Carter, A. R., Harrison, S., Gill, D., Davey Smith, G., Taylor, A. E., Howe, L. D., & Davies, N. M. (2022). Educational attainment as a modifier for the effect of polygenic scores for cardiovascular risk factors: cross-sectional and prospective analysis of UK Biobank. *International Journal of Epidemiology*, *(2022)*, [dyac002]. https://doi.org/10.1093/ije/dyac002

Original article

Educational attainment as a modifier for the effect of polygenic scores for cardiovascular risk factors: cross-sectional and prospective analysis of UK Biobank

Alice R Carter [1,2]* Sean Harrison [1,2] Dipender Gill [3,4,5,6]
George Davey Smith [1,2,7] Amy E Taylor,[1,2,7] Laura D Howe[1,2†] and
Neil M Davies [1,2,8†]

[1]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK, [2]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK, [3]Clinical Pharmacology and Therapeutics Section, Institute of Medical and Biomedical Education and Institute for Infection and Immunity, St George's, University of London, London, UK, [4]Clinical Pharmacology Group, Pharmacy and Medicines Directorate, St George's University Hospitals NHS Foundation Trust, London, UK, [5]Novo Nordisk Research Centre Oxford, Old Road Campus, Oxford, UK, [6]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK, [7]NIHR Bristol Biomedical Research Centre, University of Bristol, Bristol, UK and [8]K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway

*Corresponding author. MRC Integrative Epidemiology Unit, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, UK. E-mail: alice.carter@bristol.ac.uk

†These authors contributed equally to this work.

Abstract

Background: Understanding the interplay between educational attainment and genetic predictors of cardiovascular risk may improve our understanding of the aetiology of educational inequalities in cardiovascular disease.

Methods: In up to 320 120 UK Biobank participants of White British ancestry (mean age = 57 years, female 54%), we created polygenic scores for nine cardiovascular risk factors or diseases: alcohol consumption, body mass index, low-density lipoprotein cholesterol, lifetime smoking behaviour, systolic blood pressure, atrial fibrillation, coronary heart disease, type 2 diabetes and stroke. We estimated whether educational attainment modified genetic susceptibility to these risk factors and diseases.

Results: On the additive scale, higher educational attainment reduced genetic susceptibility to higher body mass index, smoking, atrial fibrillation and type 2 diabetes, but increased genetic susceptibility to higher LDL-C and higher systolic blood pressure. On the multiplicative scale, there was evidence that higher educational attainment increased

genetic susceptibility to atrial fibrillation and coronary heart disease, but little evidence of effect modification was found for all other traits considered.

Conclusions: Educational attainment modifies the genetic susceptibility to some cardiovascular risk factors and diseases. The direction of this effect was mixed across traits considered and differences in associations between the effect of the polygenic score across strata of educational attainment was uniformly small. Therefore, any effect modification by education of genetic susceptibility to cardiovascular risk factors or diseases is unlikely to substantially explain the development of inequalities in cardiovascular risk.

Key words: Polygenic scores, education, inequalities, cardiovascular disease, gene*environment interactions

---

Key Messages

- The role of educational attainment in modifying the effect of polygenic scores for a wide range of cardiovascular risk factors or diseases has not previously been studied.
- We explore whether educational attainment modifies the effects of polygenic susceptibility to alcohol consumption, body mass index, low-density lipoprotein cholesterol, lifetime smoking behaviour, systolic blood pressure, atrial fibrillation, coronary heart disease, type 2 diabetes and stroke.
- Effect modification by education was observed for some polygenic scores for cardiovascular risk factors, but not all.
- Effects were not always in the hypothesized direction and were dependent on the scale of analysis.
- Modification of the effect of genetic susceptibility to cardiovascular risk factors or cardiovascular disease by educational attainment is unlikely to substantially explain the development of inequalities in cardiovascular risk.

Introduction

Socioeconomically deprived individuals have a greater risk of cardiovascular disease (CVD) than less deprived individuals.[1] Most cardiovascular outcomes are multifactorial diseases with environmental and genetic aetiology.[2–4] Therefore, it is plausible that socioeconomic position (SEP) may interact with, or modify, genetic susceptibility to CVD.

Large genome-wide association studies (GWASs) have identified many genetic variants associated with liability to CVD and its risk factors.[5–7] Polygenic scores (PGSs) can subsequently be constructed, explaining substantial fractions of variation. Using UK Biobank, two studies have demonstrated that individuals with a higher Townsend deprivation index score have an accentuated risk of obesity in genetically susceptible adults.[8,9] However, previous studies in the UK and Finland did not find evidence that education modified the effect of genetic susceptibility to high body mass index (BMI) on measured BMI.[9,10]

Whilst educational attainment, a measure of SEP, has been shown to modify the association of cardiovascular risk factors on CVD,[1,11] it is unclear whether educational attainment modifies the effect of genetic susceptibility to a wide range of cardiovascular risk factors. If higher levels of education mitigate some of the genetic risk of cardiovascular risk ('gene*environment interaction'), this may contribute to educational inequalities in CVD.[12]

Where two variables are known risk factors for an outcome, evidence of effect modification is expected on both, or one of, the additive or the multiplicative scale.[13] Therefore, we carry out analyses on both scales. Identifying the magnitude and direction of any effect modification is of greatest importance for public health and in understanding the aetiology of cardiovascular inequalities.

Methods

UK Biobank

UK Biobank recruited 503 317 adults from the UK between 2006 and 2010, aged 37–73 years.[14] Participants attended baseline assessment centres involving questionnaires, interviews and anthropometric, physical and genetic measurements.[14,15] We use ≤320 120 individuals of White British ancestry (Supplementary Figure S1, available as Supplementary data at *IJE* online).

Educational attainment

Participants reported their highest qualification achieved, which was converted to the International Standard

Classification for Education (ISCED) coding for years of education (Supplementary Table S1, available as Supplementary data at *IJE* online).[16] This definition has been used previously,[17] including in UK Biobank.[18]

Cardiovascular risk factors and cardiovascular disease

Cardiovascular risk factors were included if there was evidence for them being a causal risk factor for CVD from randomized–controlled trials, Mendelian randomization studies or clinical studies (see Supplementary Table S2, available as Supplementary data at *IJE* online) with suitable GWAS summary statistics available. Additionally, we included PGSs for several CVD outcomes. In total, nine PGSs were included in analyses: alcohol consumption,[19] BMI,[20] type 2 diabetes,[21] low-density lipoprotein cholesterol (LDL-C),[22] lifetime smoking behaviour,[23] systolic blood pressure,[18] atrial

fibrillation,[5] coronary heart disease (CHD)[6] and stroke.[7] Cardiovascular risk factors were measured at baseline, whilst incident cardiovascular outcomes (atrial fibrillation, CHD, stroke and type 2 diabetes) were determined prospectively by linked mortality records and hospital inpatient records (see Supplementary Table S3, available as Supplementary data at *IJE* online). A full description of how each risk factor/outcome was measured phenotypically and genetically is presented in the Supplementary Methods (available as Supplementary data at *IJE* online).

Deriving polygenic scores

Summary statistics of the associations of the single-nucleotide polymorphisms (SNPs) with each cardiovascular risk factor/outcome were downloaded from MR-Base[24] or directly from the relevant GWAS. We used the most recent GWAS for each risk factor/outcome excluding UK Biobank participants to avoid bias by sample overlap (See Table 1).

Table 1 Summary characteristics for each GWAS used to derive external weights in polygenic scores

Phenotype	Author/consortium	Population	Sample size (cases)	Unit
Alcohol consumption	GWAS and Sequencing Consortium of Alcohol and Nicotine Use[19]	European ancestry (summary statistics excluding UK Biobank)	630 154	Drinks per week
Body mass index	Genetic Investigation of Anthropometric Traits[20]	European ancestry	339 224	SD (kg/m^2)
Low-density lipoprotein cholesterol	Global Lipids Genetics consortium[22]	European ancestry	188 578	SD (circulating lipids)
Smoking	Wootton *et al.*[23]	White British (split sample GWAS of UK Biobank; see Supplementary Methods, available as Supplementary data at *IJE* online)	318 147	SD (lifetime smoking index)
Systolic blood pressure	Carter *et al.*[18]	White British (split sample GWAS of UK Biobank; see Supplementary Methods, available as Supplementary data at *IJE* online)	318 147	SD (mm/Hg)
Atrial fibrillation	Ellinor *et al.*[5]	European ancestry	59 133 (6 707)	Log odds ratio
Coronary heart disease	CARDIoGRAMplusC4D[6]	Predominantly European (77%)	184 305 (60 801)	Log odds ratio
Type 2 diabetes	DIAbetes Genetics Replication And Meta-analysis[21]	European ancestry	159 208 (26 276)	Log odds ratio
Stroke	MEGASTROKE[7]	Predominantly European (85%)	521 612 (67 162)	Log odds ratio

SD, standard deviation; GWAS, genome-wide association study.

The 1000 Genomes Project was used to find proxy SNPs in linkage disequilibrium (LD) with SNPs not found in UK Biobank. Pruning of SNPs was carried out using the clump command in PLINK using an r^2 parameter of 0.25 and a physical-distance threshold for clumping of 500 kB. PGSs were constructed using a range of P-value thresholds: $P \le 5 \times 10^{-8}$ (genome-wide significant), ≤ 0.05 and ≤ 0.5. As the P-value threshold increases, the variance explained by the PGS typically increases. However, increasing the numbers of SNPs increases the risk of pleiotropy and false-positive effects. Pruned SNPs from each GWAS were harmonized with SNPs from UK Biobank, aligning the effect estimates and alleles. Any SNPs that could not be harmonized, palindromic SNPs (where alleles on the forward and reverse strand are read the same) or triallelic SNPs were excluded. PGSs were created by multiplying the number of effect alleles for each participant by the association of the SNP with the phenotype in the GWAS, then summed across all SNPs for each phenotype. PGSs were standardized for use in analyses and reflect a 1 SD change.

Main analyses are presented using PGSs at the genome-wide significance threshold with other thresholds presented in the supplement.

Exclusion criteria

Reverse causality can introduce bias when the temporality of the exposure and outcome is mis-specified and the outcome itself affects the exposure.[25] Although CVD in adulthood cannot alter genetic variants determined at conception, and indeed is unlikely to change educational attainment typically determined in early adulthood, a diagnosis may lead to behavioural or lifestyle changes that change the relative importance of the PGS in determining the outcome. Participants were therefore excluded if they had experienced at least one diagnosis of any of the outcomes considered before baseline (atrial fibrillation, CHD, stroke and type 2 diabetes) or any one of myocardial infarction, angina, transient ischaemic attack, peripheral arterial disease, familial hypercholesterolaemia, type 1 diabetes and chronic kidney disease. These diagnoses can all result in statins being prescribed to prevent CVD, which may lead to behaviour change and therefore reverse causality.[26] Diagnoses were ascertained through linked mortality data and hospital inpatient records using ICD-9 and ICD-10 codes (Supplementary Table S4, available as Supplementary data at *IJE* online).

Quality control of the genetic data was carried out using the Medical Research Council Integrative Epidemiology Unit quality-control pipeline, described in full previously.[27] In brief, individuals were excluded if their genetic sex differed to their gender reported at baseline or for having aneuploidy of their sex chromosomes (non-XX or -XY chromosomes). Further individuals were excluded for extreme heterozygosity or a substantial proportion of missing genetic data. Related individuals were excluded, removing those related to the greatest number of other participants until no related pairs were left.[27] This exclusion list was derived in-house using an algorithm applied to the list of all the related pairs provided by UK Biobank (third-degree or closer) (Supplementary Figure S1, available as Supplementary data at *IJE* online). Individuals were excluded if they had withdrawn from UK Biobank or were, or may be, pregnant at baseline.

Individuals were further excluded if they were missing data for education, age and sex. Individuals were excluded from specific analyses if they were missing phenotypic measurements of the risk factor/outcome under consideration (see Supplementary Figure S1, available as Supplementary data at *IJE* online).

Statistical analysis

Association of educational attainment with outcomes
Multivariable linear regression (adjusting for age and sex) was carried out to estimate the association between educational attainment and cardiovascular risk factors/outcomes.

Association between each polygenic score and observed phenotype
For each cardiovascular risk factor/outcome, we estimated the association between each PGS and the phenotype using multivariable regression, adjusting for age, sex and 40 genetic principal components to control for population structure. For continuous risk factors, measures were standardized, so estimates reflect the mean difference in SD of the phenotype, or natural log of the phenotype, per 1 SD higher PGS. For binary outcomes, estimates reflect the risk difference or odds ratio of the outcome per 1 SD higher PGS.

Effect modification by educational attainment on polygenic scores for cardiovascular risk
To test for effect modification, the linear model was stratified by years of educational attainment. To estimate the magnitude and direction of the effect modification, an interaction term was included in the linear model [e.g. PGS*education (continuous)]. Analyses were adjusted for age, sex and 40 genetic principal components. As effect modification is scale-dependent, tests of effect modification were carried out on both the additive and

multiplicative scales.[13] Additive and multiplicative effects were carried out as previously defined.[13]

Secondary analyses

All analyses were replicated for PGSs at *P*-value thresholds of ≤0.05 and ≤0.5.

Results

UK Biobank cohort

Eligible UK Biobank participants (55% female) had a mean age of 57 (SD = 8.00) years. A higher proportion of participants (33%) left school after 20 years (equivalent to obtaining a degree) compared with those who left school after 7 years (equivalent to no formal qualifications) (16%) (Table 2).

For a *P*-value of $\leq 5 \times 10^{-8}$, the PGSs explained between 0.06% (atrial fibrillation) and 14% (systolic blood pressure) of variance in the phenotypes (Supplementary Table S5, available as Supplementary data at *IJE* online).

Association between educational attainment, polygenic scores and cardiovascular risk factors use

Educational attainment was associated with all cardiovascular risk factors/outcomes, except for LDL-C, although

for all outcomes the effect was small (Supplementary Table S6, available as Supplementary data at *IJE* online). Except for alcohol consumption, higher educational attainment led to a reduction in the mean difference of all risk factors/outcomes (Supplementary Table S6, available as Supplementary data at *IJE* online).
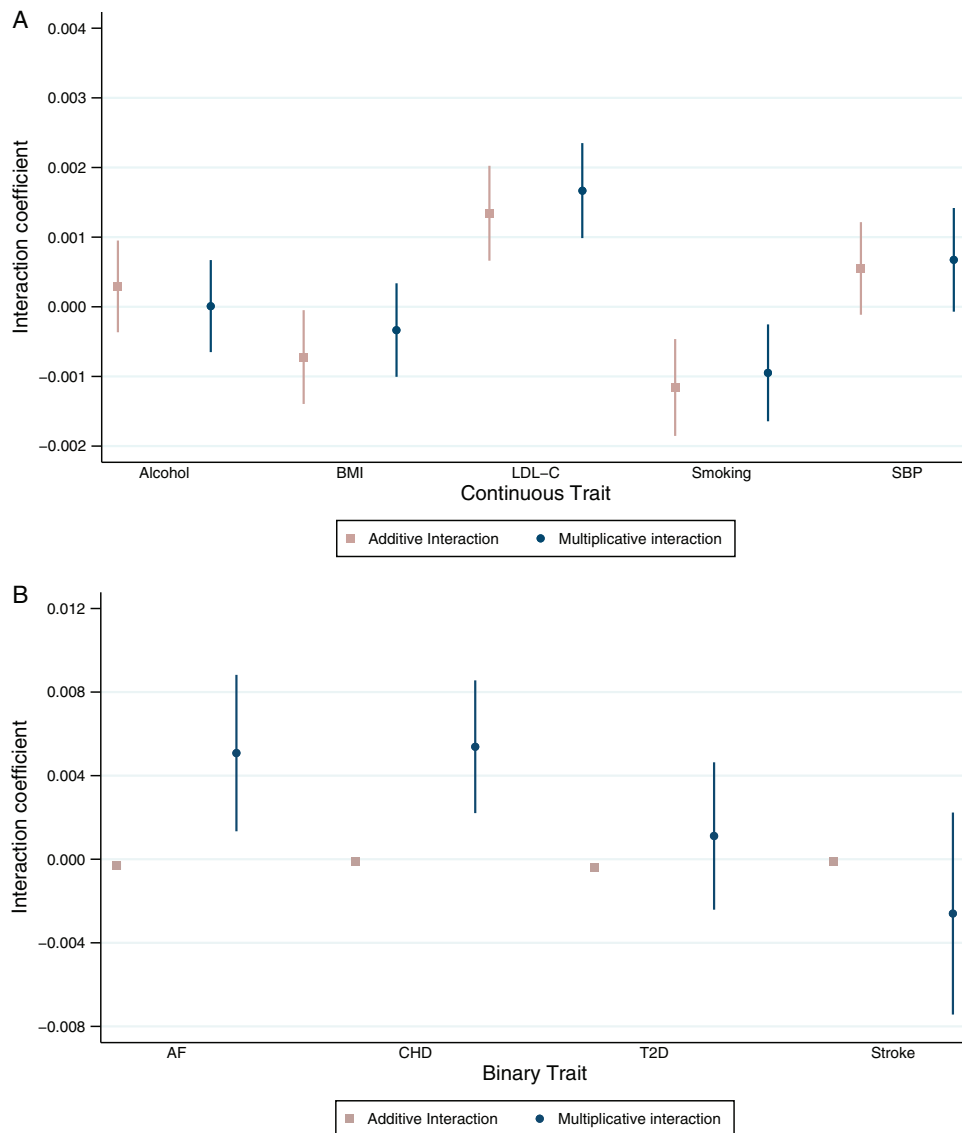
Effect modification by educational attainment of genetic susceptibility to cardiovascular risk factors

For most PGSs, there was evidence that educational attainment modified the effect of the PGS on either the additive or multiplicative scale (Figures 1–3 and Supplementary Table S7 and S8, available as Supplementary data at *IJE* online). The exception was alcohol consumption, for which there was little evidence on either scale.

On the additive scale, higher educational attainment protected against genetic susceptibility to higher BMI, smoking, atrial fibrillation and type 2 diabetes (Figures 1 and 2). For example, a 1 SD increase in PGS for smoking increased the mean difference in lifetime smoking by 0.05 SD [95% confidence interval (CI): 0.04 to 0.06] for those with 7 years of education and by 0.03 SD (95% CI: 0.02 to

Table 2 Descriptive characteristics of the main analysis sample compared with all individuals in UK Biobank at baseline who have not since withdrawn from the study

| Variable		Analysis sample		Full UK Biobank	
		(N = 320 120)		(N = 502 156)	
Continuous variables		N	Mean (SD)	N	Mean (SD)
Age		320 120	56.66 (8.00)	502 156	56.54 (8.09)
Drinks per week		318 300	8.17 (9.05)	497 917	7.79 (9.05)
Body mass index		319 201	27.3 (4.72)	499 065	27.43 (4.8)
Low-density lipoprotein cholesterol		304 700	3.61 (0.86)	468 390	3.56 (0.87)
Systolic blood pressure		292 277	138.16 (18.58)	456 647	137.79 (18.62)
Smoking (lifetime behaviour)		301 684	0.32 (0.66)	318 112	0.34 (0.67)
Categorical variables		N	Frequency (%)	N	Frequency (%)
Sex	Female	320 120	175 108 (55)	502 156	273 025 (54)
Years of education	7 years	320 120	52 012 (16)	493 033	84 648 (17)
	10 years		54 899 (17)		82 357 (17)
	13 years		17 355 (5)		26 857 (5)
	15 years		39 144 (12)		58 271 (12)
	19 years		51 418 (16)		77 668 (16)
	20 years		105 292 (33)		163 232 (33)
Atrial fibrillation (incident)	Control	316 912	307 352 (97)	495 772	480 007 (97)
	Case		9560 (3)		15 765 (3)
Coronary artery disease (incident)	Control	317 055	302 574 (95)	481 533	458 689 (95)
	Case		14 481 (5)		22 844 (5)
Type 2 diabetes (incident)	Control	316 406	305 327 (96)	492 726	472 098 (96)
	Case		11 079 (4)		20 628 (4)
Stroke (incident)	Control	320 120	314 191 (98)	497 151	487 084 (98)
	Case		5929 (2)		10 067 (2)

Figure 1 Coefficient for educational attainment as an effect modifier of polygenic susceptibility to cardiovascular risk factors or diseases on the additive and multiplicative scale. Analyses adjusted for age, sex and 40 genetic principal components. Alcohol = drinks per week; BMI = body mass index; LDL-C = low-density lipoprotein cholesterol; smoking = lifetime smoking behaviour; SBP = systolic blood pressure; AF = atrial fibrillation; CHD = coronary heart disease; T2D = type 2 diabetes. Analyses for binary outcomes on the multiplicative scale are presented as log odds ratios

0.03) for people with 20 years of education ([Figures 1](#) and [2](#) and [Supplementary Table S7](#), available as [Supplementary data](#) at *IJE* online).

Also on the additive scale, higher educational attainment increased genetic susceptibility to LDL-C and systolic blood pressure. For example, for those with 7 years of education, an increase of 1 SD in the PGS for LDL-C increased mean LDL-C by 0.19 SD (95% CI: 0.18 to 0.19) compared with 0.22 SD (95% CI: 0.22 to 0.23) for people with 20 years of education per SD increase in PGS ([Figures 1](#) and [2](#) and [Supplementary Table S7](#), available as [Supplementary data](#) at *IJE* online).

On the multiplicative scale, there was evidence that higher educational attainment increased genetic susceptibility to atrial fibrillation and CHD. For example, for a 1 SD increase in atrial fibrillation PGS, the odds ratio for atrial fibrillation in individuals with 7 years of education was 1.59 (95% CI: 1.45 to 1.57) and for people with 20 years of educational attainment the odds ratio was 1.65 (95% CI: 1.59 to 1.71) ([Figures 1](#) and [3](#) and [Supplementary Table S8](#), available as [Supplementary data](#) at *IJE* online). There was little evidence of modification by education on the multiplicative scale for all other PGSs.
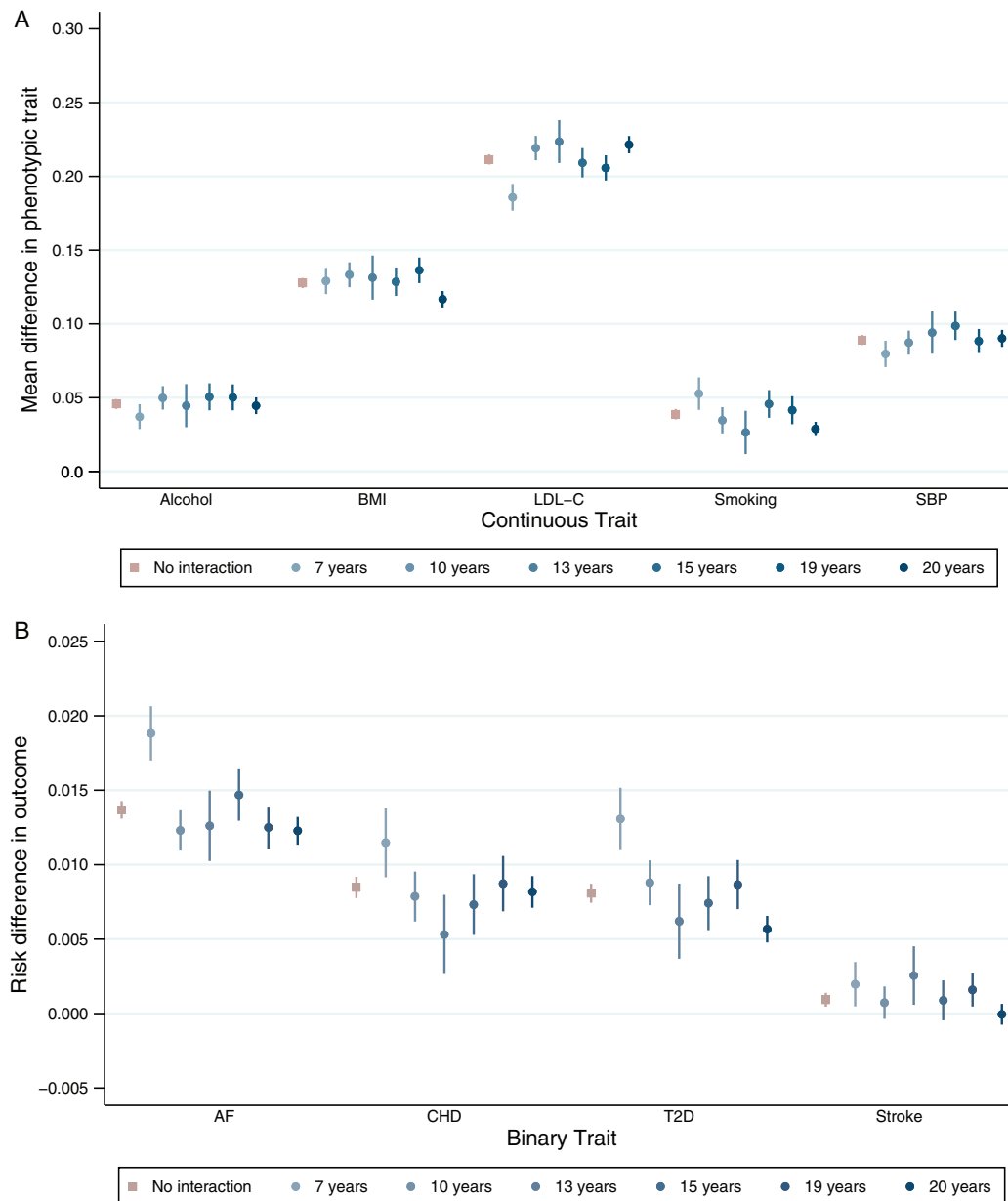
Figure 2 Association between polygenic scores for susceptibility to cardiovascular risk and phenotypic measure of each risk factor, stratified by educational attainment demonstrating effect modification on the additive scale. Analyses adjusted for age, sex and 40 genetic principal components. Alcohol (drinks per week) $P_{EM} = 0.384$; body mass index (BMI) $P_{EM} = 0.036$; low-density lipoprotein cholesterol (LDL-C) $P_{EM} = 1.12 \times 10^{-4}$; lifetime smoking behaviour $P_{EM} = 0.001$; systolic blood pressure (SBP) $P_{EM} = 0.104$; atrial fibrillation (AF) $P_{EM} = 9.03 \times 10^{-8}$; coronary heart disease (CHD) $P_{EM} = 0.103$; type 2 diabetes (T2D) $P_{EM} = 3.23 \times 10^{-10}$; stroke $P_{EM} = 0.036$. $P_{EM} = P$-value for effect modification

For all outcomes, the size of the coefficients for effect modification was small. Non-linear effects by strata of educational attainment were observed for a number of outcomes, including LDL-C, smoking, atrial fibrillation, CHD and type 2 diabetes. For some outcomes, such as with BMI, the effect modification is observed at a single level of educational attainment (Figures 2 and 3).

Secondary analyses

Analyses using more liberal P-value thresholds for the PGS were broadly consistent with the main results. Similar

directions of effect were observed, e.g. on the additive scale, a one-unit increase in educational attainment protected against genetic susceptibility to BMI and lifetime smoking behaviour (Supplementary Table S9 and S10, available as Supplementary data at *IJE* online).

Discussion

In this analysis of UK Biobank, we found evidence that educational attainment modified the risk of genetic susceptibility to some, but not all, cardiovascular risk factors/
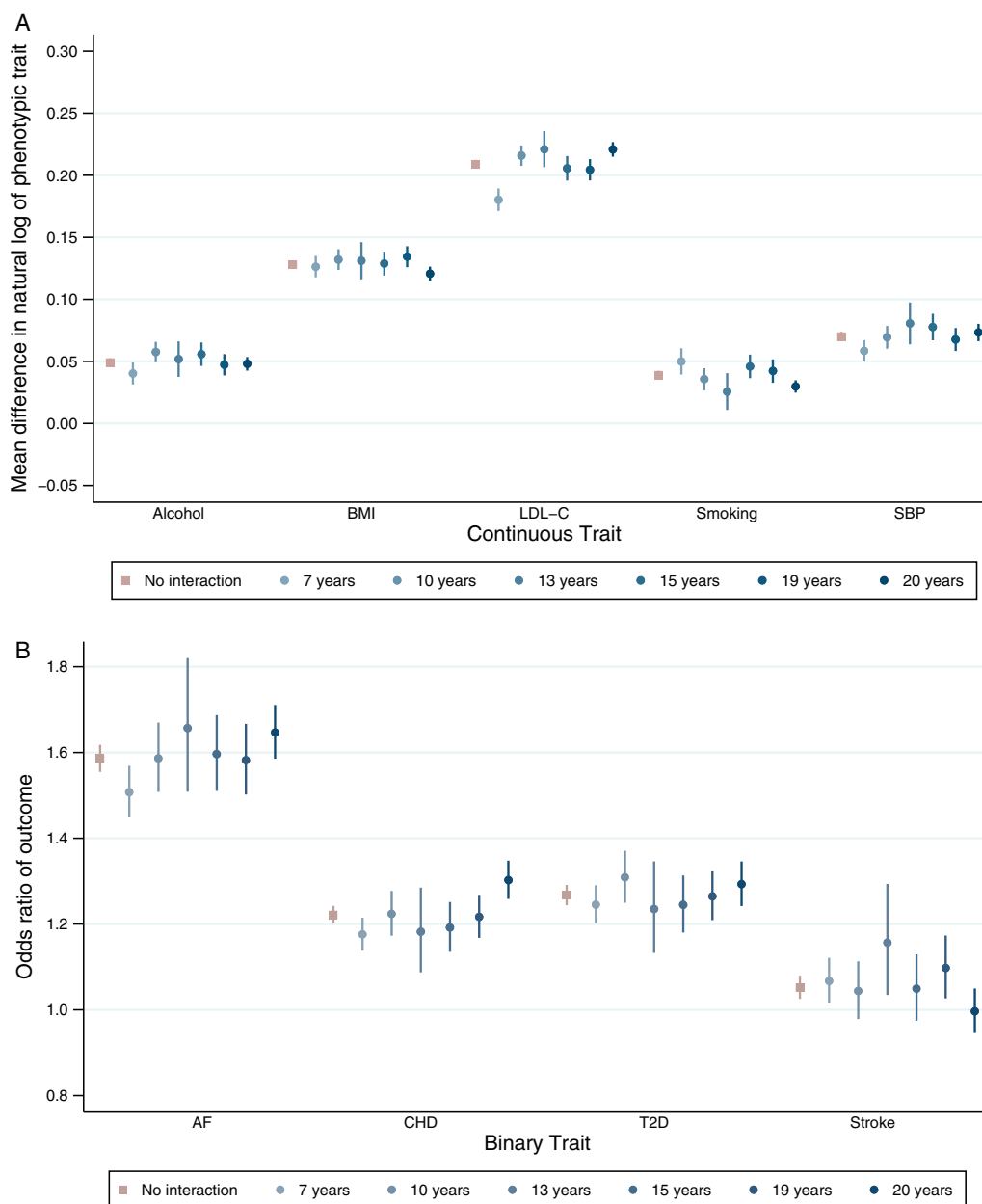
Figure 3 Association between polygenic scores for susceptibility to cardiovascular risk and phenotypic measure of each risk factor, stratified by educational attainment demonstrating effect modification on the multiplicative scale. Analyses adjusted for age, sex and 40 genetic principal components. Alcohol (drinks per week) $P_{EM} = 0.976$; body mass index (BMI) $P_{EM} = 0.330$; low-density lipoprotein cholesterol (LDL-C) $P_{EM} = 1.63 \times 10^{-6}$; lifetime smoking behaviour $P_{EM} = 0.008$; systolic blood pressure (SBP) $P_{EM} = 0.076$; atrial fibrillation (AF) $P_{EM} = 0.008$; coronary heart disease (CHD) $P_{EM} = 8.94 \times 10^{-4}$; type 2 diabetes (T2D) $P_{EM} = 0.537$; stroke $P_{EM} = 0.292$. $P_{EM} = P$-value for effect modification

outcomes. Our a priori hypothesis was that higher levels of education would mitigate genetic susceptibility to cardiovascular risk. However, in several cases, the effect modification was in the other direction, i.e. higher education accentuated genetic predisposition. Furthermore, the magnitude of the differences in associations between PGSs and cardiovascular risk factors/outcomes across levels of educational attainment was small in all cases. These results suggest that modification of the effect of PGSs by

educational attainment is unlikely to play a substantial role in the generation of educational inequalities in CVD.

Results in context

A number of studies have sought to identify the interplay between genetic susceptibility to cardiovascular risk factors with a range of lifestyle and environmental factors.[28–33] However, few have considered the role of SEP interacting

with genetic risk or investigated a wide range of cardiovascular risk factors/outcomes.

Two recent studies using UK Biobank demonstrated that a greater Townsend deprivation index score accentuated the genetic risk of obesity.[8,9] In contrast to our results, the previous literature has not found evidence that education modifies the genetic risk of obesity.[9,10] This may be related to power, where previous studies have used smaller sample sizes to estimate interactions.

These differences could also be due to the education definition used. Here, we used the ISCED years of schooling measure, whereas previous research has used age of completing full-time education[9] and highest qualification.[10]

Typically, non-linear associations were observed when stratifying by years of education, demonstrating that years of education is not a homogenous exposure. For many outcomes, including LDL-C, smoking, atrial fibrillation, CHD and type 2 diabetes, effect modification was driven by individuals with the lowest levels of education. These non-linear effects may be explained by later measures of adult SEP. Much of the variation in educational attainment is determined by early adulthood and therefore does not capture later-life factors that may be important in the development of cardiovascular inequalities, such as occupation or income.

Strengths and weaknesses, and caveats to the analysis of effect modification

There are a number of strengths in this study. Many previous analyses of gene*environment interactions in CVD rely on candidate gene studies,[32,34,35] often resulting in spurious associations.[36] We have used PGSs for nine cardiovascular risk factors/outcomes. Whilst candidate gene studies focus on a single genetic variant, or a small group of (common) genetic variants that individually explain a large(r) amount of the variance in the trait, PGSs include a large number of genetic variants, each explaining a small amount of the variation, but cumulatively explaining a large amount.[37,38] For most diseases, including CVD, polygenic inheritance of these common variants plays a greater role than rare monogenic mutations.[38,39] Therefore, the broad measure of genetic susceptibility used here is likely to represent a greater number of biological pathways for the aetiology of CVD.

We created PGSs at a range of *P*-value thresholds. At a more stringent threshold (e.g. $P \leq 5 \times 10^{-8}$), the genetic variants included are less likely to be pleiotropic (i.e. also associated with different phenotypes), but the variance explained by the PGS may be lower than with a more liberal threshold (e.g. $P \leq 0.5$). Additionally, less-stringent clumping thresholds were used to improve polygenic prediction, but this may introduce pleiotropic SNPs.

The lack of evidence for effect modification between education and the PGS for alcohol consumption observed here could be due to insufficient power to detect an interaction or because of the variable definition. Alcohol consumption was defined as drinks per week, but the type of alcoholic drink consumed may be an important factor. Additionally, alcohol consumption was self-reported by participants, which is prone to recall bias.[40,41] If this recall bias is differential by educational attainment, this may mask any effect modification between educational attainment and genetic susceptibility to alcohol consumption. Alternatively, different patterns of drinking may occur by strata of educational attainment. It has been shown that individuals of lower SEP are more likely to drink to extreme levels,[42] but individuals of higher SEP consume similar or even greater amounts of alcohol.[43]

Where effect modification was found, should different definitions of the outcome variables be used, e.g. smoking initiation as opposed to lifetime smoking behaviour, the observed evidence of effect modification may change. Similarly, the effect modification identified here may differ for alternative measures of SEP.

These results may be specific to the model used to derive PGSs. For example, a recent GWAS of systolic blood pressure demonstrated that educational attainment interacts with the genetic architecture of blood pressure.[44] If these interactions were accounted for when deriving the PGS, different results may be observed.

Low statistical power reduces the chance of detecting a true effect should one exist.[45] Although some power calculators have been developed to calculate power in gene*environment interaction analyses,[46] to our knowledge, none has been developed for use with PGSs. Additionally, power calculations rely on making assumptions about the true effect size, which is difficult to estimate in this case. Therefore, we believe it is more informative to interpret the results and likely power based on the point estimates, standard errors and width of the confidence intervals. However, it is possible that we did not detect effect modification by education on the effects of alcohol consumption due to insufficient power to detect an interaction, as demonstrated by small point estimates and wide confidence intervals.

Studies of effect modification can be biased by reverse causality and confounding. Analyses of CVD outcomes were restricted to incident cases to avoid bias from behaviour changes following a CVD diagnosis to avoid reverse causality. Genetic variants are determined at conception and therefore not affected by unmeasured later-life confounding factors. However, they can be confounded by

population structure.[47] In this analysis, we controlled for genetic principal components to minimize this bias.

It has been suggested that further to controlling for confounders, the interaction between the (i) confounders and environmental exposure and (ii) confounders and genetic exposure should be controlled for.[48,49] This avoids specification error by accounting for the covariation between the confounders and the interactions tested. However, due to the large number of principal components included as confounders in these analyses, there is not enough variation in the data to include these additional covariates. Therefore, these analyses may be biased by residual confounding.

UK Biobank participants are typically more highly educated and of a higher SEP than the UK population.[14] Therefore, evidence of effect modification by education in this sample may be due to collider bias caused by non-random selection into the study.[14,50]

These results do not specifically identify what it is about educational attainment that modifies genetic susceptibility to cardiovascular risk. For example, remaining in education may lead to an increased knowledge of the smoking harms, even if they have genetic variants increasing their susceptibility to heavier smoking.[51] Indeed, a number of alternative factors associated with educational attainment could be contributing to the effect modification. For example, parental genotype or family-level environmental factors may explain both educational attainment and differential effects of PGSs by strata of education.

Due to limited power, we have not used causal inference methods to test whether education is causal with respect to all outcomes. To understand whether effect modification by education is causal, exogenous exposures for education, such as the Raising of the School Leaving Age (RoSLA), could be used. However, it is challenging to identify sufficiently large samples that were both exposed to the RoSLA and have genotypic data.

Where educational attainment increased genetic susceptibility to CVD, such as for atrial fibrillation, effect modification may be due to differential rates of diagnosis, which may independently contribute to cardiovascular inequalities. Whilst risk factors such as BMI and smoking were measured near universally in participants at baseline, CVD was ascertained through linkage to hospital inpatient records.

Interpreting analyses of interaction and effect modification

The terms interaction and effect modification are often used interchangeably in modern epidemiology. Whilst statistically the same, the distinction can be made where an interaction is defined in terms of the effects of two causal risk factors, whereas effect modification specifies that the effect of one risk factor varies by strata of a second factor, the effect of which on the outcome is not necessarily causal.[52] We have used the term 'effect modification' throughout this analysis, where we specifically hypothesize that the effect of the PGSs varies by strata of educational attainment. This term also acknowledges that we have not explicitly tested the causal associations between (i) educational attainment and (ii) PGSs on each outcome.

Interaction and effect modification have often been dichotomized into 'biologic interaction' and 'statistical interaction'.[53,54] Biologic interaction is said to be a deviation from an additive effect of two risk factors on the risk difference of the outcome. However, this term has been criticized for being difficult to interpret and giving potentially misleading assurances about causal biological mechanisms that have not been assessed.[54]

Statistical interaction is described as the deviation from the expected effect of two joint risk factors, under the assumption the risk factors are independent on the additive or the multiplicative scale.[54] When two risk factors are associated with the outcome, there should always be evidence of an interaction on at least one scale, so we present results on both the additive and the multiplicative scales.[13] This is an important distinction from previous analyses, which have typically only reported results on the additive scale.[8,9] A full discussion of additive and multiplicative interactions can be found elsewhere.[13]

Public health relevance

To determine the public health relevance of these results, it is important to interpret the magnitude and direction of any effect modification. Coefficients for effect modification were uniformly small in this analysis and the direction of the effect across outcomes was not consistent. This indicates that any effect modification by educational attainment on the effect of genetic susceptibility to cardiovascular risk factors/outcomes is unlikely to contribute to the development of inequalities in cardiovascular risk. Although some results were scale-dependent—e.g. greater evidence of effect modification by education on genetic susceptibility to type 2 diabetes on the additive scale—the direction of effect modification was generally consistent within outcomes across both the additive and multiplicative scales. Given the small coefficients for effect modification, these differences in precision are likely driven by low power.

Conclusions

In this study, we found that educational attainment modifies genetic susceptibility to a number of cardiovascular

risk factors/outcomes. The direction of this effect was mixed and the size of the effect-modification coefficients was small. This suggests that effect modification by educational attainment on the effect of genetic susceptibility to cardiovascular risk factors/outcomes is unlikely to explain the development of inequalities in cardiovascular risk.

Ethics approval

This research was conducted using the UK Biobank resource using the approved application 10953.

Author contributions

A.R.C. designed the study, cleaned and analysed the data, interpreted the results, and wrote and revised the manuscript. S.H. assisted with data analysis, interpreted the results and critically reviewed and revised the manuscript. D.G. interpreted the results and critically reviewed and revised the manuscript. G.D.S., A.E.T., N.M.D. and L.D.H. all designed the study, interpreted the results, critically reviewed and revised the manuscript, and provided supervision for the project. N.M.D. and L.D.H. contributed equally and are joint senior authors of this manuscript. A.R.C. and N.M.D. serve as guarantors of the paper. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

A.R.C. affirms that the manuscript is an honest, accurate and transparent account of the study being reported and no important aspects of the study have been omitted.

Data availability

The data used in this study have been archived with the UK Biobank study. Summary data for external weightings of polygenic scores are available from each GWAS. The code used to derive polygenic scores is available at https://github.com/sean-harrison-bristol/UK_Biobank_PRS/blob/master/mrbase_grs_v3.01.R and the analysis code is available at github.com/alicerosecarter/gxe_cv_riskfactors.

Supplementary data

Supplementary data are available at *IJE* online.

Conflict of interest

D.G. is employed part-time by Novo Nordisk. For the remaining authors, none declared.

References

1. Rosengren A, Smyth A, Rangarajan S *et al.* Socioeconomic status and risk of cardiovascular disease in 20 low-income, middle-income, and high-income countries: the Prospective Urban Rural Epidemiologic (PURE) study. *Lancet Glob Health* 2019;**7**: e748–60.
2. Sing CF, Stengard JH, Kardia SL. Genes, environment, and cardiovascular disease. *Arterioscler Thromb Vasc Biol* 2003;**23**: 1190–96.
3. Kathiresan S, Srivastava D. Genetics of human cardiovascular disease. *Cell* 2012;**148**:1242–57.
4. Bhatnagar A. Environmental Determinants of Cardiovascular Disease. *Circ Res* 2017;**121**:162–80.
5. Ellinor PT, Lunetta KL, Albert CM *et al.* Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nat Genet* 2012; **44**:670–75.
6. Nikpay M, Goel A, Won HH *et al.* A comprehensive 1,000 genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* 2015;**47**:1121–30.
7. Malik R, Chauhan G, Traylor M *et al.*; MEGASTROKE Consortium. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet* 2018;**50**:524–37.

8. Tyrrell J, Wood AR, Ames RM *et al.* Gene-obesogenic environment interactions in the UK Biobank study. *Int J Epidemiol* 2017;**46**:559–75.

9. Rask-Andersen M, Karlsson T, Ek WE, Johansson A. Gene-environment interaction study for BMI reveals interactions between genetic factors and physical activity, alcohol consumption and socioeconomic status. *PLoS Genet* 2017;**13**: e1006977.

10. Amin V, Bockerman P, Viinikainen J *et al.* Gene-environment interactions between education and body mass: evidence from the UK and Finland. *Soc Sci Med* 2017;**195**:12–16.

11. Roskam AJ, Kunst AE, Van Oyen H *et al.*; for additional participants to the study. Comparative appraisal of educational inequalities in overweight and obesity among adults in 19 European countries. *Int J Epidemiol* 2010;**39**:392–404.

12. Flowers E, Froelicher ES, Aouizerat BE. Gene-environment interactions in cardiovascular disease. *Eur J Cardiovasc Nurs* 2012; **11**:472–78.

13. VanderWeele TJ, Knol MJ. A tutorial on interaction. *Epidemiol Methods* 2014;**3**:33–72.

14. Fry A, Littlejohns TJ, Sudlow C *et al.* Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. *Am J Epidemiol* 2017;**186**:1026–34.

15. Collins R. What makes UK Biobank special? *Lancet* 2012;**379**: 1173–74.

16. Okbay A, Beauchamp JP, Fontana MA *et al.*; LifeLines Cohort Study. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 2016;**533**:539–42.

17. Lee JJ, Wedow R, Okbay A, Social Science Genetic Association Consortium *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* 2018;**50**:1112–21.

18. Carter AR, Gill D, Davies NM *et al.* Understanding the consequences of education inequality on cardiovascular disease: Mendelian randomisation study. *BMJ* 2019;**365**:l1855.

19. Liu M, Jiang Y, Wedow R *et al.*; HUNT All-In Psychiatry. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* 2019;**51**:237–44.

20. Locke AE, Kahali B, Berndt SI *et al.*; The LifeLines Cohort Study. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015;**518**:197–206.

21. Scott RA, Scott LJ, Magi R *et al.* An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes* 2017; **66**:2888–902.

22. Willer CJ, Schmidt EM, Sengupta S *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013;**45**: 1274–83.

23. Wootton RE, Richmond RC, Stuijfzand BG *et al.* Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: a Mendelian randomisation study. *Psychol Med* 2020;**50**: 2435–43.

24. Hemani G, Zheng J, Elsworth B *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 2018;**7**:e34408.

25. Kopec JA, Esdaile JM. Bias in case-control studies: a review. *J Epidemiol Community Health* 1990;**44**:179–86.

26. NICE (2019). Lipid modification—CVD prevention. https://cks. nice.org.uk/lipid-modification-cvd-prevention (17 November 2020, date last accessed).

27. Mitchell R, Hemani G, Dudding T, Corbin L, Harrison S, Paternoster L. UK Biobank Genetic Data: MRC-IEU Quality Control, version 2. 2019.

28. Qi Q, Chu AY, Kang JH *et al.* Sugar-sweetened beverages and genetic risk of obesity. *N Engl J Med* 2012;**367**:1387–96.

29. Reddon H, Gerstein HC, Engert JC *et al.* Physical activity and genetic predisposition to obesity in a multiethnic longitudinal study. *Sci Rep* 2016;**6**:18672.

30. Qi Q, Li Y, Chomistek AK *et al.* Television watching, leisure time physical activity, and the genetic predisposition in relation to body mass index in women and men. *Circulation* 2012; **126**:1821–27.

31. Qi Q, Chu AY, Kang JH *et al.* Fried food consumption, genetic risk, and body mass index: gene-diet interaction analysis in three US cohort studies. *BMJ* 2014;**348**:g1610.

32. Lindi VI, Uusitupa MI, Lindstrom J *et al.*; Finnish Diabetes Prevention Study. Association of the Pro12Ala polymorphism in the PPAR-gamma2 gene with 3-year incidence of type 2 diabetes and body weight change in the Finnish Diabetes Prevention Study. *Diabetes* 2002;**51**:2581–86.

33. Hamrefors V, Hedblad B, Hindy G *et al.* Smoking modifies the associated increased risk of future cardiovascular disease by genetic variation on chromosome 9p21. *PLoS One* 2014;**9**: e85893.

34. Lahoz C, Schaefer EJ, Cupples LA *et al.* Apolipoprotein E genotype and cardiovascular disease in the Framingham Heart Study. *Atherosclerosis* 2001;**154**:529–37.

35. Schmidt B, Frölich S, Dragano N *et al.* Socioeconomic status interacts with the genetic effect of a chromosome 9p21.3 common variant to influence coronary artery calcification and incident coronary events in the Heinz Nixdorf Recall Study (risk factors, evaluation of coronary calcium, and lifestyle). *Circ Cardiovasc Genet* 2017;**10**:e001441.

36. Duncan LE, Keller MC. A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *Am J Psychiatry* 2011;**168**:1041–49.

37. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res* 2007;**17**:1520–28.

38. Khera AV, Chaffin M, Aragam KG *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* 2018;**50**:1219–24.

39. Gibson G. Rare and common variants: twenty arguments. *Nat Rev Genet* 2012;**13**:135–45.

40. Del Boca FK, Darkes J. The validity of self-reports of alcohol consumption: state of the science and challenges for research. *Addiction* 2003;**98**(Suppl 2):1–12.

41. Ekholm O. Influence of the recall period on self-reported alcohol intake. *Eur J Clin Nutr* 2004;**58**:60–63.

42. Lewer D, Meier P, Beard E, Boniface S, Kaner E. Unravelling the alcohol harm paradox: a population-based study of social gradients across very heavy drinking thresholds. *BMC Public Health* 2016;**16**:599.

43. Beard E, Brown J, West R, Kaner E, Meier P, Michie S. Associations between socio-economic factors and alcohol

consumption: a population survey of adults in England. *PLoS One* 2019;**14**:e0209442.

44. de Las Fuentes L, Sung YJ, Noordam R *et al.*; LifeLines Cohort Study. Gene-educational attainment interactions in a multi-ancestry genome-wide meta-analysis identify novel blood pressure loci. *Mol Psychiatry* 2021;**26**:2111–25.

45. Button KS, Ioannidis JP, Mokrysz C *et al.* Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 2013;**14**:365–76.

46. Moore CM, Jacobson SA, Fingerlin TE. Power and sample size calculations for genetic association studies in the presence of genetic model misspecification. *Hum Hered* 2019;**84**:256–71.

47. Haworth S, Mitchell R, Corbin L *et al.* Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nat Commun* 2019;**10**:333.

48. Keller MC. Gene x environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. *Biol Psychiatry* 2014;**75**:18–24.

49. Domingue BW, Trejo S, Armstrong-Carter E, Tucker-Drob EM. Interactions between polygenic scores and environments: methodological and conceptual challenges. *SocScience* 2020;**7**:365–86.

50. Munafo MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: when selection bias can substantially influence observed associations. *Int J Epidemiol* 2018;**47**:226–35.

51. Siahpush M, McNeill A, Hammond D, Fong GT. Socioeconomic and country variations in knowledge of health risks of tobacco smoking and toxic constituents of smoke: results from the 2002 International Tobacco Control (ITC) Four Country Survey. *Tob Control* 2006;**15**(Suppl 3):iii65–70.

52. VanderWeele TJ. On the distinction between interaction and effect modification. *Epidemiology* 2009;**20**:863–71.

53. Rothman KJ. Synergy and antagonism in cause-effect relationships. *Am J Epidemiol* 1974;**99**:385–88.

54. Lawlor DA. Biological interaction: time to drop the term? *Epidemiology* 2011;**22**:148–50.