The Causal Effects of Health Conditions and Risk Factors on Social and Socioeconomic Outcomes: Mendelian Randomization in UK Biobank

Sean Harrison*, PhD [1,2], Alisha R Davies, PhD [3], Matt Dickson, PhD [4], Jessica Tyrrell, PhD [5], Michael J Green, PhD [6], Srinivasa Vittal Katikireddi, PhD [6], Desmond Campbell, PhD [6], Marcus Munafò, PhD [7], Padraig Dixon, PhD [1,2], Hayley E Jones, PhD [2], Frances Rice, PhD [8], Neil M Davies^, PhD [1,2,9], Laura D Howe^, PhD [1,2]

* Corresponding author (email: sean.harrison@bristol.ac.uk)

^ denotes equal contribution

1. MRC Integrative Epidemiology Unit (IEU), Population Health Sciences, Bristol Medical School, University of Bristol, Bristol
2. Population Health Sciences, Bristol Medical School, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol
3. Research and Evaluation Division, Public Health Wales NHS Trust, Capital Quarter No.2, Tyndall Street, Cardiff
4. Institute for Policy Research, University of Bath, Bath
5. University of Exeter Medical School, RILD Building, RD&E Hospital Wonford, Barrack Road, Exeter
6. MRC/CSO Social and Public Health Sciences Unit, University of Glasgow, 200 Renfield Street, Glasgow
7. UK Centre for Tobacco and Alcohol Studies, School of Experimental Psychology, University of Bristol, Bristol
8. Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff
9. K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Norway.

Key Words:

Health, socioeconomic, social, economic, health risk factors, health conditions, Mendelian Randomization, UK Biobank

Abstract

Background

To estimate the causal effect of health conditions and risk factors on social and socioeconomic outcomes in UK Biobank. Evidence on socioeconomic impacts is important to understand because it can help governments, policy-makers and decision-makers allocate resources efficiently and effectively.

Methods

We used Mendelian randomization to estimate the causal effects of eight health conditions (asthma, breast cancer, coronary heart disease, depression, eczema, migraine, osteoarthritis, type 2 diabetes) and five health risk factors (alcohol intake, body mass index [BMI], cholesterol, systolic blood pressure, smoking) on 19 social and socioeconomic outcomes in 336,997 men and women of white British ancestry in UK Biobank, aged between 39 and 72 years. Outcomes included annual household income, employment, deprivation (measured by the Townsend deprivation index [TDI]), degree level education, happiness, loneliness, and 13 other social and socioeconomic outcomes.

Results

Results suggested that BMI, smoking and alcohol intake affect many socioeconomic outcomes. For example, smoking was estimated to reduce household income (mean difference = -£22,838, 95% confidence interval (CI): -£31,354 to -£14,321), the chance of owning accommodation (absolute percentage change [APC] = -20.8%, 95% CI: -28.2% to -13.4%), being satisfied with health (APC = -35.4%, 95% CI: -51.2% to -19.5%), and of obtaining a university degree (APC = -65.9%, 95% CI: -81.4% to -50.4%), while also increasing deprivation (mean difference in TDI = 1.73, 95% CI: 1.02 to 2.44, approximately 216% of a decile of TDI). There was evidence that asthma decreased household income, the chance of obtaining a university degree and the chance of cohabiting, and migraine reduced the chance of having a weekly leisure or social activity, especially in men. For other associations, estimates were null.

Conclusions

Higher BMI, alcohol intake and smoking were all estimated to adversely affect multiple social and socioeconomic outcomes. Effects were not detected between health conditions and socioeconomic outcomes using Mendelian randomization, with the exceptions of depression, asthma and migraines. This may reflect true null associations, selection bias given the relative health and age of participants in UK Biobank, and/or lack of power to detect effects.

Key Messages

- Studies have shown associations between poor health and adverse social (e.g. wellbeing, social contact) and socioeconomic (e.g. educational attainment, income, employment) outcomes, but there is also strong evidence that social and socioeconomic factors influence health.
- These bidirectional relationships, as well as confounding, make it difficult to establish whether health conditions and health risk factors have causal effects on social and socioeconomic outcomes.
- Mendelian randomization is a technique that uses genetic variants robustly related to an exposure of interest (here, health conditions and risk factors for poor health) as a proxy for the exposure, and is typically less prone to both reverse causation and confounding, allowing us to estimate more causal effects of health conditions and risk factors on social and socioeconomic outcomes.
- This study suggests causal effects of higher BMI, smoking and alcohol use on a range of social and socioeconomic outcomes, implying that population-level improvements in these risk factors may, in addition to the well-known health benefits, have social and socioeconomic benefits for individuals and society.
- There was evidence that asthma increased deprivation, decreased household income and the chance of having a university degree, migraine reduced the chance of having a weekly leisure or social activity, especially in men, and depression increased loneliness and decreased happiness, but little evidence for causal effects of cholesterol, systolic blood pressure or breast cancer on any social and socioeconomic outcome.

1. Background

Poor health has the potential to affect an individual's ability to engage with society (1–4). For example, illnesses or adverse health behaviours could influence the ability to attend and concentrate at school or work and hence affect educational attainment, employment, and income. Illness and health behaviours may also affect an individual's ability to maintain wellbeing and an active social life. From an individual perspective, maintaining good health can therefore have considerable social and socioeconomic benefits (5). Similarly from a population perspective, improving population health could lead to a happier and more productive population (6).

Understanding the causal impacts of health on social and socioeconomic outcomes can help demonstrate the potential broader benefits of investing in effective health policy, thereby strengthening the case for cross-governmental action to improve health and its wider determinants at the population-level (7). Furthermore, patients require accurate information about how their lives might be affected by their health, for example on returning to work after cancer (8). However, studying the social and socioeconomic consequences of ill health ('social drift') is challenging because of social causation, i.e. the strong role of social and socioeconomic circumstances in disease causation. Social causation means that associations between health and social and socioeconomic outcomes are likely to be severely biased by confounding and reverse causality. Methodological approaches strengthening causal inference in this field are therefore essential.

Mendelian randomization is a technique that uses genetic variants robustly related to an exposure of interest (here, health conditions and risk factors for poor health) as proxies for the exposure (instrumental variables) (9,10). Since genetic variants are randomly allocated at conception, results from Mendelian randomization studies are much less likely to suffer from confounding and reverse causality than traditional observational studies (11). In this paper, we apply Mendelian randomization within a large study of UK individuals aged between 39 and 72 years to estimate the causal effects of health conditions and risk factors with the greatest burden on UK adults on a range of social (e.g. social contact, wellbeing, and cohabitation status) and socioeconomic (e.g. education, employment, income) outcomes.

2. Methods

Population

UK Biobank is a population-based health research resource consisting of approximately 500,000 people, who were recruited between the years 2006 and 2010 from 22 centres across the UK (12). Participants provided medical history and socioeconomic information via questionnaires, interviews and anthropometric measures at recruitment. Medical data from hospital episode statistics (HES) and the cancer registry have been linked to participants. The study design, participants and quality control methods have been described in detail previously (13–15). UK Biobank received ethics approval from the Research Ethics Committee (REC reference for UK Biobank is 11/NW/0382).

We restricted analyses to unrelated individuals of white British ancestry. Full details of inclusion criteria and genotyping are in **Supplementary Information 1**. After exclusions, 336,997 participants remained in the dataset.

Measures of Health Conditions and Risk Factors (Exposures)

We used the Global Burden of Disease Study 2010 (GBD) (16) to identify health conditions and risk factors that contributed 100 or more disability-adjusted life years lost per 100,000 adults in the UK. From this list, we restricted our analysis to health conditions and risk factors with known genetic determinants and a prevalence of ≥2% among UK Biobank participants. This resulted in the inclusion of eight health conditions: asthma, breast cancer, coronary heart disease, depression, eczema, migraine, osteoarthritis and type 2 diabetes; and five risk factors: alcohol consumption, BMI, cholesterol, smoking, systolic blood pressure (**Supplementary Figure 1** and **Supplementary Table 1**).

Except for depression, we categorised a participant as having a health condition if they reported the condition at the baseline visit, or if they had the corresponding HES or cancer registry ICD-9 or ICD-10 code for the health condition before the baseline visit (ICD codes and specific questions used shown in **Supplementary Table 1**).

We coded depression as in Tyrrell (17), where participants were considered to have depression if they self-reported seeing a GP or psychiatrist for nerves, anxiety or depression and reported at least a 2-week duration of depression or unenthusiasm, or had the relevant ICD-9 or ICD-10 codes for depression. Participants were considered to not have depression if they did not report ever visiting a GP or psychiatrist for nerves, anxiety or depression, did not self-report having depression and did not have an ICD code for depression. Only 10 centres asked the questions related to depression, so only participants from these centres were considered in the depression analyses.

The measurement of health risk factors is described in **Box 1**.

Box 1*: Measurement of health risk factors at baseline (except smoking variables)*

Alcohol intake

We estimated the average weekly intake of alcoholic units (10ml of pure alcohol) for all participants based on the average reported intake of six different types of alcoholic beverage. The nominal number of units we assigned per drink for each type of alcoholic beverage are listed below:

- Red wine: 125 ml (6/bottle), 14% = 1.75 units
- Champagne/white wine: 125 ml (6/bottle), 14% = 1.75 units
- Beer/cider: 1 pint, 3.5% = 2 units
- Spirits: 25 ml (25 standard measures in a normal sized bottle), 40% = 1 unit
- Fortified wine: 60 ml (12/bottle), 20% = 1.2 units
- Other: Unknown, example is an alcopop = 1 unit

We removed self-reported former drinkers, participants with a very high number of units per week (>200 units), and participants who did not report they were never drinkers but who answered none of the questions about weekly alcohol intake, leaving 252,585 participants (75%).

Body mass index

BMI was estimated as measured weight in kilograms divided by measured height in metres squared.

Cholesterol

Cholesterol was measured by UK Biobank at baseline (measured by CHO-POD analysis on a Beckman Coulter AU5800).

Smoking

We used two measures of self-reported smoking.

Lifetime smoking index*:* a composite (continuous) measure of relevant smoking variables with a simulated half-time constant representing the decreasing effect of smoking on health outcomes over time. This variable was created by Wootton and colleagues and used in a paper studying smoking and depression/schizophrenia (18).

Smoking initiation: a binary measure indicating whether participants had ever versus never smoked participants, based on whether the lifetime smoking index value had a non-zero value.

Systolic blood pressure

Systolic blood pressure was measured using an automated device, and two measurements were taken a few moments apart. If the standard automated device could not be employed, two manual readings were taken instead.

Polygenic Risk Scores (Instrumental Variables)

We searched previous genome-wide association studies (GWAS) for single nucleotide polymorphisms (SNPs) with strong evidence of associations for each health condition and risk factor, defined as having a P value at genome-wide significance ($P \leq 5 \times 10^{-8}$) (further details in **Supplementary Information 2** and **Supplementary Tables 2 and 3**). The polygenic risk scores (PRS) for each health condition and risk factor were then calculated as the sum of the effect alleles for all SNPs associated with the health condition or risk factor, with each SNP weighted by the regression coefficient from the GWAS from which the SNP was identified.

Covariates

Age, sex and UK Biobank recruitment centre were reported at the baseline assessment, and genetic principal components (used to control for population stratification (19)) were derived by UK Biobank.

Social and Socioeconomic Measures (Outcomes)

We selected social and socioeconomic outcomes measured at the UK Biobank baseline assessment centre. Where possible, we dichotomised outcomes to simplify interpretability and comparability across outcomes. **Box 2** contains a list of all outcomes; **Supplementary Information 3** and **Supplementary Table 4** give further information on how each outcome was measured.

We considered breast cancer, coronary heart disease, osteoarthritis, cholesterol or systolic blood pressure unlikely to have plausible causal effects on the chance of obtaining a university degree given that these health conditions usually occur later in life; the Mendelian randomization effect estimates for these associations were thus used as negative controls (i.e. where no effect should be expected) (20,21).

***Box 2**: List of all social and socioeconomic measures (outcomes)*

Socioeconomic Outcomes

- Average household income before tax, with each category assigned the mid-point of the range (and open-ended categories a nominal value) to allow for continuous analysis*:
  - <£18,000 = £15,000
  - £18,000 to £30,999 = £24,500
  - £31,000 to £51,999 = £41,500
  - £52,000 to £100,000 = £76,000
  - >£100,000 = £150,000
- Deprivation, measured using the Townsend Deprivation Index (TDI) of current address*
- Current employment status, coded as three separate outcomes:
  - Non-employed, not retired (versus employed or retired)
  - Non-employed (versus employed, retired excluded)
  - Retired (versus still employed, other non-employed excluded)
- Job class, coded as skilled versus unskilled (22)
- Degree status, coded as degree-level education versus lower
- Owner-occupied accommodation versus renting

Social Outcomes

Measures of social contact:

- Having someone to confide in weekly or more frequently versus less frequently
- Friend/family visits weekly or more frequently versus less frequently
- Cohabiting with partner or spouse versus not cohabiting
- Participation in any leisure/social activity versus none

Measures of happiness and wellbeing:

- Lonely/isolated versus not lonely/isolated
- Extremely/very/moderately happy versus not
- Extremely/very/moderately happy with family relationship versus not
- Extremely/very/moderately happy with financial situation versus not
- Extremely/very/moderately happy with friendships versus not
- Extremely/very/moderately happy with health versus not
- Extremely/very/moderately happy with work/job versus not

*Income and deprivation were both dichotomised as additional analyses so the results could be included in plots of all results comparing across outcomes: ≥£52,000 versus <£52,000 for income, most deprived third of TDI versus two least deprived thirds for deprivation

Main Mendelian Randomization Analysis

We used Mendelian randomization to estimate the causal association between each health condition and risk factor and each outcome, using the PRS as an instrumental variable, with age at baseline assessment, sex, UK Biobank recruitment centre and 40 genetic principal components as covariates. We used the ivreg2 package in Stata (version 15.1) with robust standard errors, and tested for weak instrument bias (using Kleibergen-Paap Wald rk F statistics) to assess whether the PRS were sufficiently predictive of the exposures (23). This Mendelian randomization analysis estimates mean and risk differences for continuous and binary outcomes respectively using additive structural mean models (24–26). Mean differences are interpreted as the average change in the outcome over all participants for having the exposure, and risk differences are interpreted as the absolute percentage point change in proportion of participants with the outcome for having the exposure (as in a linear probability model). For health conditions, we are measuring the effects of genetic liability to the health condition (27). The analysis of breast cancer as an exposure was restricted to women. Despite the limitations of an approach based on statistical significance (28), the number of results generated in these analyses necessitated a decision about which results to present in the main paper. Therefore, in the main table of results, we report results with a P value less than 0.0026 (a Bonferroni-corrected P value of 0.05 divided by 19 outcomes, with no correction for multiple exposures), while full results are reported in supplementary tables. However, we considered the public health implications of all effect estimates when interpreting results.

To compare the Mendelian randomization results with associations from non-genetic analysis, we estimated the multivariable adjusted associations between the exposures and outcomes using linear regression, with age, sex, recruitment centre and 40 genetic principle components as covariates, i.e. observational analyses without genetic variables. These are linear probability models for binary outcomes (rather than logistic regression models), which were necessary to be able to compare with the Mendelian randomization analyses, as they are equivalent to additive structural mean models. We also performed endogeneity tests (29) to test whether the Mendelian randomization and multivariable adjusted association estimates differed, where a low P value indicates there was evidence the Mendelian randomization and multivariable effects were different.

Sensitivity Analyses

The robustness of Mendelian randomization analyses is reliant on the assumption that the SNPs, and therefore PRS, do not affect the outcome except through the exposure, i.e. the SNPs are not pleiotropic. We tested this assumption by conducting sensitivity Mendelian randomization analyses, including inverse-variance weighted (IVW), MR Egger (an indicator of directional pleiotropy), weighted median, weighted mode and simple mode analyses (30–32). We also measured Cochran's Q statistic from the IVW analyses (a measure of heterogeneity in the effects of individual SNPs on the outcome), an indicator of pleiotropy (33) or problems with modelling assumptions (34).

From these analyses, we determined: a) whether the results were consistent with the main Mendelian randomization analysis, which would indicate the results of the main analysis were robust, and b) whether there was evidence of pleiotropy from both the Egger regression constant term and Cochran's Q statistic. We also visually inspected plots of the sensitivity Mendelian randomization analyses, which would indicate possible bias in the results of the main analysis. Sensitivity Mendelian randomization analyses could only be performed when there were three or more SNPs included in each PRS.

We also conducted split-sample GWAS and Mendelian randomization analysis using UK Biobank data, in which we randomly split UK Biobank into halves, and for each half conducted a GWAS for each

health condition and risk factor using the MRC IEU UK Biobank GWAS pipeline (35). The results of the two GWAS were used to create PRS for the other half of UK Biobank avoiding sample overlap (36), and we repeated the Mendelian randomization analysis with the two PRS separately, then combined the two results with fixed-effect meta-analysis to give a single estimate. The split-sample analysis a) allowed us to analyse lifetime smoking, as this has only been generated in UK Biobank, and thus no previous GWAS could have been used to inform the PRS, b) allowed us to potentially increase the size and power of the GWAS, possibly improving the predictive ability of the PRS and c) guaranteed homogeneity of the GWAS and analysis populations, which removes the potential bias from using data from an external GWAS to inform the creation of the PRS, for example, through differences in populations giving different effects of SNPs. We also performed sensitivity Mendelian randomization sensitivity analyses on each split to check the robustness of the split-sample results.

Supplementary Table 5 shows a summary of all PRS created and used in the split-sample analyses, and all GWAS significant SNPs from the split-sample GWAS are detailed in **Supplementary Table 6**.

Secondary Analyses

We conducted secondary analyses to check the robustness of results, looking at whether: a) results are different by sex and deprivation at birth, b) results for household income are affected by household size (income equivalisation), c) results for employment outcomes are different when restricting to working age participants, d) results for household income are different when restricting to participants who have not retired and e) results for smoking are robust when only looking at the SNP rs1051730, known to affect smoking heaviness (37). Additionally, we estimated the correlation between each of the PRS in both the main analyses and within each split in the split-sample analyses, to determine whether any of the PRS share genetic information. Further information for the secondary analyses and results are in **Supplementary Information 4.**

Patient and Public Involvement

This study was conducted using UK Biobank. Details of patient and public involvement in the UK Biobank are available online (www.ukbiobank.ac.uk/about-biobank-uk/ and https://www.ukbiobank.ac.uk/wp-content/uploads/2011/07/Summary-EGF-consultation.pdf?phpMyAdmin=trmKQlYdjjnQIgJ%2CfAzikMhEnx6). No patients were specifically involved in setting the research question or the outcome measures, nor were they involved in developing plans for recruitment, design, or implementation of this study. No patients were asked to advise on interpretation or writing up of results. There are no specific plans to disseminate the results of the research to study participants, but the UK Biobank disseminates key findings from projects on its website.

Data and Code Availability

The empirical dataset will be archived with UK Biobank and made available to individuals who obtain the necessary permissions from the study's data access committees. The code used to clean and analyse the data is available here: https://github.com/sean-harrison-bristol/Effects-of-Health-Conditions-and-Risk-Factors-on-Socioeconomic-Outcomes

3. Results

Summary demographics, including prevalence of health conditions, risk factors and all outcomes, are presented in **Table 1**. The mean age of participants was 56.9 years (standard deviation: 8.0 years), mean household income (estimated from household income category midpoints) was £44,409 (standard deviation: £33,181), and 46% of participants were male. Results from the main Mendelian randomization analysis are displayed in a heat map of the P values, where the P value of each analysis is displayed in a cell, with the colour of the cell increasing in intensity as the P value of the analysis decreases, **Figure 1**. **Table 2** shows results from the main Mendelian randomization, split-sample Mendelian randomization and multivariable adjusted analyses for all outcomes where the main or split-sample Mendelian randomization analysis had a P value less than 0.0026. All health conditions (except osteoarthritis) and risk factors in the main Mendelian randomization analysis had a low risk of weak instrument bias, and 75% of regressions had F-statistics above 1000.

Forest plots showing the results for the main Mendelian randomization, split-sample Mendelian randomization and multivariable adjusted analyses for health conditions and risk factors on household income are shown in **Figures 2 and 3**, although there was evidence of heterogeneity between SNPs in sensitivity Mendelian randomization analyses for some exposures on income (Cochran's Q statistic P < 0.01 for alcohol intake, BMI, breast cancer, depression, smoking initiation, systolic blood pressure), indicating possible pleiotropy. As such, results for income for these exposures should be interpreted with some caution, although there was little evidence of directional pleiotropy from MR Egger analyses of these exposures on income. As additional examples, the main Mendelian randomization, split-sample Mendelian randomization and multivariable adjusted analyses for health conditions and risk factors on loneliness are shown in **Figures 4 and 5**; plots for all other analyses are presented in **Supplementary Materials**.

3.1 Health Conditions

Asthma

In the main Mendelian randomization analysis, asthma was estimated to reduce household income (mean difference = -£13,474, 95% confidence interval (CI): -£18,749 to -£8,199), the chance of obtaining a university degree (absolute percentage change [APC] = -17.1%, 95% CI: -25.4% to -8.7%), and the chance of cohabiting (APC = -11.0%, 95% CI: -17.9% to -4.0%). There was little evidence asthma affected other outcomes. Split-sample Mendelian randomization analysis estimates similarly showed detrimental estimates of asthma on obtaining a university degree and income, but not on cohabiting, and there was only evidence of pleiotropy in sensitivity Mendelian randomization analyses for obtaining a university degree. The multivariable adjusted association estimates tended to be weaker than the Mendelian randomization estimates, and in some cases (e.g. the chance of obtaining a university degree) in the opposite direction.

Depression

In the main Mendelian randomization analysis, depression was estimated to reduce satisfaction with health (APC = -29.1%, -44.6% to -13.6%), financial situation (APC = -26.4%, 95% CI: -41.9% to -10.9%) and family relationships (APC = -19.3%, 95% CI: -30.4% to -8.1%), and, as expected, reduce the chance of being happy (APC = -19.1%, 95% CI: -28.4% to -9.8%) and increase the chance of being lonely (APC = 58.7%, 95% CI: 38.5% to 78.9%). CIs were wide, but the point estimates were consistent with depression being detrimental for almost all socioeconomic outcomes, including household income (mean difference = -£19,540, 95% CI: -£37,635 to -£1,445). Depression was excluded from the split-sample analyses as no GWAS-significant SNPs were found in either split. There was evidence of heterogeneity in SNP effects for most outcomes, but no evidence of directional pleiotropy from Egger

regression. Multivariable adjusted association estimates tended to be weaker than Mendelian randomization estimates.

Eczema

In the main Mendelian randomization analysis, eczema was estimated to reduce household income (mean difference = -£46,965, 95% CI: -£71,028 to -£22,902). However, this was not observed in the split-sample Mendelian randomization analysis (mean difference = £-12,545, 95% CI: £-30,268 to £5,177) or multivariable adjusted analysis (mean difference = £158, 95% CI: £-544 to £859). CIs for all other outcomes were very wide.

Migraine

In the main Mendelian randomization analysis, migraines were estimated to reduce the chance of having a weekly leisure or social activity (APC = -47.9%, 95% CI: -71.1% to -24.7%). This estimate was smaller in the split-sample Mendelian randomization (APC = -26.3% 95% CI: -57.7% to 5.2%) and multivariable regression analyses (APC = -2.9%, 95% CI: -3.8% to -2.0%). When "Pub or social club" was removed from the weekly leisure and social activity outcome, the main Mendelian randomization effect estimate was substantially reduced (APC = -23.4%, 95% CI: -47.9% to 1.2%), while looking only at going to a pub or social club weekly showed a stronger effect (APC = -68.5%, 95% CI: -90.8% to -46.1%). The CIs in Mendelian randomization analyses were wide for all other outcomes. There was no evidence of pleiotropy.

Type 2 Diabetes

In the main and split-sample Mendelian randomization analyses, there were no strong associations for type 2 diabetes with any outcome. Directions of effects were inconsistent across outcomes. Multivariable adjusted association estimates tended to be larger than Mendelian randomization estimates, and associations were apparent with several outcomes, most notably satisfaction with health (APC for multivariable adjusted association estimate = -19.1%, 95% CI: -20.1% to -18.2%).

Other Health Conditions

The CIs in Mendelian randomization analyses for breast cancer, coronary heart disease and osteoarthritis were very wide for all outcomes, and as such, these analyses were inconclusive. For breast cancer and coronary heart disease, there was no clear pattern of the direction of effects across outcomes, and CIs were wide. The CIs for osteoarthritis were very wide for all outcomes. As expected, given life course temporal relationships, there was little evidence from the main or split-sample Mendelian randomization analyses that breast cancer, coronary heart disease or osteoarthritis were associated with the chance of obtaining a university degree (included as negative controls). In the multivariable adjusted analysis, breast cancer was not associated with the chance of obtaining a university degree, while coronary heart disease and osteoarthritis were (APC = -8.1%, 95% CI: -9.0% to -7.1% and APC = -6.3%, 95% CI: -6.9% to -5.6%, respectively), indicating, together with the null estimates from the Mendelian randomization analyses, possible social causation of the health conditions, rather than vice versa. Osteoarthritis was excluded from the sensitivity Mendelian randomization analysis as there were fewer than 3 GWAS-significant SNPs in the osteoarthritis GWAS.

In the multivariable adjusted analysis, breast cancer was only associated with increased chances of being non-employed and retired and a decreased satisfaction with health, whereas coronary heart disease and osteoarthritis were negatively associated with all economic outcomes and most social outcomes, though not satisfaction with friendships or work nor with weekly friend visits.

3.2 Risk Factors

Alcohol Intake

All results are expressed for a 5 units per week increase in alcohol intake.

In the main Mendelian randomization analysis, alcohol was estimated to reduce household income (mean difference = -£2,446, 95% CI: -£3,362 to -£1,530) and the chance of owning accommodation (APC = -1.8, -2.4% to -1.2%), and increase deprivation (mean difference in TDI = 0.18, 95% CI: 0.11 to 0.25, approximately 23% of a decile of TDI). In the split-sample Mendelian randomization analysis, alcohol was estimated to reduce the chance of cohabiting (APC = -1.5%, 95% CI: -2.4% to -0.6%) and owning accommodation (APC = -1.2%, 95% CI: -1.7% to -0.6%) and increase deprivation (mean difference in TDI = 0.14, 95% CI: 0.08 to 0.19, approximately 18% of a decile of TDI). There was no evidence of causal effects on other outcomes. There was evidence of heterogeneity in SNP effects for being happy, household income and receiving a university degree, but no evidence of directional pleiotropy in Egger regression. The multivariable adjusted analysis estimated that alcohol increased (rather than reduced) household income (mean difference = £442, 95% CI: £400 to £484, P value from endogeneity test = $1.6x10^{-10}$), and no associations were seen with other outcomes.

Body Mass Index

All results are expressed for a 5 kg/m^2 increase in BMI.

In the main Mendelian randomization analysis, BMI was estimated to be detrimental for all socioeconomic outcomes: BMI was estimated to reduce household income (mean difference = -£2,777, 95% CI: -£3,6923 to -£1,863), and the chance of owning accommodation (APC = -1.6%, 95% CI: -2.4% to -0.8%), being satisfied with health (APC = -5.2%, -6.8% to -3.5%), obtaining a university degree (APC = -2.9%, 95% CI: -4.4% to -1.5%), and having a skilled job (APC = -2.3%, 95% CI: -3.5% to -1.0%), and increase deprivation (mean difference in TDI = 0.25, 95% CI: 0.17 to 0.33, approximately 31% of a decile of TDI) and the chance of being lonely (APC = 2.4%, 95% CI: 1.4% to 3.5%). In the split-sample analysis, effects of BMI were estimated to be more detrimental than in the main analysis for the above associations, and additionally to increase the chance of being non-employed, both when including and excluding retired participants (APC = 1.5%, 95% CI: 0.8% to 2.1% and APC = 2.3%, 95% CI: 1.3% to 3.2%, respectively), and reduce the chance of being satisfied with financial situation (APC = -3.1%, 95% CI: -4.5% to -1.6%) and having a weekly leisure or social activity (APC = -3.0%, 95% CI: -4.2% to -1.9%).

There was evidence of heterogeneity in SNPs for most outcomes, but evidence of directional pleiotropy in Egger regression only for obtaining a university degree. The multivariable adjusted associations between BMI and socioeconomic outcomes were generally consistent with the Mendelian randomization estimates.

Cholesterol

All results are expressed for a 1 mmol/litre increase in cholesterol.

In the main and split-sample Mendelian randomization analyses, there was no evidence of effects of cholesterol on any outcome. In the multivariable adjusted analyses, cholesterol was beneficial for all socioeconomic outcomes and most social contact and wellbeing outcomes, which, together with the null estimates from the Mendelian randomization analyses, confounding or reverse causation in the multivariable adjusted association estimates. Cholesterol was excluded from the sensitivity Mendelian randomization analysis as there were fewer than 3 GWAS-significant SNPs in the cholesterol GWAS.

Lifetime Smoking

All results are expressed for a one standard deviation increase in the continuous lifetime smoking index value. We did not perform a main Mendelian randomization analysis, as there was no previous GWAS for lifetime smoking.

In the split-sample Mendelian randomization analysis, smoking was estimated to reduce household income (mean difference = -£7,585, 95% CI: -£10,155 to -£5,014), the chance of cohabiting (APC = -5.4%, 95% CI: -8.8% to -2.0%), owning accommodation (APC = -8.6%, 95% CI: -10.79% to -6.4%), having a skilled job (APC = -8.6%, 95% CI: -12.71% to -4.5%), obtaining a university degree (APC = -15.9%, 95% CI: -20.7% to -11.1%), and being satisfied with one's financial situation (APC = -10.0%, 95% CI: -14.7% to -5.3%) and health (APC = -8.4%, 95% CI: -13.3% to -3.6%). Lifetime smoking was also estimate to increase deprivation (mean difference in TDI = 0.98, 95% CI: 0.76 to 1.19, approximately 123% of a decile of TDI) and the chance of being non-employed, both with retired participants included and excluded (APC = 4.2%, 95% CI: 2.1% to 6.2% and APC = 5.9%, 95% CI: 2.9% to 8.9% respectively). There was little evidence smoking affected other social outcomes. There was evidence of heterogeneity in SNPs for obtaining a university degree, but no other outcomes, and no evidence of directional pleiotropy in Egger regression. Multivariable adjusted analyses showed smaller estimates for all outcomes.

Smoking Initiation

In the main Mendelian randomization analysis, smoking initiation was estimated to reduce household income (mean difference = -£22,838, 95% CI: -£31,354 to -£14,321), the chance of owning accommodation (APC = -20.8%, 95% CI: -28.2% to -13.4%), being satisfied with health (APC = -35.4%, 95% CI: -51.2% to -19.5%), and of obtaining a university degree (APC = -65.9%, 95% CI: -81.4% to -50.4%), and to increase deprivation (mean difference in TDI = 1.73, 95% CI: 1.02 to 2.44, approximately 216% of a decile of TDI). All effects were also seen in the split-sample analysis. Smoking initiation was also estimated to increase the chance of having a skilled job (APC = -37.0%, 95% CI: -50.0% to -23.9%), and reduce the chance of being non-employed, both including and excluding retired participants (APC = 13.3%, 95% CI: 6.3% to 20.2% and APC = 19.0%, 95% CI: 9.0% to 29.0% respectively), and of having weekly friend visits (APC = 19.8%, 95% CI: 9.2% to 30.5%), but only in the main Mendelian randomization analysis. Additionally, smoking initiation was estimated to reduce the chance of being satisfied with one's financial situation (APC = -22.7%, 95% CI: -36.0% to -8.9%) in the split-sample Mendelian randomization analysis, with a similar effect size in the main Mendelian randomization analysis. CIs were wide for all outcomes. There was evidence of heterogeneity in SNP effects for most outcomes, but no evidence of directional pleiotropy from Egger regression. Multivariable adjusted association estimates tended to be closer to the null than the MR analyses.

Systolic BP

All results are expressed for a 10-mmHg increase in systolic blood pressure.

In the main and split-sample Mendelian randomization analyses, there was no evidence of effects of systolic BP on any outcome.

3.3 Further Analyses

Full results from main Mendelian randomization, sensitivity Mendelian randomization, split-sample Mendelian randomization, and split-sample sensitivity Mendelian randomization analyses are shown in **Supplementary Tables 7-10**, with secondary and sensitivity analyses results in **Supplementary Tables 11 and 12**. For all health conditions and risk factors, forest plots showing results for the main Mendelian randomization, split-sample Mendelian randomization and multivariable adjusted analyses (presented both as each exposure on social and socioeconomic outcomes, and for each outcome on

health conditions and risk factors) are available in **Supplementary Materials**, along with forest plots of SNPs and plots showing IVW, MR Egger, simple mode, weighted median and weighted mode Mendelian randomization analyses. There was little evidence of correlation between any PRS in the main analysis (all R^2 values below 0.01), however, for the split-sample PRS, there was evidence of correlations between asthma and eczema (r = 0.17 for both splits combined), and between smoking initiation and lifetime smoking (r = 0.37 for both splits combined), **Supplementary Table 13**.

4. Discussion

We estimate the putative causal effects of a variety of health conditions and risk factors on socioeconomic and social outcomes using Mendelian randomization, a genetically-informed methodology typically less affected by confounding and reverse causality than observational analyses that adjust for measured confounders (11). Our results indicate that higher BMI, greater alcohol intake and smoking all negatively affect socioeconomic outcomes, and depression negatively affects many social outcomes. We do not observe an effect of cholesterol or systolic BP on any outcome, which may reflect effective treatments for high cholesterol and hypertension protecting participants from adverse consequences. For breast cancer, coronary heart disease, migraine and osteoarthritis, the confidence intervals for all Mendelian randomization analyses were all very wide, meaning that it is not possible to draw firm conclusions about the social and socioeconomic consequences of these conditions from our analyses. However, we estimated that migraine reduced the chance of going to a pub or social club weekly, possibly as alcohol increases the risk of migraines (38,39).

Potential reasons for adverse effects of high BMI, alcohol use and smoking on social and socioeconomic outcomes include increased disease burden, social stigma (e.g. bias against obese people, smokers etc.), or behaviours which make employment, retention of employment, or social interaction challenging. Our previous analyses of UK Biobank have shown evidence of effects of BMI on social and socioeconomic outcomes in both Mendelian randomization and non-genetic within-sibling analyses (40). Here, we build on these previous analyses by including a broader set of social and socioeconomic outcomes, conducting additional sensitivity and secondary analyses, and facilitating comparisons across a range of health conditions and risk factors.

Higher genetic propensities towards asthma and eczema were estimated to reduce household income (mean difference = -£13,519, 95% CI: -£18,794 to -£8,243 for asthma, and mean difference = -£46,987, 95% CI: -£71,048 to -£22,925 for eczema). However, it is possible these estimates are susceptible to bias from pleiotropy, given the extreme size of the effects. Asthma and eczema share many genetic loci, along with inflammatory bowel disease and other autoimmune conditions (41). Therefore, the Mendelian randomization results for eczema and asthma may reflect an underlying genetic predisposition toward autoimmune condition susceptibility, rather than asthma or eczema specifically. This would not be detectable with Mendelian randomization sensitivity analyses if all SNPs included in the PRS were affecting autoimmune susceptibility rather than the conditions themselves (directional unbalanced pleiotropy). Additionally, the PRS for smoking initiation may capture impulsivity and risk taking as well as a propensity to smoke.

For some health conditions (asthma, breast cancer, eczema, migraine), we saw little evidence for observational (multivariable adjusted) associations with both socioeconomic and social outcomes, despite prior evidence often showing strong associations. For example, breast cancer has been associated with lower income (42), but there was no observational association between breast cancer and household income in UK Biobank. This could result from selection bias in UK Biobank (43), with participants potentially liable to have less severe/advanced forms of the condition or quicker recovery than all breast cancer patients across a population, and also to have greater financial support and better employment conditions than the general population. The effects of health conditions may also diminish over time; there is some evidence that the negative effect on income amongst breast cancer survivors reduces over time (42). It is therefore possible our study does not have the correct time frame to capture the effects of each health condition, or that well-functioning insurance markets and pension provision could mitigate socioeconomic effects of health conditions, at least within this generally affluent UK population (44). Additionally, if a participant developed any health condition

after baseline, we would only know if the participant had a hospital episode which mentioned the condition.

There was evidence that depression was detrimental to multiple social outcomes, including reduced happiness, reported satisfaction rates, and increased loneliness. Given these are common features of depression this result was expected and gives us confidence the PRS for depression was suitably predictive of depression.

Strengths and Limitations

The main strengths of this analysis are that Mendelian randomization analyses are generally less affected by confounding and reverse causation than multivariable adjusted (observational) analyses (45), and that UK Biobank is a very large sample with sufficient data to enable us to examine multiple health exposures and multiple socioeconomic and social outcomes. For some associations, there were marked differences between the Mendelian randomization and multivariable adjusted association estimates, which could result from reverse causation or confounding in the multivariable association adjusted estimates. For example, coronary heart disease was associated with a decreased chance of obtaining a university degree (APC = -8.1%, 95% CI: -9.0% to -7.1%) in the multivariable adjusted analysis, which is implausible given coronary heart disease usually occurs later in life than attending university, and this association was not seen in the Mendelian randomization analysis. Additionally, the SNPs contributing to the PRS were drawn from GWAS that excluded UK Biobank to avoid biases caused by sample overlap (36) and all reached genome-wide significance. Finally, the results from the main and split-sample analyses were largely consistent across exposures and outcomes, reducing the possibility of bias from differences in SNP effects between the GWAS and UK Biobank populations.

However, Mendelian randomization rests on assumptions that cannot be proven to be true (45). Assessing pleiotropy was difficult or impossible for many exposures, due to the low number of SNPs and wide CIs, but there was evidence for heterogeneity between SNPs for some associations (e.g. for income), and directional pleiotropy from Egger regression for a limited number of associations (e.g. for BMI on obtaining a university degree). As the outcomes were social and socioeconomic, not biological, the exclusion restriction assumption would be strong for any genetic variant (i.e. that the genetic variant affects the outcome only through the exposure). For example, we cannot assume that a SNP associated with income affects any health condition or risk factor solely through income. We therefore did not perform bi-directional Mendelian randomization (46), and so cannot rule out reverse causation for any analysis.

The PRS represent lifetime exposure to or risk for the health condition or risk factor, and interventions to reduce the exposure or risk of the exposure at different time points in a person's life may have different effects; effects at specific points in life cannot be explored with the methodology used in this paper. As we used linear prediction models for all analyses, some effect estimates may also be impossibly large (i.e. over 100%), which could occur when precision is very low, though this was rare. Although Mendelian randomization is generally less affected by confounding and reverse causality than multivariable regression analyses, an important potential source of bias in these analyses is family-level effects. Recent evidence suggests that assortative mating and dynastic effects can lead to bias in Mendelian randomization effect estimates (47), with estimates of the effect of BMI on educational attainment being consistent with the null in within-family Mendelian randomization models using data from UK Biobank and the Norwegian HUNT study. In our previous analysis of UK Biobank (40), within-family Mendelian randomization models in UK Biobank alone were too imprecise to draw conclusions about whether the estimated effects of BMI on social and socioeconomic outcomes are robust to potential confounding by family-level factors. Since BMI is the exposure for

which we have greatest statistical power (due to the strength of the genetic instrumental variable), we have not repeated the within-family analyses for our other exposures as power will be extremely limited. However, as more datasets are available that include genetic information for multiple family members, examination of whether these effects can be detected with a within-family Mendelian randomization design will be a high priority.

UK Biobank, while large, is not representative of the UK population as participants tend to be wealthier and healthier compared to the country as a whole, which may impart bias to our analyses (48). It is likely this biased some estimates towards the null, as wealthier and healthier people may be more resistant to any detrimental effects of health conditions and risk factors. Additionally, there is evidence of a geographic structure in the UK Biobank genotype data that cannot be accounted for using adjustment for principal components, which may also have biased our analyses (49). Some outcomes were dichotomised, which may have reduced our ability to detect associations (e.g. satisfaction with health).

For health conditions, the uncertainty around the Mendelian randomization effect estimates was large. As many health conditions had small associations with outcomes on multivariable adjusted analyses, this often meant the Mendelian randomization estimates were larger than the observed estimates or had a different sign, but this can be explained by the imprecision in the Mendelian randomization estimates. The uncertainty is due in part to the relatively poor ability of the PRS to predict some health conditions. There were minimal differences in prevalence between UK Biobank and the UK for most health conditions studied (apart from migraine and depression, which were less and more prevalent in UK Biobank respectively), but it is possible the health conditions were milder or better-managed in UK Biobank participants compared with the population as a whole (50). Therefore, null results should be interpreted as a lack of evidence for a causal effect, not evidence of a lack of a causal effect.

5. Conclusion

The results of this study imply that higher BMI, smoking and alcohol consumption are likely detrimental to socioeconomic outcomes. While the prevalence of smoking is decreasing in the UK (51), the average BMI has risen and is continuing to rise worldwide (52). Reducing average BMI levels, and further reducing smoking and alcohol intake, in addition to health benefits, may also improve socioeconomic outcomes for individuals and populations.

There was little evidence of causal effects of health conditions on socioeconomic outcomes, which may reflect true absence of causal effects or bias due to the characteristics of UK Biobank participants, or the low precision of our estimates for health condition effects.

Funding and Acknowledgements

This work was supported by the Health Foundation as part of a project entitled 'social and economic consequences of health: causal inference methods and longitudinal, intergenerational data', which is part of the Health Foundation's Efficiency Research Programme. LDH is funded by a Career Development Award from the UK Medical Research Council (MR/M020894/1). MG, SVK and DC work for the Health Foundation project entitled 'Causal effects of alcohol and mental health problems on employment outcomes: Harnessing UK Biobank and linked administrative data'. The Health Foundation is an independent charity committed to bringing about better health and health care for people in the UK. The Medical Research Council (MRC) and the University of Bristol support the MRC Integrative Epidemiology Unit [MC_UU_12013/1, MC_UU_12013/9, MC_UU_00011/1]. The Economics and Social Research Council (ESRC) support NMD via a Future Research Leaders grant [ES/N000757/1] and a Norwegian Research Council Grant number 295989. PD acknowledges support from a MRC Skills Development Fellowship (MR/P014259/1). SVK acknowledges funding from a NHS Research Scotland Senior Clinical Fellowship (SCAF/15/02). The MRC/CSO Social & Public Health Sciences Unit, University of Glasgow is supported by the Medical Research Council (MC_UU_12017/13 & MC_UU_12017/15) and the Scottish Government Chief Scientist Office (SPHSU13 & SPHSU15). HEJ acknowledges support from an MRC Career Development Award in Biostatistics (MR/M014533/1). The authors would like to thank Iyas Daghlas for his insightful comments on the effect of alcohol on migraines. No funding body has influenced data collection, analysis or its interpretation. This publication is the work of the authors, who serve as the guarantors for the contents of this paper.

Conflicts of Interest

The authors declare they have no conflicts of interest.

Author Contributions

LDH, ARD, NMD, MD, HEJ and FR obtained funding for this study. SH cleaned and analysed the data and wrote the first draft. All authors contributed to study design, interpreted the results and revised the manuscript.

Transparency Statement

Transparency statement: The lead author (the manuscript's guarantor) affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

References

1. Garland A, Jeon SH, Stepner M, Rotermann M, Fransoo R, Wunsch H, et al. Effects of cardiovascular and cerebrovascular health events on work and earnings: A population-based retrospective cohort study. Ann Intern Med. 2019;191(1):E3–10.

2. Hamood R, Hamood H, Merhasin I, Keinan-Boker L. Work Transitions in Breast Cancer Survivors and Effects on Quality of Life. Journal of Occupational Rehabilitation. 2018;1–14.

3. Wright C, Kipping R, Hickman M, Campbell R, Heron J. Effect of multiple risk behaviours in adolescence on educational attainment at age 16 years: A UK birth cohort study. BMJ Open. 2018;8(7).

4. Howe LD, Kanayalal R, Beaumont R, Davies AR, Frayling TM, Harrison S, et al. Effects of body mass index on relationship status, social contact, and socioeconomic position: Mendelian Randomization study in UK Biobank. bioRxiv [Internet]. 2019 Jan 1;524488. Available from: http://biorxiv.org/content/early/2019/01/18/524488.abstract

5. Garland A. Labor Market Outcomes: Expanding the List of Patient-centered Outcomes in Critical Care. Am J Respir Crit Care Med. 2017;196(8):946–7.

6. Johnson P, Stoye G, Sturrock D. Chief Medical Officer annual report 2018: better health within reach [Internet]. 2018. Available from: https://www.ifs.org.uk/publications/13786

7. de Leeuw E. Engagement of Sectors Other than Health in Integrated Health Governance, Policy, and Action. Annu Rev Public Health. 2017;38(1):329–49.

8. Wells M, Williams B, Firnigl D, Lang H, Coyle J, Kroll T, et al. Supporting “work-related goals” rather than “return to work” after cancer? A systematic review and meta-synthesis of 25 qualitative studies. Vol. 22, Psycho-Oncology. 2013. p. 1208–19.

9. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Smith GD. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Stat Med. 2008;27(8):1133–63.

10. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet [Internet]. 2014;23(R1):R89-98. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25064373%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4170722

11. Smith GD, Lawlor DA, Harbord R, Timpson N, Day I, Ebrahim S. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. PLoS Med. 2007;4(12):1985–92.

12. Allen NE, Sudlow C, Peakman T, Collins R. UK biobank data: Come and get it. Science Translational Medicine. 2014;6(224).

13. Collins R. What makes UK Biobank special? Vol. 379, The Lancet. 2012. p. 1173–4.

14. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203–9.

15. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLoS Med. 2015;12(3).

16. Murray CJL, Richards M a, Newton JN, Fenton K a, Anderson HR, Atkinson C, et al. UK health performance : fi ndings of the Global Burden of Disease Study 2010. Lancet [Internet].

2013;381(13):997–1020. Available from: http://dx.doi.org/10.1016/S0140-6736(13)60355-4%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/23668584

17. Tyrrell J, Mulugeta A, Wood AR, Zhou A, Beaumont RN, Tuke MA, et al. Using genetics to understand the causal influence of higher BMI on depression. Int J Epidemiol [Internet]. 2018; Available from: https://academic.oup.com/ije/advance-article/doi/10.1093/ije/dyy223/5155677

18. Wootton RE, Richmond RC, Stuijfzand BG, Lawn RB, Sallis HM, Taylor GMJ, et al. Causal effects of lifetime smoking on risk for depression and schizophrenia: Evidence from a Mendelian randomisation study. bioRxiv. 2018;1–27.

19. Price AL, N.j.patterson, R.m.plenge, M.e.weinblatt, N.a.shadick. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet [Internet]. 2006;38(8):904–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16862161

20. Gage SH, Munafò MR, Davey Smith G. Causal Inference in Developmental Origins of Health and Disease (DOHaD) Research. Annu Rev Psychol. 2016;67(1):567–85.

21. Craig P, Katikireddi SV, Leyland A, Popham F. Natural Experiments: An Overview of Methods, Approaches, and Contributions to Public Health Intervention Research. Annu Rev Public Health. 2017;38(1):39–56.

22. Tyrrell J, Jones SE, Beaumont R, Astley CM, Lovell R, Yaghootkar H, et al. Height, body mass index, and socioeconomic status: Mendelian randomisation study in UK Biobank. BMJ. 2016;352.

23. Kleibergen F, Paap R. Generalized reduced rank tests using the singular value decomposition. J Econom. 2006;133(1):97–126.

24. Harbord RM, Didelez V, Palmer TM, Meng S, Sterne JAC, Sheehan NA. Severity of bias of a simple estimator of the causal odds ratio in Mendelian randomization studies. Stat Med. 2013;32(7):1246–58.

25. Clarke PS, Windmeijer F. Instrumental variable estimators for binary outcomes. Vol. 107, Journal of the American Statistical Association. 2012. p. 1638–52.

26. Clarke PS, Windmeijer F. Identification of causal effects on binary outcomes using structural mean models. Biostatistics. 2010;11(4):756–70.

27. Burgess S, Labrecque JA. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. Eur J Epidemiol. 2018;33(10):947–52.

28. Sterne JAC, Smith GD, Cox DR. Sifting the evidence—what's wrong with significance tests? BMJ. 2001;322(7280):226.

29. Hayashi F. Econometrics. Princeton University Press. 2000. 233-234 p.

30. Haycock PC, Burgess S, Wade KH, Bowden J, Relton C, Smith GD. Best (but oft-forgotten) practices: The design, analysis, and interpretation of Mendelian randomization studies. Vol. 103, American Journal of Clinical Nutrition. 2016. p. 965–78.

31. Burgess S, Scott RA, Timpson NJ, Smith GD, Thompson SG. Using published data in Mendelian randomization: A blueprint for efficient identification of causal risk factors. Eur J Epidemiol. 2015;30(7):543–52.

32. Pierce BL, Burgess S. Efficient design for mendelian randomization studies: Subsample and 2-sample instrumental variable estimators. Am J Epidemiol. 2013;178(7):1177–84.

33. Greco M F Del, Minelli C, Sheehan NA, Thompson JR. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. Stat Med. 2015;34(21):2926–40.

34. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. Hum Mol Genet. 2018;27(R2):R195–208.

35. Elsworth B, Mitchell R, Raistrick C, Paternoster L, Hemani G, Gaunt T. MRC IEU UK Biobank GWAS pipeline version 2. 2019.

36. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. Genet Epidemiol. 2016;40(7):597–608.

37. Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature. 2008;452(7187):638–42.

38. Dresler T, Caratozzolo S, Guldolf K, Huhn JI, Loiacono C, Niiberg-Pikksööt T, et al. Understanding the nature of psychiatric comorbidity in migraine: A systematic review focused on interactions and treatment implications. J Headache Pain. 2019;20(1).

39. Daghlas I, Guo Y, Chasman DI. Effect of genetic liability to migraine on coronary artery disease and atrial fibrillation: a Mendelian randomization study. Eur J Neurol [Internet]. 2019;n/a(n/a). Available from: https://doi.org/10.1111/ene.14111

40. Howe LD, Kanayalal R, Beaumont RN, Davies AR, Frayling T, Harrison S, et al. Effects of body mass index on relationship status, social contact, and socioeconomic position: Mendelian Randomization study in UK Biobank. Int J Epidemiol.

41. Paternoster L, Standl M, Waage J, Baurecht H, Hotze M, Strachan DP, et al. Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. Nat Genet. 2015;47(12):1449–56.

42. Jensen LS, Overgaard C, Bøggild H, Garne JP, Lund T, Overvad K, et al. The long-term financial consequences of breast cancer: A Danish registry-based cohort study. BMC Public Health. 2017;17(1).

43. Munafò MR, Tilling K, Taylor AE, Evans DM, Smith GD. Collider scope: When selection bias can substantially influence observed associations. Int J Epidemiol. 2018;47(1):226–35.

44. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants with Those of the General Population. Am J Epidemiol. 2017;186(9):1026–34.

45. Davies NM, Holmes M V., Davey Smith G. Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. BMJ. 2018;362.

46. Zheng J, Baird D, Borges M-C, Bowden J, Hemani G, Haycock P, et al. Recent Developments in Mendelian Randomization Studies. Curr Epidemiol Reports. 2017;4(4):330–45.

47. Brumpton B, Sanderson E, Hartwig FP, Harrison S, Vie GÅ, Cho Y, et al. Within-family studies for Mendelian randomization: avoiding dynastic, assortative mating, and population stratification biases. bioRxiv [Internet]. 2019;602516. Available from: https://www.biorxiv.org/content/10.1101/602516v1?rss=1&utm_source=dlvr.it&utm_medium=twitter

48. Hughes RA, Davies NM, Smith GD, Tilling K. Selection bias in instrumental variable analyses. bioRxiv [Internet]. 2017;192237. Available from:

https://www.biorxiv.org/content/early/2017/09/22/192237

49. Haworth S, Mitchell R, Corbin L, Wade KH, Dudding T, Budu-Aggrey A, et al. Common genetic variants and health outcomes appear geographically structured in the UK Biobank sample: Old concerns returning and their implications. bioRxiv [Internet]. 2018;294876. Available from: https://www.biorxiv.org/content/early/2018/04/11/294876

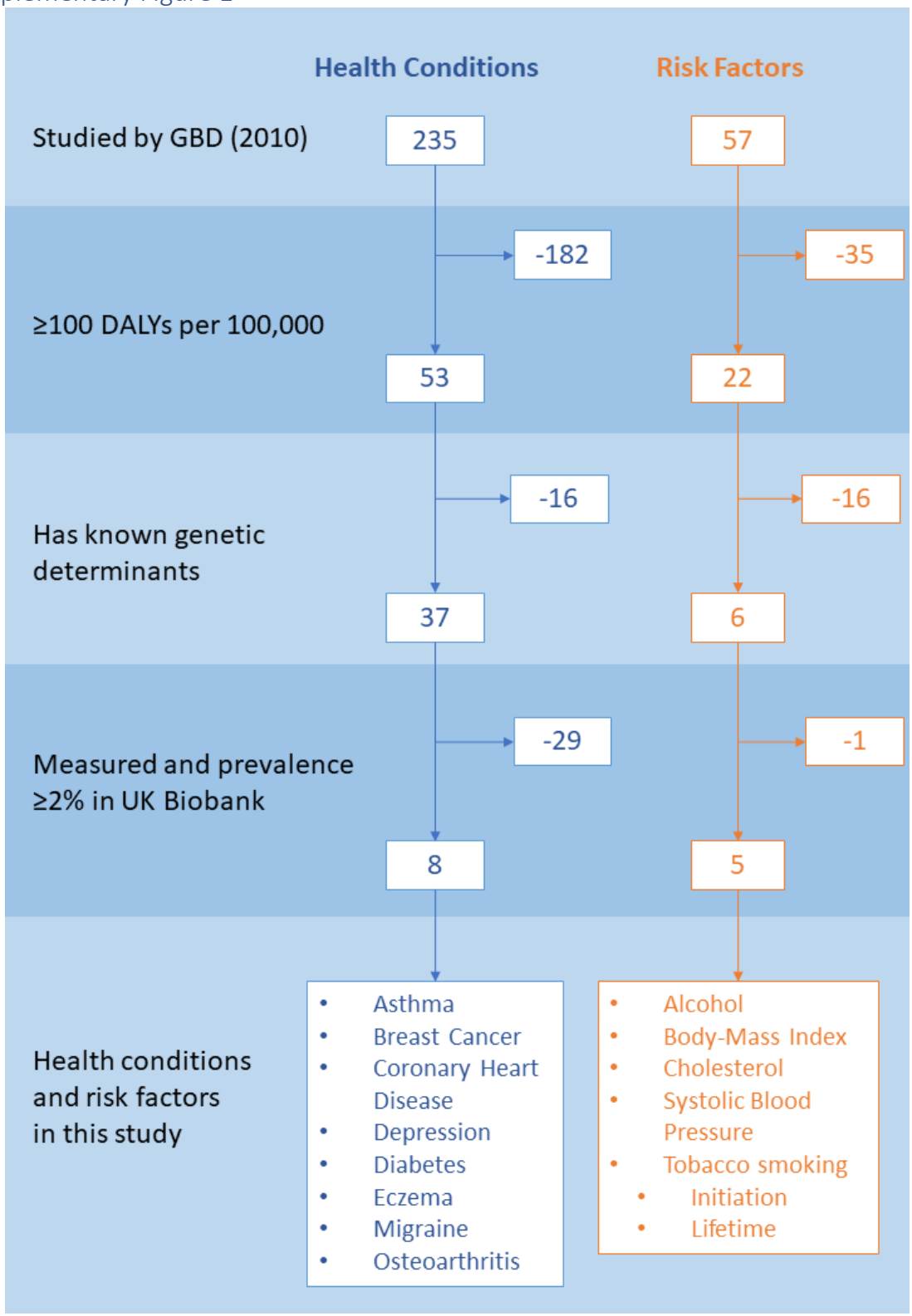
50. David Batty G, Gale C, Kivimaki M, Dreary I, Bell S. Generalisability of Results from UK Biobank: Comparison With a Pooling of 18 Cohort Studies. medRxiv (preprint) [Internet]. 2019; Available from: https://www.medrxiv.org/content/10.1101/19004705v1

51. Simpson CR, Hippisley-Cox J, Sheikh A. Trends in the epidemiology of smoking recorded in UK general practice. Br J Gen Pract. 2010;60(572):187–92.

52. NCD Risk Factor Collaboration, Lewington S, Clarke R, Qizilbash N, Peto R, Collins R, et al. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: A pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults. Lancet. 2017;(2627–2642).

Supplementary Information

Supplementary Figure 1



Supplementary Figure 1: Flow chart showing how health conditions and risk factors were chosen for inclusion in this study

1. Inclusion Criteria and Genotyping

Inclusion Criteria

We restricted analyses to individuals of white British ancestry, as defined by participants who self-reported as "White British" and who had very similar ancestral backgrounds according to the principal component analysis (n=409,703), as described by Bycroft (1). We excluded individuals with sex-mismatch (derived by comparing genetic sex and reported sex) or individuals with sex-chromosome aneuploidy from the analysis (n=814). We estimated kinship coefficients using the KING toolset (2) and identified 107,162 pairs of related individuals (1). We applied an in-house algorithm to this list and preferentially removed the individuals related to the greatest number of other individuals until no related pairs remain. This resulted in the exclusion of 79,448 individuals. Additionally, two individuals were removed due to them relating to a very large number (>200) of individuals, and 135 individuals were excluded as they withdrew from the study. After exclusions, 336,997 participants remained.

Genotyping

The full data release contains the cohort of successfully genotyped samples (n=488,377). 49,979 individuals were genotyped using the UK BiLEVE array and 438,398 using the UK Biobank axiom array. Pre-imputation QC, phasing and imputation are described elsewhere (1). In brief, prior to phasing, multiallelic SNPs or those with MAF ≤1% were removed. Phasing of genotype data was performed using a modified version of the SHAPEIT2 algorithm (3). Genotype imputation to a reference set combining the UK10K haplotype and HRC reference panels (4) was performed using IMPUTE2 algorithms (5). The analyses presented here were restricted to autosomal variants within the HRC site list using a graded filtering with varying imputation quality for different allele frequency ranges. Therefore, rarer genetic variants are required to have a higher imputation INFO score (Info>0.3 for minor allele frequency (MAF) >3%; Info>0.6 for MAF 1-3%; Info>0.8 for MAF 0.5-1%; Info>0.9 for MAF 0.1- 0.5%) with MAF and Info scores having been recalculated on an in house derived 'European' subset.

Further information on the MRC-IEU quality control of UK Biobank genetic data is available online (6).

2. Genome Wide Association Study Search and Polygenic Risk Score Generation

We searched MR-Base (7) and the NHGRI-EBI Catalog of published genome-wide association studies (GWAS) (https://www.ebi.ac.uk/gwas) to find suitable GWAS. If we found no suitable GWAS for any health condition or risk factor, we conducted an online search for a previous GWAS. Only European populations were considered, and if we found multiple GWAS for the same health condition or risk factor, we selected the GWAS with the most participants. We searched for proxy SNPs with an R^2 above 0.8 (using genotype data from European individuals (CEU) from phase 3 (version 5) of the 1000 Genomes project (8)) for any SNP missing from UK Biobank. SNPs from each GWAS were clumped using an R^2 threshold of 0.001 and a window of 10,000 kilo-bases.

We selected GWAS-significant SNPs for smoking initiation and alcohol intake from GSCAN (9), excluding the UK Biobank and 23andMe datasets.

To estimate the strength of the association between the PRS and the health condition or risk factor it proxied, we regressed each health condition or risk factor against its PRS using linear or logistic regression (as appropriate) with no covariables to estimate the R^2 (or pseudo-R^2) value.

The Wray et al. GWAS (2018 (11)) included the pilot sample of UK Biobank, and therefore we only created a PRS for participants not in the pilot sample (241,868 of 336,997 participants, 72%), to preserve independence of the GWAS and analysis samples. This reduced the number of participants with a depression phenotype and PRS to 94,131 of the potential 336,997 participants (28%).

There was no previous GWAS for lifetime smoking, as this was only measured in UK Biobank. As such, lifetime smoking was not considered in the main Mendelian randomization analysis, only in the split-sample Mendelian randomization analysis.

3. Outcome Definitions

Household Income and Deprivation

Household income was recoded from categories to numerical values taking the mid-point of the range, or a nominal value for the open-ended categories, and analysed as a continuous variable:

- <£18,000 = £15,000
- £18,000 to £30,999 = £24,500
- £31,000 to £51,999 = £41,500
- £52,000 to £100,000 = £76,000
- >£100,000 = £150,000

As household income and deprivation (measured using the Townsend Deprivation Index [TDI]) were the only continuous outcomes, we created binary variables for both to allow for comparison between all outcomes, especially on plots. For household income, we compared those with a total household income above and below £52,000 (i.e. upper two categories of household income versus bottom three categories). For deprivation, we split the participants into tertiles of TDI, and compared the most deprived tertile with the remaining two tertiles. These results are included in forest plots of results, but not included in **Table 2**.

Employment

Job class was coded as skilled versus unskilled as in the Tyrrell (2016, (12)) Mendelian randomization analysis of height, BMI and socioeconomic status, where a skilled job was defined as ones in the following categories:

1. Managers and Senior Officials
2. Professional Occupations
3. Associate Professional and Technical Occupations

4. Administrative and Secretarial Occupations
5. Skilled Trades Occupations

Unskilled jobs were defined as ones in the following categories:

6. Personal Service Occupations
7. Sales and Customer Service Occupations
8. Process, Plant and Machine Operatives
9. Elementary Occupations

Current employment status was coded as:

1. Non-employed, not retired (versus employed or retired)
2. Non-employed (versus employed, retired excluded)
3. Retired (versus still employed, other non-employed excluded)

Degree Status

Degree status was coded as having a college or university degree. We did not consider professional qualifications to be equivalent to degree-level education.

Social Outcomes

We dichotomised the frequency of confiding and friend/family visits, comparing weekly or more frequently with less frequently.

We dichotomised all satisfaction outcomes, comparing satisfied (extremely/very/moderately happy) with not satisfied (extremely/very/moderately unhappy).

We defined cohabiting with partner as positive if the participant's response to the question, "How are the other people who live with you related to you?" included "husband, wife or partner".

4.1 Secondary Analyses

Interactions with Sex and Deprivation at Birth

As both sex and deprivation at birth could modify the effect of health conditions and risk factors on outcomes, we repeated the main Mendelian randomization analyses separately in men and women and within thirds of TDI at birth. We then compared the stratum-specific estimates to determine whether there was statistical evidence of differences by sex or between the top and bottom thirds of TDI at birth (13). We considered any P value lower than 0.01 to be indicative of an interaction.

Current deprivation of place of birth was estimated from east and north co-ordinates of birthplace and the Index of Multiple Deprivation taken from the February 2017 Office of National Statistics Postcode Directory (http://ons.maps.arcgis.com/home/item.html?id=dfa0ff74981b4b228d2030d852f0b14a) (17). Co-ordinates of birth were matched with the nearest postcode, although UK Biobank rounded the co-ordinates, so matching is not necessarily precise.

This approach has the strong assumption that the deprivation of a postcode at the birth of a participant is equivalent to the deprivation of the same postcode in 2017, and also that postcode of place of birth has been accurately recorded and is a reliable proxy for deprivation of the participant at birth. As such, the results from this analysis should be treated with caution.

Results – Interactions with Sex

Supplementary Table 11 contains all results for interactions from secondary analyses.

There was evidence of an interaction of migraine and sex on the outcome of having a weekly leisure or social activity (females: APC = -22.1%, 95% confidence interval [CI]: -43.4% to -0.7%; males: APC = -141%, 95% CI: -219% to -63.6%, P for difference = 0.0037). When "Pub or social club" was removed from the weekly leisure and social activity outcome, evidence for an interaction was substantially reduced (females: APC = -15.5%, 95% CI: -38.1% to 7.0%; males: APC = -49.6%, 95% CI: -129% to 30.2%, P for difference = 0.42). However, when going to a pub or social club weekly was the outcome, there was stronger evidence of an interaction of migraine and sex (females: APC = -31.3%, 95% CI: -49.7% to -12.9%; males: APC = -205%, 95% CI: -291% to -120%, P for difference = 0.000094).

There was also evidence of an interaction of BMI and sex on the outcome of cohabiting (females, difference for a 5 kg/m^2 increase in BMI = -2.9%, 95% CI: -4.5% to -1.3%; males = 0.6%, 95% CI: -1.2% to 2.5%, P for difference = 0.0047). Additionally, there was evidence of an interaction of systolic blood pressure and sex on the outcome of being happy (females, difference for a 10 mmHg increase in systolic blood pressure = -0.2%, 95% CI: -1.0% to 0.6%; males = 2.0%, 95% CI: 0.6% to 3.5%, P for difference = 0.0063).

There was little evidence of interactions by sex for other associations.

Results – Interactions with Deprivation at Birth

There was little evidence of interactions by deprivation at birth for any association.

Correlations of PRS

We estimated the correlations of all PRS used in the main analysis, as well as within each split of the split-sample analyses, to determine whether there was evidence of shared genetic information being used in multiple PRS. This would indicate, for instance, whether there was evidence that the PRS for smoking and alcohol intake shared genetic information, for example through a propensity towards risk-taking behaviour. For the main analysis, we used simple correlation. For the split-sample analysis, we used simple correlation within each split, then combined the results using fixed-effect meta-analysis. The results for the main analysis, each split and the splits combined is presented in **Supplementary Table 13**. We considered a correlation coefficient of 0.1 (corresponding to an R^2 of 0.01) or higher to indicate the presence of an association between the PRS.

Results – Correlations of PRS

There was little evidence of associations between any PRS in the main analysis.

There was evidence in the split-sample analysis for correlations between asthma and eczema (correlation coefficient = 0.17 for both splits combined), and, as expected, between smoking initiation and lifetime smoking (correlation coefficient = 0.37 for both splits combined). However, we detected no other associations.

These results indicate we have little evidence that the PRS we used share genetic information, except for asthma and eczema.

4.2 Sensitivity Analyses

Supplementary Table 12 contains all results for sensitivity analyses.

Equivalisation of Household Income

We equivalised household income before tax (household income divided by the number in household) to explore whether household size contributed to effects on household income.

In the main Mendelian randomization analysis, alcohol intake, asthma, BMI, eczema and smoking initiation were estimated to reduce household income. The following health conditions and risk factors were all estimated to reduce equivalised household income: asthma (mean difference = -

£4,265, 95% CI: -£6,831 to -£1,699), eczema (mean difference = -£17,533, 95% CI: -£29,070 to -£5,997), smoking initiation (mean difference = -£10,063, 95% CI: -£14,190 to -£5,936) and alcohol intake (mean difference for a 5 unit increase = -£755, 95% CI: -£1,193 to -£318). BMI (mean difference for a 5 kg/m^2 increase = -£575, 95% CI: -£1,020 to -£130) was still estimated to be detrimental to equivalised household income, but with a larger P value. No other health condition or risk factor was estimated to materially affect equivalised household income.

These results indicate household size may account for some of the estimated effect of BMI on household income, though BMI was still estimated to be detrimental to equivalised household income.

Restriction of Household Income to Participants Who Have Not Retired

We restricted household income to participants who had not retired to explore whether retirement contributed to effects on household income

In the main Mendelian randomization analysis, alcohol intake, asthma, BMI, eczema and smoking initiation were estimated to reduce household income. When restricted to participants who had not retired, alcohol intake (mean difference for a 5 unit increase = -£2,566, 95% CI: -£3,777 to -£1,356), asthma (mean difference = -£13,842, 95% CI: -£20,813 to -£6,870), BMI (mean difference for a 5 kg/m^2 increase = -£2,806, 95% CI: -£4,054 to -£1,558), and smoking initiation (mean difference = -£25,464, 95% CI: -£37,888 to -£13,040) were all estimated to reduce equivalised household income with P values below 0.0026. Eczema (mean difference = -£36,346, 95% CI: -£64,000 to -£8,692) was still estimated to be reduce household income with a similarly large effect to the main Mendelian randomization analysis, but with a larger P value. No other health condition or risk factor was estimated to materially change equivalised household income.

These results indicate that early retirement did not account for the estimated effects of health conditions and risk factors on household income.

Restriction of Employment Outcomes to Working Age

We restricted current employment status outcomes to participants of less than 65 years of age to explore whether early retirement contributed to effects on employment.

In the main Mendelian randomization analysis, smoking initiation was estimated to increase the chance of being non-employed versus employed, both with and without retired participants included in the analysis. When restricted to participants of working age, smoking initiation was still estimated to reduce the chance of being non-employed, both including and excluding retired participants (APC = 16.9%, 95% CI: 8.3% to 25.5% and APC = 20.1%, 95% CI: 9.7% to 30.6% respectively).

These results indicate that early retirement did not account for the estimated effect of smoking initiation on employment outcomes.

Smoking Heaviness SNP: rs1051730

We analysed the effect of rs1051730 on all outcomes to examine whether the effects of lifetime smoking were replicated, as smoking SNPs may be pleiotropic, for instance for SNPs affecting impulsivity. Rs1051730 is a SNP in the nicotinic acetylcholine receptor alpha 3 subunit CHRNA3 gene, and has been used as a conservative proxy for smoking heaviness (14). We analysed subgroups to better estimate the effects of the SNP, analysing: 1) all participants, 2) ever smokers, 3) current smokers, 4) former smokers, and 5) never smokers. Analyses were conducted for all outcomes with rs1051730 as the exposure using linear and logistic regression (as appropriate), and with rs1051730 as an instrumental variable and lifetime smoking (for all participants and those that have smoked) or smoking initiation (for all participants only) as exposures in Mendelian randomization analysis.

rs1051730 as a Proxy for Lifetime Smoking

In the split-sample Mendelian randomization analysis, lifetime smoking was estimated to reduce household income, the chance of cohabiting, owning accommodation, having a skilled job, receiving a university degree, and being satisfied with one's financial situation and health, and increase deprivation and the chance of being lonely and being non-employed, both with retired participants included and excluded.

There were no estimates with P values below 0.0026 when using rs1051730 as an instrumental variable for lifetime smoking in any group of participants (all, ever smokers, current smokers, former smokers). However, CIs were very wide, and some effect sizes were of a similar magnitude to the main Mendelian randomization analysis. For example, lifetime smoking was estimated to reduce household income in all groups (e.g. mean difference for ever smokers = -£5,800, 95% CI: -£10,984 to -£618) and the chance of cohabiting in all groups (e.g. APC for ever smokers = -4.1%, 95% CI: -11.7% to 3.5%). However, other effects were estimated to be inconsistent with the main Mendelian randomization analysis, for example lifetime smoking was estimated to reduce deprivation in subgroups of smokers using rs1051730 as the instrumental variable (mean difference for ever smokers = -0.33, 95% CI: -0.86 to 0.19, approximately 41% of a decile of TDI).

rs1051730 as a Proxy for Smoking Initiation

In the main and split-sample Mendelian randomization analyses, smoking initiation was estimated to reduce household, the chance of owning accommodation, being satisfied with health, and of receiving a university degree, and increase deprivation.

There were no estimates with P values below 0.0026 using rs1051730 as an instrumental variable for smoking initiation in all participants. However, CIs were very wide, and most effect sizes were of a similar magnitude to the main Mendelian randomization analysis. For example, smoking initiation was estimated to reduce household income (mean difference = -£15,415, 95% CI: -£33,693 to £2,863) and increase deprivation (mean difference in TDI = 0.85, 95% CI: -0.73 to 2.43, approximately 106% of a decile of TDI).

rs1051730 Alone

There were no associations with P values below 0.0026 in any group of participants (all, ever smokers, current smokers, never smokers, former smokers) between rs1051730 and any outcome when analysed using linear or logistic regression (as appropriate). However, CIs were very wide. When restricting to P values of less than 0.05, rs1051730 was estimated to increase satisfaction with friendship in current smokers (APC = 23.1%, 95% CI: 5.6% to 40.6%) and ever smokers (APC = 9.2%, 95% CI: 0.1% to 18.4%), reduce household income in current smokers (mean difference = -£572, 95% CI: -£1,082 to -£62) and ever smokers (mean difference = -£286, 95% CI: -£566 to -£6), and reduce weekly friend visits in never smokers (APC = -1.7%, 95% CI: -3.2% to -0.2%).

Overall, the estimates from sensitivity analyses using rs1051730 have CIs that are too wide to confirm whether the main and split-sample Mendelian randomization estimates for lifetime smoking and smoking initiation are driven by smoking heaviness or other factors. However, many of the estimates using rs1051730 are consistent with the estimates using the PRS for lifetime smoking and smoking initiation, increasing our confidence in those results, especially for household income.

References

1. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. bioRxiv [Internet]. 2017;166298. Available from: https://www.biorxiv.org/content/early/2017/07/20/166298

2. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010;26(22):2867–73.

3. O'Connell J, Sharp K, Shrine N, Wain L, Hall I, Tobin M, et al. Haplotype estimation for biobank-scale data sets. Nat Genet. 2016;48(7):817–20.

4. Huang J, Howie B, McCarthy S, Memari Y, Walter K, Min JL, et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. Nat Commun. 2015;6.

5. Howie B, Marchini J, Stephens M. Genotype Imputation with Thousands of Genomes. G3: Genes|Genomes|Genetics [Internet]. 2011;1(6):457–70. Available from: http://g3journal.org/lookup/doi/10.1534/g3.111.001198

6. Mitchell, R., Hemani, G., Dudding, T., Corbin, L., Harrison, S., Paternoster L. UK Biobank Genetic Data: MRC-IEU Quality Control, version 2 - Datasets - data.bris [Internet]. data.bris. 2018. Available from: https://data.bris.ac.uk/data/dataset/1ovaau5sxunp2cv8rcy88688v

7. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife [Internet]. 2018;7:e34408. Available from: https://elifesciences.org/articles/34408

8. Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;135(V):0–9.

9. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Vol. 51, Nature Genetics. 2019. p. 237–44.

10. Okbay A, Baselmans BML, De Neve JE, Turley P, Nivard MG, Fontana MA, et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. Nat Genet. 2016;48(6):624–33.

11. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet. 2018;50(5):668–81.

12. Tyrrell J, Jones SE, Beaumont R, Astley CM, Lovell R, Yaghootkar H, et al. Height, body mass index, and socioeconomic status: Mendelian randomisation study in UK Biobank. BMJ. 2016;352.

13. Altman DG, Bland JM. Interaction revisited: The difference between two estimates. BMJ. 2003;326(7382):219.

14. Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature. 2008;452(7187):638–42.

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	5
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	NA
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	8-9
Bias	9	Describe any efforts to address potential sources of bias	8-9
Study size	10	Explain how the study size was arrived at	5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	8-9
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	8-9
		(*b*) Describe any methods used to examine subgroups and interactions	8-9
		(*c*) Explain how missing data were addressed	NA
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	NA
		(*e*) Describe any sensitivity analyses	8-9

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	5,SAp1
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	NA
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	10
		(b) Indicate number of participants with missing data for each variable of interest	Table1
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	NA
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	Table1
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Table2, 10-14
		(*b*) Report category boundaries when continuous variables were categorized	Box 2, SAp3
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	SAp4
Discussion			
Key results	18	Summarise key results with reference to study objectives	15-17
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	15-17
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	15-17
Generalisability	21	Discuss the generalisability (external validity) of the study results	15-17
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	19

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Figures

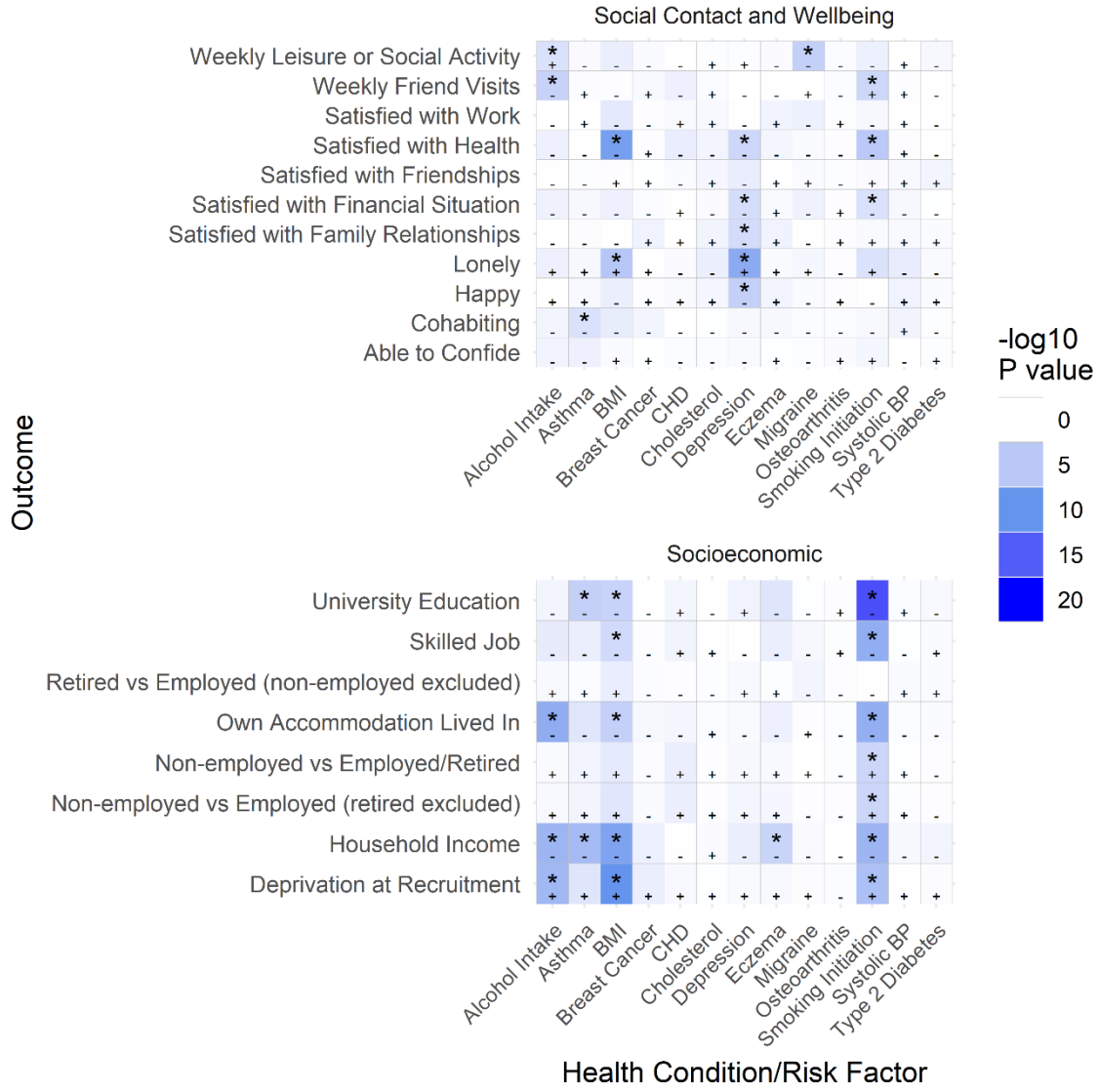


Figure 1: Heat map of results from the main analysis. Each cell shows the P value of the main analysis result for the indicated exposure and outcome, with the colour of the cell increasing in intensity as the P value of the analysis decreases. Starred results are below the Bonferroni-corrected P value threshold ($P < 0.0026$), negative effect directions are denoted with a minus symbol (-), and positive effect directions are denoted with a plus symbol (+).

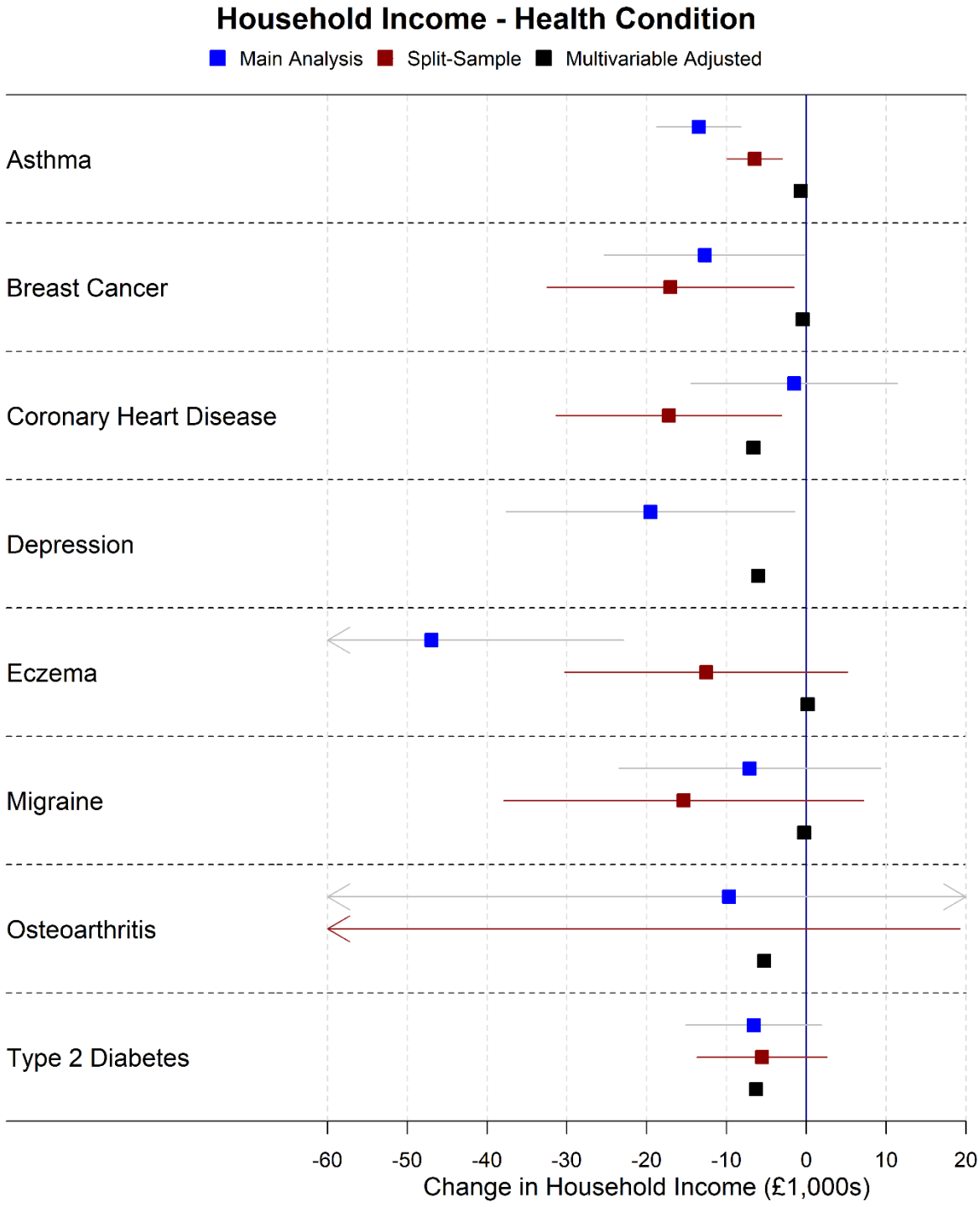


Figure 2: Forest plot showing effects of health conditions on household income for the main Mendelian randomization, split-sample Mendelian randomization and multivariable adjusted analyses (note: confidence intervals are so narrow for the multivariable adjusted analyses they cannot be seen)

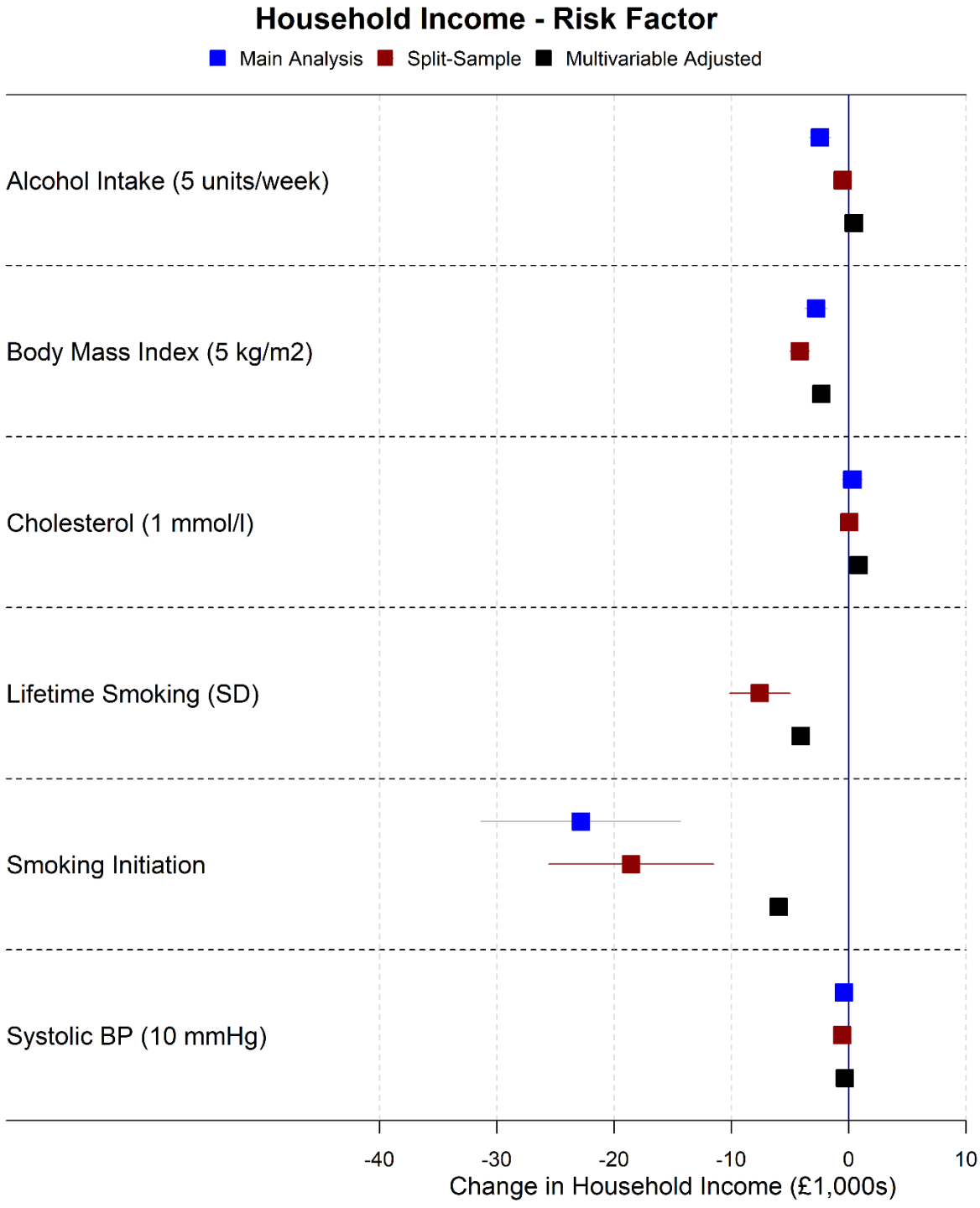


Figure 3: Forest plot showing effects of risk factors on household income for the main Mendelian randomization, split-sample Mendelian randomization and multivariable adjusted analyses (note: confidence intervals are so narrow they cannot be seen for most associations)

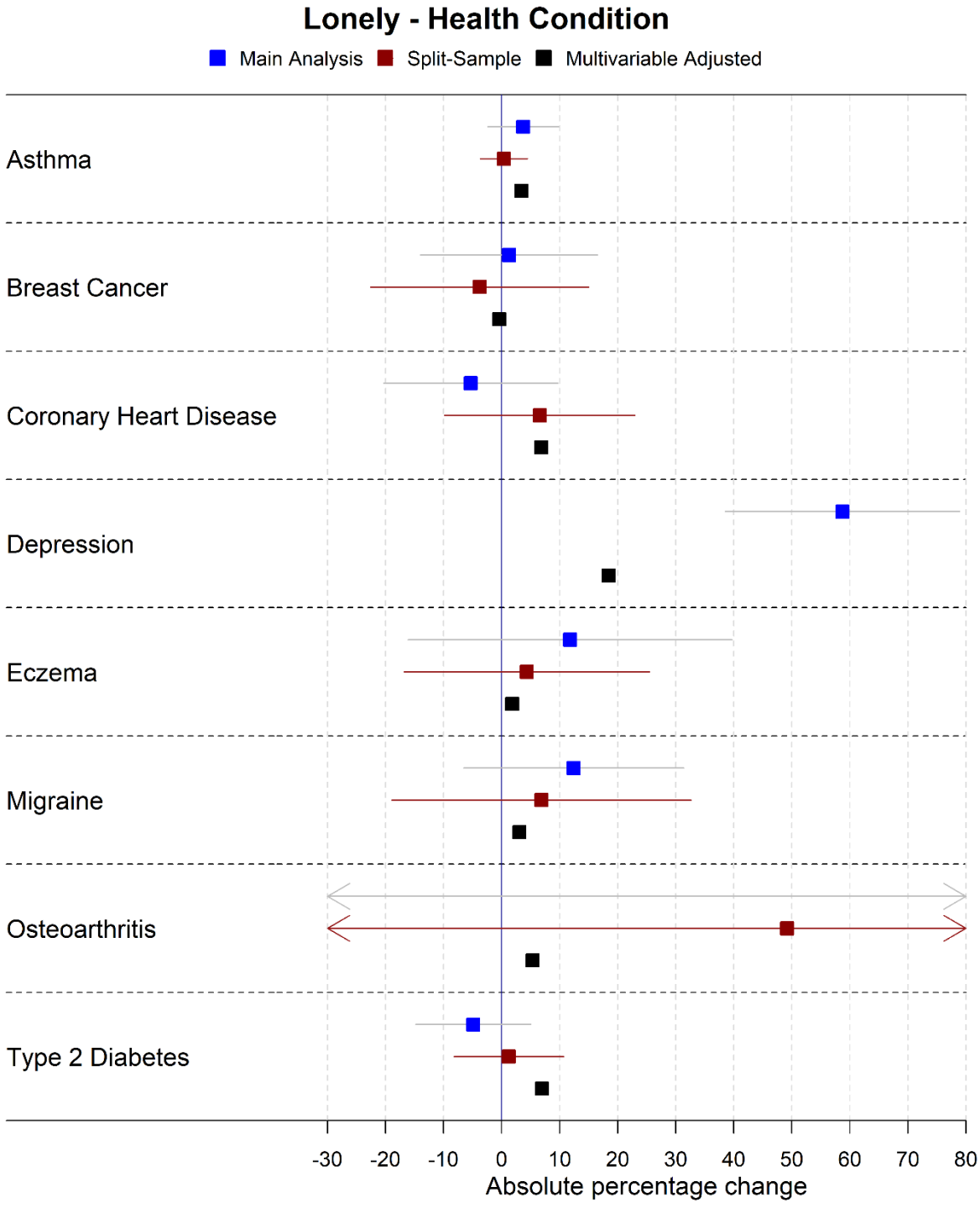


Figure 4: Forest plot showing effects of health conditions on being lonely for the main Mendelian randomization, split-sample Mendelian randomization and multivariable adjusted analyses (note: confidence intervals are so narrow for the multivariable adjusted analyses they cannot be seen)

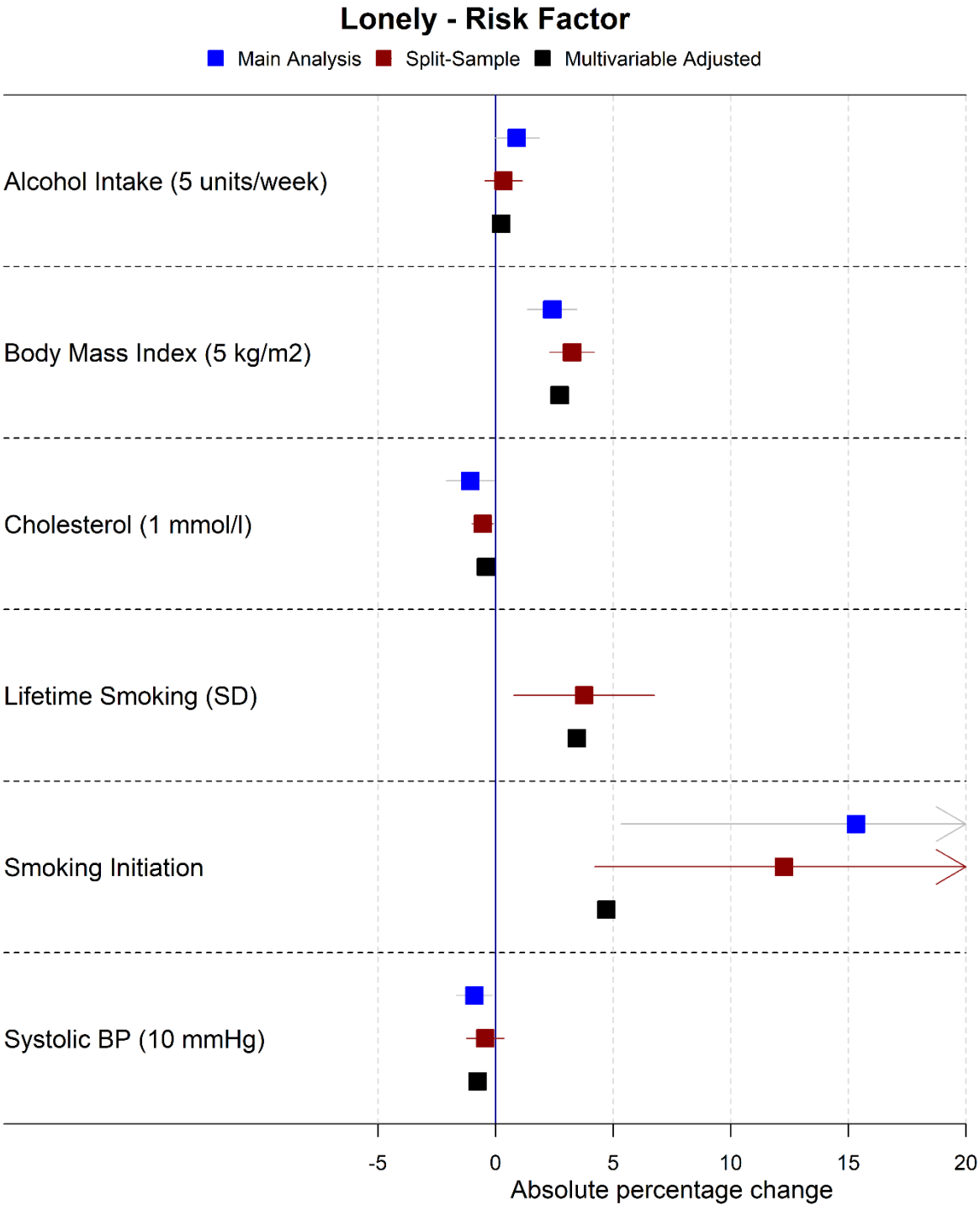


Figure 5: Forest plot showing effects of risk factors on being lonely for the main Mendelian randomization, split-sample Mendelian randomization and multivariable adjusted analyses (note: confidence intervals are so narrow for the multivariable adjusted analyses they cannot be seen)

Variable	All	N	Men	N	Women	N
N	336,997		155,714		181,283	
Age at recruitment, years [Mean (SD)]	56.9 (8.00)	336,997	57.1 (8.09)	155,714	56.7 (7.91)	181,283
Health Conditions						
Asthma [N (%)]	42,832 (12.71)	336,997	18,333 (11.77)	155,714	24,499 (13.51)	181,283
Breast cancer [N (%)]	7,625 (2.26)	336,997	74 (0.05)	155,714	7,551 (4.17)	181,283
Coronary heart disease [N (%)]	16,055 (4.76)	336,997	11,351 (7.29)	155,714	4,704 (2.59)	181,283
Depression* [N (%)]	19,088 (20.28)	94,131	6,841 (15.14)	45,184	12,247 (25.02)	48,947
Eczema [N (%)]	8,685 (2.58)	336,997	3,961 (2.54)	155,714	4,724 (2.61)	181,283
Migraine [N (%)]	10,603 (3.15)	336,997	2,359 (1.51)	155,714	8,244 (4.55)	181,283
Osteoarthritis [N (%)]	36,683 (10.89)	336,997	14,404 (9.25)	155,714	22,279 (12.29)	181,283
Type 2 diabetes** [N (%)]	15,140 (4.51)	335,454	9,349 (6.04)	154,820	5,791 (3.21)	180,634
Risk Factors						
Alcohol intake per week, units of alcohol [Mean (SD)]	18.8 (16.51)	252,578	23.7 (18.76)	126,820	13.8 (12.00)	125,758
Body mass index, kg/m2 [Mean (SD)]	27.4 (4.75)	335,916	27.8 (4.22)	155,193	27.0 (5.13)	180,723
Cholesterol, mmol/l [Mean (SD)]	5.7 (1.14)	321,282	5.5 (1.13)	148,546	5.9 (1.13)	172,736
Ever smoked [N (%)]	98,996 (29.48)	335,829	51,812 (33.39)	155,154	47,184 (26.12)	180,675
Lifetime tobacco smoking [Mean (SD)]	0.3 (0.68)	335,829	0.4 (0.72)	155,154	0.3 (0.63)	180,675
Systolic blood pressure, mmHg [Mean (SD)]	140.2 (19.66)	336,684	143.2 (18.52)	155,633	137.6 (20.24)	181,051
Outcomes						
Socioeconomic						
Average total household income before tax [Mean (SD)]	£44,409 (33,180.94)	290,457	£46,508 (34,101.10)	139,975	£42,458 (32,179.04)	150,482
<£18,000 [N (%)]	63,004 (21.69)	63,004	27,238 (19.46)	27,238	35,766 (23.77)	35,766
£18,000 to £30,999 [N (%)]	74,531 (25.66)	74,531	34,451 (24.61)	34,451	40,080 (26.63)	40,080
£31,000 to £51,999 [N (%)]	76,966 (26.50)	76,966	38,213 (27.30)	38,213	38,753 (25.75)	38,753
£52,000 to £100,000 [N (%)]	60,264 (20.75)	60,264	31,615 (22.59)	31,615	28,649 (19.04)	28,649
>£100,000 [N (%)]	15,692 (5.40)	15,692	8,458 (6.04)	8,458	7,234 (4.81)	7,234
Townsend deprivation index (TDI) at recruitment [Mean (SD)]	-1.6 (2.93)	336,600	-1.5 (2.99)	155,531	-1.6 (2.88)	181,069
Non-employed [N (%)]	25,448 (7.61)	334,514	10,139 (6.56)	154,556	15,309 (8.51)	179,958
Non-employed (retired excluded) [N (%)]	25,448 (11.77)	216,241	10,139 (9.83)	103,134	15,309 (13.53)	113,107
Retired [N (%)]	118,273 (38.27)	309,066	51,422 (35.61)	144,417	66,851 (40.60)	164,649
Skilled job [N (%)]	181,138 (82.60)	219,290	88,921 (84.16)	105,656	92,217 (81.15)	113,634
Degree level education [N (%)]	106,750 (38.57)	276,784	51,813 (40.48)	127,983	54,937 (36.92)	148,801
Own accommodation lived in [N (%)]	304,492 (91.47)	332,904	139,559 (90.79)	153,708	164,933 (92.04)	179,196
Social						
Able to confide (weekly or more frequently) [N (%)]	245,029 (74.84)	327,392	107,078 (70.99)	150,834	137,951 (78.13)	176,558
Frequency of friend/family visits (weekly or more frequently) [N (%)]	264,355 (78.90)	335,071	114,730 (74.17)	154,690	149,625 (82.95)	180,381
Cohabiting [N (%)]	249,951 (74.55)	335,271	120,960 (78.09)	154,895	128,991 (71.51)	180,376
Leisure/social activity [N (%)]	234,303 (69.70)	336,170	108,670 (69.96)	155,338	125,633 (69.47)	180,832
Lonely or isolated [N (%)]	58,573 (17.64)	332,073	22,158 (14.43)	153,549	36,415 (20.40)	178,524
Happy [N (%)]	106,155 (95.67)	110,958	49,098 (95.23)	51,556	57,057 (96.05)	59,402
Satisfied with family relationship [N (%)]	103,620 (93.93)	110,312	47,863 (93.59)	51,143	55,757 (94.23)	59,169
Satisfied with financial situation [N (%)]	96,704 (87.24)	110,843	44,488 (86.37)	51,507	52,216 (88.00)	59,336
Satisfied with friendships [N (%)]	106,855 (97.01)	110,144	49,012 (96.13)	50,985	57,843 (97.78)	59,159
Satisfied with health [N (%)]	96,555 (86.99)	110,997	44,828 (86.87)	51,602	51,727 (87.09)	59,395
Satisfied with work/job [N (%)]	68,536 (91.05)	75,277	31,898 (89.60)	35,602	36,638 (92.35)	39,675

*Depression was restricted to those not in the pilot sample

**Participants with type 1 diabetes were excluded from all type 2 diabetes analyses