



OPEN ACCESS



# Understanding the consequences of education inequality on cardiovascular disease: mendelian randomisation study

Alice R Carter,<sup>1,2</sup> Dipender Gill,<sup>3</sup> Neil M Davies,<sup>1,2</sup> Amy E Taylor,<sup>2,4</sup> Taavi Tillmann,<sup>5</sup> Julien Vaucher,<sup>6,7</sup> Robyn E Wootton,<sup>1,4,8</sup> Marcus R Munafò,<sup>1,4,8,9</sup> Gibran Hemani,<sup>1,2</sup> Rainer Malik,<sup>10</sup> Sudha Seshadri,<sup>11,14</sup> Daniel Woo,<sup>15</sup> Stephen Burgess,<sup>1,16,17</sup> George Davey Smith,<sup>1,2,4</sup> Michael V Holmes,<sup>1,18,20</sup> Ioanna Tzoulaki,<sup>3,21,22</sup> Laura D Howe,<sup>1,2</sup> Abbas Dehghan<sup>3,21</sup>

For numbered affiliations see end of the article.

Correspondence to: A Carter [alice.carter@bristol.ac.uk](mailto:alice.carter@bristol.ac.uk) (or @alicerosecarter on Twitter ORCID 0000-0003-2817-4195)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2019;365:l1855 <http://dx.doi.org/10.1136/bmj.l1855>

Accepted: 25 March 2019

## ABSTRACT

### OBJECTIVES

To investigate the role of body mass index (BMI), systolic blood pressure, and smoking behaviour in explaining the effect of education on the risk of cardiovascular disease outcomes.

### DESIGN

Mendelian randomisation study.

### SETTING

UK Biobank and international genome-wide association study data.

### PARTICIPANTS

Predominantly participants of European ancestry.

### EXPOSURE

Educational attainment, BMI, systolic blood pressure, and smoking behaviour in observational analysis, and randomly allocated genetic variants to instrument these traits in mendelian randomisation.

### MAIN OUTCOMES MEASURE

The risk of coronary heart disease, stroke, myocardial infarction, and cardiovascular disease (all subtypes; all measured in odds ratio), and the degree to which this is mediated through BMI, systolic blood pressure, and smoking behaviour respectively.

### RESULTS

Each additional standard deviation of education (3.6 years) was associated with a 13% lower risk of coronary heart disease (odds ratio 0.86, 95% confidence interval 0.84 to 0.89) in observational analysis and a 37% lower risk (0.63, 0.60 to 0.67) in mendelian randomisation analysis. As a proportion of the total risk reduction, BMI was estimated to mediate 15% (95% confidence interval 13% to 17%) and 18% (14% to 23%) in the observational and mendelian randomisation estimates, respectively. Corresponding estimates were 11% (9% to 13%) and 21% (15% to 27%) for systolic blood pressure and 19% (15% to 22%) and 34% (17% to 50%) for smoking behaviour.

All three risk factors combined were estimated to mediate 42% (36% to 48%) and 36% (5% to 68%) of the effect of education on coronary heart disease in observational and mendelian randomisation analyses, respectively. Similar results were obtained when investigating the risk of stroke, myocardial infarction, and cardiovascular disease.

### CONCLUSIONS

BMI, systolic blood pressure, and smoking behaviour mediate a substantial proportion of the protective effect of education on the risk of cardiovascular outcomes and intervening on these would lead to reductions in cases of cardiovascular disease attributable to lower levels of education. However, more than half of the protective effect of education remains unexplained and requires further investigation.

## Introduction

Cardiovascular disease is the leading cause of mortality worldwide, accounting for over 17 million deaths annually.<sup>1</sup> Recent studies suggest that socioeconomic risk factors, such as education, play a causal role in the aetiology of cardiovascular disease.<sup>2-4</sup> Tillmann and colleagues found that an additional 3.6 years of education reduced the risk of coronary heart disease by approximately one third.<sup>2</sup> However, educational opportunities are not equitable throughout populations and education is inherently difficult to intervene on. Therefore, understanding the risk factors that might be driving the adverse later life outcomes associated with lower levels of education would provide the opportunity to reduce inequalities with interventions.

Existing studies suggest that body mass index (BMI), systolic blood pressure, and smoking behaviour partly explain differences in the risk of cardiovascular disease related to educational attainment.<sup>5-7</sup> However, these studies have relied on observational mediation analyses that might suffer from biases. Traditional methods use one snapshot of a risk factor, which could incompletely capture a person's lifetime exposure.<sup>8</sup> For example, systolic blood pressure measured at one time point will suffer from measurement error due to day-to-day fluctuations and will not capture changes across the life course. This measurement error can lead to an underestimation of mediation.<sup>8</sup> Furthermore, other biases such as unmeasured confounding cannot be addressed by using observational methods.<sup>9</sup>

Mendelian randomisation uses genetic variants as instruments to estimate the effect of an exposure

## WHAT IS ALREADY KNOWN ON THIS TOPIC

Lower levels of education are causally related to a higher risk of cardiovascular disease

## WHAT THIS STUDY ADDS

Body mass index, systolic blood pressure, and smoking behaviour mediate the effect of education

Cardiovascular disease attributable to lower levels of education can be reduced by intervening on these risk factors

on an outcome of interest,<sup>10</sup> exploiting the random allocation of genetic variants to infer causal effects that are robust to non-differential measurement error and confounding.<sup>10</sup> Two-step mendelian randomisation for mediation analysis, unlike traditional observational mediation analysis approaches, is both sensitive to the causal effects of the mediator and corrects for its measurement error.<sup>11</sup> Recent genome-wide association study meta-analyses have identified a number of genetic variants for educational attainment that could be used as instrumental variables.<sup>12 13</sup>

Mendelian randomisation has previously been used to show the causal effects of education on BMI, systolic blood pressure, and smoking behaviour and also the effects of BMI and smoking behaviour on cardiovascular disease.<sup>14-19</sup> Although the results from these studies suggest that BMI, systolic blood pressure, and smoking behaviour are likely to explain some of the protective mechanisms of education on cardiovascular disease, they alone do not quantify the mediated effect. In this study, we investigated the role of BMI, systolic blood pressure, and lifetime smoking behaviour in mediating the causal effect of educational attainment on the risk of cardiovascular disease by using three complementary approaches: multivariable regression, one-sample mendelian randomisation, and two-sample mendelian randomisation. BMI, systolic blood pressure, and smoking behaviour were selected as intermediate risk factors based on the literature, implicating them as both being affected by education and as risk factors for cardiovascular disease, with availability of data across all three complementary methods. We consider the three risk factors both individually and simultaneously. Understanding the mechanisms by which education affects cardiovascular health could have powerful applications for public health policy. For this, it is important to understand the population-level implications of changes to BMI, systolic blood pressure, and smoking behaviour on inequalities in the risk of cardiovascular disease.

## Methods

### Overall study design

This study used multivariable regression of observational data, one-sample mendelian randomisation of individual level genetic data, and two-sample mendelian randomisation of summary level genetic data to investigate whether BMI, systolic blood pressure, and lifetime smoking behaviour explain the protective effect of education on the risk of coronary heart disease, stroke, myocardial infarction, and cardiovascular disease (all subtypes).

### Data sources

#### UK Biobank

The UK Biobank recruited 503 317 British adults between 2006 and 2010. Participants attended assessment centres involving questionnaires, interviews, anthropometric, physical, and genetic measurements.<sup>20 21</sup> In the observational analysis, we included 217 013 white British participants, with

complete data on genotypes, age, sex, educational attainment, cardiovascular outcomes, BMI, systolic blood pressure, smoking behaviour, socioeconomic status (as measured by Townsend Deprivation Index at birth), and place of birth. Supplementary figure 1 shows the exclusion criteria for the main UK Biobank analyses, and supplementary figure 2 shows the exclusion criteria for the genome-wide association studies carried out for systolic blood pressure and smoking. White British participants were defined by using both self-reported questionnaire data and similar genetic ancestry to the European ancestry principal components computed from the 1000 genomes project.<sup>22</sup>

The supplementary methods show the data on educational attainment (as measure by highest achieved qualifications), BMI, systolic blood pressure, smoking behaviour, and all covariates. A continuous lifetime measure of smoking behaviour, incorporating smoking initiation, duration, heaviness, and cessation is used in this analysis. Full details are presented elsewhere.<sup>23</sup> Participants reported their highest qualification and age of leaving school if they did not have a degree. These were converted to the International Standard Classification for Education coding of educational attainment (supplementary table 1).<sup>13</sup> We used available follow-up data where baseline data were missing. To account for the effects of antihypertensive treatment, participants who were taking antihypertensive drugs had 10 mm Hg added to their measured systolic blood pressure.<sup>24</sup> Supplementary table 2 shows that cardiovascular disease diagnoses (including diagnoses of coronary heart disease, stroke, and myocardial infarction) and events were ascertained through linkage mortality data and hospital episode statistics, with cases defined according to ICD-9 (international classification of diseases, ninth revision) and ICD-10 codes (international classification of diseases, 10th revision).<sup>25</sup> We excluded participants who had experienced a cardiovascular disease event before the baseline assessment (prevalent cases) and we only considered first event, incident cases after the assessment centre.

### Genome-wide association study meta-analyses

In the two-sample mendelian randomisation analysis, we obtained summary genetic associations from genome-wide association study data for each respective phenotype. For education, this was the Social Science Genetic Association Consortium genome-wide association study meta-analysis of years of schooling in 1 131 881 participants of European ancestry,<sup>13</sup> with summary data made available for 766 345 of these participants. Instruments were selected as the 1271 independent (pairwise  $r^2 < 0.1$ ) genome-wide-wide significant single nucleotide polymorphisms after analysing data from the full sample.<sup>13</sup> We obtained genetic estimates for BMI from the Genetic Investigation of Anthropometric Traits consortium's 2018 genome-wide association study meta-analysis of 681 275 participants of European decent.<sup>26</sup> Genetic association estimates for systolic blood pressure and smoking

behaviour were estimated from a genome-wide association study of 318 147 white British participants in the UK Biobank (see supplementary methods). Instruments for BMI, systolic blood pressure, and smoking behaviour were identified as the lead single nucleotide polymorphisms in loci reaching genome-wide significance after clumping summary estimates from the largest available genome-wide association study for linkage disequilibrium threshold  $r^2 < 0.001$  and distance  $> 10\,000$  kb, by using a 1000 genomes European reference panel through the TwoSampleMR package (default settings of the 'clump\_data' command) in the statistical software R.<sup>27</sup> For coronary heart disease, we used publicly available genetic association estimates from the CARDIoGRAMplusC4D 1000 Genomes-based genome-wide association study meta-analysis of 60 801 cases and 123 504 controls.<sup>28</sup> The definition for coronary heart disease was broad and inclusive, considering acute coronary syndrome, myocardial infarction, angina with one or angiographic stenoses of greater than 50%, and chronic stable angina. Details of the genome-wide association study for stroke and myocardial infarction, and further details for other genome-wide association study are provided in the supplementary methods. All genetic association estimates used in each two-sample mendelian randomisation analyses are provided in supplementary tables 4 to 18.

### Statistical analysis

#### *Effect of education on cardiovascular disease*

In observational analysis of UK Biobank data, we used multivariable logistic regression to estimate the association of education with cardiovascular disease and its subtypes. We adjusted all analyses using UK Biobank data for potential confounders: age, sex, place of birth, birth distance from London, and Townsend Deprivation Index at birth. These confounders were determined a priori, with place of birth and birth distance from London included to control for population structure in the UK Biobank.<sup>3 29</sup>

In two-sample mendelian randomisation analysis, the effects of education on cardiovascular subtypes were investigated using ratio method mendelian randomisation with standard errors derived using the delta method.<sup>30</sup> We used fixed-effect inverse-variance weighted meta-analysis to pool mendelian randomisation estimates across individual single nucleotide polymorphisms.<sup>31</sup>

#### *Mediation by BMI, systolic blood pressure, and smoking behaviour*

In multivariable observational and one-sample mendelian randomisation analyses, when investigating the degree to which the effects of education on cardiovascular disease and its subtypes are mediated through each risk factor (BMI, systolic blood pressure, and smoking behaviour) individually, we used the product of coefficients method to estimate the indirect effect (that is, the effect of education on cardiovascular disease that goes through the risk factor).<sup>32</sup> This

involved first estimating the effect of education on each risk factor individually, then multiplying this with the effect of that risk factor on the outcome after adjusting for education. We estimated the proportion of the overall effect of education on cardiovascular disease subtypes that was mediated by each risk factor by dividing the indirect effect by the total effect. Standard errors were derived by bootstrapping in the observational and one-sample mendelian randomisation analysis and by using the delta method in the two-sample mendelian randomisation analysis.

We used multivariable linear regression in the observational analysis to estimate the association of education with each risk factor after adjusting for confounders (as in the total effects models). We then estimated the effect of each risk factor on the individual cardiovascular disease subtypes with additional adjustment for self-reported educational attainment.<sup>33</sup> The two estimates were multiplied together to estimate the indirect effect (of education, through the risk factor).

We used the inverse-variance weighted mendelian randomisation approach for the two-sample mendelian randomisation to estimate the effect of education on each risk factor. We used regression-based multivariable mendelian randomisation to estimate the effect of each risk factor on the risk of the considered cardiovascular disease subtypes, adjusting for genetic effect of the instruments on education.<sup>34</sup> We estimated the indirect effect of education on the risk of each cardiovascular disease subtype through the considered risk factor by multiplying results from these two mendelian randomisation analyses.

Methods for investigating the role of all three risk factors together are presented in the supplementary methods.

#### *One-sample mendelian randomisation and sensitivity analyses*

One-sample mendelian randomisation analysis was carried out in the UK Biobank. Full details of the genetic variants used are in the supplementary methods and supplementary tables 19 to 21. In brief, a weighted allele score was created based on genome-wide significant single nucleotide polymorphisms identified in recent genome-wide association study meta-analyses for the considered exposures (that is, educational attainment, BMI, systolic blood pressure, and smoking behaviour).<sup>12 23 35</sup> These genome-wide association study estimates were selected from studies that did not include UK Biobank participants, so as to avoid participant overlap, and therefore, in some cases, the genome-wide association study and subsequent instruments differed from the genome-wide association study studies used for the two-sample mendelian randomisation described previously. Table 1 shows a summary of all phenotypes and genome-wide association study data used. The variants in each instrument were harmonised for consistent directions of association and each single nucleotide polymorphism in the genetic score was weighted by its relative effect size in the respective genome-wide association study,

Table 1 | Summary of phenotypes and genome-wide association study (GWAS) data used as instrumental variables across analyses

Exposure	Multivariable observational analysis*	Mendelian randomisation	
		One-sample	Two-sample
Educational attainment	Self-reported highest qualification mapped to ISCED years of schooling	Weighted allele score, using genome-wide significance SNPs (n=74) and $\beta$ weights from Okbay and colleagues, 2016 <sup>58</sup>	Individual SNPs from Lee and colleagues, 2018 <sup>13</sup> (n=1271)
BMI	Measured weight and height	Weighted allele score, using genome-wide significance SNPs (n=77) and $\beta$ weights from Locke and colleagues, 2015 <sup>35</sup>	Individual SNPs from Yengo and colleagues, 2018 <sup>26</sup> (n=360)
Systolic blood pressure	Median of two automated blood pressure measurements	Weighted allele score, using genome-wide significance SNPs (n=65 and 55 sample 1 and 2 respectively) from a split sample GWAS in the UK Biobank†	Individual SNPs from SBP GWAS carried out as part of this work on full UK Biobank sample† (n=191)
Smoking behaviour	Estimate of lifetime smoking behaviour using self-report data	Weighted allele score, using genome-wide significance SNPs (n=18 and 15 sample 1 and 2 respectively) from a split sample GWAS in UK Biobank† <sup>23</sup>	Individual SNPs from Wootton and colleagues, 2018 using full UK Biobank sample <sup>23</sup> (n=126)

\*All in the UK Biobank.

†Full methods in the supplementary material.

ISCED=International Standard Classification for Education; SNPs=single nucleotide polymorphisms; GWAS=genome-wide association study

with effects combined in an additive model. Analyses using the UK Biobank were replicated using the risk difference scale. A range of sensitivity analyses were carried out, including exploring the assumption of no pleiotropy within mendelian randomisation analyses by using mendelian randomisation-Egger and weighted median mendelian randomisation.<sup>34 36</sup> One-sample mendelian randomisation analyses and multivariable observational associations between education and cardiovascular disease were repeated after stratifying for sex and age below or above the median (39-57 years compared with 58-72 years). On a subsample of participants with dietary recall data, an observational model was run considering diet and exercise as mediators in addition to BMI, systolic blood pressure, and smoking behaviour mediators. Full details of sensitivity analyses are provided in the supplementary methods.

#### Statistical software and ethical approval

We performed the analysis by using Stata version 14 (StataCorp LP) and R version 3.4.3 (The R Foundation for Statistical Computing). We used the *mrr* package for Stata and the *TwoSampleMR* package for R to facilitate MR analyses.<sup>27 37</sup>

#### Patient and public involvement

No patients or participants were involved in setting the research question or the outcome measures, nor were they involved in developing plans for design or implementation of the study. No patients were asked to advise on interpretation or writing up of results. There are no plans to disseminate the results of the research to study participants or the relevant patient community.

## Results

### UK Biobank cohort description

The UK Biobank sample used in the observational and one-sample mendelian randomisation analysis was comparable to the participants in UK Biobank as a whole, although the UK Biobank is not representative of the wider UK population (participants are typically more educated and of a higher socioeconomic status as compared with the general population).<sup>21</sup> In the analysis sample, 38% of participants had over 19 years of education, which is equivalent to a vocational

qualification or degree and only 17% of participants left school with no formal qualifications after seven years (supplementary table 3). The standard deviation of educational attainment was 3.6 years, BMI was 4.69, and systolic blood pressure was 18.68 mm Hg. For lifetime smoking behaviour, one standard deviation increase is equivalent to, for example, an individual smoking 20 cigarettes a day for 15 years and stopping 17 years ago, or an individual smoking 60 cigarettes a day for 13 years and stopping 22 years ago.<sup>23</sup>

### Effect of education on the risk of coronary heart disease and stroke

Figure 1 shows that in observational analyses, one standard deviation of higher education was associated with a 14% lower risk of coronary heart disease with an odds ratio of 0.86 (95% confidence interval 0.84 to 0.89). Two-sample mendelian randomisation analysis indicated a stronger protective effect with an odds ratio of 0.63 (0.60 to 0.67).

Figure 1 shows that similar protective associations were found for the effect of education on other cardiovascular disease outcomes. In observational analyses, a one standard deviation higher education was associated with an 11% lower risk of stroke, with an odds ratio of 0.89 (95% confidence interval 0.85 to 0.93). In two-sample mendelian randomisation analyses the protective effect was stronger, with an odds ratio of 0.71 (0.68 to 0.75).

All three approaches also provided consistent evidence for a protective effect of education with risk of cardiovascular disease and its subtypes (supplementary figures 6 and 7).

### Effect of education on BMI, systolic blood pressure, and smoking behaviour

Figure 2 and supplementary figures 8 to 11 show that in all methods, a longer time in education was associated with lower BMI, systolic blood pressure, and smoking behaviour.

### Effect of BMI, systolic blood pressure, and smoking behaviour on the risk of cardiovascular disease outcomes

Figure 3 shows that both observational and two-sample mendelian randomisation analyses consistently



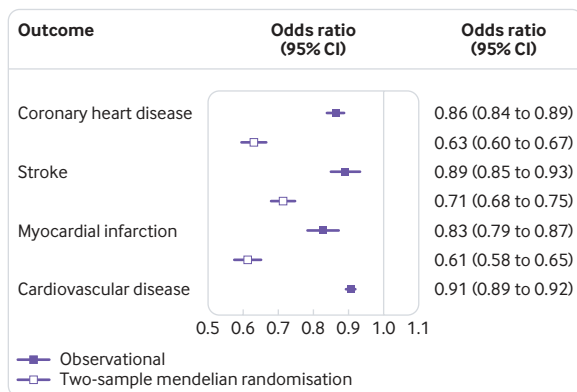


Fig 1 | The effect of a one standard deviation increase in education on the risk of cardiovascular disease outcomes

supported an increased risk of coronary heart disease with higher BMI, systolic blood pressure, and smoking behaviour, after adjusting for education. The effect of each risk factor was less consistent in the one-sample mendelian randomisation and estimates had wide confidence intervals (supplementary figure 12).

#### Mediation by BMI, systolic blood pressure, and smoking behaviour

Figure 4 shows that in the observational analysis, the proportion of the effect of education on the risk of coronary heart disease mediated by BMI was 15% (95% confidence interval 13% to 17%), 11% for systolic blood pressure (9% to 13%), and 19% for smoking behaviour (15% to 22%). In the two-sample mendelian randomisation analysis, the percentage mediated by BMI was 18% (14% to 23%), 21% for systolic blood pressure (15% to 26%), and 34% for smoking behaviour (95% confidence interval 17% to 50%). In observational analyses, combining all three risk factors explained 42% (36% to 48%) of the effect of education on the risk of coronary heart disease. In two-sample mendelian randomisation, combining all three risk factors explained 36% (5% to 68%) of the effect of education on coronary heart disease.

Similar results were found for other cardiovascular disease subtypes in multivariable observational analyses. Smoking behaviour consistently mediated

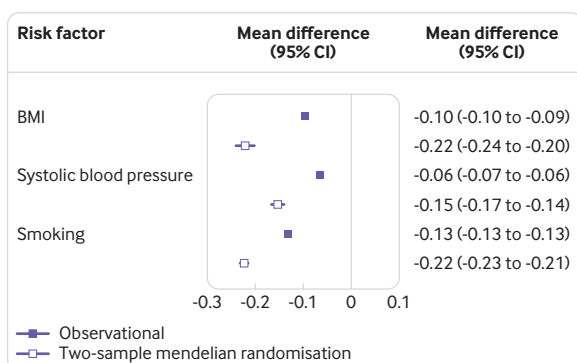


Fig 2 | Estimates for the association between one standard deviation higher education and intermediate risk factors

around 20% of the association. BMI explained between 10% and 17% of the association between education and cardiovascular disease and its subtypes; systolic blood pressure explained between 8% and 18% of the association. In two-sample mendelian randomisation analyses, smoking behaviour explained up to 34% of the association between education and cardiovascular disease subtypes; BMI estimated up to 18% and systolic blood pressure up to 28% of the association. One-sample mendelian randomisation analyses estimated similar amounts of the association explained by systolic blood pressure and smoking behaviour but was less consistent for BMI (supplementary figure 13).

#### Sensitivity analyses

Table 2 and supplementary figures 14 to 17 show that results from sensitivity analyses were comparable, but produced less precise estimates with wider confidence intervals. Unadjusted and age and sex adjusted models were also consistent with the main fully adjusted models for multivariable analyses (supplementary tables 22 and 23). Analyses stratified by age and separately by sex were consistent with the non-stratified main results, although confidence intervals were wide in the mendelian randomisation (supplementary figures 14 and 15). The effects of each mediator individually, and combined, estimated on the risk difference scale and by using the difference method in individual data were consistent with the main analyses on the log odds ratio scale (supplementary figure 16). Including diet and exercise measures in addition to BMI, systolic blood pressure, and smoking behaviour did not change the amount of the association that education to cardiovascular disease (all subtypes) explained (supplementary figure 17).

#### Discussion

Our observational and genetic analyses support that the effect of education on the risk of cardiovascular disease is mediated by approximately up to one third through any of BMI, systolic blood pressure, or smoking behaviour. When investigating all three risk factors together, it was estimated that around 40% of the association between education and cardiovascular disease is explained by the three risk factors combined, both in observational and mendelian randomisation analyses. Note that over half of the effects of education remain unexplained in these analyses. Our main analysis did not consider the contributions of exercise, diet, health system factors, lipid profile, and glycaemic traits.<sup>38-44</sup> However, these risk factors are likely to be inter-related with the risk factors already considered in our analysis. For example, much of the effect of diet and activity on cardiovascular disease is likely to act through BMI and systolic blood pressure, and therefore the cumulative effect of BMI, systolic blood pressure, and smoking behaviour together is likely to be capturing some of their effects. Indeed, in a sensitivity analysis including diet and exercise alongside BMI, systolic blood pressure, and smoking behaviour, they explain no more of the association between education

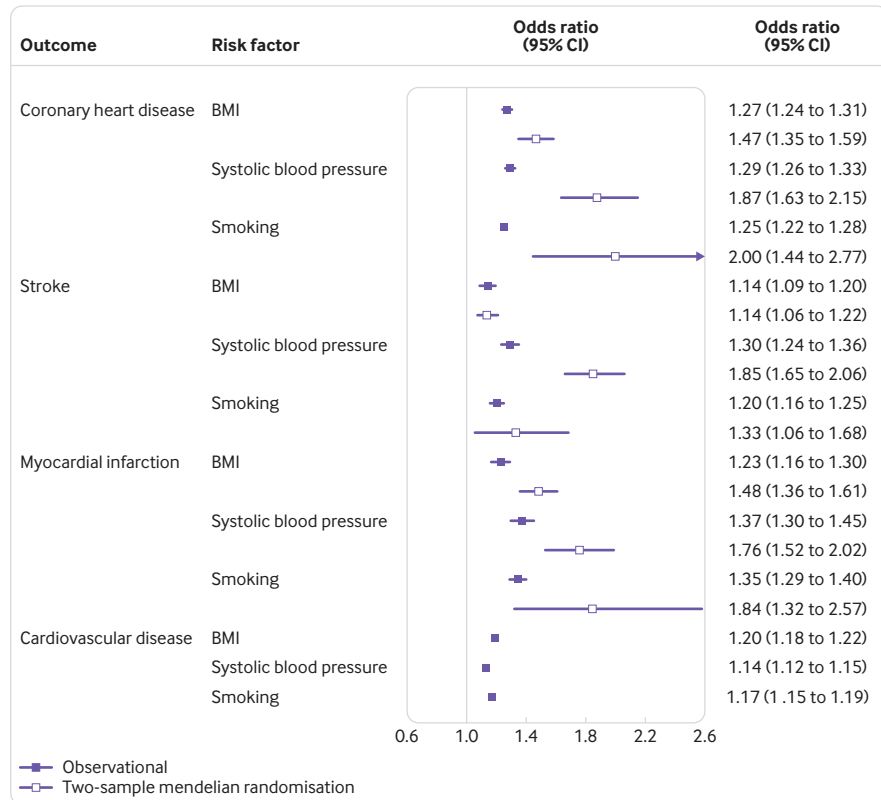


Fig 3 | Associations of a one standard deviation higher risk factor on the risk of cardiovascular disease outcomes

and cardiovascular disease than the three mediators in our main analysis.

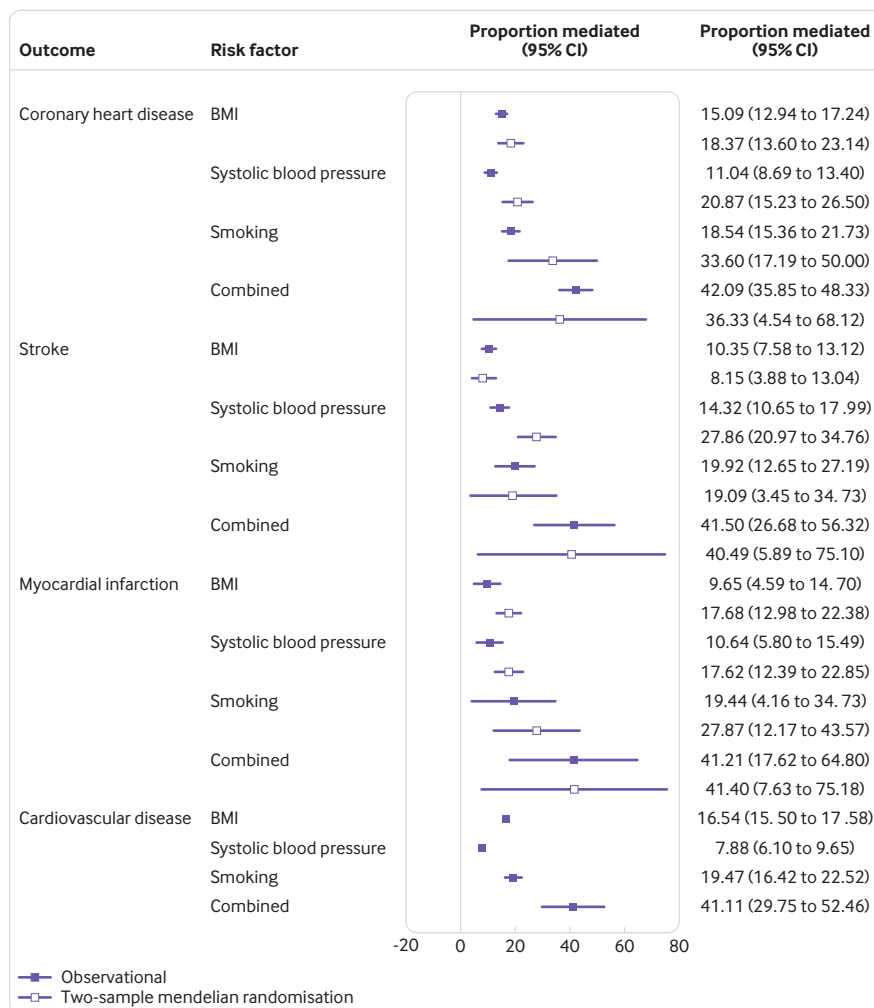
We have triangulated evidence across three distinct approaches. Although the point estimates vary, along with the mediation results, all three approaches indicate the same conclusions. Our mendelian randomisation estimates are much larger in magnitude than the observational results. In mendelian randomisation, the genetic instruments used to proxy the exposure and mediators estimate their lifetime effect, rather than one snapshot, which could explain the larger estimates in mendelian randomisation. Additionally, this could be owing to bias from negative confounding or measurement error in multivariable analyses. Cases recruited to the case-control studies included in two-sample analyses might represent a more extreme phenotype than in cohort studies such as the UK Biobank. The two-sample mendelian randomisation estimates are more precise than the one-sample mendelian randomisation results from the UK Biobank, likely related to the larger sample sizes and the number of cases.

### Findings in context

Mendelian randomisation studies have previously investigated the causal effects of education on coronary heart disease, BMI, systolic blood pressure, and smoking behaviour,<sup>2 14-16</sup> with others further estimating the effects of BMI and smoking behaviour on cardiovascular disease.<sup>18 19</sup> Our current study makes several notable advances. We have used the most

recent genome-wide association study of educational attainment to optimise the power of our two-sample mendelian randomisation analysis. With the larger sample size, the instruments selected from this study explained approximately 12% of the variance in education, as compared with the 3% accounted for in the previous iteration.<sup>12 13</sup> Similarly, by leveraging the power of the UK Biobank and recent large-scale genome-wide association study meta-analyses, we were able to study additional cardiovascular outcomes, including stroke and myocardial infarction. In addition to the overall effects of the considered risk factors on cardiovascular disease, we were able to estimate the proportion of the effect of education that they mediate by using network (or two-step) mendelian randomisation, a recently developed method.<sup>11 32</sup> Genetic instruments for smoking behaviour are limited and are typically related to binary measures that would introduce severe bias in mendelian randomisation.<sup>45</sup> The development of a genome-wide association study for the continuous measure of lifetime smoking behaviour has allowed us to include this in a mediation model.<sup>23</sup>

Several studies have used observational multivariable regression methods to support mediating roles of BMI, systolic blood pressure, and smoking behaviour in the pathway between education and the risk of cardiovascular disease,<sup>5 6 46 47</sup> with consistent results obtained by using various measures of education, including time spent in schooling and academic qualifications. In an analysis of Dutch participants, Kershaw and colleagues attributed almost 27% of the



**Fig 4 | Estimates for the effect of education on cardiovascular disease outcomes explained by the risk factors. Combined estimates refer to the effect of BMI, systolic blood pressure, and smoking behaviour considered together in one model**

association between education and coronary heart disease to smoking behaviour, with 10% and 5% attributed to obesity and hypertension respectively.<sup>46</sup> Similarly, Dégano and colleagues found 7% and 14% of the association between education and cardiovascular disease could be explained by BMI and hypertension respectively.<sup>6</sup> However, they did not find any evidence of smoking behaviour mediating the association. Veronesi and colleagues analysed their data stratified by sex, but consistently found mediating effects of systolic blood pressure and smoking behaviour in both men and women.<sup>47</sup> The findings in our study show that observational estimates underestimate the mediating role of smoking behaviour, BMI, and systolic blood pressure compared with mendelian randomisation, likely due to measurement error in the mediators that bias observational estimates towards the null, but do not affect mendelian randomisation estimates to the same degree.<sup>11</sup> Given the importance of measurement error as a source of bias in mediation analysis,<sup>8</sup> the mendelian randomisation approach offers favourable opportunities for understanding mediation.

### Strengths and limitations

The major strength of our work is that it allowed for assessment of the causal role of mediators by using mendelian randomisation, an approach that is robust to non-differential measurement error in the mediator. We have used multiple data sources and approaches, each with different potential sources of biases, to thus improve the reliability of our findings through triangulation.<sup>48</sup> Furthermore, the mediated effects estimated were consistent across the two approaches and in the statistical sensitivity analyses. The imprecision in the one-sample mendelian randomisation analysis showed the need for very large sample sizes to achieve enough statistical power when estimating mediation in a mendelian randomisation framework. The results were complemented by the two-sample mendelian randomisation approach, which had greater statistical power, but might be susceptible to alternative sources of bias, including those related to participant overlap in the samples used to obtain genetic association estimates for the exposures and outcomes.<sup>49</sup> Existing genome-wide association study

Table 2 | Mendelian randomisation (MR) sensitivity analyses

Analysis	Two-sample		One-sample	
	Estimate (95% CI)	P value	Estimate (95% CI)	P value
<b>Education-coronary heart disease</b>				
IVW	0.63 (0.60 to 0.67)	<0.001	0.51 (0.26 to 1.00)	0.05
MR-Egger	0.68 (0.54 to 0.85)	0.001	0.54 (0.12 to 2.34)	0.41
MR-Egger intercept	NA	0.37	NA	0.93
Weighted median	0.62 (0.57 to 0.67)	<0.001	0.98 (0.96 to 0.99)	0.33
<b>Education-stroke</b>				
IVW	0.71 (0.68 to 0.75)	<0.001	0.46 (0.30 to 0.71)	<0.001
MR-Egger	0.72 (0.60 to 0.87)	0.001	0.57 (0.22 to 1.47)	0.25
MR-Egger intercept	NA	0.76	NA	0.60
Weighted median	0.71 (0.66 to 0.76)	<0.001	0.99 (0.98 to 1.00)	0.002
<b>Education-myocardial infarction</b>				
IVW	0.61 (0.58 to 0.65)	<0.001	0.18 (0.08 to 0.38)	<0.001
MR-Egger	0.67 (0.52 to 0.85)	0.001	0.20 (0.04 to 1.03)	0.05
MR-Egger intercept	NA	0.39	NA	0.88
Weighted median	0.59 (0.54 to 0.65)	<0.001	0.99 (0.98 to 1.00)	0.002
<b>Education-cardiovascular disease</b>				
IVW	NA	NA	0.64 (0.51 to 0.82)	<0.001
MR-Egger	NA	NA	0.57 (0.34 to 0.95)	0.03
MR-Egger intercept	NA	NA	NA	0.59
Weighted median	NA	NA	0.96 (0.94 to 0.98)	<0.001
<b>Education-BMI</b>				
IVW	-0.22 (-0.24 to -0.20)	<0.001	-0.36 (-0.49 to -0.23)	<0.001
MR-Egger	-0.28 (-0.49 to -0.07)	0.009	-0.15 (-0.41 to 0.12)	0.29
MR-Egger intercept	NA	0.99	NA	0.08
Weighted median	-0.27 (-0.30 to -0.23)	<0.001	-0.51 (-0.62 to -0.39)	<0.001
<b>Education-systolic blood pressure</b>				
IVW	-0.15 (-0.17 to -0.14)	<0.001	-0.14 (-0.24 to -0.04)	0.005
MR-Egger	-0.13 (-0.21 to -0.05)	0.002	-0.10 (-0.30 to 0.11)	0.37
MR-Egger intercept	NA	0.32	NA	0.65
Weighted median	-0.18 (-0.21 to -0.16)	<0.001	-0.12 (-0.21 to -0.03)	0.008
<b>Education-smoking</b>				
IVW	-0.32 (-0.33 to -0.31)	<0.001	-0.37 (-0.47 to -0.28)	<0.001
MR-Egger	-0.29 (-0.36 to -0.22)	<0.001	-0.40 (-0.60 to -0.20)	<0.001
MR-Egger intercept	NA	NA	NA	0.73
Weighted median	-0.35 (-0.37 to -0.33)	<0.001	-0.37 (-0.46 to -0.29)	<0.001

IVW=inverse-variance weighted; NA=not applicable

Coronary heart disease, stroke, and myocardial infarction are in odds ratio units. BMI, systolic blood pressure, and smoking behaviour are in standard deviation units. In one-sample analyses, the weighted median was estimated on the risk difference scale and converted to odds ratio by using linear combinations.

meta-analyses of systolic blood pressure have adjusted for BMI as a covariate, which could introduce collider bias,<sup>50 51</sup> and for this reason we performed a genome-wide association study of systolic blood pressure in the UK Biobank to select instruments, without adjusting for BMI. We also applied a 'split sample' systolic blood pressure genome-wide association study approach on unrelated individuals in the UK Biobank for use in individual-level data (one-sample) mendelian randomisation to avoid overlapping populations in the genetic association estimates for the exposure and outcome,<sup>52</sup> and any associated bias.<sup>49 53</sup> To this end, the one-sample mendelian randomisation entirely avoided any population overlap when obtaining genetic estimates for the exposures and the outcomes.

For all cardiovascular disease subtypes and individual risk factors considered, the largest effects of education were consistently seen with the mendelian randomisation approaches, with smaller effects seen in the analysis of observational data. Measurement error in a mediator leads to an underestimation of the proportion mediated, so the discrepancy between our observational

and mendelian randomisation analyses could be attributable to mendelian randomisation analyses suffering less bias from measurement error.<sup>8</sup> BMI is accurately measured and has little daily variation—and correspondingly the estimates of the proportion of effect mediated by BMI in the observational and mendelian randomisation analyses are similar (15% and 18% respectively). In contrast, systolic blood pressure and lifetime smoking behaviour are difficult to measure accurately—and the estimated proportion mediated is smaller in the observational analysis than the mendelian randomisation (11% and 1% respectively for systolic blood pressure, and 19% and 34% respectively for smoking behaviour). Measurement error could also be introduced by participants over-reporting traits perceived to be 'desirable' such as education and underreporting traits perceived to be 'undesirable' such as smoking behaviour.<sup>54</sup> The estimates for all three risk factors together were more similar between observational and mendelian randomisation estimates, although for all models, the confidence intervals were wide. Note that although mendelian randomisation is



more robust to measurement error, the instruments might not necessarily be capturing all aspects of the exposure phenotype under consideration. For example, the instruments for systolic blood pressure capture average systolic blood pressure but might not necessarily reflect the variability in blood pressure.

Estimates from mendelian randomisation analyses are robust to reverse causation bias, owing to the random allocation of genetic instruments at conception (and thus before development of the outcome under consideration). Although Tyrrell and colleagues have previously used mendelian randomisation to suggest an effect of BMI on education,<sup>55</sup> and the relationship between education and BMI might be bi-directional, in our analyses where we focused on one direction of effect (that from education to BMI) the use of a large number of strong instruments for education makes it unlikely that the effect estimates derived from mendelian randomisation are biased due to reverse causation.

Another limitation of the mendelian randomisation approach is that estimates can be biased by pleiotropic pathways where the instrument is associated with the outcome through a phenotype independent of the exposure under consideration. To investigate this possibility, we additionally performed mendelian randomisation-Egger and weighted median sensitivity analyses that are more robust to such pleiotropy,<sup>34 36 56</sup> which produced results consistent with those from the main mendelian randomisation analyses. If we assume that the genetic variants have a monotonic effect on the exposure, mendelian randomisation estimates will reflect the average effect of the exposure on the outcome for all individuals whose exposure was affected by the genetic instrument. We found little evidence of heterogeneity in the effect of the exposures. This suggests that the effects of the single nucleotide polymorphisms on the exposure might be similar across the population, in which case the mendelian randomisation estimate could be a relatively unbiased estimate of the average effect in the population.

Analyses in the UK Biobank were carried out on white, British participants, potentially limiting the generalisability of our results to other populations and ethnicities. However, two-sample mendelian randomisation analyses were not exclusive to white European participants (although proportions were low for other populations) and produced consistent results to one-sample mendelian randomisation analyses. The UK Biobank is not representative of the UK population as a whole and is subject to healthy volunteer bias.<sup>56 57</sup>

When estimating the indirect effects of a mediator on a binary outcome, the product of the coefficients method results in the least amount of bias,<sup>33</sup> and as such we used this approach for estimating the effects of education through each risk factor individually. However, this method cannot currently be used to consider multiple mediators simultaneously in a mendelian randomisation model. For this reason,

we used the difference method for estimating the effect of education through the three risk factors collectively with mendelian randomisation. Although such an approach could introduce a theoretical risk of bias when investigating a binary outcome, individual level data analyses in the UK Biobank were also carried out on a linear risk difference scale to minimise any related issues. Estimates for the effect of education through the risk factors collectively were consistent between different scales in these analyses, and as such we would not expect any potential biases to alter the interpretation of our results.

### Clinical and public health implications

Past policies that increase the duration of compulsory education have improved health and such endeavours must continue.<sup>4</sup> However, intervening directly on education is difficult to achieve without social and political reforms. The findings of this study have notable implications for policymakers as they identify potential strategies for reducing education inequalities in health. Furthermore, they also produce quantitative estimates of this, allowing specific consideration of the potential impact to public health. That BMI, systolic blood pressure, and smoking behaviour together explain less than half of the overall effect of education is an important finding of this work. Further research identifying the other related factors and the interplay between them will be key to reducing social inequalities in cardiovascular disease. Furthermore, work investigating more diverse populations will be necessary to support the extrapolation of these findings outside of the considered contexts.

### Conclusion

By using distinct analytical methods, including genetic approaches that can draw causal inference, our results suggest that interventions aimed at reducing BMI, systolic blood pressure, and smoking behaviour in European populations would lead to reductions in cases of cardiovascular disease attributable to lower levels of education. Importantly, over half of the effect of education on the risk of cardiovascular disease is not mediated through these risk factors and further work is required towards investigating this.

### AUTHOR AFFILIATIONS

<sup>1</sup>MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

<sup>2</sup>Population Health Sciences, Bristol Medical School, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, UK

<sup>3</sup>Department of Biostatistics and Epidemiology, School of Public Health, Imperial College London, London, UK

<sup>4</sup>National Institute for Health Research Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol, Bristol, UK

<sup>5</sup>Institute for Global Health, University College London, London, UK

<sup>6</sup>Lausanne University Hospital, Lausanne, Switzerland

<sup>7</sup>Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland

<sup>8</sup>School of Experimental Psychology, University of Bristol, Bristol, UK

<sup>9</sup>UK Centre for Tobacco and Alcohol Studies, School of Psychological Science, University of Bristol, Bristol, UK

<sup>10</sup>Institute for Stroke and Dementia Research, University Hospital of Ludwig-Maximilians University, Munich, Germany

<sup>11</sup>Glenn Biggs Institute for Alzheimer's & Neurodegenerative Diseases, University of Texas Health Sciences Centre, San Antonio, TX, USA

<sup>12</sup>University of Texas Health Sciences Center, San Antonio, TX, USA

<sup>13</sup>Department of Neurology, Boston University School of Medicine, Boston, MA, USA

<sup>14</sup>The Framingham Heart Study, Framingham, MA, USA

<sup>15</sup>Department of Neurology and Rehabilitation Medicine, University of Cincinnati College of Medicine, Cincinnati, OH, USA

<sup>16</sup>MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

<sup>17</sup>Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

<sup>18</sup>MRC Population Health Research Unit at the University of Oxford, Oxford, UK

<sup>19</sup>Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford, Oxford, UK

<sup>20</sup>National Institute for Health Research, Oxford Biomedical Research Centre, Oxford University Hospital, Oxford, UK

<sup>21</sup>MRC-PHE Centre for Environment, School of Public Health, Imperial College London, London, UK

<sup>22</sup>Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece

**Contributors:** ARC and DG contributed equally to this project and are joint first authors. AD and LDH contributed equally to this project and are joint senior authors. ARC and DG devised the project, analysed and cleaned the data, interpreted results, wrote and revised the manuscript. ARC primarily carried out analyses using the UK Biobank. DG primarily carried out two-sample mendelian randomisation analyses. AET, NMD, TT, JV, SB, GDS, MVH, IT, LDH, and AD devised the project, interpreted the results, and revised the manuscript. TT, JV, GH, SB, and GDS contributed to the design of the project and critically revised the manuscript. REW, MRM, RM, SS, and DW provided data, and critically reviewed and revised the manuscript. All authors had full access to the data in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis. ARC and DG are the guarantors. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

**Funding:** No funding body has influenced data collection, analysis or its interpretations. This work was carried out using the computational facilities of the Advanced Computing Research Centre and the Research Data Storage Facility of the University of Bristol. This research was conducted using the UK Biobank Resource using application 10953. ARC is funded by an MRC PhD stipend at the MRC IEU, University of Bristol (MC\_UU\_00011/1). ARC, NMD, REW, MRM, GH, GDS, and LDH work in a unit that receives core funding from the MRC and University of Bristol (MC\_UU\_00011/1/MC\_UU\_00011/7). DG is funded by the Wellcome 4i Clinical PhD Programme at Imperial College London. The Economics and Social Research Council (ESRC) support NMD via a Future Research Leaders grant (ES/N000757/1). AET, REW, MRM and GDS are supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at University Hospitals Bristol NHS Foundation and the University of Bristol. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health. TT is funded by an NIHR Academic Clinical Lectureship. This work was supported by the National Heart, Lung, and Blood Institute's Framingham Heart Study (contracts N01-HC-25195 and HHSN268201500001) and by grants from the National Institute on Ageing: R01AG033193, U01AG049505, U01AG058589, U01AG52409, R01AG054076 (SS). SS was also supported by additional grants from the National Institute on Ageing (R01AG049607, R01AG033040, RFAG1059421, RF1AG061872) and the National Institute of Neurological Disorders and Stroke (R01-NS017950, NS100605). The content is solely the responsibility of the authors and does not necessarily represent the official views of the US National Institutes of Health. DW acknowledges NIH Funding (NS036695 and NS030678). MVH works in a unit that receives funding from the UK Medical Research Council and is supported by a British Heart Foundation Intermediate Clinical Research Fellowship

(FS/18/23/33512) and the National Institute for Health Research Oxford Biomedical Research Centre. LDH is funded by a Career Development Award from the UK Medical Research Council (MR/M020894/1).

**Competing interests:** All authors have completed the ICMJE uniform disclosure form at [www.icmje.org/coi\\_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf). ARC, DG, TT, JV, REW, GH, RM, SS, SB, GDS, MVH, IT, and AD declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work. MRM reports grants from Pfizer and non-financial support from GlaxoSmithKline, outside the submitted work. NMD reports grants from ESRC, grants from MRC, during the conduct of the study; grants from GRAND/Pfizer for unrelated research, outside the submitted work. AET reports grants from Pfizer, outside the submitted work. LDH reports grants from MRC, during the conduct of the study. DW reports grants from NIH, during the conduct of the study.

**Ethical approval:** Not required. Project approval was obtained from the UK Biobank (study ID: 10953) and data will be returned to them for archiving.

**Data sharing:** Statistical code is available on request from the corresponding author at [alice.carter@bristol.ac.uk](mailto:alice.carter@bristol.ac.uk). All summary statistics for two-sample mendelian randomisation are available online from each genome-wide association study consortia.

The manuscript's guarantors (ARC and DG) affirm that the 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 originally planned have been explained.

This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

- Roth GA, Johnson C, Abajobir A, et al. Global, Regional, and National Burden of Cardiovascular Diseases for 10 Causes, 1990 to 2015. *J Am Coll Cardiol* 2017;70:1-25. doi:10.1016/j.jacc.2017.04.052
- Tillmann T, Vaucher J, Okbay A, et al. Education and coronary heart disease: mendelian randomisation study. *BMJ* 2017;358:j3542. doi:10.1136/bmj.j3542
- Davies NM, Dickson M, Davey Smith G, van den Berg GJ, Windmeijer F. The causal effects of education on health outcomes in the UKUK Biobank. *Nat Hum Behav* 2018;2:117-25. doi:10.1038/s41562-017-0279-y
- Di Chiara T, Scaglione A, Corrao S, Argano C, Pinto A, Scaglione R. Association between low education and higher global cardiovascular risk. *J Clin Hypertens (Greenwich)* 2015;17:332-7. doi:10.1111/jch.12506
- Nordahl H, Rod NH, Frederiksen BL, et al. Education and risk of coronary heart disease: assessment of mediation by behavioral risk factors using the additive hazards model. *Eur J Epidemiol* 2013;28:149-57. doi:10.1007/s10654-012-9745-z
- Dégano IR, Marrugat J, Grau M, et al. The association between education and cardiovascular disease incidence is mediated by hypertension, diabetes, and body mass index. *Sci Rep* 2017;7:12370. doi:10.1038/s41598-017-10775-3
- Lynch JW, Kaplan GA, Cohen RD, Tuomilehto J, Salonen JT. Do cardiovascular risk factors explain the relation between socioeconomic status, risk of all-cause mortality, cardiovascular mortality, and acute myocardial infarction? *Am J Epidemiol* 1996;144:934-42. doi:10.1093/oxfordjournals.aje.a008863
- Blakely T, McKenzie S, Carter K. Misclassification of the mediator matters when estimating indirect effects. *J Epidemiol Community Health* 2013;67:458-66. doi:10.1136/jech-2012-201813
- Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. *Int J Epidemiol* 2013;42:1511-9. doi:10.1093/ije/dyt127
- Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;32:1-22. doi:10.1093/ije/dyg070
- Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int J Epidemiol* 2012;41:161-76. doi:10.1093/ije/dyr233
- Okbay A, Beauchamp JP, Fontana MA, et al. Lifelines Cohort Study. Genome-wide-wide association study identifies 74 loci

- associated with educational attainment. *Nature* 2016;533:539-42. doi:10.1038/nature17671
- 13 Lee JJ, Wedow R, Okbay A, et al, 23andMe Research Team, COGENT (Cognitive Genomics Consortium), Social Science Genetic Association Consortium. 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. doi:10.1038/s41588-018-0147-3
  - 14 Böckerman P, Viinikainen J, Pulkki-Råback L, et al. Does higher education protect against obesity? Evidence using Mendelian randomization. *Prev Med* 2017;101:195-8. doi:10.1016/j.ypmed.2017.06.015
  - 15 Hageaars SP, Gale CR, Deary IJ, Harris SE. *Cognitive ability and physical health: a Mendelian randomization study*. Sci Rep-Uk, 2017; 7. doi:10.1038/s41598-017-02837-3
  - 16 Gage SH, Bowden J, Davey Smith G, Munafò MR. Investigating causality in associations between education and smoking: a two-sample Mendelian randomization study. *Int J Epidemiol* 2018;47:1131-40. doi:10.1093/ije/dyy131
  - 17 Holmes MV, Lange LA, Palmer T, et al. Causal effects of body mass index on cardiometabolic traits and events: a Mendelian randomization analysis. *Am J Hum Genet* 2014;94:198-208. doi:10.1016/j.ajhg.2013.12.014
  - 18 Lyall DM, Celis-Morales C, Ward J, et al. Association of Body Mass Index With Cardiometabolic Disease in the UKUK Biobank: A Mendelian Randomization Study. *JAMA Cardiol* 2017;2:882-9. doi:10.1001/jamacardio.2016.5804
  - 19 Åsvold BO, Bjørngaard JH, Carslake D, et al. Causal associations of tobacco smoking with cardiovascular risk factors: a Mendelian randomization analysis of the HUNT Study in Norway. *Int J Epidemiol* 2014;43:1458-70. doi:10.1093/ije/dyu113
  - 20 Collins R. What makes UK Biobank special? *Lancet* 2012;379:1173-4. doi:10.1016/S0140-6736(12)60404-8
  - 21 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. doi:10.1093/aje/kwx246
  - 22 Mitchell R, Hemani, G, Dudding, T, Paternoster, L. UK Biobank Genetic Data: MRC-IEU Quality Control, Version 1. 2017. <https://data.bris.ac.uk/data/dataset/3074krb6t2frj29yh2b03x3wxj>
  - 23 Wootton RE, Richmond RC, Stuijzand BG, et al. Causal effects of lifetime smoking on risk for depression and schizophrenia: Evidence from a Mendelian randomisation study. *bioRxiv* 2018. doi:10.1101/381301
  - 24 Tobin MD, Sheehan NA, Scurrah KJ, Burton PR. Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. *Stat Med* 2005;24:2911-35. doi:10.1002/sim.2165
  - 25 Wade KH, Carslake D, Sattar N, Davey Smith G, Timpson NJ. BMI and Mortality in UKUK Biobank: Revised Estimates Using Mendelian Randomization. *Obesity (Silver Spring)* 2018;26:1796-806. doi:10.1002/oby.22313
  - 26 Yengo L, Sidorenko J, Kempner KE, et al, GIANT Consortium. Meta-analysis of genome-wide association studies for height and body mass index in 700000 individuals of European ancestry. *Hum Mol Genet* 2018;27:3641-9. doi:10.1093/hmg/ddy271
  - 27 Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 2018;7:7. doi:10.7554/eLife.34408
  - 28 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. doi:10.1038/ng.3396
  - 29 Haworth S, Mitchell R, Corbin L, et al. Apparent latent structure within the UKUK Biobank sample has implications for epidemiological analysis. *Nat Commun* 2019;10:333. doi:10.1038/s41467-018-08219-1
  - 30 Thompson JR, Minelli C, Del Greco M F. Mendelian Randomization using Public Data from Genetic Consortia. *Int J Biostat* 2016;12. doi:10.1515/ijb-2015-0074
  - 31 Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013;37:658-65. doi:10.1002/gepi.21758
  - 32 Burgess S, Daniel RM, Butterworth AS, Thompson SG, EPIC-InterAct Consortium. Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways. *Int J Epidemiol* 2015;44:484-95. doi:10.1093/ije/dyu176
  - 33 VanderWeele TJ. Mediation Analysis: A Practitioner's Guide. *Annu Rev Public Health* 2016;37:17-32. doi:10.1146/annurev-publhealth-032315-021402
  - 34 Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015;44:512-25. doi:10.1093/ije/dyv080
  - 35 Locke AE, Kahali B, Berndt SI, et al, LifeLines Cohort Study, ADIPOGen Consortium, AGEN-BMI Working Group, CARDIOGRAMplusC4D Consortium, CKDGen Consortium, GLGC, ICBP, MAGIC Investigators, MuTHER Consortium, MiGen Consortium, PAGE Consortium, ReproGen Consortium, GENIE Consortium, International Endogene Consortium. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015;518:197-206. doi:10.1038/nature14177
  - 36 Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol* 2016;40:304-14. doi:10.1002/gepi.21965
  - 37 Spiller W, Davies NM, Palmer TM. Software application profile: mrrobust—a tool for performing two-sample summary Mendelian randomization analyses. *Int J Epidemiol* 2018;1-7. doi:10.1093/ije/dyy195
  - 38 Luepker RV, Rosamond WD, Murphy R, et al. Socioeconomic status and coronary heart disease risk factor trends. The Minnesota Heart Survey. *Circulation* 1993;88:2172-9. doi:10.1161/01.CIR.88.5.2172
  - 39 Garrison RJ, Gold RS, Wilson PWF, Kannel WB. Educational attainment and coronary heart disease risk: the Framingham Offspring Study. *Prev Med* 1993;22:54-64. doi:10.1006/pmed.1993.1004
  - 40 Mayer OJr, Šimon J, Heidrich J, Cokkinos DV, De Bacquer D, EUEUROASPIRE II Study Group. Educational level and risk profile of cardiac patients in the EUEUROASPIRE II substudy. *J Epidemiol Community Health* 2004;58:47-52. doi:10.1136/jech.58.1.47
  - 41 Jacobsen BK, Thelle DS. Risk factors for coronary heart disease and level of education. The Tromsø Heart Study. *Am J Epidemiol* 1988;127:923-32. doi:10.1093/oxfordjournals.aje.a114895
  - 42 Matthews KA, Kelsey SF, Meilahn EN, Kuller LH, Wing RR. Educational attainment and behavioral and biologic risk factors for coronary heart disease in middle-aged women. *Am J Epidemiol* 1989;129:1132-44. doi:10.1093/oxfordjournals.aje.a115235
  - 43 Lynch J, Davey Smith G, Harper S, Hillemeier M. Is income inequality a determinant of population health? Part 2. U.S. National and regional trends in income inequality and age- and cause-specific mortality. *Milbank Q* 2004;82:355-400. doi:10.1111/j.0887-378X.2004.00312.x
  - 44 Loucks EB, Gilman SE, Howe CJ, et al. Education and coronary heart disease risk: potential mechanisms such as literacy, perceived constraints, and depressive symptoms. *Health Educ Behav* 2015;42:370-9. doi:10.1177/1090198114560020
  - 45 Burgess S, Labrecque JA. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. *Eur J Epidemiol* 2018;33:947-52. doi:10.1007/s10654-018-0424-6
  - 46 Kershaw KN, Droomers M, Robinson WR, Carnethon MR, Daviglius ML, Monique Verschuren WM. Quantifying the contributions of behavioral and biological risk factors to socioeconomic disparities in coronary heart disease incidence: the MORGEN study. *Eur J Epidemiol* 2013;28:807-14. doi:10.1007/s10654-013-9847-2
  - 47 Veronesi G, Ferrario MM, Kuulasmaa K, et al. Educational class inequalities in the incidence of coronary heart disease in Europe. *Heart* 2016;102:958-65. doi:10.1136/heartjnl-2015-308909
  - 48 Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol* 2016;45:1866-86.
  - 49 Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol* 2016;40:597-608. doi:10.1002/gepi.21998
  - 50 Warren HR, Evangelou E, Cabrera CP, et al. International Consortium of Blood Pressure (ICBP) 1000G Analyses, BIOS Consortium, Lifelines Cohort Study, Understanding Society Scientific group, CHD Exome+ Consortium, ExomeBP Consortium, T2D-GENES Consortium, GoT2DGenes Consortium, Cohorts for Heart and Ageing Research in Genome Epidemiology (CHARGE) BP Exome Consortium, International Genomics of Blood Pressure (iGen-BP) Consortium, UKUK Biobank CardioMetabolic Consortium BP working group. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat Genet* 2017;49:403-15. doi:10.1038/ng.3768
  - 51 Evangelou E, Warren HR, Mosen-Ansorena D, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet* 2018;50:1412-25. doi:10.1038/s41588-018-0205-x
  - 52 Taylor AE, Davies NM, Ware JJ, VanderWeele T, Davey Smith G, Munafò MR. Mendelian randomization in health research: using appropriate genetic variants and avoiding biased estimates. *Econ Hum Biol* 2014;13:99-106. doi:10.1016/j.ehb.2013.12.002
  - 53 Haycock PC, Burgess S, Wade KH, Bowden J, Relton C, Davey Smith G. Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. *Am J Clin Nutr* 2016;103:965-78. doi:10.3945/ajcn.115.118216

- 54 Pulcu E. Self-report distortions of puffing topography in daily smokers. *J Health Psychol* 2016;21:1644-54. doi:10.1177/1359105314560917
- 55 Tyrrell J, Jones SE, Beaumont R, et al. Height, body mass index, and socioeconomic status: mendelian randomisation study in UKUK Biobank. *BMJ* 2016;352:i582. doi:10.1136/bmj.i582
- 56 Gkatzionis A, Burgess S. Contextualizing selection bias in Mendelian randomization: how bad is it likely to be? *Int J Epidemiol* 2018;1-11. doi:10.1093/ije/dyy202
- 57 Hughes RA, Davies NM, Davey Smith G, Tilling K. Selection Bias When Estimating Average Treatment Effects Using One-sample Instrumental Variable Analysis. *Epidemiol* 2019;30:350-7. doi:10.1097/EDE.0000000000000972
- 58 Okbay A, Baselmans BM, De Neve JE, et al. LifeLines Cohort Study. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide-wide analyses. *Nat Genet* 2016;48:624-33. doi:10.1038/ng.3552

**Supplementary materials:** Supplementary methods, figures, and tables