HEDG

HEALTH, ECONOMETRICS AND DATA GROUP

THE UNIVERSITY of York

WP 21/12

The Effect of Heavy Smoking on Early Retirement: An Instrumental Variable Approach

Alessio Gaggero; Olesya Ajnakina and Ruth A. Hackett

August 2021

http://www.york.ac.uk/economics/postgrad/herc/hedg/wps/

The Effect of Heavy Smoking on Early Retirement: An Instrumental Variable Approach

Alessio Gaggero[1], Olesya Ajnakina[2,3], Ruth A. Hackett[4]

[1] Departamento de Métodos Cuantitativos para la Economía y la Empresa, Facultad de Ciencias Económicas y Empresariales, Universidad de Granada, Granada, Spain. alessiogaggero@go.ugr.es

[2] Department of Biostatistics & Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, University of London, London, UK. olesya.ajnakina@kcl.ac.uk

[3] Department of Behavioural Science and Health, Institute of Epidemiology and Health Care, University College London, London, UK

[4] Health Psychology Section, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. ruth.hackett@kcl.ac.uk

***Corresponding author:** Alessio Gaggero. Departamento de Métodos Cuantitativos para la Economía y la Empresa, Facultad de Ciencias Económicas y Empresariales, Universidad de Granada, Granada, Spain. **Phone**: +34 958248344. **Email**: alessiogaggero@go.ugr.es.

Word count: 6024
Number of text pages of manuscript (including tables): 18
Number of tables: 6

Funding: The English Longitudinal Study of Ageing is funded by the National Institute on Aging (Grant: RO1AG7644) and by a consortium of UK government departments coordinated by the Economic and Social Research Council. O.A. is further funded by the National Institute for Health Research (PDF-2018-11-ST2-020).
Role of the funding sources: The funders had no role in the design or preparation of this paper.
Declarations of interest: none

Abstract

The extent to which heavy smoking and early retirement are causally related remains to be determined. To overcome the endogeneity of heavy smoking behaviour, we employ a novel approach by exploiting Mendelian Randomisation and use genetic predisposition to heavy smoking, as measured with a polygenic risk score (PGS), as an instrumental variable. A total of 3578 participants from the English Longitudinal Study of Ageing (mean age 64.41 years) had data on smoking behaviour, employment and a heavy smoking PGS. Heavy smoking was indexed as smoking at least 20 cigarettes a day. Early retirement was classified as retiring before state pension age. Our results show that being a heavy smoker increases significantly the probability of early retirement ($\beta = 0.635$, standard error = 0.209, $p < 0.001$). Results were robust to a battery of robustness checks and a falsification test. Overall, our findings support a causal pathway from heavy smoking to early retirement.

Keywords: Smoking; Early Retirement; Polygenic Risk Scores; Instrumental variable; Mendelian Randomisation

1. Introduction and Motivation

Smoking is widely known to be damaging to health. It greatly increases the risk of various cancers, cardiovascular and respiratory diseases (ASH, 2020; Murray, 2014), and remains the leading preventable cause of morbidity and mortality in the United Kingdom (UK) (ASH, 2020). Although there is no risk-free level of smoke exposure, research suggests that heavy smokers have a particularly elevated risk of negative smoking-related health outcomes (Murray, 2014).

There is increasing interest in the link between health and labour market outcomes, as observational research from Europe and North America indicates that poor health is a key predictor of labour force exit (Fisher et al., 2016; Round, 2017). Pooled evidence from 29 longitudinal studies suggests that poor health is a risk factor for exit from paid employment through disability pension, unemployment and early retirement (van Rijn et al., 2014). Early retirement represents a considerable economic challenge as governments pursue policies to extend working lives in an attempt to improve the financial sustainability of pension schemes (OECD, 2019). Therefore, there is a need to understand the health-related determinants of early retirement. As health behaviours, such as smoking, are modifiable (Hackett et al., 2018; West et al., 2015), it is crucial to understand how they influence early retirement from the labour force.

In the UK, an estimated 7.8% of those aged 65 and older smoke and these smokers are less likely to be economically active than non-smokers (Office for National Statistics, 2020). Emerging longitudinal evidence suggests that smoking may be a predictor of early exit from the labour force. Occupational cohort data indicates that there is a dose-response relationship between the number of cigarettes smoked per day and early disability retirement in German construction workers (Claessen et al., 2010; Rothenbacher et al., 1998). Evidence from more representative Scandinavian general population cohorts further suggests that smoking is associated with increased risk of being granted a disability pension (Haukenes et al., 2013; Husemoen et al., 2004; Lallukka et al., 2015). Heavy smoking was the most robust predictor of disability retirement in one study (Lallukka et al., 2015) but similar associations were only seen in those aged under 60 (Husemoen et al., 2004) and in women only (Haukenes et al., 2013) in other studies. While data from 11 European studies found no overall association between current smoking and exit from the labour market with a disability pension over 4-year follow-up (Robroek et al., 2013). Sensitivity analyses found that smoking was only significantly related with labour market exit in Scandinavia, but not in other regions.

Although these studies have collected data across several years and have assessed smoking and the subsequent occurrence of retirement, the possibility of reverse causality, whereby retirement could influence smoking behaviour, cannot be ruled out. A recent review that included 14 longitudinal studies investigating the influence of retirement on smoking had mixed findings, with decreased smoking as well as no effect on smoking mainly reported (Xue et al., 2020). In contrast, 2 studies reported retirement was associated with an increased probability of smoking (Xue et al., 2020), including analysis of almost 12,000 individuals from the US-based Health and Retirement Study (Ayyagari, 2016).

Another limitation of multivariable conditional correlational analyses is that they are subject to residual confounding. For example, the association between smoking and labour market exit may be due to confounding by socio-economic status (Pietikäinen et al., 2011). The findings of earlier studies have been attenuated by the inclusion of indicators of socio-economic status, such as education and occupation (Bengtsson and Nilsson, 2018; Haukenes et al., 2013; Husemoen et al., 2004; Lallukka et al., 2015). However, some important confounders, such as household income which is a robust indicator of socio-economic status in older adults (Steptoe et al., 2013), were not accounted for in these studies. In turn, the omitted variable may have also contributed to the potential confounding.

Another possibility is that there may be confounding by other health behaviours as health behaviours tend to cluster (Mawditt et al., 2016), such that those who smoke are more likely to be physically inactive than those who do not smoke. Two previous studies (Eriksen et al., 1998; Lallukka et al., 2015) found that current smoking was only predictive of disability pension risk in those who were less physically active. This offers the possibility that smoking is an indicator of a number of other health-related factors that increase the likelihood of exit from the labour market, rather than a causal predictor of retirement in of itself.

To resolve these issues, investigators have turned to more sophisticated approaches to analyse the available observational data. Genetic factors influence smoking behaviour (Maes et al., 2004; Tobacco and Genetics Consortium, 2010) and several studies have attempted to account for unobserved factors related to family background by using sibling and twin pair designs. The largest study to date estimated the impact of smoking on disability retirement in over 80,000 Swedish sisters (Bengtsson and Nilsson, 2018). The authors found a strong association between current smoking and disability retirement. This was attenuated but remained significant when accounting for sibling effects. The relationship between heavy smoking and disability retirement was also assessed in the study. Significant associations were

detected in analyses accounting for sibling effects but reverted towards the null when education and occupation were included in models.

Earlier work using the Finnish Twins Cohort had the advantage of including both male and female participants and may have been more fully able to account for in the influence of genetic and shared environmental factors in the link between smoking and disability retirement (Korhonen et al., 2015; Koskenvuo et al., 2011; Ropponen et al., 2013). One of these studies investigated smoking and disability retirement for any condition (Korhonen et al., 2015) rather than for retirement due to particular diseases (Koskenvuo et al., 2011; Ropponen et al., 2013). This analysis of almost 22,000 participants found a dose response relationship between the number of cigarettes smoked and the likelihood of receiving a disability pension in men and women. This association was replicated in analyses within twin pairs discordant for the outcome (i.e., one twin got the pension, and the other did not). This provides additional evidence that heavy smoking could be a causal contributor to the chance of receiving a disability pension.

Although studies using a familial design are likely to come closer to causal estimates of the impact of smoking on disability retirement, we cannot be certain all relevant factors have been accounted for (e.g., aspects of the home environment that might influence smoking behaviour or access and use of social (pension) services). Further those who are part of a twin pair or have siblings may differ from those who are only children.

Technological advances, such as the advent of genome-wide association studies (GWAS) has made it possible to investigate genetic variants across the entire genome for association with various traits measured on unrelated individuals (Dudbridge, 2013). GWAS aim to identify individual genetic markers which are significantly linked with a trait of interest. However, many complex traits are polygenic in nature, implying their onset cannot be attributed to the independent contributions of individual genetic markers, but rather to the combined additive effects of multiple common genetic traits (So and Sham, 2017). Heavy smoking is one such 'polygenic' trait (Tobacco and Genetics Consortium, 2010). This has led to the creation of polygenic scores (PGS), which reflect the mathematical aggregate of risk conferred by many genetic variants of small effect into a single continuous score that represents an individual load for the common variants associated with a particular trait.

PGS of robust genetic variants associated with heavy smoking have been developed in the English Longitudinal Study of Ageing (ELSA; Ajnakina and Steptoe, 2019). This offers the possibility of a using Mendelian Randomisation approach (Davey Smith and Hemani, 2014),

whereby genetic predisposition to heavy smoking can be used an instrumental variable to test the inferred causality of heavy smoking on early retirement.

The current study will take this novel form of instrumental variable (IV) analysis that is beginning to be applied in the field of health economics (Dixon et al., 2020). We will use heavy smoking-related genetic variants (PGS) as an unconfounded IV for heavy smoking behaviour. The rationale for our approach is that as different smoking-associated genotypes are randomly allocated at conception, the associations of smoking-related variants with early retirement should be free of confounding and reverse causation. In principle this approach will avoid the limitations of both multivariable conditional correlational analyses (smoking predicting early retirement adjusting for confounders) and analyses accounting for familial effects. We will test the hypothesis that heaving smoking is a casual factor leading to early retirement, using PGS as an instrument in a community dwelling sample of middle-aged and older adults living in England

Employing ELSA data, our two-stage least squares estimates reveal that heavy smokers are significantly more likely to retire before achieving state pension age than their light smoking counterparts. Specifically, our estimates suggest that individuals who smoke more than 10 cigarettes a day have a 61% increased probability of retiring before state pension age than those who smoke less than this. These estimates are robust to the inclusion of various control variables (e.g., age, income, and education), the inclusion of individual smoking genes, as well as different definitions of being a "heavy smoker". Notably, to strengthen the validity of our approach, we propose a falsification test using a sample of non-smokers which supports the causal interpretation of our findings.

The paper proceeds as follows: Section 2 introduces the data source and includes descriptive statistics. Section 3 describes the methods. Section 4 presents the results and Section 5 is the conclusion.

2. Data

2.1 Study cohort

In this study, we use data from ELSA, which is a large-scale longitudinal panel study of people aged 50 and over, and their partners, living in private households in England. The original sample was drawn in 2002-03 (Wave 1) from households that had previously responded to the Health Survey for England (HSE). Every two years, the sample is surveyed to measure changes in their health status, in their economic conditions, and in their social

circumstances. The latest available wave of data collection took place in 2018-19 (Wave 9). In order to increase the power of our study, we pooled every available wave of data (Wave 1 - Wave 9).

Importantly for our investigation, ELSA provides PGS data for a number of behavioural, emotional and health-related phenotypes. Specifically, ELSA participants of European ancestry were genotyped in 2013/14, using the Illumina HumanOmni2.5 Bead-Chips (HumanOmni2.5-4v1, HumanOmni2.5-8v1.3).[1] Principal components analysis was performed to investigate population structure, and ten principal components were retained to account for any ancestry differences in genetic structures (Price et al., 2006). A total of 7183 samples and 1,372,240 genetic variants remained after quality control.

In this study, we are specifically interested in the PGS for heavy smoking behaviour, which was informed by the GWAS summary statistics from the Tobacco and Genetics (TAG) Consortium (Tobacco and Genetics Consortium, 2010). The calculation of this score has been described in detail elsewhere (Ajnakina et al., 2020; Ajnakina and Steptoe, 2019). In brief, this PGS represents the weighted sum of cumulative genetic risk for heavy smoking, calculated by aggregating multiple individual loci associated with number of cigarettes per day across the human genome and weighting them by their corresponding effects sizes derived from the TAG summary statistics. The resulting continuous PGS for number of cigarettes per day can be thought to measure the genetic predisposition towards heavy or light smoking. To ease interpretation of the results, the PGS was standardised ($N\sim(0,1)$), and its distribution is depicted in Figure A.1 in the Appendix. In sensitivity analyses, we additionally assessed CHRNA5 rs16969968 and CHRNA3 rs10417309 allele carrier status as IVs, as a large genetic component of heavy smoking has been attributed to these variants (Leung et al., 2015).

Our outcome variable of interest is early retirement. We defined this variable as a dummy which equals to unity (1) if the individual retired before reaching state pension age, and zero otherwise. In line with UK pension regulations, for ELSA waves before 2010 (Waves 1-4) we

[1] A full quality control protocol has been described in https://www.elsa-project.ac.uk/genetics. Briefly, individuals with suspected non-European ancestry and heterozygosity scores >3 standard deviations from the mean were removed. Furthermore, initial quality control measures were conducted to test for duplicates and missingness of more than 2% of the genotype data. Single Nucleotide Polymorphisms (SNPs) with a call rate of <98%, a minor allele frequency of <0.01%, and Hardy-Weinberg Equilibrium p values of <10−4 were excluded. Non-autosomal markers were also removed, as well as regions known to contain clusters of highly correlated SNPs, as these can bias the analyses.

considered state pension age to be 65 and 60 for the sample of men and women, respectively, and considered state pension age to be 65 for both men and women after 2010 (Waves 5-9).

The key regressor of interest is a dummy which identifies heavy smokers. In line with the World Health Organization definition of heavy smoking we coded this variable to take the value of one for individuals who reported smoking at least twenty cigarettes per day.[2] We included a number of covariates in our analyses. Age was measured in years and was entered as both a continuous variable as and as a quadratic term (age^2) to account for the non-linear influence of age on early retirement. Gender was coded as binary (male/female). Education was coded as a binary variable based on whether an individual had obtained a higher education qualification. Marital status was entered as a binary variable (married vs divorced, separated, widowed, never married). Family size was measured as the number of people belonging to the household. Finally, household income was measured using the log-yearly equivalised disposable real household income deflated using the Consumer Price Index with baseline 2005 = 100.

[TABLE 1 HERE]

2.2 Characteristics of the sample

The characteristics of the sample of interest are summarised in Table 1. As we were specifically interested in smoking behaviour, namely heavy versus light smoking, we restricted the sample to always smokers. After excluding non-smokers and observations with missing values on the variables used in the analysis, our final study sample size was 3578 observations.

Table 1 shows that in our sample of smokers, 31 percent reported being a heavy smoker, and that on average the participants smoked around 0.7 packs of cigarettes per day. In terms of sociodemographic characteristics, the average age in the sample was 64.4 years (standard deviation (SD) = 8.57).[3] The majority of the sample were female (66 percent) and married (56 percent), and the average household size is around 2 (mean= 1.94; SD= 0.94). Few participants in the sample (8 percent) reported having obtained higher education. Finally, the (log-equivalised) household income was 5.45 (SD= 0.71).[4]

[2] In Table A.6, in the Appendix, we show that the results are consistent when using different cut points for heavy smoking (between 12 and 20 cigarettes a day)

[3] Individuals aged below 50 are partners of main participants.

[4] Table A.1. in the Appendix reports the difference between heavy and light smokers on these key variables.

3. Methods

In what follows, we show the effect of heavy smoking on early retirement, as a benchmark, by means of the naïve estimator, and then we present estimated effects of the IV approach that overcomes the potential omitted variable bias.

3.1 Naïve Estimator

Let R_i be the outcome variable of interest taking the value of one if individual i has retired before reaching state pension age and let H_i be an indicator function taking the value of one for heavy smokers. We first estimate the effect of heavy smoking on early retirement by means of the following naïve regression model:

$$R_i = \alpha + \beta H_i + \boldsymbol{X_i'}\gamma + \varepsilon_i \quad (1)$$

In this context, β is the main term of interest, as it measures the effect of heavy smoking on early retirement. $\boldsymbol{X_i'}$ is a set of covariates that might affect the early retirement as described in Section 2. Additionally, we include a set of regional and time fixed effects to account for regional differences in retirement and macroeconomic differences between waves. Finally, ε_i is the random error term.

However, the reliability of these estimates rests on the conditional independence assumption, that is, conditional on the vector of covariates we control for, $\boldsymbol{X_i'}$, the error term is uncorrelated with the regressor of interest, namely, $E[\varepsilon_i|H_i, \boldsymbol{X}_i] = 0$. This may be too strong an assumption. In the context of our study, for example, it is plausible that because of the impact of smoking on health, heavy smokers have systematically poorer health. In this case a naïve comparison of early retirement could lead to an overestimation of the true causal effect. On the other hand, the effect may be underestimated if heavy smokers are those more work orientated and less likely to early retire. In the following analyses we address this issue.

3.2 Instrumental Variable (IV) Approach

In an ideal experiment, given a sample of smokers, we would randomly assign a group of individuals to a heavy smoking regime (≥ 20 cigarettes per day) and assign the remaining to a light smoking regime (<20 cigarettes per day), and observe differences in early retirement between the two groups. While such experiment is not feasible for ethical reasons, our IV approach mimics this ideal experiment by exploiting a Mendelian Randomisation framework.

The key idea of our approach is that, at conception, different heavy smoking-associated genotypes are randomly allocated to offspring. Accordingly, individuals can be characterised by their level of genetic predisposition toward heavy smoking.

[FIGURE 1 HERE]

In Figure 1, we present graphical evidence of the relationship between the heavy smoking PGS and actual heavy smoking behaviour. The Figure shows a strong and positive relationship between the PGS and the number of cigarettes smoked per day. This observation is in line with the idea that a great level of these genetic variants (as indexed by the PGS) are strongly associated with the number of cigarettes smoked per day.

The main assumption an IV must satisfy is exclusion restriction, namely the instrument should not have any direct effect on the outcome variable of interest but should only have an effect through the main regressor. In our case, exclusion restriction is threatened if the same genetic variants associated with heavy smoking have either a direct effect on early retirement or have an effect toward other lifestyle behaviours that can, in turn, affect the decision to retire before reaching state pension age.

To address this concern, in Table 2 we provide evidence that this is unlikely to be the case by running a series of models of PGS on key psychosocial factors and lifestyle behaviours that could potentially be linked to early retirement, namely: (1) Quality of life, as measured by the Control, Autonomy, Self-Realization and Pleasure (CASP) 19 scale (Hyde et al., 2003); (2) Physical function, as measured using a 10-item Body Mobility Index (Steel et al., 2002) ; (3) depressive symptoms, as measured by the Centre for Epidemiological Studies – Depression (CES-D) scale (Radloff, 1977); Executive function, as measured with a verbal fluency test: (5) drinking behaviour, defined as whether an individual reported drinking alcohol every day; and finally (6) sedentary behaviour, as measured by whether the individual reports not doing any type of physical activity on a regular basis. The results in Table 2 report that the estimated coefficients are all statistically indistinguishable from 0. This is consistent with the idea that, the heavy smoking genetic variants (indexed by the PGS) only affect heavy smoking behaviour, and no other psychosocial and behavioural factors assessed here.

[TABLE 2 HERE]

Overall, these pieces of evidence suggest that the PGS is a valid instrument that can be used in our setting, and in what follows we provide further evidence of this. To see our identification strategy analytically, let P_i denote the PGS of individual i; we estimate a standard two-stage least squares (2SLS) with the following model:

$$H_i = \mu + P_i\pi + \boldsymbol{X_i'}\Omega + \upsilon_i \quad (2)$$

$$R_i = \alpha + \beta\widehat{H}_i + \boldsymbol{X_i}'\gamma + \varepsilon_i \quad (3)$$

Where $\widehat{H}_i$ is the heavy smoking dummy as predicted from Equation (2), namely the first stage. As above, $\boldsymbol{X_i'}$ is a vector of covariates and γ its vector of associated coefficients. These include the same control variables as before. Importantly, in addition we now include a set of 10 genetic principal components, in order to take into account the potential confounding effect of genetic ancestry. Finally, υ_i is the error term, uncorrelated with ε_i, namely $E[\varepsilon_i, \upsilon_i | Z_i, X_i]$= 0. We cluster the standard error at the individual level because some individuals appear in the regression in multiple time periods, but results are consistent when using different cluster types.

4. Results

4.1. Main Results

Table 3 shows naïve estimates of the effect of heavy smoking on early retirement based on Equation (1). In Column (1), we report the unconditional effect of heavy smoking on early retirement. In Column (2), we include the full set of covariates as described above. In Columns (3) and (4), we include time and region fixed effects, respectively. According to the naïve estimator, the results suggest that on average, heavy smokers are less likely to retire before reaching state pension age than their light smoking counterparts. As mentioned above, however, these estimates are likely to suffer from omitted variable bias.

[TABLE 3 HERE]

In Table 4 we report estimates coming from our IV approach. In Panel A we report estimates of the effect of the PGS on the probability of being a heavy smoker, and in Panel B we report our IV estimates on the effect of heavy smoking behaviour on early retirement. As above, we show that our results are robust to the inclusion (or exclusion) of different control variables

from the model. Specifically, for each Panel, in Column (1), we report the unconditional effect of heavy smoking on early retirement (unadjusted analyses). In Columns (2), we include the full set of covariates as described above. In Columns (3), we include time and region fixed effects, and in Column (4) we include the principal components to account for genetic ancestry.

Consistent with Figure 1, Panel A of Table 4 supports a strong and significant first stage. Specifically, the estimated coefficient in Column (4) implies that if an individual were to be reassigned a PGS one standard deviation larger, the probability of being a heavy smoker would increase by 3.8 percent (β= 0.038, standard error = 0.08, $p < 0.001$). This is in line with the notion, that individuals with a higher genetic predisposition towards heavy smoking are more likely to be classified as heavy smokers than their light smoking counterparts.[5]

In Panel B of Table 4 we report the main findings of the paper.[6] The Table reports a strong effect of heavy smoking behaviour on the probability of early retirement. The estimated coefficient in Column (4), our preferred specification, implies that heavy smokers are around 63.5 percent more likely to retire before state pension age than their light smoking counterparts (β= 0.635, standard error = 0.209, $p < 0.001$). These findings suggest that the naïve estimates were suffering from a significant omitted variable bias. Overall, our findings provide evidence of a strong and negative effect of heavy smoking behaviour on early retirement.

[TABLE 4 HERE]

4.2. Sensitivity Analysis and Placebo

Our results are robust to a battery of checks, including the inclusion/exclusion of a number of different covariates, as well as different cut-points used to classify heavy smokers (e.g., >10 cigarettes a day rather than >20 cigarettes a day, as in our main analysis, see Table A.6 in the appendix). In addition, In Table 5 we report two further robustness checks in Panel B and C, while in Panel A we present the benchmark estimates (the results from our main analysis) for reference. In Panel B, we present the sensitivity analyses of our estimates when we include additional genes that have been linked to smoking, namely the CHRNA3 and the CHRNA5 gene, as instruments in the analysis.[7] As can be seen in the results are slightly attenuated but remain similar. In Panel C, we show that our results are directionally consistent when we

[5] Table A.2 in the Appendix reports the full table.

[6] Table A.3 in the Appendix reports the full table

[7] Table A.4 in the Appendix reports the full table.

consider the number of packs of cigarettes smoked per day as main regressor of interest, as opposed to the binary indicator for heavy smoking (≥20 or more cigarettes per day vs < 20 cigarettes per day).[8]

[TABLE 5 HERE]

In order to further strengthen the validity of our results, we conclude this section by reporting the results of a falsification test. As mentioned above, the main threat to our identification is that exclusion restriction criteria are not satisfied, namely that the heavy smoking PGS could have an effect on early retirement through channels other than heavy smoking behaviour. A natural falsification test is to look at the sample of non-smokers. Specifically, if exclusion restriction is satisfied, we should not observe any direct effect of the PGS on early retirement for the sample of non-smokers. Table 6 reports the results of this exercise. In each Column of the Table, the estimated coefficients are all statistically indistinguishable from 0 and suggesting that exclusion restriction criteria is satisfied in our setting and, consequently, support a causal interpretation of our findings.

[TABLE 6 HERE]

5. Conclusion

To the best of our knowledge, this is the first study to investigate the link between heavy smoking and early retirement using a Mendelian Randomisation approach. We used genetic predisposition to heavy smoking (indexed by a PGS) as an IV to test the inferred causality of heavy smoking on early retirement. We found in our community dwelling sample of adults living in England that heavy smokers are significantly more likely to retire before reaching state pension age than their light smoking counterparts. Specifically, our results suggest that middle aged and older adults who smoke 20 or more cigarettes a day have a 63.5% increased likelihood of early retirement than those who smoke less than this. These findings were robust to adjustment for range of covariates including age, gender, marital status, household size, education, household income and genetic ancestry. Sensitivity analyses applying different cut-

[8] Table A.5 in the Appendix reports the full table.

points for heavy smoking and the inclusion of individual variants related to smoking behaviour did not change the pattern of results.

The rationale behind our chosen method was that as different smoking-related genetic variants are randomly allocated at conception, the association between the heavy smoking PGS and early retirement should be free of reverse causation and confounding. Our heavy smoking PGS, met the assumptions required for an IV analysis. Firstly, the PGS was strongly and positively associated with heavy smoking behaviour in the sample. Secondly, our outcome, retirement before state pension age, cannot plausibly affect the allocation of smoking-related genetic variants. This meant our IV analysis was more protected from reverse causality than conventional correlational analyses (i.e., heavy smoking behaviour predicting early retirement). The final assumption an IV must satisfy is exclusion restriction, namely that except for its association with the risk factor of interest, there is no other pathway linking the IV with the outcome of interest. In the current study, exclusion restriction would be threatened if the heavy smoking PGS effected health behaviours (other than heavy smoking) or psychosocial factors that could, in turn, influence the decision to take early retirement. This assumption was also met, as our PGS was not associated with health behaviours such as alcohol consumption and sedentary behaviour or psychosocial factors such as depression and poor quality of life, which could influence the decision to retire early. Further, the observation from our natural falsification test, that the heavy smoking PGS does not impact early retirement for non-smokers, supports the notion that the exclusion restriction assumption was met.

Our results add to existing observational evidence that has assessed whether smoking behaviour is a predictor of labour market exit. The majority (Claessen et al., 2010; Haukenes et al., 2013; Husemoen et al., 2004; Lallukka et al., 2015; Rothenbacher et al., 1998), but not all (Robroek et al., 2013) previous studies report a positive association between the number of cigarettes smoked daily and early exit from the labour force. However, it is possible that these findings may be at least partially confounded by socio-economic status (Pietikäinen et al., 2011), as attenuation by education and occupation was commonly reported (Haukenes et al., 2013; Husemoen et al., 2004; Lallukka et al., 2015). Confounding by other health behaviours might also be an issue, as two earlier studies (Eriksen et al., 1998; Lallukka et al., 2015) found that smoking was only associated with disability pension risk in participants who were physically inactive. This might suggest that smoking is an indicator of other health-related factors (Mawditt et al., 2016), that increase the probability of early exit from the labour market, rather than a causal predictor in of itself. The results of the current analysis are less likely to be influenced by confounding due to measures of socioeconomic status and physical inactivity

due to our novel IV methodology. Our findings were robust to adjustment for education and household income and our IV (heavy smoking PGS) was only associated with early retirement through smoking behaviour and was not associated with sedentary activity.

Several studies have specifically assessed heavy smoking in relation to retirement with a disability pension (Haukenes et al., 2013; Husemoen et al., 2004; Lallukka et al., 2015). A study of over 6000 Finnish adults found that heavy smoking women and men (classified as those smoking more than 15 or 20 cigarettes per day, respectively) had an increased risk of disability retirement (Lallukka et al., 2015). However, in other heavy smoking studies similar associations were only observed in those aged under 60 (Husemoen et al., 2004) and in women alone (Haukenes et al., 2013). The results of our current analysis were robust to adjustment for age and gender and are less likely to be impacted by this type of confounding due to the IV approach undertaken. Taken together, the findings of the current study add to the existing observational literature by demonstrating that the association between heavy smoking behaviour and early retirement is likely to be causal.

We are not the first study to use unobserved genetic factors in an attempt to improve understanding of the relationship between smoking and labour market exit. A Swedish study of 80,000 sisters (Bengtsson and Nilsson, 2018) found an association between current smoking and disability retirement when taking sibling effects into account. The authors also assessed the relationship between heavy smoking and disability retirement. Significant associations were detected when accounting for sibling effects but reverted towards the null when education and occupation were included in models. The authors state that the measures of socio-economic status and marital status in their study could have been influenced by smoking. The results of the current study are less likely to suffer from this issue, as the genetic variants for heavy smoking allocated at conception which inform our analyses are unlikely to influence subsequent socio-economic status or marital status. Another possibility is that an omitted variable may have contributed to the reported confounding. For example, household income which is a reliable indicator of socio-economic status in older adults (Steptoe et al., 2013) was not controlled for in this Swedish study (Bengtsson and Nilsson, 2018). In the present analysis, our results were robust to adjustment for education, marital status, and household income, and were less likely to have been influenced by omitted variable bias due to the IV approach employed.

Our findings are in agreement with earlier work from the Finnish Twins Cohort, whereby an analysis of almost 22,000 men and women found a dose response relationship between the number of cigarettes smoked and the probability of receiving a disability pension

(Korhonen et al., 2015). This result was replicated in analyses within twin pairs discordant for the outcome (i.e., one twin got the pension, and the other did not), which suggests that heavy smoking is likely a causal contributor to the chance of receiving a disability pension. The results of the present study add to this casual evidence by demonstrating an association between heavy smoking and early retirement in unrelated individuals.

Our study needs to be assessed in light of its strengths, policy relevance and weaknesses. In terms of strengths, firstly, studies examining smoking and early retirement in the context of genetics are limited, and thus our results contribute to an important, yet sparse, literature. Secondly, Mendelian Randomisation is a powerful control for confounding and reverse causation, which often impede studies with a prospective observational design. Thirdly, our results reflect lifelong exposure to heavy smoking genetic variants rather than the temporary effect of current light or heavy smoking captured in observational studies, based on the assumption that the association between genetic variations and the relative effect of heavy smoking does not change with age (Bengtsson and Nilsson, 2018; Dixon et al., 2020; Holmes and Smith, 2017).

Our results are likely to be policy relevant as smoking is a leading cause of poor health (ASH, 2020; Murray, 2014) and in tandem early retirement represents a considerable economic challenge for the financial sustainability of pension schemes (OECD, 2019). The rationale behind the increased statutory pension age is that longer life expectancy will enable people to extend their working lives (OECD, 2019). However, as poor health is a key predictor of labour market exit (Fisher et al., 2016; Round, 2017) there is a need to understand health-related determinants of early retirement, particularly those of which are amenable to change. Our study provides novel evidence that heavy smoking is likely a causal factor influencing retirement before state pension age. As smoking is a potentially modifiable behaviour (Hackett et al., 2018; West et al., 2015), this suggests that policies targeting reductions in smoking could plausibly reduce rates of early retirement (amongst other impacts on morbidity and mortality).

Our results reflect the average lifetime effects of heavy smoking (randomly determined at conception) rather than current smoking, so allow for the possibility that individuals could quit smoking as they age. It is possible that workers may be willing to change their health behaviours if they expect to retire later (Bertoni et al., 2018) and policy research on extending working lives recommends that health interventions be integrated to the workplace (Round, 2017). Most adults spend a large proportion of their time at work, and activities that start before workers are considering retirement are thought to be particularly effective as they allow healthy habits to develop and be maintained (Loeppke et al., 2013). Our findings also add justification

for policymakers' attempts to reduce smoking through smoking bans in public places (such as restaurants, bars, theatres, and cinemas) where adults may spend their leisure time.

This study is not without limitations. Although the ELSA sample is representative of the English population, the wider generalisability of genetic studies across populations is limited (Martin et al., 2019). This is because the method for computing PGS depends on summary statistics from GWAS focused almost exclusively on participants of European ancestry. By design, PGSs do not capture other structural variants beyond common genetic markers of relatively small effects, such as rare variants, poorly tagged or multiple independent variants, gene-by-gene interactions and gene-environment correlation (Reynolds and Finkel, 2015). Further, our measure of retirement before state pension age was based on self-report data rather than linkage with official records, although self-reported and register pension data are reported to have good concordance (Svedberg et al., 2010). Similar studies should be replicated with larger sample sizes, across multiple ancestries and should include linkage to pension records to improve our understanding of the association between heavy smoking and early retirement.

In summary, to our knowledge this is the first study to adopt a MR approach to examine the association between heavy smoking and early retirement. Although our study does not provide a definitive answer to the complexities in the relationship between smoking and retirement before state pension age, it adds a novel component to an emerging literature using genetically sensitive designs and suggests that the relationship between heavy smoking and early retirement is likely causal.

CRediT authorship contribution statement

Alessio Gaggero: Conceptualization; Formal analysis**;** Roles/Writing - original draft; Writing - review & editing

Olesya Ajnakina: Writing - review & editing

Ruth A Hackett: Roles/Writing - original draft; Writing - review & editing

References

Ajnakina, O., Cadar, D., Steptoe, A., 2020. Interplay between Socioeconomic Markers and Polygenic Predisposition on Timing of Dementia Diagnosis. J. Am. Geriatr. Soc. 68, 1529–1536. https://doi.org/10.1111/jgs.16406

Ajnakina, O., Steptoe, A., 2019. The English Longitudinal Study of Ageing (ELSA) Polygenic Scores.

ASH, 2020. Action on Smoking and Health (ASH) Factsheet: Smoking Statistics.

Ayyagari, P., 2016. The Impact of Retirement on Smoking Behavior. East. Econ. J. 42, 270–287. https://doi.org/10.1057/eej.2014.51

Bengtsson, T., Nilsson, A., 2018. Smoking and early retirement due to chronic disability. Econ. Hum. Biol. 29, 31–41. https://doi.org/10.1016/j.ehb.2017.12.005

Bertoni, M., Brunello, G., Mazzarella, G., 2018. Does postponing minimum retirement age improve healthy behaviors before retirement? Evidence from middle-aged Italian workers. J. Health Econ. 58, 215–227. https://doi.org/10.1016/j.jhealeco.2018.02.011

Claessen, H., Arndt, V., Drath, C., Brenner, H., 2010. Smoking habits and occupational disability: a cohort study of 14,483 construction workers. Occup. Environ. Med. 67, 84–90. https://doi.org/10.1136/oem.2009.046318

Davey Smith, G., Hemani, G., 2014. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum. Mol. Genet. 23, R89-98. https://doi.org/10.1093/hmg/ddu328

Dixon, P., Hollingworth, W., Harrison, S., Davies, N.M., Davey Smith, G., 2020. Mendelian Randomization analysis of the causal effect of adiposity on hospital costs. J. Health Econ. 70, 102300. https://doi.org/10.1016/j.jhealeco.2020.102300

Dudbridge, F., 2013. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 9, e1003348. https://doi.org/10.1371/journal.pgen.1003348

Eriksen, W., Natvig, B., Rutle, O., Bruusgaard, D., 1998. Smoking as a predictor of long-term work disability in physically active and inactive people. Occup. Med. Oxf. Engl. 48, 315–320. https://doi.org/10.1093/occmed/48.5.315

Fisher, G.G., Chaffee, D.S., Sonnega, A., 2016. Retirement Timing: A Review and Recommendations for Future Research. Work Aging Retire. 2, 230–261. https://doi.org/10.1093/workar/waw001

Hackett, R.A., Moore, C., Steptoe, A., Lassale, C., 2018. Health behaviour changes after type 2 diabetes diagnosis: Findings from the English Longitudinal Study of Ageing. Sci. Rep. 8, 16938. https://doi.org/10.1038/s41598-018-35238-1

Haukenes, I., Riise, T., Haug, K., Farbu, E., Maeland, J.G., 2013. Smokers' increased risk for disability pension: social confounding or health-mediated effects? Gender-specific analyses of the Hordaland Health Study cohort. J Epidemiol Community Health 67, 758–764. https://doi.org/10.1136/jech-2012-202182

Holmes, M.V., Smith, G.D., 2017. Revealing the effect of CETP inhibition in cardiovascular disease. Nat. Rev. Cardiol. 14, 635–636. https://doi.org/10.1038/nrcardio.2017.156

Husemoen, L.L.N., Osler, M., Godtfredsen, N.S., Prescott, E., 2004. Smoking and subsequent risk of early retirement due to permanent disability. Eur. J. Public Health 14, 86–92. https://doi.org/10.1093/eurpub/14.1.86

Hyde, M., Wiggins, R.D., Higgs, P., Blane, D.B., 2003. A measure of quality of life in early old age: the theory, development and properties of a needs satisfaction model (CASP-19). Aging Ment. Health 7, 186–194. https://doi.org/10.1080/1360786031000101157

Korhonen, T., Smeds, E., Silventoinen, K., Heikkilä, K., Kaprio, J., 2015. Cigarette smoking and alcohol use as predictors of disability retirement: A population-based cohort study.

Drug Alcohol Depend. 155, 260–266. https://doi.org/10.1016/j.drugalcdep.2015.06.047
Koskenvuo, K., Broms, U., Korhonen, T., Laitinen, L.A., Huunan-Seppälä, A., Keistinen, T., Autti-Rämö, I., Kaprio, J., Koskenvuo, M., 2011. Smoking strongly predicts disability retirement due to COPD: the Finnish Twin Cohort Study. Eur. Respir. J. 37, 26–31. https://doi.org/10.1183/09031936.00008910
Lallukka, T., Rahkonen, O., Lahelma, E., Lahti, J., 2015. Joint associations of smoking and physical activity with disability retirement: a register-linked cohort study. BMJ Open 5, e006988. https://doi.org/10.1136/bmjopen-2014-006988
Leung, T., Bergen, A., Munafò, M.R., De Ruyck, K., Selby, P., De Luca, V., 2015. Effect of the rs1051730–rs16969968 variant and smoking cessation treatment: a meta-analysis. Pharmacogenomics 16, 713–720. https://doi.org/10.2217/pgs.15.34
Loeppke, R.R., Schill, A.L., Chosewood, L.C., Grosch, J.W., Allweiss, P., Burton, W.N., Barnes-Farrell, J.L., Goetzel, R.Z., Heinen, L., Hudson, T.W., Hymel, P., Merchant, J., Edington, D.W., Konicki, D.L., Larson, P.W., 2013. Advancing Workplace Health Protection and Promotion for an Aging Workforce. J. Occup. Environ. Med. 55, 500–506. https://doi.org/10.1097/JOM.0b013e31829613a4
Maes, H.H., Sullivan, P.F., Bulik, C.M., Neale, M.C., Prescott, C.A., Eaves, L.J., Kendler, K.S., 2004. A twin study of genetic and environmental influences on tobacco initiation, regular tobacco use and nicotine dependence. Psychol. Med. 34, 1251–1261. https://doi.org/10.1017/s0033291704002405
Martin, A.R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B.M., Daly, M.J., 2019. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat. Genet. 51, 584–591. https://doi.org/10.1038/s41588-019-0379-x
Mawditt, C., Sacker, A., Britton, A., Kelly, Y., Cable, N., 2016. The clustering of health-related behaviours in a British population sample: Testing for cohort differences. Prev. Med. 88, 95–107. https://doi.org/10.1016/j.ypmed.2016.03.003
Murray, R., 2014. The role of smoking in the progressive decline of the body's major systems.
OECD, 2019. Pensions at a Glance 2019: OECD and G20 Indicators. OECD Publishing, Paris.
Office for National Statistics, 2020. Adult Smoking Habits in the UK.
Pietikäinen, S., Silventoinen, K., Svedberg, P., Alexanderson, K., Huunan-Seppälä, A., Koskenvuo, K., Koskenvuo, M., Kaprio, J., Ropponen, A., 2011. Health-related and sociodemographic risk factors for disability pension due to low back disorders: a 30-year prospective Finnish Twin Cohort Study. J. Occup. Environ. Med. 53, 488–496. https://doi.org/10.1097/JOM.0b013e31821576dd
Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., Reich, D., 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38, 904–909. https://doi.org/10.1038/ng1847
Radloff, L.S., 1977. The CES-D Scale A Self-Report Depression Scale for Research in the General Population. Appl. Psychol. Meas. 1, 385–401. https://doi.org/10.1177/014662167700100306
Reynolds, C.A., Finkel, D., 2015. A Meta-analysis of Heritability of Cognitive Aging: Minding the "Missing Heritability" Gap. Neuropsychol. Rev. 25, 97–112. https://doi.org/10.1007/s11065-015-9280-2
Robroek, S.J.W., Schuring, M., Croezen, S., Stattin, M., Burdorf, A., 2013. Poor health, unhealthy behaviors, and unfavorable work characteristics influence pathways of exit from paid employment among older workers in Europe: a four year follow-up study. Scand. J. Work. Environ. Health 39, 125–133. https://doi.org/10.5271/sjweh.3319
Ropponen, A., Korhonen, T., Svedberg, P., Koskenvuo, M., Silventoinen, K., Kaprio, J., 2013. Persistent smoking as a predictor of disability pension due to musculoskeletal

diagnoses: A 23year prospective study of Finnish twins. Prev. Med. 57, 889–893. https://doi.org/10.1016/j.ypmed.2013.10.001

Rothenbacher, D., Arndt, V., Fraisse, E., Zschenderlein, B., Fliedner, T.M., Brenner, H., 1998. Early retirement due to permanent disability in relation to smoking in workers of the construction industry. J. Occup. Environ. Med. 40, 63–68. https://doi.org/10.1097/00043764-199801000-00012

Round, A., 2017. Extending working lives: A devolved, lifecourse approach to enabling work beyond state pension age.

So, H.-C., Sham, P.C., 2017. Improving polygenic risk prediction from summary statistics by an empirical Bayes approach. Sci. Rep. 7, 1–11. https://doi.org/10.1038/srep41262

Steel, N., Huppert, F.A., McWilliams, B., Melzer, D., 2002. Physical and cognitive function, in: The Health, Wealth and Lifestyles of the Older Population in England. Institute for Fiscal Studies, pp. 249–300.

Steptoe, A., Breeze, E., Banks, J., Nazroo, J., 2013. Cohort profile: the English Longitudinal Study of Ageing. Int. J. Epidemiol. 42, 1640–1648. https://doi.org/10.1093/ije/dys168

Svedberg, P., Ropponen, A., Lichtenstein, P., Alexanderson, K., 2010. Are self-report of disability pension and long-term sickness absence accurate? Comparisons of self-reported interview data with national register data in a Swedish twin cohort. BMC Public Health 10, 763. https://doi.org/10.1186/1471-2458-10-763

Tobacco and Genetics Consortium, 2010. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat. Genet. 42, 441–447. https://doi.org/10.1038/ng.571

van Rijn, R.M., Robroek, S.J.W., Brouwer, S., Burdorf, A., 2014. Influence of poor health on exit from paid employment: a systematic review. Occup. Environ. Med. 71, 295–301. https://doi.org/10.1136/oemed-2013-101591

West, R., Raw, M., McNeill, A., Stead, L., Aveyard, P., Bitton, J., Stapleton, J., McRobbie, H., Pokhrel, S., Lester-George, A., Borland, R., 2015. Health-care interventions to promote and assist tobacco cessation: a review of efficacy, effectiveness and affordability for use in national guideline development. Addict. Abingdon Engl. 110, 1388–1403. https://doi.org/10.1111/add.12998

Xue, B., Head, J., McMunn, A., 2020. The Impact of Retirement on Cardiovascular Disease and Its Risk Factors: A Systematic Review of Longitudinal Studies. The Gerontologist 60, e367–e377. https://doi.org/10.1093/geront/gnz062

Tables and Figures

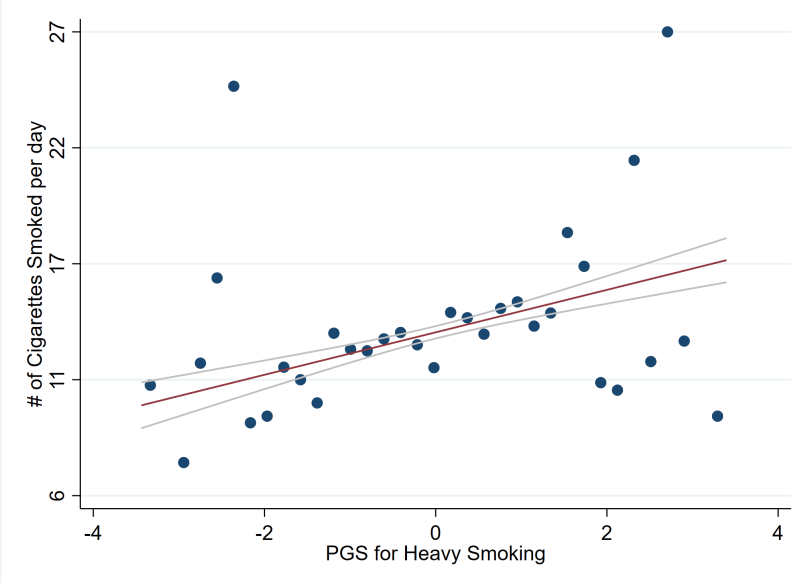
5th July 2021

Table 1: Summary Statistics of Participant Characteristics

	Mean	S.D.	Min	Max
Key Variables:				
Early Retired [0,1]	0.31	0.46	0	1
Heavy Smoker [0,1]	0.31	0.46	0	1
Packs Smoked per day	0.70	0.42	0	3
PGS for Heavy Smoking	0.05	0.97	-3	3
Socio-Demographics:				
Years of Age	64.41	8.57	40	89
Female [0,1]	0.66	0.47	0	1
Higher Education [0,1]	0.08	0.27	0	1
Married [0,1]	0.56	0.50	0	1
Household Size	1.94	0.94	1	11
Log of Household Income	5.45	0.71	-2	10
Observations	3578			

Source: English Longitudinal Study of Ageing (ELSA), Wave 1-9.
Note: The table reports summary statistics of the main variables of interest.

Figure 1: Graphical Evidence



Note: The Figure reports the raw relationship between the polygenic score and smoking behaviour. Slope coefficient: 0.96; Standard Error 0.146.

Table 2: Effect of Polygenic Scores on Different Outcomes

	(1) **Quality of Life**	(2) **Body Mobility Index**	(3) **CES Depression**	(4) **Executive Function**	(5) **Drinker [0,1]**	(6) **Sedentary [0,1]**
PGS for Heavy Smoking	-0.081	0.018	-0.003	0.036	0.002	0.009
	(0.168)	(0.048)	(0.037)	(0.039)	(0.008)	(0.009)
Covariates:						
Years of Age	0.635**	0.064	-0.101*	0.024	-0.023*	-0.038***
	(0.271)	(0.069)	(0.058)	(0.058)	(0.012)	(0.013)
Years of Age Sq.	-0.005**	-0.000	0.001*	-0.001	0.000**	0.000***
	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Female [0,1]	0.215	0.704***	0.488***	-0.086	-0.064***	0.067***
	(0.334)	(0.094)	(0.074)	(0.079)	(0.016)	(0.017)
Higher Education [0,1]	1.446**	-0.553***	-0.232*	0.402***	0.194***	-0.065**
	(0.602)	(0.153)	(0.129)	(0.140)	(0.034)	(0.028)
Married [0,1]	2.499***	-0.451***	-0.618***	0.062	0.006	-0.085***
	(0.405)	(0.108)	(0.087)	(0.088)	(0.017)	(0.020)
Household Size	-0.317	0.051	-0.005	-0.007	-0.027***	0.033***
	(0.231)	(0.059)	(0.046)	(0.048)	(0.008)	(0.011)
Log of Household Income	1.949***	-0.265***	-0.295***	0.262***	0.038***	-0.038***
	(0.272)	(0.070)	(0.057)	(0.059)	(0.011)	(0.012)
Time FE	✓	✓	✓	✓	✓	✓
Region FE	✓	✓	✓	✓	✓	✓
Principal Components	✓	✓	✓	✓	✓	✓
Observations	2949	3578	3534	3223	3216	2717

Source: English Longitudinal Study of Ageing (ELSA), Wave 1-9.

Note: The table reports estimates of the effect of the polygenic score on different key factors and lifestyle behaviours that can be potentially linked to early retirement, namely: (1) Quality of life, as measured by the Control, Autonomy, Self-Realization and Pleasure (CASP) 19; (2) Physical function, as measured by a body mobility index; (3) Mental health, as measured by the Center for Epidemiological Studies – Depression (CES-D) index; Executive function, as measured by the verbal fluency test: (5) Drinking behaviour, defined as whether an individual reported drinking every day; and finally (6) Sedentary behaviour, as measured by whether the individual reports not doing any type of physical activity. Standard errors are clustered at the individual level. ** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*

Table 3: Naïve Estimates: The Effect of Heavy Smoking on Early Retirement

	(1)	(2)	(3)	(4)
Heavy Smoker [0,1]	-0.055***	-0.038***	-0.038***	-0.022*
	(0.016)	(0.014)	(0.014)	(0.013)
Covariates:				
Years of Age		0.086***	0.084***	0.038***
		(0.010)	(0.010)	(0.008)
Years of Age Sq.		-0.000***	-0.000***	-0.000***
		(0.000)	(0.000)	(0.000)
Female [0,1]		-0.330***	-0.332***	-0.344***
		(0.014)	(0.014)	(0.014)
Higher Education [0,1]		0.045*	0.049**	0.050**
		(0.024)	(0.024)	(0.023)
Married [0,1]		-0.001	-0.005	0.028**
		(0.016)	(0.016)	(0.014)
Household Size		-0.002	-0.000	-0.018**
		(0.008)	(0.008)	(0.007)
Log of Household Income		0.005	0.008	-0.047***
		(0.009)	(0.009)	(0.010)
Time FE			✓	✓
Region FE				✓
Observations	3578	3578	3578	3578

Source: English Longitudinal Study of Ageing (ELSA), Wave 1-9.

Note: The table reports (naïve) estimates of the effect of heavy smoking behaviour on early retirement. Specifically, Column (1) reports the unconditional effect of heavy smoking on early retirement. In Columns (2), we include the full set of covariates as described in the manuscript, namely age, and its square, gender, marital status, household size, education, and log of household income. In Columns (3), we include time fixed effects, and in Column (4) we include region fixed effects. Standard errors are clustered at the individual level. ** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*

Table 4: IV Estimates: The Effect of Heavy Smoking on Early Retirement

	(1)	(2)	(3)	(4)
Panel A: First Stage				
PGS for Heavy Smoking	0.039***	0.039***	0.037***	0.038***
	(0.008)	(0.008)	(0.008)	(0.008)
Observations	3578	3578	3578	3578
Panel B: IV Estimates				
Heavy Smoker [0,1]	0.863***	0.723***	0.636***	0.635***
	(0.276)	(0.223)	(0.212)	(0.209)
Time FE		✓	✓	✓
Region FE			✓	✓
Principal Components				✓
Observations	3578	3578	3578	3578

Source: English Longitudinal Study of Ageing (ELSA), Wave 1-9.

Note: The table reports instrumental variable (IV) estimates of the effect of heavy smoking behaviour on early retirement. In Column (1), we report the unconditional effect of heavy smoking on early retirement. In Columns (2), we include the full set of covariates, namely age, and its square, gender, marital status, household size, education, and log of household income. In Columns (3), we include time and region fixed effects (FE), and in Column (4) we include the principal components to account for genetic ancestry. Standard errors are clustered at the individual level. ** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*

Table 5: IV Estimates - Sensitivity Analysis

	(1)	(2)	(3)	(4)
Panel A: Benchmark				
Heavy Smoker [0,1]	0.863***	0.723***	0.636***	0.635***
	(0.276)	(0.223)	(0.212)	(0.209)
Observations	3578	3578	3578	3578
Panel B: Additional Genetic Instruments				
Heavy Smoker [0,1]	0.691***	0.623***	0.568***	0.595***
	(0.246)	(0.205)	(0.195)	(0.198)
Observations	3578	3578	3578	3578
Panel C: Outcome Packs of Cigarettes Smoked per day				
Packs Smoked p/week	0.711***	0.606***	0.526***	0.542***
	(0.204)	(0.169)	(0.158)	(0.163)
Time FE		✓	✓	✓
Region FE			✓	✓
Principal Components				✓
Observations	3578	3578	3578	3578

Source: English Longitudinal Study of Ageing (ELSA), Wave 1-9.

Note: The table reports the sensitivity analysis of the main findings of the study. In Panel A, we report the benchmark estimates. In Panel B, we include in the analysis additional genetic instruments that have been linked to smoking, namely the CHRNA3 and the CHRNA5. Lastly, in Panel C, we consider packs of cigarettes smoked per day as main regressor of interest, as opposed to the binary indicator for heavy smokers. As above, in Column (1), we report the unconditional effect of heavy smoking on early retirement. In Columns (2), we include the full set of covariates as described in the manuscript, namely age, and its square, gender, marital status, household size, education, and log of household income. In Columns (3), we include time and region fixed effects (FE), and in Column (4) we include the principal components to account for genetic ancestry. Standard errors are clustered at the individual level. ** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*

Table 6: Falsification test

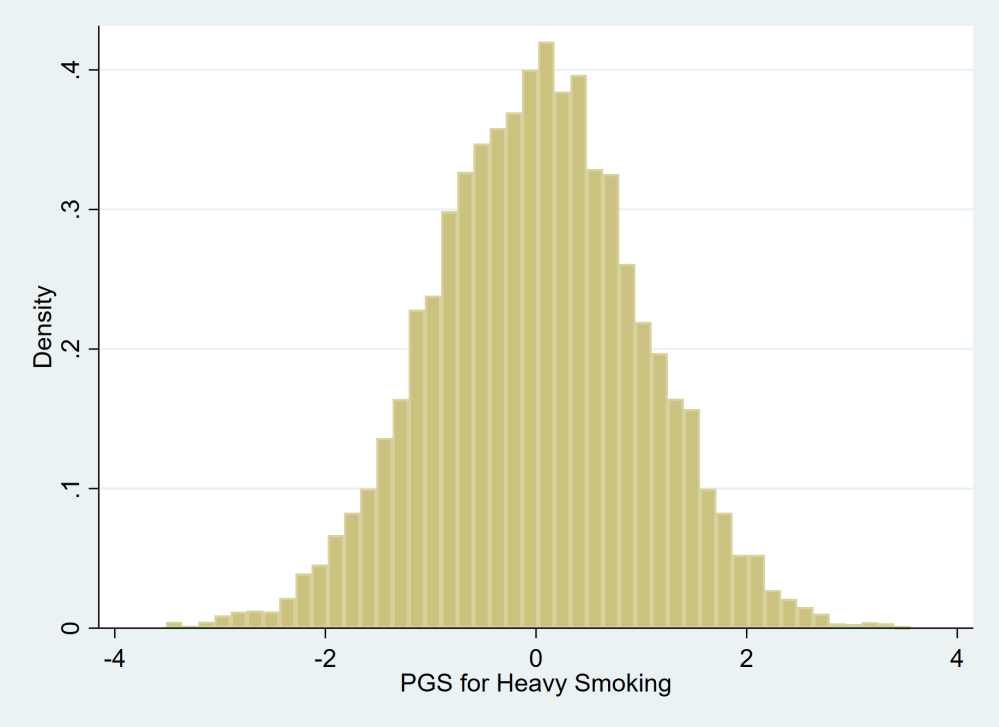
	(1)	(2)	(3)	(4)
PGS for Heavy Smoking	-0.002	-0.000	-0.001	-0.001
	(0.003)	(0.002)	(0.002)	(0.002)
Covariates:				
Years of Age		0.127***	0.082***	0.082***
		(0.003)	(0.002)	(0.002)
Years of Age Sq.		-0.001***	-0.000***	-0.000***
		(0.000)	(0.000)	(0.000)
Female [0,1]		-0.354***	-0.361***	-0.361***
		(0.004)	(0.004)	(0.004)
Higher Education [0,1]		0.048***	0.056***	0.057***
		(0.006)	(0.006)	(0.006)
Married [0,1]		0.043***	0.056***	0.056***
		(0.006)	(0.005)	(0.005)
Household Size		-0.014***	-0.019***	-0.019***
		(0.003)	(0.003)	(0.003)
Log of Household Income		0.034***	-0.036***	-0.036***
		(0.004)	(0.004)	(0.004)
Time FE		✓	✓	✓
Region FE			✓	✓
Principal Components				✓
Observations	37491	37491	37491	37491

Source: English Longitudinal Study of Ageing (ELSA), Wave 1-9.

Note: The table reports estimates of the effect of the heavy smoking polygenic score (PGS) on early retirement for the sample of non-smokers. In Column (1), we report the unconditional effect of heavy smoking on early retirement. In Columns (2), we include the full set of covariates as described in the manuscript, namely age, and its square, gender, marital status, household size, education, and log of household income. In Columns (3), we include time and region fixed effects (FE), and in Column (4) we include the principal components to account for genetic ancestry. Standard errors are clustered at the individual level. ** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*

Appendix

Figure A.1: Density of the Polygenic Score for Heavy Smoking Behaviour



Source: English Longitudinal Study of Ageing (ELSA), Wave 1-9.

Note: The figure depicts the distribution of the polygenic score of heavy smoking behaviour.

Table A.1: Summary Statistics - Heavy vs Light Smokers

	(1) **Heavy Smokers**	(2) **Light Smokers**	(3) ***p*-value**
Key Variables:			
Early Retired [0,1]	0.27	0.32	0.001
	(0.443)	(0.468)	
Packs Smoked per day	1.19	0.48	0.000
	(0.360)	(0.223)	
PGS for Heavy Smoking	0.17	-0.00	0.000
	(0.964)	(0.963)	
Socio-Demographics:			
Years of Age	62.59	65.21	0.000
	(7.609)	(8.848)	
Female [0,1]	0.59	0.69	0.000
	(0.493)	(0.462)	
Higher Education [0,1]	0.08	0.07	0.224
	(0.278)	(0.260)	
Married [0,1]	0.56	0.55	0.744
	(0.497)	(0.497)	
Household Size	2.04	1.89	0.000
	(1.084)	(0.867)	
Log of Household Income	5.42	5.46	0.195
	(0.760)	(0.681)	
Observations	1102	2476	3578

Source: English Longitudinal Study of Ageing (ELSA), Wave 1-9.

Note: The table compares characteristics of heavy and light smokers. Specifically, in Columns (1) and (2), we compare characteristics of the two sub-groups, and in Column (3) we report the p-values from a two sided tests which compares the means of the two groups. Heavy smokers is defined as smoking at least 20 cigarettes per day.

Table A.2: First Stage - The effect of the Polygenic Score (PGS) for Heavy Smoking on Heavy Smoking Behaviour - Full Table

	(1)	(2)	(3)	(4)
Instrument:				
PGS for Heavy Smoking	0.039***	0.039***	0.037***	0.038***
	(0.008)	(0.008)	(0.008)	(0.008)
Covariates:				
Years of Age		0.022**	0.030***	0.029***
		(0.011)	(0.011)	(0.011)
Years of Age Sq.		-0.000***	-0.000***	-0.000***
		(0.000)	(0.000)	(0.000)
Female [0,1]		-0.107***	-0.106***	-0.104***
		(0.017)	(0.017)	(0.017)
Higher Education [0,1]		0.019	0.025	0.028
		(0.030)	(0.029)	(0.030)
Married [0,1]		-0.054***	-0.057***	-0.059***
		(0.018)	(0.018)	(0.018)
Household Size		0.032***	0.033***	0.033***
		(0.010)	(0.010)	(0.010)
Log of Household Income		-0.022*	-0.012	-0.013
		(0.011)	(0.012)	(0.012)
Time FE		✓	✓	✓
Region FE			✓	✓
Principal Components				✓
Observations	3578	3578	3578	3578

Source: English Longitudinal Study of Ageing (ELSA), Wave 1-9.

Note: The figure reports first stage estimates, namely the effect of the polygenic score (PGS) for heavy smoking on heavy smoking behaviour. In Column (1), we report the unconditional estimate. In Columns (2), we include the full set of covariates as described in the manuscript, namely age, and its square, gender, marital status, household size, education, and log of household income. In Columns (3), we include time and region fixed effects (FE), and in Column (4) we include the principal components to account for genetic ancestry. Standard errors are clustered at the individual level. ** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*

Table A.3: IV Estimates: The Effect of Heavy Smoking on Early Retirement - Full Table

	(1)	(2)	(3)	(4)
Key Regressor:				
Heavy Smoker [0,1]	0.863***	0.723***	0.636***	0.635***
	(0.276)	(0.223)	(0.212)	(0.209)
Covariates:				
Years of Age		0.067***	0.017	0.019
		(0.014)	(0.013)	(0.012)
Years of Age Sq.		-0.000***	0.000	-0.000
		(0.000)	(0.000)	(0.000)
Female [0,1]		-0.249***	-0.274***	-0.276***
		(0.030)	(0.029)	(0.028)
Higher Education [0,1]		0.031	0.034	0.036
		(0.034)	(0.030)	(0.031)
Married [0,1]		0.043*	0.068***	0.069***
		(0.025)	(0.022)	(0.022)
Household Size		-0.026*	-0.039***	-0.038***
		(0.014)	(0.012)	(0.012)
Log of Household Income		0.023*	-0.038***	-0.038***
		(0.014)	(0.013)	(0.013)
Time FE		✓	✓	✓
Region FE			✓	✓
Principal Components				✓
Observations	3578	3578	3578	3578

Source: English Longitudinal Study of Ageing (ELSA), Wave 1-9.

Note: The figure reports instrumental variable IV estimates of the effect of heavy smoking behaviour on early retirement. In Column (1), we report the unconditional effect of heavy smoking on early retirement. In Columns (2), we include the full set of covariates as described in the manuscript, namely age, and its square, gender, marital status, household size, education, and log of household income. In Columns (3), we include time and region fixed effects (FE), and in Column (4) we include the principal components to account for genetic ancestry. Standard errors are clustered at the individual level. ** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*

Table A.4: IV Estimates: The Effect of Heavy Smoking on Early Retirement - Accounting for specific smoking genes - Full Table

	(1)	(2)	(3)	(4)
Key Regressor:				
Heavy Smoker [0,1]	0.691***	0.623***	0.568***	0.595***
	(0.246)	(0.205)	(0.195)	(0.198)
Covariates:				
Years of Age		0.069***	0.019	0.021*
		(0.013)	(0.012)	(0.012)
Years of Age Sq.		-0.000***	-0.000	-0.000
		(0.000)	(0.000)	(0.000)
Female [0,1]		-0.260***	-0.281***	-0.280***
		(0.028)	(0.027)	(0.027)
Higher Education [0,1]		0.033	0.036	0.037
		(0.032)	(0.029)	(0.030)
Married [0,1]		0.037	0.064***	0.067***
		(0.023)	(0.021)	(0.022)
Household Size		-0.023*	-0.036***	-0.037***
		(0.013)	(0.011)	(0.011)
Log of Household Income		0.021	-0.039***	-0.038***
		(0.013)	(0.013)	(0.013)
Time FE		✓	✓	✓
Region FE			✓	✓
Principal Components				✓
Observations	3578	3578	3578	3578

Source: English Longitudinal Study of Ageing (ELSA), Wave 1-9.

Note: The figure reports instrumental variable IV estimates of the effect of heavy smoking behaviour on early retirement. In Column (1), we report the unconditional effect of heavy smoking on early retirement. In Columns (2), we include the full set of covariates as described in the manuscript, namely age, and its square, gender, marital status, household size, education, and log of household income. In Columns (3), we include time and region fixed effects (FE), and in Column (4) we include the principal components to account for genetic ancestry. Standard errors are clustered at the individual level. ** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*

Table A.5: IV Estimates: The Effect of Heavy Smoking on Early Retirement - Outcome: Packs of Cigarette Smoked per day - Full Table

	(1)	(2)	(3)	(4)
Packs Smoked p/week	0.711***	0.606***	0.526***	0.542***
	(0.204)	(0.169)	(0.158)	(0.163)
Covariates:				
Years of Age		0.067***	0.020*	0.022**
		(0.013)	(0.011)	(0.011)
Years of Age Sq.		-0.000***	-0.000	-0.000
		(0.000)	(0.000)	(0.000)
Female [0,1]		-0.262***	-0.286***	-0.285***
		(0.024)	(0.023)	(0.024)
Higher Education [0,1]		0.037	0.042	0.045
		(0.031)	(0.027)	(0.028)
Married [0,1]		0.034	0.060***	0.061***
		(0.020)	(0.018)	(0.019)
Household Size		-0.011	-0.025***	-0.025***
		(0.010)	(0.008)	(0.008)
Log of Household Income		0.019	-0.039***	-0.040***
		(0.013)	(0.013)	(0.013)
Time FE		✓	✓	✓
Region FE			✓	✓
Principal Components				✓
Observations	3578	3578	3578	3578

Source: English Longitudinal Study of Ageing (ELSA), Wave 1-9.

Note: The figure reports instrumental variable IV estimates of the effect of heavy smoking behaviour on the number of packs of cigarettes smoked per day. In Column (1), we report the unconditional effect of heavy smoking on early retirement. In Columns (2), we include the full set of covariates as described in the manuscript, namely age, and its square, gender, marital status, household size, education, and log of household income. In Columns (3), we include time and region fixed effects (FE), and in Column (4) we include the principal components to account for genetic ancestry. Standard errors are clustered at the individual level. ** p < 0.1, ** p < 0.05, *** p < 0.01.*

Table A.6: IV Estimates: The Effect of Heavy Smoking on Early Retirement - Different Cut-points for Heavy Smoker

	(1) **Cut-point 10 Cigarettes**	(2) **Cut-point 12 Cigarettes**	(3) **Cut-point 14 Cigarettes**	(4) **Cut-point 16 Cigarettes**	(5) **Cut-point 18 Cigarettes**	(6) **Cut-point 20 Cigarettes**
Heavy Smoker [0,1]	0.608***	0.508***	0.597***	0.596***	0.619***	0.635***
	(0.190)	(0.158)	(0.198)	(0.193)	(0.201)	(0.209)
Covariates:						
Years of Age	0.032***	0.036***	0.032***	0.026**	0.020	0.019
	(0.011)	(0.010)	(0.011)	(0.012)	(0.012)	(0.012)
Years of Age Sq.	-0.000	-0.000*	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female [0,1]	-0.336***	-0.312***	-0.300***	-0.290***	-0.281***	-0.276***
	(0.016)	(0.019)	(0.022)	(0.024)	(0.026)	(0.028)
Higher Education [0,1]	0.047	0.061**	0.048	0.041	0.041	0.036
	(0.029)	(0.028)	(0.030)	(0.030)	(0.030)	(0.031)
Married [0,1]	0.060***	0.043**	0.041**	0.060***	0.064***	0.069***
	(0.019)	(0.017)	(0.019)	(0.020)	(0.021)	(0.022)
Household Size	-0.028***	-0.031***	-0.033***	-0.035***	-0.036***	-0.038***
	(0.009)	(0.009)	(0.010)	(0.011)	(0.011)	(0.012)
Log of Household Income	-0.041***	-0.049***	-0.046***	-0.038***	-0.038***	-0.038***
	(0.012)	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)
Time FE	✓	✓	✓	✓	✓	✓
Region FE	✓	✓	✓	✓	✓	✓
Principal Components	✓	✓	✓	✓	✓	✓
Observations	3578	3578	3578	3578	3578	3578

Source: English Longitudinal Study of Ageing (ELSA), Wave 1-9.

Note: The figure reports instrumental variable IV estimates of the effect of heavy smoking behaviour on early retirement, by considering different definitions for heavy smokers. Standard errors are clustered at the individual level. ** p < 0.1, ** p < 0.05, *** p < 0.01.*