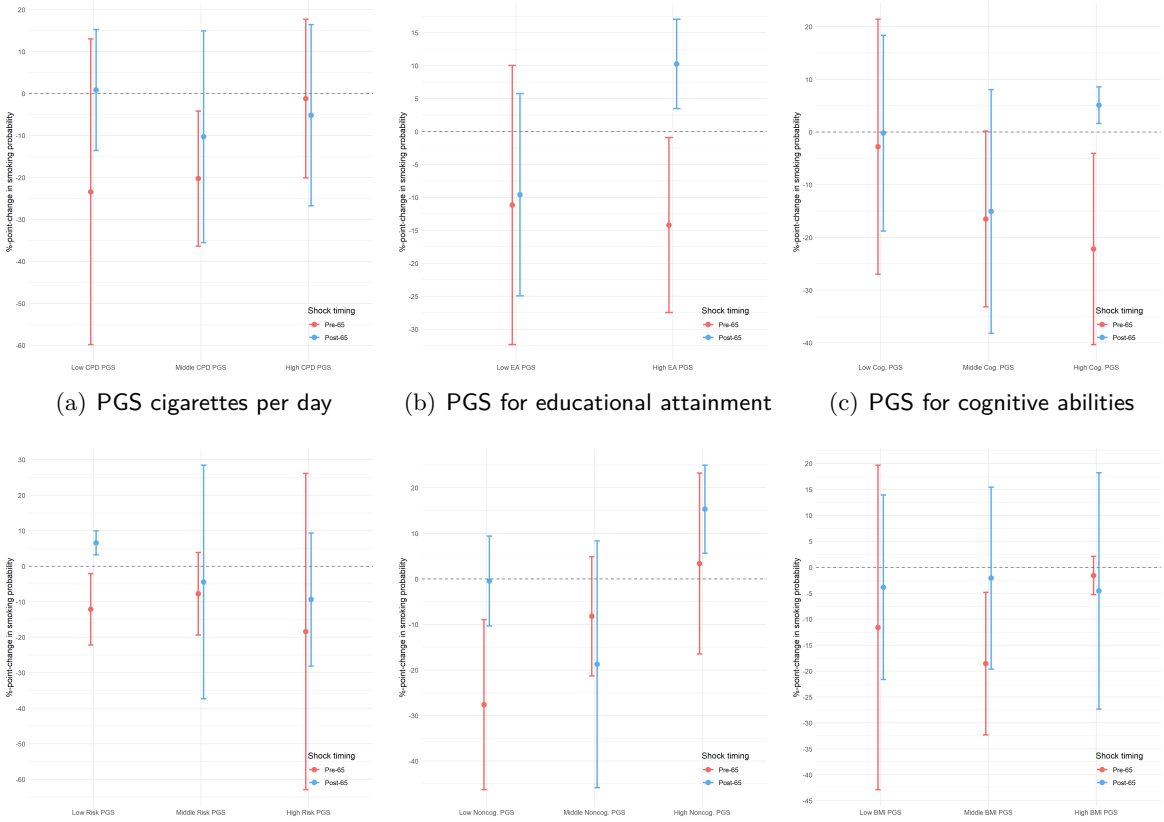
Low PGS = polygenic score below the median; high PGS = polygenic score above the median, constructed using the summary statistics from the Tobacco and Genetics Consortium (The Tobacco and Genetics Consortium et al., 2010) (PGS publicly available from the HRS website). Pre-65: Health shock since the last survey reported at ages 60-64. Post-65: Health shock since the last survey reported at ages 67-70. *P $<$ 0.1; **P $<$ 0.05; ***P $<$ 0.01. Robust standard errors in parentheses are clustered at the individual level. The covariance matrix used for calculating standard errors is shown in Appendix Table 7. Effects are estimated using the coefficients in Table 3 and following the derivation described in C.5.
Data used: HRS study sample, n = 5,854.

Figure 13: Effect of a Health Shock on the Smoking Probability in the Pre-65 Uninsured Subgroup, Stratified by Timing of the Shock and Genetic Type (older PGS)



Notes: Low PGS = polygenic score below the median; high PGS = polygenic score above the median. Pre-65: Health shock since the last survey reported at ages 60-64. Post-65: Health shock since the last survey reported at ages 67-70. Estimates and standard errors are shown in Panel A of Table 2. Effects are estimated using the coefficients in Table 3 and following the derivation described in C.5. Bars show 95% confidence intervals, standard error clustered at the individual level.
Data used: HRS study sample, n = 5,854.

Figure 14: Other Polygenic Scores as Proxy for Moral Hazard Heterogeneity



(a) PGS cigarettes per day (b) PGS for educational attainment (c) PGS for cognitive abilities

(d) PGS for risk tolerance (e) PGS for non-cognitive skills (f) PGS for BMI

Notes: The figures report the effect of suffering a health shock on the probability of smoking for the pre-65 uninsured subgroup, stratified by timing of the shock (before and after the age of 65) and different polygenic scores. Pre-65: Health shock since the last survey reported at ages 60-64. Post-65: Health shock since the last survey reported at ages 67-70. Effects are estimated using a combination of the coefficients from equation 2 where g_i is replaced by the different polygenic scores reported in the sub-figure captions, following the derivation described in C.5. Bars show 95% confidence intervals, standard error clustered at the individual level.
Data used: HRS study sample, n = 5,854.

D.3.6 Relaxing the Criteria for Inclusion in the Pre-65 Uninsured Group

Table 14: Summary of Statistical Results for the Pre-65 Uninsured Subgroup (Using Different Definitions of the Pre-65 Uninsured Status Indicator)

	100% uninsured		*66% uninsured*		*33% uninsured*	
Effect of health shock on smoking probability						
	Low PGS	High PGS	Low PGS	High PGS	Low PGS	High PGS
Pre 65	-0.165**	-0.108	-0.147**	-0.055	-0.055	-0.09*
	(0.069)	(0.083)	(0.059)	(0.06)	(0.068)	(0.048)
Post 65	0.09***	-0.13	0.086***	-0.091	0.052	-0.102**
	(0.026)	(0.089)	(0.021)	(0.078)	(0.036)	(0.05)
Effect of health insurance on effect of health shock						
	Low PGS	High PGS	Low PGS	High PGS	Low PGS	High PGS
Post 65 - Pre 65	0.256***	-0.023	0.233***	-0.036	0.107	-0.012
	(0.079)	(0.121)	(0.067)	(0.098)	(0.076)	(0.069)
Differential effect of health insurance by genetic group						
	High PGS	- low PGS	High PGS	- low PGS	High PGS	- low PGS
Post 65 - Pre 65	-0.279*		-0.269**		-0.119	
	(0.144)		(0.118)		(0.103)	

Main study results, for comparison.
Pre-65 uninsured status indicator set to 1 for respondents who were uninsured in at least 66% of all pre-65 observations.
Pre-65 uninsured status indicator set to 1 for respondents who were uninsured in at least 33% of all pre-65 observations.
$^{*}P < 0.1$; $^{**}P < 0.05$; $^{***}P < 0.01$. Robust standard errors in parentheses were clustered at the individual level.
Low PGS = lowest tertile of the polygenic score distribution; high PGS = upper two tertiles of the polygenic score distribution.
Pre-65: Health shock since the last survey reported at ages 60-64. Post-65: Health shock since the last survey reported at ages 67-70.
Data used: HRS study sample, n = 5,854.

D.3.7 Including Individuals with Shocks Reported at Ages 65 and 66

Table 15: Summary of Statistical Results for the Pre-65 Uninsured Subgroup (Including Individuals Reporting a Health Shock when Aged 65 or 66 in the study Sample)

	Analytic Sample		*Including reported at 65/66*	
Effect of health shock on smoking probability				
	Low PGS	High PGS	Low PGS	High PGS
Pre 65	-0.165**	-0.108	-0.166**	-0.113
	(0.069)	(0.083)	(0.069)	(0.084)
Post 65	0.09***	-0.13	0.091***	-0.213**
	(0.026)	(0.089)	(0.025)	(0.089)
Effect of health insurance on effect of health shock				
	Low PGS	High PGS	Low PGS	High PGS
Post 65 - Pre 65	0.256***	-0.023	0.257***	-0.1
	(0.079)	(0.121)	(0.079)	(0.121)
Differential effect of health insurance by genetic group				
	High PGS	- low PGS	High PGS	- low PGS
Post 65 - Pre 65	-0.279*		-0.356**	
	(0.144)		(0.145)	

Main study results, for comparison.
study sample additionally includes individuals who reported experiencing a health shock since the last survey wave when interviewed at ages 65 or 66.
*P $<$ 0.1; **P $<$ 0.05; ***P $<$ 0.01. Robust standard errors in parentheses were clustered at the individual level.
Low PGS = lowest tertile of the polygenic score distribution; high PGS = upper two tertiles of the polygenic score distribution.
Pre-65: Health shock since the last survey reported at ages 60-64. Post-65: Health shock since the last survey reported at ages 67-70.
Data used: study sample, but skipping Step 6 in Section C.2.1.

D.3.8 Using Medicare Enrollment Status instead of age 65

Medicare enrollment refers to the RAND variable `GOVMR`, which indicates whether the respondent is covered by Medicare in a given wave. For details on the survey questions and construction of this variable, see the RAND HRS documentation.

Table 16: Summary of Statistical Results for the Pre-65 Uninsured Subgroup (Using Medicare Enrollment Status Instead of Medicare Eligibility Status)

Effect of health shock on smoking probability		
	Low PGS	High PGS
Without Medicare	-0.136**	-0.109
	(0.066)	(0.083)
With Medicare	0.097***	-0.124
	(0.028)	(0.089)
Effect of health insurance on effect of health shock		
	Low PGS	High PGS
With - Without Medicare	0.233***	-0.016
	(0.077)	(0.121)
Differential effect of health insurance by genetic group		
	High PGS	- low PGS
With - Without Medicare	-0.248*	
	(0.144)	

*P $<$ 0.1; **P $<$ 0.05; ***P $<$ 0.01. Robust standard errors in parentheses, clustered at the individual level. Low PGS = lowest tertile of the polygenic score distribution; high PGS = upper two tertiles of the polygenic score distribution. Effects are estimated using the coefficients in Table 3 and following the derivation described in C.5, but replacing the post-65 indicator with Medicare enrollment status. *Data used*: HRS study sample with the additional restriction of a non-missing Medicare enrollment status.

D.4 Confounders

Figure 15 and Tables 17-19 report the results for the effect of the shock on potential confounding variables. There is no systematic differential effect of the health shock before or after the age of 65 on the reported income or the probability of retirement of persistently uninsured individuals. The absence of any discernible effect suggests that neither income nor retirement are potential mechanisms behind the observed change in smoking responses.

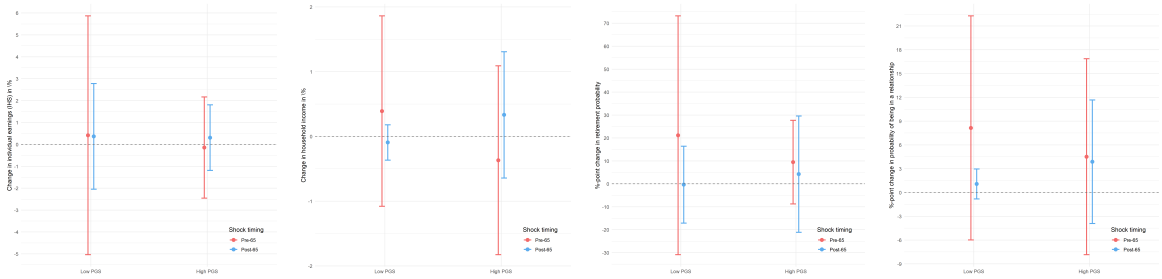
Table 17: Summary of Placebo Checks for the Pre-65 Uninsured Subgroup for Income and Wealth

	Effect of health shock on...					
	...individual earnings		**...household earnings**		**...household wealth**	
	Low PGS	High PGS	Low PGS	High PGS	Low PGS	High PGS
Pre 65	0.403	-0.11	0.389	-0.369	-5.42	-1.687
	(2.771)	(1.178)	(0.751)	(0.744)	(3.861)	(2.017)
Post 65	0.016	0.59	-0.087	0.328	-1.26	-0.326
	(1.314)	(0.786)	(0.148)	(0.497)	(0.853)	(2.021)
	Effect of health insurance on effect of health shock					
	Low PGS	High PGS	Low PGS	High PGS	Low PGS	High PGS
Post 65 - Pre 65	-0.387	0.7	-0.476	0.697	4.159	1.361
	(3.092)	(1.434)	(0.791)	(0.903)	(3.954)	(2.866)
	Differential effect of health insurance by genetic group					
	High PGS	- low PGS	High PGS	- low PGS	High PGS	- low PGS
Post 65 - Pre 65	1.087		1.173		-2.798	
	(3.408)		(1.201)		(4.884)	

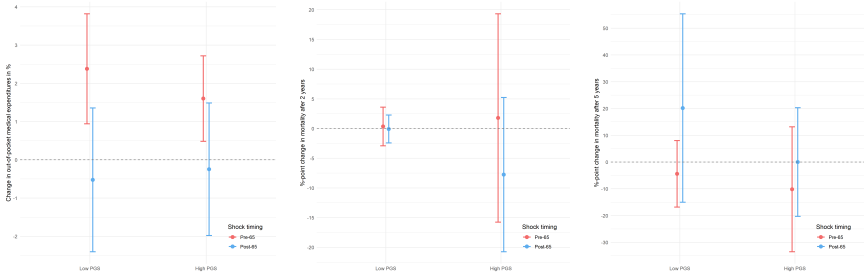
Table 18: Summary of Placebo Checks for the Pre-65 Uninsured Subgroup for Retirement, Relationship Status and Out-of-Pocket Medical Expenditures

	Effect of health shock on...					
	...retirement		**...relationship status**		**...OOP medical expenditures**	
	Low PGS	High PGS	Low PGS	High PGS	Low PGS	High PGS
Pre 65	0.212	0.094	0.081	0.044	2.384***	1.607***
	(0.265)	(0.093)	(0.073)	(0.063)	(0.736)	(0.571)
Post 65	0.015	0.022	0.01	0.041	0.26	-1.27
	(0.093)	(0.13)	(0.01)	(0.04)	(0.658)	(0.873)
	Effect of health insurance on effect of health shock					
	Low PGS	High PGS	Low PGS	High PGS	Low PGS	High PGS
Post 65 - Pre 65	-0.197	-0.071	-0.071	-0.003	-2.123**	-2.878***
	(0.28)	(0.16)	(0.07)	(0.074)	(1.008)	(1.056)
	Differential effect of health insurance by genetic group					
	High PGS	- low PGS	High PGS	- low PGS	High PGS	- low PGS
Post 65 - Pre 65	0.125		0.068		-0.754	
	(0.322)		(0.102)		(1.457)	

Figure 15: Testing for Potential Confounders



(a) Earned Income (IHS) (b) Total Household Income (IHS) (c) Probability of Retiring (d) Relationship status



(e) Medical expenditure (f) 2 Year Mortality (g) 5 Year Mortality

Notes: The figures report the effect of suffering a health shock for the pre-65 uninsured subgroup, stratified by timing of the shock (before and after the age of 65) and having a high or low polygenic score, on various possible confounders. Effects are estimated using a combination of the coefficients from equation 2 where the outcome Y_{it} is replaced by the different confounders reported in the sub-figure captions, following the derivation described in C.5. Low PGS = lowest tertile of the polygenic score distribution; high PGS = upper two tertiles of the polygenic score distribution. IHS: inverse hyperbolic sine (similar to log). Pre-65: Health shock since the last survey reported at ages 60-64. Post-65: Health shock since the last survey reported at ages 67-70. Estimates and standard errors are shown in Panel A of Table 2. Effects are estimated using the coefficients in the last column of Table 3 and following the derivation described in C.5. Bars show 95% confidence intervals, standard error clustered at the individual level.
Data used: HRS study sample, n = 5,854.

E The Model

In this section, we introduce a model to explore the theoretical basis and the testable empirical consequences of heterogeneity in moral hazard.[16]

What drives heterogeneity in moral hazard? In the context of health insurance, moral hazard is usually considered to be a reaction to insurance coverage: an increase in insurance coverage leads to changes in health behaviors, such as increased usage of medical care (Einav and Finkelstein, 2018) or smoking. Hence moral hazard can be considered as a sort of price sensitivity of the agent (as in Einav et al., 2013). Knowing they can afford to go to the doctor if they fall sick, the insured agents might be more inclined to engage in immediately rewarding behavior which is harmful in the long run, such as smoking.[17] Heterogeneity in this response to health care coverage can

[16]This model was initially drafted by Regina Seibel, who was working on the project as a research assistant. An abridged version of the model is reported here with her permission.

[17]This change in risky health behaviors due to the anticipation of being able to afford health care in the future has been dubbed *ex-ante* moral hazard. *Ex-post* moral hazard refers to an increase in the use of medical care—such as doctor visits or medicines—following health insurance coverage.

Table 19: Summary of Placebo Checks for the Pre-65 Uninsured Subgroup for Mortality within 2 and 5 years

	Effect of health shock on mortality within...			
	...2 years		**...5 years**	
	Low PGS	High PGS	Low PGS	High PGS
Pre 65	0.004	0.015	-0.046	-0.113
	(0.015)	(0.09)	(0.061)	(0.119)
Post 65	-0.001	-0.074	0.204	0.009
	(0.011)	(0.065)	(0.179)	(0.103)
	Effect of health insurance on effect of health shock			
	Low PGS	High PGS	Low PGS	High PGS
Post 65 - Pre 65	-0.005	-0.089	0.25	0.122
	(0.025)	(0.113)	(0.191)	(0.154)
	Differential effect of health insurance by genetic group			
	High PGS	- low PGS	High PGS	- low PGS
Post 65 - Pre 65	-0.085		-0.129	
	(0.116)		(0.245)	

then be driven by social factors, like exposure to family or peers who behave similarly (Chatterjee et al., 2018; Hoffmann, 2017), or biological factors, like genetic propensity to smoke.

Following these insights, we model the demand side of a health insurance market with agents who are heterogeneous in two dimensions: exogenous health risk and a moral hazard parameter governing the behavioral response towards being insured.

E.1 Setup

Consider two time periods: in period 1, all agents are healthy[18] and must make two decisions: whether to be insured, and how much effort to invest in reducing the probability of a health shock. More effort leads to a lower probability of falling sick, given an individual baseline level of risk. In period 2 the health risk realizes, with a probability mediated by the above-mentioned health-enhancing effort, and individuals can either be sick (S) or healthy (H). In case of sickness, insured agents receive the agreed-upon coverage $1-\tau$ and pay the premium p.[19] Uninsured agents pay the full medical treatment. In the healthy state of the world, insured agents still pay the premium while uninsured keep their full income for consumption.

[18] In other words, we do not consider the case of pre-existing conditions.

[19] To start off with a model as simple as possible, we abstract from the supply side and just assume that the insurer offers a fixed contract with premium p and coverage rate $1-\tau$. A special case would be complete insurance with $\tau = 0$.

Utility depends on consumption c and the health state:

$$u(c; H) > u(c; S)$$

Utility is increasing and concave in consumption, i.e. $u'(c;\cdot) > 0,\ u''(c;\cdot) \leq 0$. Utility from no consumption at all is 0 irrespective of the state of health, $u(0;\cdot) = 0$ and utility in consumption satisfies the Inada conditions[20]. The agent's consumption is equal to the money available for consumption goods, i.e. a fixed income y which he receives in every period, less potentially insurance premia p and medical expenses m in case of sickness. Exerting health-enhancing effort μ is reducing his utility in the period it is exerted, which we phrase as health-enhancing effort cost. We assume effort costs to be additively separable from consumption utility and to be increasing and convex in the effort, i.e. $e'(\mu;\cdot) > 0,\ e''(\mu;\cdot) > 0, e(0,\cdot) = 0$. The effort costs are governed by the agent's moral hazard type g. This type can be interpreted as some characteristics of the individual, which make it harder for him to refrain from harmful behavior such as smoking and eating unhealthy food or similarly to engage in health-enhancing behavior such as exercising. In the current setting, we consider the moral hazard type g to be proxied by an agent's genotype, but it could also be interpreted as self-discipline or another individual characteristic. Higher predisposition to unhealthy behavior g leads to higher effort costs given the same level of effort, $\forall\mu$: $e(\mu; g') > e(\mu, g)$ if $g' > g$. Regardless of whether the agent knows his genetic make-up, we assume that he is at least partly aware of his general type in the sense that he knows how easy it is for him to exert self-discipline.

The benefit of health-enhancing effort is the reduced probability of getting sick in the second period, which depends linearly on effort $\lambda = \lambda_0 - \mu$. The ex-ante risk type of the agent λ_0 is equivalent to this probability, if no effort is exerted. The linear specification naturally bounds the maximal effort to $\mu \in [0, \lambda_0]$, which raises the question of boundary solutions, which we will address later when going through the steps of the model.[21]

E.2 The optimal health-enhancing effort

Since the game is finite, we solve it by backward induction. There is no decision to be made in the second period, the uncertainty about the health state just unfolds and the agent consumes what is left from his budget after paying his potential bills on insurance and medical expenditures.[22] In the first period, the agent decides about whether to be insured I or not, as well as the health-enhancing effort μ he is willing to exert. Hence,

[20] Inada conditions: $\lim_{c\to 0} u'(c;\cdot) = \infty,\ \lim_{c\to\infty} u'(c;\cdot) = 0$

[21] Another way to model the problem would be to allow for a more general functional form $\lambda(\mu)$. It is reasonable to assume risk to be decreasing in effort, i.e. $\lambda'(\mu)$. However, assumptions about convexity or concavity of the function are less straighforward and would require further supporting evidence. One could find examples supporting both increasing and decreasing marginal effects of effort on risk. Note that the results are robust to slight convexity or concavity. Extreme concavity might rule out an interior solution — the problem becomes a choice between no effort and maximal effort — while extreme convexity renders Proposition 1 is not as clear-cut.

[22] One could think additionally modeling the decision about the size of the medical expenditures, for example a choice between an expensive or a cheap treatment. This is what Einav et al. (2013) consider as moral hazard. Insured agents are more likely to choose the expensive treatment.

he maximizes the present value of his utility stream:

$$\max_{I,\mu} I \cdot V_I + (1-I) \cdot V_{NI}$$
$$V_I(\mu;\lambda_0,g) = u(y-p;H) - e(\mu;g) + \beta\left[(\lambda_0-\mu)u(y-p-\tau m;S) + (1-(\lambda_0-\mu))u(y-p;H)\right]$$
$$V_{NI}(\mu;\lambda_0,g) = u(y;H) - e(\mu;g) + \beta\left[(\lambda_0-\mu)u(y-m;S) + (1-(\lambda_0-\mu))u(y;H)\right]$$

where V_I is the value function when insured and V_{NI} the value function when not insured, β is the time discount factor, y is income, p is the insurance premium, m is the medical expenditure, τ is the coverage rate, H and S are the healthy and sick state respectively, μ is the amount of health-enhancing effort, g is the (genetic) moral hazard type, and λ_0 is the health risk type.

Proposition 1 *The agent exerts at least as much health-enhancing effort μ if he is insured compared to if he is not insured.*

Proof: Assuming an interior solution, if insured, the agent chooses optimal health-enhancing effort μ_I^* according to:

$$\frac{e'(\mu_I^*;g)}{\delta} = u(y-p;H) - u(y-p-\tau m;S)$$

If uninsured, the FOC can be rearranged to:

$$\frac{e'(\mu_{NI}^*;g)}{\delta} = u(y;H) - u(y-m;S)$$

Now irrespective of the specific form of the utility function,[23] $\mu_{NI}^* > \mu_I^*$. To see this more clearly consider subtracting (4) from (5):

$$\frac{e'(\mu_{NI};g) - e'(\mu_I;g)}{\delta} = [u(y;H) - u(y-m;S)] - [u(y-p;H) - u(y-p-\tau m;S)] \tag{24}$$

Effort in the uninsured state is larger than in the insured state, if the RHS of this expression is larger than zero (since effort costs are increasing and convex), which can be rearranged to:

$$u(y;H) - u(y-p;H) > u(y-m;S) - u(y-p-\tau m;S) \tag{25}$$

The LHS of this inequality is always positive, since $y > y-p$, while the RHS must be negative, since $m > p + \tau m$, otherwise, no agent ever chooses the insurance.[24] Consequently, the above inequality will always hold in an interior optimum.

Now, let's consider the boundary solutions. Since the marginal gains of exerting one extra unit of effort are constant, this consideration is rather simple. Suppose marginal cost at zero effort is higher than marginal gain in the case of no insurance: $\frac{e'(0)}{\delta} > u(y;H) - u(y-m;S) \Rightarrow \mu_{NI} = 0$. Then, we know that it must also be larger than

[23] In particular, irrespective of the interaction between health and income in determining the agent's utility, i.e. utility being sub-modular or supermodular in health and consumption.

[24] If $m \leq p + \tau m$, the insurance renders the agent weakly worse off in the sick state, while making him strictly worse off in the healthy state, since he has to pay the premium. A rational agent, irrespective of his idiosyncratic health risk, would never choose such an insurance.

marginal gain in case of insurance, i.e. $\mu_I = 0$. In this case, the health-enhancing effort is the same in both cases and the agent does not gain anything from being given the possibility to exert it. If the marginal cost at zero effort is below marginal gain in case of no insurance, but above marginal gain in case of insurance, the effort will be positive if not insured and thus higher than if not insured.
On the other bound, if marginal cost at exerting λ_0 effort is lower than marginal gain if insured, then the agent will exert maximal effort in both cases. Finally, the effort will again be higher in case of no insurance compared to insurance, if marginal costs $e'(\lambda_0)$ is lower than marginal gain in case of no insurance, but higher in case of insurance, $\mu_I < \mu_{NI} = \lambda_0$. □

For the remaining part, we will assume an interior solution.

Proposition 2 *Higher moral hazard parameter g leads to*

- *a smaller difference in health-enhancing effort μ between an agent having insurance and an agent not having insurance, all else equal.*
- *lower levels of effort when being insured and when not being insured respectively, compared to an agent with lower g and all else equal.*

Proof: As already mentioned, we assume that $\forall\mu$: $e(\mu; g') > e(\mu, g)$ if $g' > g$ and $\forall g : e(0, g) = 0$. Therefore it must be true that effort costs are supermodular in effort and genetic predisposition, i.e. $\frac{\partial e'(\mu;g)}{\partial g} > 0$. Consider (4) and (5) for two different values of genetic predisposition $g' > g$. Since the RHS of both equations is independent of g, it follows from the assumptions that $\mu_I(g) > \mu_I(g')$ and $\mu_{NI}(g) > \mu_{NI}(g')$. Moreover, note that the RHS of equation (24) is independent of the moral hazard parameter g. Hence, it must hold in equilibrium that:

$$e'(\mu_{NI}(g); g) - e'(\mu_I(g); g) = e'(\mu_{NI}(g'); g') - e'(\mu_I(g'); g')$$

But since $\mu_{NI}(g) > \mu_I(g)$, we know that $e'(\mu_{NI}(g); g) - e'(\mu_I(g); g) < e'(\mu_{NI}(g'); g') - e'(\mu_I(g'); g')$. Consequently, it must hold that $\mu_{NI}(g) - \mu_I(g) > \mu_{NI}(g') - \mu_I(g')$. □

Intuitively, Proposition 2 means, that agents for whom it is harder to engage in healthy or disengage in unhealthy behavior, e.g. starting to exercise or quit smoking, will react less to an increasing benefit to do so. Thus, a higher moral hazard parameter g coincides with less moral hazard because there is less leeway for the individual to adjust effort to insurance coverage. A higher moral hazard parameter also implies a lower effort of the agent, in both states of the world. This agent is *more* risky from the perspective of the insurance *ex-ante as well as ex-post* despite showing less reaction to insurance coverage.

Our empirical results outlined above can be considered an empirical counterpart to this proposition, leveraging the occurrence of a health shock to make the situation more salient to the individual.

Proposition 3 *Given genetic predisposition g, an agent with higher risk type λ_0 has a higher net benefit from being insured.*

Proof: Let us denote the present values of the optimal health-enhancing effort level given the moral hazard type g as $V_I(\mu_I(g); \lambda_0, g) = v_I^*(\lambda_0, g)$, $V_{NI}(\mu_{NI}(g); \lambda_0, g) = v_{NI}^*(\lambda_0, g)$. In period 1, the agent will choose to be insured, if $v_I^*(\lambda_0, g) > v_{NI}^*(\lambda_0, g)$.

Both expressions consist of parts that are dependent on the optimal health preventing effort and those that are not :

$$\begin{aligned} v_I^*(\lambda_0, g) = & \underbrace{u(y-p;H) + \delta\left[\lambda_0 u(y-p-\tau m;S) + (1-\lambda_0)u(y-p;H)\right]}_{G_I(\lambda_0)} \\ & \underbrace{-e(\mu_I(g);g) + \mu_I \delta\left[u(y-p;H) - u(y-p-\tau m;S)\right]}_{B_I(g)} \\ v_{NI}^*(\lambda_0, g) = & \underbrace{u(y;H) + \delta\left[\lambda_0 u(y-m;S) + (1-\lambda_0)u(y;H)\right]}_{G_{NI}(\lambda_0)} \\ & \underbrace{-e(\mu_{NI}(g);g) + \mu_{NI} \delta\left[u(y;H) - u(y-m;S)\right]}_{B_{NI}(g)} \end{aligned}$$

From equation (25), it follows that $v_I^*(\lambda_0, g) - v_{NI}^*(\lambda_0, g) = G_I(\lambda_0) - G_{NI}(\lambda_0) + B_I(g) - B_{NI}(g)$ is linearly increasing in health risk λ_0. □

Proposition 4 *Higher genetic predisposition g leads to higher adverse selection on risk type.*

Proof: For the lowest risk type $\lambda_0 = 0$ there is no effort he can exert to improve his health prospect, and he will never choose to be insured:

$$v_I^*(0, g) = u(y-p;H) + \delta u(y-p;H) < u(y;H) + \delta u(y;H) = v_{NI}^*(0, g)$$

Note that a risk type $\tilde{\lambda}_0$, who is indifferent between insurance and no insurance with zero health-enhancing effort, is *not* choosing the insurance with optimal effort: $v_I^*(\tilde{\lambda}_0, g) - v_{NI}^*(\tilde{\lambda}_0, g) < G_I(\tilde{\lambda}_0) - G_{NI}(\tilde{\lambda}_0) = 0$ Hence there exists a new threshold $\hat{\lambda}_0(g) > \tilde{\lambda}_0$, above which all agents choose to be insured given genetic predisposition.

From Proposition 4, we know that a higher moral hazard parameter g leads to a higher net benefit in being insured. To evaluate how the threshold risk type $\tilde{\lambda}_0(g)$ is changing in the moral hazard parameter, we can use the implicit function theorem on equation (8):

$$\frac{d\hat{\lambda}_0(g)}{dg} = -\frac{\overbrace{\frac{\partial(G_I(\hat{\lambda}_0(g), y) - G_{NI}(\hat{\lambda}_0(g), y))}{\partial \lambda}}^{>0}}{\underbrace{\frac{\partial(e(\mu_{NI}(g), g) - e(\mu_I(g), g))}{\partial g}}_{>0}}^{0} < 0$$

Consequently, if we have two groups of agents with genotype $g' > g$, then adverse selection will be less severe for the group with the higher moral hazard parameter $\hat{\lambda}_0(g_1) < \hat{\lambda}_0(g_2)$ and the lower reaction of effort to insurance. □

Note that because we did not further specify the insurance contract offered by the insurer, nothing prevents some thresholds to be above 1, i.e. no agent of a given genetic predisposition might choose the insurance. We assume that without health-enhancing

effort, the highest risk type $\lambda_0 = 1$ would always like to choose the insurance, i.e.[25]:

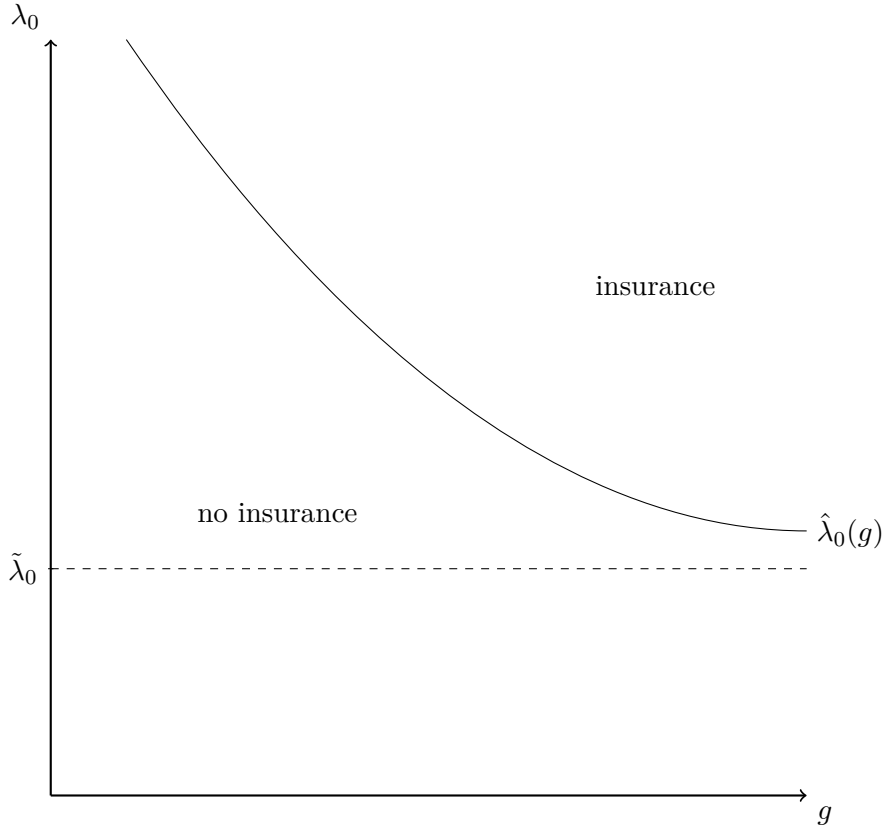
$$G_I(1) - G_{NI}(1) = u(y-p;H) + \delta u(y-p-\tau m;S) - u(y;H) - \delta u(y-m;S) > 0$$

This is equivalent to stating that there exists a $\tilde{\lambda}_0 \in (0,1)$. Now allowing for health-enhancing effort, it will depend on g, whether insurance is preferred to no insurance:

$$\begin{aligned} v_I^*(1,g) =& u(y-p;H) + \delta u(y-p-\tau m;S) + B_I(g) \\ & \gtrless u(y;H) + \delta u(y-m;S) + B_{NI}(g) = v_{NI}^*(1,g) \end{aligned}$$

Finally, if $\hat{\lambda}_0(g) \leq 1$, $\forall \lambda_0 \in \left[\hat{\lambda}_0(g), 1\right]$ the agent is choosing the insurance and $\forall \lambda_0 \in \left[0, \hat{\lambda}_0(g)\right)$ the agent is not choosing the insurance, all else equal. This is the usual adverse selection result: Agents with a higher risk of falling sick inflict more costs to the insurance, but also have a higher valuation for insurance. If $\hat{\lambda}_0(g) > 1$, which might happen for a very low level of g, we have the extreme case that no agent of a given moral hazard parameter g is choosing to be insured.

The following picture summarizes the relationship:



Over a population with independently distributed risk and moral hazard types we should expect to see agents of low risk type ($\lambda_0 > \lambda_0'$) choosing an insurance only if they have high costs of self-discipline ($g > g'$). Agents with high exogenous health risks should choose an insurance irrespective of their moral hazard type. Consequently,

[25] This is an assumption that needs to be justified by looking at the firm side later and solve for the market equilibrium of the problem

within a self-selected population with insurance we should observe a positive correlation of risk type and degree of moral hazard (or a negative correlation of risk type and moral hazard parameter g), despite advantageous selection on moral hazard, i.e. despite the fact that within each risk group, agents with the largest moral hazard potential do not choose the insurance.

While exogenous risk type λ_0 is an interesting variable to look at theoretically, it is neither the realized risk of the insured or uninsured agent. To compare the realized risk with and without health-enhancing effort, it is useful to divide the indirect utility when insured and uninsured into its components. We define $\hat{\lambda}(g) - \mu_{NI}(g)$ as the threshold ex-ante risk of an insured agent, which is what researchers usually consider the exogenous risk type on which adverse selection occurs.

$$\begin{aligned} v_I^*(\lambda_0, g) =& u(y-p;H) + \delta\left[(\lambda_0 - \mu_{NI}(g))u(y-p-\tau m;S) + (1-(\lambda_0-\mu_{NI}(g)))u(y-p;H)\right] \\ & - e(\mu_I(g);g) + (\mu_{NI}(g) - \mu_I(g)\delta\left[u(y-p-\tau m;S) - u(y-p;H)\right] \\ v_{NI}^*(\lambda_0, g) =& u(y;H) + \delta\left[(\lambda_0 - \mu_{NI}(g))u(y-m;S) + (1-(\lambda_0-\mu_{NI}(g)))u(y;H)\right] \\ & - e(\mu_{NI}(g);g) \end{aligned}$$

From this reformulation, it is easy to see that the relatively higher attractiveness of no insurance does not only result from the lower realized health risk $\lambda_0 - \mu_{NI}(g)$ compared to exogenous risk type λ_0. Also, the behavioral reaction to being covered[26] makes insurance *less* attractive. Moral hazard drives a wedge between the health risk if an agent is covered and the health risk if he is not covered, which can be interpreted as the agent's differential attempt to prevent the bad state of the world himself. Consequently, it must be true that $\hat{\lambda}_0(g) - \mu_{NI}(g) < \tilde{\lambda}_0$, but $\hat{\lambda}_0(g) - \mu_I(g) > \tilde{\lambda}_0$, so health-enhancing effort leads to an increase in realized risk among the insured agents given the same contract. This, in general, makes the adverse selection problem more severe. An insurer offering a fixed contract thus should increase premia in a world with a behavioral response of the agent compared to one without a behavioral response. which would exclude more agents from being insured. Notice, however, that not being insured has the positive side effect of the agents partly reduce risk themselves and staying healthy more likely.

[26] Note that $\mu_{NI}\delta(u(y-p;H) - u(y-p-\tau m;S)) - e(\mu_{NI}(g);g) < \mu_I\delta(u(y-p;H) - u(y-p-\tau m;S)) - e(\mu_I(g);g)$, since $\mu_I(g)$ is the maximum of the problem of the insured agent.