bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

What explains the effect of education on cardiovascular disease? Applying Mendelian randomization to identify the consequences of education inequality

Authors and affiliations

Alice R Carter [1,2]*, Dipender Gill [3]*, Neil M Davies [1,2], Amy E Taylor [2,4], Taavi Tillmann [5], Julien Vaucher [6,7], Robyn E Wootton [1,8], Marcus R Munafò [1,8,9], Gibran Hemani [1,2], Rainer Malik [10], Sudha Seshadri [11-14], Daniel Woo [5], Stephen Burgess [16-18], George Davey Smith [1,2], Michael V Holmes [19-22], Ioanna Tzoulaki [3,23,24], Laura D Howe [1,2†] & Abbas Dehghan [3,23†]

*These authors contributed equally

[†]These authors contributed equally

Affiliations

[1]Medical Research Council Integrative Epidemiology Unit, University of Bristol, BS8 2BN, United Kingdom.

[2]Population Health Sciences, Bristol Medical School, University of Bristol, Barley House, Oakfield Grove, Bristol, BS8 2BN, United Kingdom.

[3]Department of Biostatistics and Epidemiology, School of Public Health, Imperial College London, London, UK.

[4]NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol, Bristol, UK

[5]Institute for Global Health, University College London, UK.

[6]Lausanne University Hospital, Lausanne, Switzerland

[7]Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland

[8]School of Experimental Psychology, University of Bristol, Bristol BS8 1TU, UK

[9]UK Centre for Tobacco and Alcohol Studies, School of Psychological Science, University of Bristol, BS8 1TU, United Kingdom.

[10]Institute for Stroke and Dementia Research, University Hospital of Ludwig-Maximilians University, Munich, Germany.

[11]Glenn Biggs Institute for Alzheimer's & Neurodegenerative Diseases

[12]University of Texas Health Sciences Center, San Antonio, TX 78229

[13]Boston University School of Medicine

bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[14]The Framingham Heart Study, Framingham, MA 01702

[15]Department of Neurology and Rehabilitation Medicine, University of Cincinnati College of Medicine, Cincinnati, OH 45267-0525, USA.

[16]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

[17]Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom.

[18]Medical Research Council Integrative Epidemiology Unit, University of Bristol, BS8 2BN, United Kingdom.

[19]Medical Research Council Population Health Research Unit at the University of Oxford, Oxford, UK.

[20]Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford, Oxford, UK.

[21]Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK.

[22]National Institute for Health Research, Oxford Biomedical Research Centre, Oxford University Hospital, Oxford, UK.

[23]MRC-PHE Centre for Environment, School of Public Health, Imperial College London, London, UK.

[24]Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece.

Funding

ARC is funded by an MRC PhD stipend at the MRC IEU, University of Bristol (R120971-102). 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]. 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 HHSN268201500001I) and by grants from the National Institute on Aging: R01AG033193, U01AG049505, U01AG058589, U01AG52409, R01AG054076 (SS). SS was also supported by additional grants from the National Institute on Aging (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

bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Development Award from the UK Medical Research Council (MR/M020894/1). LDH works in a unit that receives core funding from the MRC and University of Bristol.

Corresponding authors

Ms Alice Carter

Medical Research Council Integrative Epidemiology Unit, Population Health Sciences, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, United Kingdom

+44 (0) 1173310134
alice.carter@bristol.ac.uk

Dr Dipender Gill

Department of Biostatistics and Epidemiology, School of Public Health, Medical School Building, St Mary's Hospital, Norfolk Place, Imperial College London, United Kingdom, W2 1PG

+44 (0) 7904843810

dipender.gill@imperial.ac.uk

Acknowledgements

No funding body has influenced data collection, analysis or its interpretations. This publication is the work of the authors, who serve as the guarantors for the contents of this paper. This work was carried out using the computational facilities of the Advanced Computing Research Centre - http://www.bris.ac.uk/acrc/ and the Research Data Storage Facility of the University of Bristol - http://www.bris.ac.uk/acrc/storage/. This research was conducted using the UK Biobank Resource using application 10953.

Conflicts of interest

All authors declare no conflicts of interest.

Author contributions

ARC and DG devised the project, analyzed and cleaned the data, interpreted results, wrote and revised the manuscript. AET, NMD, TT, JV, SB, GDS, MVH, JT, 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. RW, MM, RM, SS and DW provided data, and critically reviewed and revised the manuscript.

Word count: 4056

bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Key Points

Question: What is the role of body mass index, systolic blood pressure and smoking in mediating the effect of education on cardiovascular disease risk?

Finding: We find consistent evidence that body mass index, systolic blood pressure and smoking mediate the effect of education, explaining up to 18%, 27% and 33% respectively. Including all three risk factors in a model together explains around 40% of the effect of education.

Meaning: Intervening on body mass index, systolic blood pressure and smoking would lead to reductions in cases of CVD attributable to lower levels of education. Over half of the effect of education on risk of cardiovascular disease is not mediated through these risk factors.

bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Importance: Lower levels of education are causally related to higher cardiovascular risk, but the extent to which this is driven by modifiable risk factors also associated with education is unknown.

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

Design: Multivariable regression analysis of observational data and Mendelian randomization (MR) analysis of genetic data.

Setting: UK Biobank and international genome-wide association study consortia.

Participants: Predominantly individuals of European ancestry.

Main outcomes and measures: The effects of education (per 1-standard deviation increase, equivalent to 3.6 years) on coronary heart disease, cardiovascular disease (all subtypes), myocardial infarction and stroke risk (all measured in odds ratio, OR), and the degree to which this is mediated through body mass index, systolic blood pressure and smoking.

Results: Each additional standard deviation of education associated with 13% lower risk of coronary heart disease (OR 0.87, 95% confidence interval [CI] 0.84 to 0.89) in observational analysis and 37% lower risk (OR 0.63, 95% CI 0.60 to 0.67) in Mendelian randomization analysis. As a proportion of the total risk reduction, body mass index mediated 15% (95% CI 13% to 17%) and 18% (95% CI 14% to 23%) in the observational and Mendelian randomization estimates, respectively. Corresponding estimates for systolic blood pressure were 11% (95% CI 9% to 13%) and 21% (95% CI 15% to 27%), and for smoking, 19% (15% to 22%) and 33% (95% CI 17% to 49%). All three risk factors combined mediated 42% (95% CI 36% to 48%) and 36% (95 % CI 16% to 63%) of the effect of education on coronary heart disease in observational and Mendelian randomization respectively. Similar results were obtained when investigating risk of stroke, myocardial infarction and all-cause cardiovascular disease.

Conclusions and relevance: BMI, SBP and smoking mediate a substantial proportion of the protective effect of education on risk of cardiovascular outcomes and intervening on these would lead to reductions in cases of CVD attributable to lower levels of education. However, more than half of the protective effect of education remains unexplained and requires further investigation.

bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Cardiovascular disease (CVD) is the leading cause of mortality worldwide, accounting for over 17 million deaths annually (1). Insights into aetiological mechanisms have improved prevention through the modification of risk factors (1, 2). Recent studies have suggested that socioeconomic risk factors such as education play a causal role in the aetiology of CVD (3-5). Education may prevent CVD, in part through its effects on modifiable risk factors (6-8), and understanding this relationship can be used to inform public health policy.

Existing studies suggest that body mass index (BMI), systolic blood pressure (SBP) and smoking behaviour at least partly explain differences in CVD risk related to educational attainment (6-8). However, these studies have relied on observational mediation analyses that may suffer from biases. Traditional methods use a single snapshot of any given risk factor, which may thus incompletely capture a person's lifetime exposure (9). For example, blood pressure measured at a single timepoint 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 (9). Furthermore, other biases such as unmeasured confounding cannot be addressed using observational methodologies (10).

Mendelian randomization (MR) uses genetic variants as instruments to estimate the effect of an exposure on an outcome of interest (11), exploiting the random allocation of genetic variants to infer causal effects that are robust to non-differential measurement error and confounding from environmental factors (11). Two-step Mendelian randomization was introduced for mediation analysis; unlike traditional observational mediation analysis approaches, it is both sensitive to the causal effects of the mediator and corrects for measurement error in the mediator (12). Recent genome-wide association study (GWAS) meta-analyses have identified a number of genetic variants for educational attainment that may be used as instrumental variables (13, 14). In this study, we investigated the role of BMI, SBP and lifetime smoking in mediating the causal effect of educational attainment on CVD risk using three complementary approaches: multivariable regression of observational data, one-sample MR and two-sample MR. BMI, SBP and smoking were selected as intermediate risk factors based on previous literature implicating them as both being affected by education and as risk factors for CVD, with suitable data available across all three complementary methods used.

bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

Overall study design

This study used complementary approaches to investigate whether lower BMI, SBP and lifetime smoking explain the protective effect of education on risk of cardiovascular disease. Multivariable regression of observational data, one-sample MR of individual level genetic data and two-sample MR of summary level genetic data investigating coronary heart disease (CHD) myocardial infarction (MI), stroke risk and CVD (all subtypes) were carried out.

Data sources

UK Biobank

The UK Biobank recruited 503,317 UK adults between 2006 and 2010. Participants attended baseline assessment centres involving questionnaires, interviews, anthropometric, physical and genetic measurements (15, 16). In the observational analysis, we included 217,013 White British individuals, with complete data on genotypes, age, sex, educational attainment, cardiovascular outcomes, BMI, smoking status, blood pressure, socioeconomic status (as measured by Townsend Deprivation Index at birth [TDI]) and place of birth. Exclusion criteria and flowchart are available in the Supplementary Methods (Supplementary Figures 1 and 2). Individuals of White British descent were defined using both self-reported questionnaire data with similar genetic ancestry based on the genetic principal components (PC) (17).

Data from the baseline assessment centre on highest qualifications completed and smoking, clinic measurements of BMI and SBP, and all covariate measures were used for the analyses (see Supplementary Methods). A continuous lifetime measure of smoking, incorporating smoking initiation, duration, heaviness and cessation is used in this analysis. Details of the derivation of this and applications have been detailed elsewhere (18). Participants reported their highest qualification and age of leaving school if they did not have a degree at the baseline assessment centre. These were converted to the International Standard Classification for Education (ISCED) coding of educational attainment (Supplementary Table 1) (14). Available follow-up data were used where baseline data were missing. To account for the effects of anti-hypertensive treatment, participants who were taking antihypertensive medication had 10mmHg added to their measured SBP (19). CVD diagnoses (including diagnoses of stroke, MI and CHD) and events were ascertained through linkage mortality data and hospital episode statistics (HES), with cases defined according to ICD-9 and ICD-10 codes (Supplementary Table 2) (20). Individuals who had experienced a CVD event prior to the baseline assessment (prevalent cases) were excluded and only first event, incident cases following

bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the assessment centre were considered. Date of diagnoses are provided by HES data, which was linked with the date of assessment centre provided by UK Biobank.

GWAS meta-analyses

In the two-sample MR analysis, we obtained summary genetic associations from GWAS data for each respective phenotype. For education, this was the Social Science Genetic Association Consortium (SSGAC) GWAS meta-analysis of years of schooling in 1,131,881 individuals of European ancestry (14), with summary data made available for 766,345 of these participants. Instruments were selected as the 1,271 independent (pairwise r^2<0.1) genome-wide significant SNPs after analysing data from the full sample (14). Education was assessed at or above the age of 30 years, with comparability between studies heterogeneous in their educational systems maximised by mapping major educational qualifications on to one of seven categories of the ISCED (13, 14). We obtained genetic estimates for BMI from the Genetic Investigation of ANthropometric Traits (GIANT) consortium's 2018 GWAS meta-analysis of 681,275 individuals of European decent (21). Genetic association estimates for SBP and smoking were estimated from a GWAS of 318,417 White British individuals in UK Biobank, protocol details for which are available in the Supplementary Methods. Instruments for BMI, SBP and smoking were identified as the lead SNPs in loci reaching genome-wide significance after clumping summary estimates from the largest available GWAS for linkage disequilibrium (LD) r^2<0.001 using a 1000 genomes European reference panel through the TwoSampleMR package of the statistical software R (22). For risk of CHD, we used publicly available genetic association estimates from the CARDIoGRAMplusC4D 1000 Genomes-based GWAS meta-analysis of 60,801 cases and 123,504 controls (23). Participants were of European, east Asian, south Asian, Hispanic and African American ancestry, and adjustment was made for population stratification using the genomic control method (23). The definition for CHD 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 studies used to obtain genetic association estimates for MI and stroke are provided in the Supplementary Methods. All genetic association estimates used in each two-sample MR analyses are provided in Supplementary Tables 7-21.

Statistical Analysis

Effect of education on cardiovascular disease

In the observational analysis of UK Biobank data, multivariable logistic regression was used to estimate the association of education with CVD and its subtypes. Analyses were adjusted for potential confounders; age, sex, place of birth, birth distance from London, and TDI. Full details of

bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

derived variables (birth distance from London and TDI at birth) are available in the supplementary methods.

In the two-sample MR analysis, the effects of education on CVD subtypes were investigated using ratio method MR with standard errors derived using the delta method (24). Fixed-effect inverse-variance weighted (IVW) meta-analysis was used to pool MR estimates across individual SNPs and calculate an overall estimate (25).

Mediation by body mass index, systolic blood pressure and smoking

When investigating the degree to which the effects of education on CVD and its subtypes are mediated through each risk factor (BMI, SBP and smoking) individually, the product of coefficients method was used to estimate the indirect effect (26). 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, to estimate the mediated (indirect) effect (i.e. the effect of education on CVD that goes through the risk factor). We estimated the proportion of the overall effect of education on CVD 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 MR analysis and using the delta method in the two-sample MR analysis.

When investigating the role of all three risk factors together on the association between education and CVD, we used the difference method (27). This involved estimating the total effect of education on each CVD subtype, as described above. We estimated the direct effect of education on each CVD subtype controlling for all three risk factors together, using either multivariable regression or multivariable MR, in observational and MR analyses respectively. The direct effect was then taken away from the total effect to find the indirect effect, which was divided by the total effect to give an estimate of the amount of mediation.

In the observational analysis, multivariable linear regression was used to estimate the association of education with each risk factor after adjusting for confounders (as in the total effects models). The effect of each risk factor on the individual CVD subtypes was then estimated with additional adjustment for self-reported educational attainment (27). The two estimates were multiplied together to estimate the indirect effect (of education, through the risk factor). Considering all three risk factors together, a multivariable logistic model for the effect of education on CVD (and subtypes) adjusting for all three risk factors was used to estimate the direct effect of education independently of the risk factors. This was taken away from the total effect to estimate the indirect effect of education through the three risk factors collectively.

bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

For the two-sample MR, the IVW MR approach was used to estimate the effect of education on each risk factor and regression-based multivariable MR was used to estimate the effect of each risk factor on risk of the considered CVD subtypes, adjusting for genetic effect of the instruments on education (28). The indirect effect of education on risk of each CVD subtype through the considered risk factor was estimated by multiplying results from these two MR analyses. To estimate the total effect of education mediated through all three risk factors collectively, the direct effect of education after adjusting for the three risk factors together was estimated using MVMR and taken away from the total effect of education previously estimated using IVW MR; this indirect effect was divided by the total effect to estimate the proportion mediated.

One-sample MR and sensitivity analyses

One sample MR analysis was carried out in UK Biobank. Analyses using UK Biobank were replicated using the risk difference scale. A range of sensitivity analyses were carried out, including to explore the assumption of no pleiotropy within MR analyses. Full details are provided in the Supplementary Methods.

Statistical software and ethical approval

Analysis was performed using Stata version 14 (StataCorp LP) and R version 3.4.3 (The R Foundation for Statistical Computing). The mrrobust package for Stata and the TwoSample MR package for R were used to facilitate MR analyses (22, 29). The code used for all analyses is available on request and details of genetic variants used are provided in the Supplementary Tables. Ethical approval was not sought for publicly available data because all participating studies had already obtained relevant authorisation. Project approval was obtained from UK Biobank (study ID: 10953).

bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

UK Biobank Cohort Description

The UK Biobank sample used in the observational and one-sample MR analysis was comparable to the participants in UK Biobank as a whole. In the analysis sample 38% of individuals had over 19 years of education, equivalent to a vocational qualification or degree. Comparatively, only 17% of individuals left school with no formal qualifications after seven years. Full details are available in Supplementary Table 3. The standard deviations (SD) of educational attainment was 3.6 years, BMI was 4.69 kg/m2 and SBP was 18.68 mm Hg. For lifetime smoking, 1-SD 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. For further explanation of this, see Wootton and colleagues (18).

Effect of education on risk of CHD and stroke

In observational analyses, 1-SD higher education was associated with a 13% lower risk of CHD with an odds ratio (OR) of 0.87 (95% CI 0.84 to 0.89). The two-sample MR analysis indicated a stronger protective effect, with OR 0.63 (95% CI 0.60-0.67) (Figure 1).

Similar protective associations were found for the effect of education on other CVD subtypes (Figure 1). In observational analyses, a 1-SD higher education was associated with an 11% lower risk of stroke, with an OR of 0.89 (95% CI 0.85, 0.94). In two-sample MR analyses the protective effect was stronger with OR 0.71 (0.68, 0.75).

One sample-MR analyses also provided consistent evidence for a protective effect of a 1-SD increase in education with CVD risk and its subtypes (Supplementary Figure 3).

Effect of education on BMI, SBP and smoking

In all methods, a longer time in education was associated with lower ORs of BMI, SBP and smoking (Figure 2 and Supplementary Figure 4). Both multivariable observational analyses and two-sample MR found the largest association of education with smoking, where each additional 1-SD of education associated with 0.13 SD-units (95% CI 0.13 to 0.13) lower lifetime smoking in observational multivariable analyses, and 0.32 (95% CI 0.31 to 0.33) in two-sample MR. For each additional 1-SD of education, BMI was lower by 0.10 SD-units (95% CI 0.10 to 0.09) in observational analyses and 0.22 SD-units (95% CI 0.20 to 0.24) in two-sample MR. The smallest association was seen for SBP, where 1-SD higher education was associated with 0.06 SD-units (95% CI 0.06 to 0.07)

bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

lower SBP in multivariable observational analyses, and 0.15 SD-units (95% CI 0.14 to 0.16) in two-sample MR analyses (Figure 2).

Effect of BMI, SBP and smoking on risk of CVD subtypes

Both observational and two-sample MR analyses consistently found evidence to support an increased risk of CHD with higher BMI, SBP and smoking (Figure 3). The strongest risk factor for CHD was SBP, where a 1-SD increase in SBP increased the risk of CHD with an OR of 1.28 (95% CI 1.25 to 1.32) in multivariable observational analyses and 1.90 (95% CI 1.64 to 2.19) in two-sample MR analyses. A 1-SD higher BMI was associated with an OR of 1.26 (95% CI 1.22 to 1.30) for CHD in observational analyses and 1.47 (95% CI 1.35 to 1.59) in two-sample MR. A 1-SD increase in lifetime smoking was associated with a CHD OR of 1.23 (95% CI 1.20 to 1.26) in multivariable observational analyses and 1.61 (95% CI 1.28 to 2.02) in two-sample MR analyses.

For all other outcomes, multivariable observational and two-sample MR analyses consistently found evidence to support an increased risk of CVD, stroke and MI with higher BMI, SBP and smoking (Supplementary Figure 6). The strongest risk factor for stroke was also SBP, with a 1-SD increase associated with OR 1.36 (95% CI 1.29, 1.44) in multivariable observational analyses and 1.87 (95% CI 1.67, 2.09) in two-sample MR analyses.

As expected, the effect of each risk factor was less consistent in the one-sample MR and estimates had wide confidence intervals (Supplementary Figure 5).

Mediation by BMI, SBP and smoking

In the observational analysis, the proportion of the effect of education on CHD risk mediated by BMI was 15% (95% CI 13% to 17%), 11% for SBP (95% CI 9% to 13%) and 19% for smoking (95% CI 15% to 22%) (Figure 4). In the two-sample MR analysis, the percentage mediated by BMI was 18% (95% CI 14% to 23%), 21% by SBP (95% CI 15% to 27%) and 33% by smoking (95% CI 17% to 49%) (Figure 4, Supplementary Table 22). In observational analyses, combining all three risk factors together explained 42% (95% CI 36% to 48%) of the effect of education on risk of CHD (Figure 4). In two-sample MR the combined effect of all three risk factors on CHD as 36% (95% CI 5% to 68%).

Similar results were found for other CVD subtypes in multivariable observational analyses. Smoking consistently mediated around 20% of the association. BMI explained between 10% and 17% of the association between education and CVD and its subtypes, whilst SBP explained between 8% and 18%. In two-sample MR analyses, smoking explained up to 28% of the association between education and CVD subtypes, whilst BMI estimated up to 18% and SBP up to 28% of the association.

bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

One-sample MR analyses estimated similar amounts of the association explained by SBP and smoking but was less consistent for BMI (Supplementary figure 7).

Sensitivity analyses

Results from sensitivity analyses were comparable but produced less precise estimates with wider confidence intervals (Supplementary Figure 2-8, Supplementary Tables 22 and 23). Unadjusted and age and sex adjusted models were also consistent with the main fully adjusted models for multivariable analyses (Supplementary Tables 24 and 25). The effects of each mediator individually, and combined, estimated on the risk difference scale and using the difference method in individual data were consistent with main analyses on the log OR scale (Supplementary Figure 8).

bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

Our observational and genetic analyses provide complementary evidence that the effect of education on risk of CVD is mediated by up to one third of the effect through any of BMI, SBP or smoking. When investigating all three risk factors together, it was estimated that around 40% of the association between education and CVD is explained by the three risk factors, both in observational and MR analyses. However, it is important that over half of the effects of education remain unexplained in these analyses. This portion of the pathway may be due to other lifestyle behaviours associated with lower levels of education, such as access to healthcare, health-seeking behaviours and general lifestyle (30-36). Alternatively, biological factors, including lipid and glycaemic traits, may also play a part.

Findings in context

A number of studies have used observational multivariable regression methods to support mediating roles of BMI, SBP and smoking in the pathway between education and CVD risk (7, 8, 37, 38), with consistent results obtained using various measures of education, including time spent in schooling and academic qualifications. In an analysis of Dutch individuals, Kershaw *et al*, attributed almost 27% of the association between education and CHD to be due to smoking, with 10% and 5% attributed to obesity and hypertension respectively (37). Similarly, Dégano *et al* found 7% and 14% of the association between education and CVD could be explained by BMI and hypertension respectively (8). However, they did not find any evidence of smoking mediating the association. Veronesi *et al* analysed their data stratified by sex, but consistently found mediating effects of SBP and smoking in both males and females (38). The findings in our study show that observational estimates underestimate the mediating role of smoking, BMI and SBP compared with MR, likely due to measurement error in the mediators that bias observational estimates towards the null, but do not affect MR estimates to the same degree (12). Given the importance of measurement error as a source of bias in mediation analysis (9), the MR approach offers favorable opportunities for understanding mediation.

Strengths and limitations

The major strength of our work is that it allowed for us to assess the causal role of mediators using MR, 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 (39). Furthermore, the mediated effects estimated were consistent across the two approaches and in statistical sensitivity analyses. The imprecision in the one-sample MR analysis demonstrated the need for very large sample sizes to

bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

achieve sufficient statistical power when estimating mediation in an MR framework. The results were complemented by the two-sample MR approach, which had greater statistical power, but may 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 (40). Existing SBP GWAS meta-analyses have adjusted for BMI as a covariate, which could introduce collider bias (41, 42), and for this reason we performed a GWAS of SBP in UK Biobank to select instruments, without adjusting for BMI. We also applied a 'split sample' SBP GWAS approach for use in individual-level data (one-sample) MR to avoid overlapping populations in the genetic association estimates for the exposure and outcome (43), and any associated bias (40, 44).

For all CVD subtypes and individual risk factors considered, the largest effects of education were consistently seen with the MR approaches, with smaller effects seen in the analysis of observational data. Furthermore, MR is not affected by measurement error in the same way as observational analysis, and particular discrepancies to the observational analysis for the proportion of the effect of educational attainment mediated by each the risk factors. For example, BMI is precisely measured and has little daily variation – and correspondingly the estimates of the proportion of effect mediated by BMI in the observational and MR analyses are similar (15% and 18% respectively). In contrast, SBP and lifetime smoking are difficult to measure precisely – and the estimated proportion mediated is smaller in the observational analysis than the MR (11% vs. 21% for SBP, and 19% vs. 33% for smoking). Non-differential measurement error in a mediator leads to an underestimation of the proportion mediated, so this discrepancy between our observational and MR analyses may be attributable to genetic variants instrumenting a lifetime exposure and therefore MR analyses suffering less from measurement error (9). The estimates for all three risk factors together were more similar between observational and MR estimates, although for all models, the confidence intervals were wide.

A limitation of the UK Biobank is that it is not representative of the UK population and is subject to healthy volunteer bias (16). Selection bias can induce biases in the results of instrumental variable analyses (45). However, this analysis shows that the estimates from UK Biobank are consistent with those from large-scale GWAS consortia that meta-analyse across multiple studies. This suggests that if selection bias is present, it is unlikely to invalidate our conclusions (46). Furthermore, data on circulating lipid levels and glycaemic traits were limited in this population, and we were therefore unable to investigate whether the effect of education on cardiovascular risk was also mediated through these pathways.

bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A limitation of the methods used for mediation analysis is that different methods have been used when considering the risk factors individually and jointly.

When estimating the indirect effects of a mediator on a binary outcome, the product of coefficients method results in the least amount of bias (27), 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 an MR model. For this reason, we used the difference method for estimating the effect of education through the three considered risk factors collectively with MR. Although such an approach may introduce theoretical risk of bias when investigating a binary outcome, individual level data analyses in 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

Intervening on education is difficult to achieve without social and political reform. Targeting BMI, SBP and smoking could therefore help to reduce inequalities in CVD risk related to educational attainment. However, these three risk factors together explain less than half of the overall effect of education. Further research identifying the other related factors and the interplay between them will be key to reducing social injustice.

Conclusion

Using distinct analytical methodologies, including genetic approaches that are able draw causal inference, our results suggest that interventions aimed at reducing BMI, SBP and smoking in the population would lead to reductions in cases of CVD attributable to lower levels of education. Importantly, over half of the effect of education on risk of cardiovascular disease is not mediated through these risk factors and further work is required towards investigating this.

bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Roth GA, Johnson C, Abajobir A, Abd-Allah F, Abera SF, Abyu G, et al. Global, Regional, and National Burden of Cardiovascular Diseases for 10 Causes, 1990 to 2015. J Am Coll Cardiol. 2017;70(1):1-25.
2. Holland C, Cooper Y, Shaw R, Pattison H, Cooke R. Effectiveness and uptake of screening programmes for coronary heart disease and diabetes: a realist review of design components used in interventions. BMJ Open. 2013;3(11):e003428.
3. Tillmann T, Vaucher J, Okbay A, Pikhart H, Peasey A, Kubinova R, et al. Education and coronary heart disease: mendelian randomisation study. BMJ. 2017;358:j3542.
4. Davies NM, Dickson M, Davey Smith G, van den Berg GJ, Windmeijer F. The causal effects of education on health outcomes in the UK Biobank. Nature Human Behaviour. 2018;2(2):117-25.
5. 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(5):332-7.
6. 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? American Journal of Epidemiology. 1996;144(10):934-42.
7. Nordahl H, Rod NH, Frederiksen BL, Andersen I, Lange T, Diderichsen F, 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(2):149-57.
8. Degano IR, Marrugat J, Grau M, Salvador-Gonzalez B, Ramos R, Zamora A, et al. The association between education and cardiovascular disease incidence is mediated by hypertension, diabetes, and body mass index. Sci Rep. 2017;7(1):12370.
9. Blakely T, McKenzie S, Carter K. Misclassification of the mediator matters when estimating indirect effects. J Epidemiol Community Health. 2013;67(5):458-66.
10. Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. Int J Epidemiol. 2013;42(5):1511-9.
11. Smith GD, Ebrahim S. Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32(1):1-22.
12. 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(1):161-76.
13. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature. 2016;533(7604):539-42.
14. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet. 2018.
15. Collins R. What makes UK Biobank special? Lancet. 2012;379(9822):1173-4.
16. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. Am J Epidemiol. 2017;186(9):1026-34.
17. Mitchell R, Hemani, G, Dudding, T, Paternoster, L. UK Biobank Genetic Data: MRC-IEU Quality Control, Version 1. 2017.
18. Wootton RE, Richmond RC, Stuijfzand BG, Lawn RB, Sallis HM, Taylor GMJ, et al. Causal effects of lifetime smoking on risk for depression and schizophrenia: Evidence from a Mendelian randomisation study. bioRxiv. 2018.
19. 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(19):2911-35.

bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

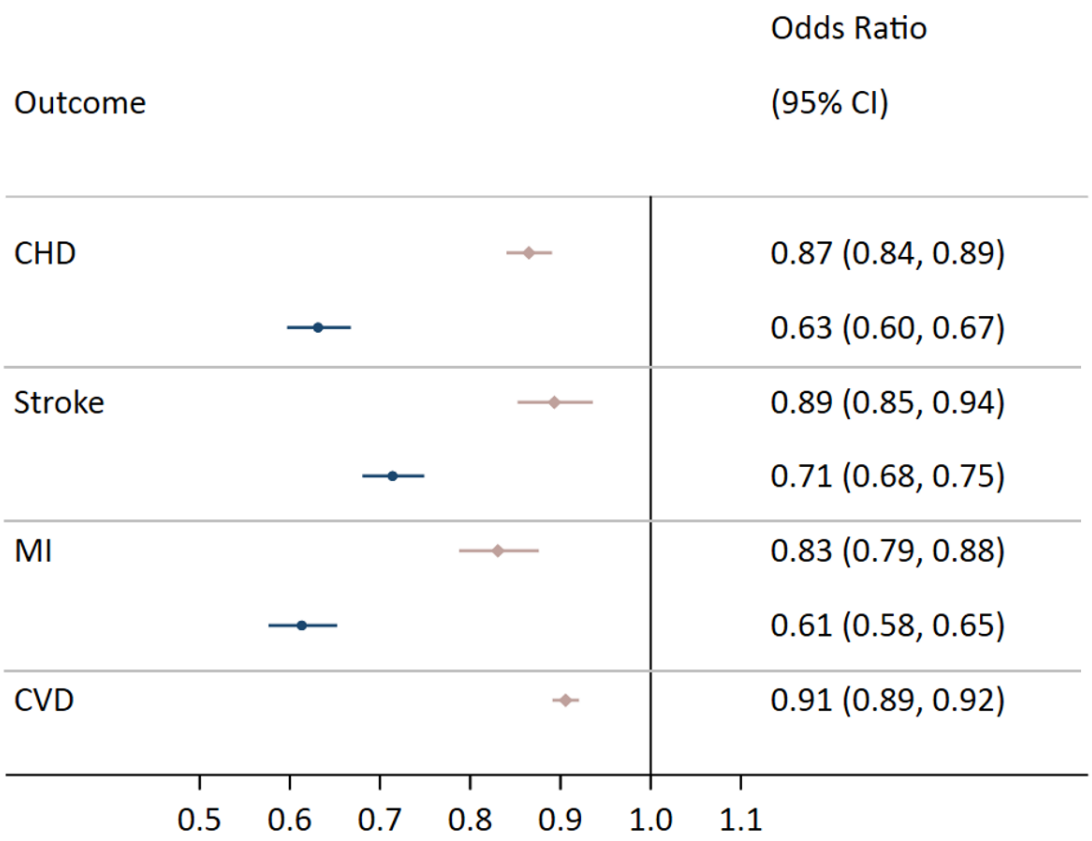
20. Wade KH, Carslake D, Sattar N, Davey Smith G, Timpson NJ. BMI and Mortality in UK Biobank: Revised Estimates Using Mendelian Randomization. Obesity (Silver Spring). 2018;26(11):1796-806.
21. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. Hum Mol Genet. 2018;27(20):3641-9.
22. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7.
23. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015;47(10):1121-30.
24. Thompson JR, Minelli C, Del Greco MF. Mendelian Randomization using Public Data from Genetic Consortia. Int J Biostat. 2016;12(2).
25. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37(7):658-65.
26. Burgess S, Daniel RM, Butterworth AS, Thompson SG, Consortium E-IA. Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways. Int J Epidemiol. 2015;44(2):484-95.
27. VanderWeele TJ. Mediation Analysis: A Practitioner's Guide. Annu Rev Public Health. 2016;37:17-32.
28. 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(2):512-25.
29. Spiller W, Davies NM, Palmer TM. Software application profile: mrrobust—a tool for performing two-sample summary Mendelian randomization analyses. International Journal of Epidemiology. 2018:dyy195-dyy.
30. Luepker RV, Rosamond WD, Murphy R, Sprafka JM, Folsom AR, McGovern PG, et al. Socioeconomic status and coronary heart disease risk factor trends. The Minnesota Heart Survey. Circulation. 1993;88(5):2172.
31. Garrison RJ, Gold RS, Wilson PWF, Kannel WB. Educational Attainment and Coronary Heart Disease Risk: The Framingham Offspring Study. Preventive Medicine. 1993;22(1):54-64.
32. Mayer O, Šimon J, Heidrich J, Cokkinos DV, De Bacquer D. Educational level and risk profile of cardiac patients in the EUROASPIRE II substudy. J Epidemiol Community Health. 2004;58(1):47-52.
33. Jacobsen BK, Thelle DS. Risk factors for coronary heart disease and level of education. The Tromso Heart Study. American Journal of Epidemiology. 1988;127(5):923-32.
34. 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(6):1132-44.
35. Lynch J, Smith GD, Harper S, Hillemeier M. Is income inequality a determinant of population health? Part 2. US national and regional trends in Income inequality and age-and cause-specific mortality. Milbank Quarterly. 2004;82(2):355-400.
36. Loucks EB, Gilman SE, Howe CJ, Kawachi I, Kubzansky LD, Rudd RE, et al. Education and Coronary Heart Disease Risk Potential Mechanisms Such as Literacy, Perceived Constraints, and Depressive Symptoms. Health Education & Behavior. 2015;42(3):370-9.
37. Kershaw KN, Droomers M, Robinson WR, Carnethon MR, Daviglus 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(10):807-14.
38. Veronesi G, Ferrario MM, Kuulasmaa K, Bobak M, Chambless LE, Salomaa V, et al. Educational class inequalities in the incidence of coronary heart disease in Europe. Heart. 2016;102(12):958-65.

bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

39. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. Int J Epidemiol. 2016;45(6):1866-86.
40. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. Genet Epidemiol. 2016;40(7):597-608.
41. Warren HR, Evangelou E, Cabrera CP, Gao H, Ren M, Mifsud B, et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. Nat Genet. 2017;49(3):403-15.
42. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nat Genet. 2018.
43. Taylor AE, Davies NM, Ware JJ, VanderWeele T, Smith GD, Munafo MR. Mendelian randomization in health research: using appropriate genetic variants and avoiding biased estimates. Econ Hum Biol. 2014;13:99-106.
44. 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(4):965-78.
45. Hughes RA, Davies NM, Davey Smith G, Tilling K. Selection bias in instrumental variable analyses. bioRxiv. 2017.
46. Gkatzionis A, Burgess S. Contextualizing selection bias in Mendelian randomization: how bad is it likely to be? arXiv. 2018.

bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1: The effect of a 1-SD increase in education on the risk of cardiovascular disease and its subtypes. Observational multivariable estimates are plotted in pink and two-sample MR estimates

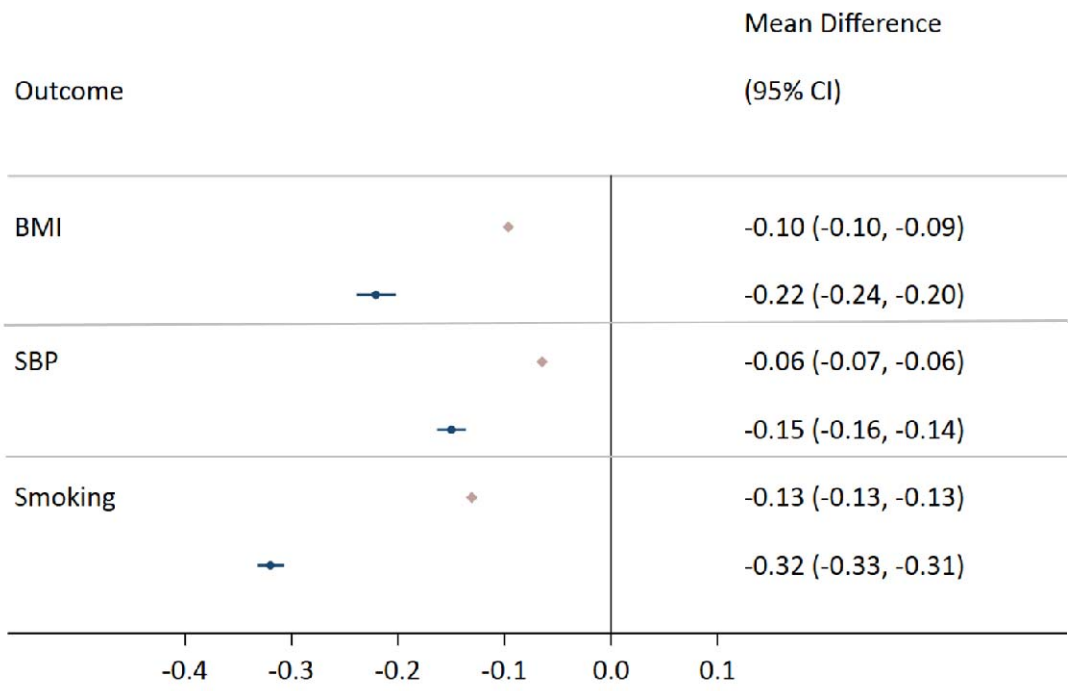


plotted in navy.

Multivariable analyses adjusted for: age, sex, place of birth and Townsend deprivation index at birth. BMI, SBP and smoking were measured in 1-SD units. CVD (All subtypes) was not available for analysis in two-sample MR analysis.

bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

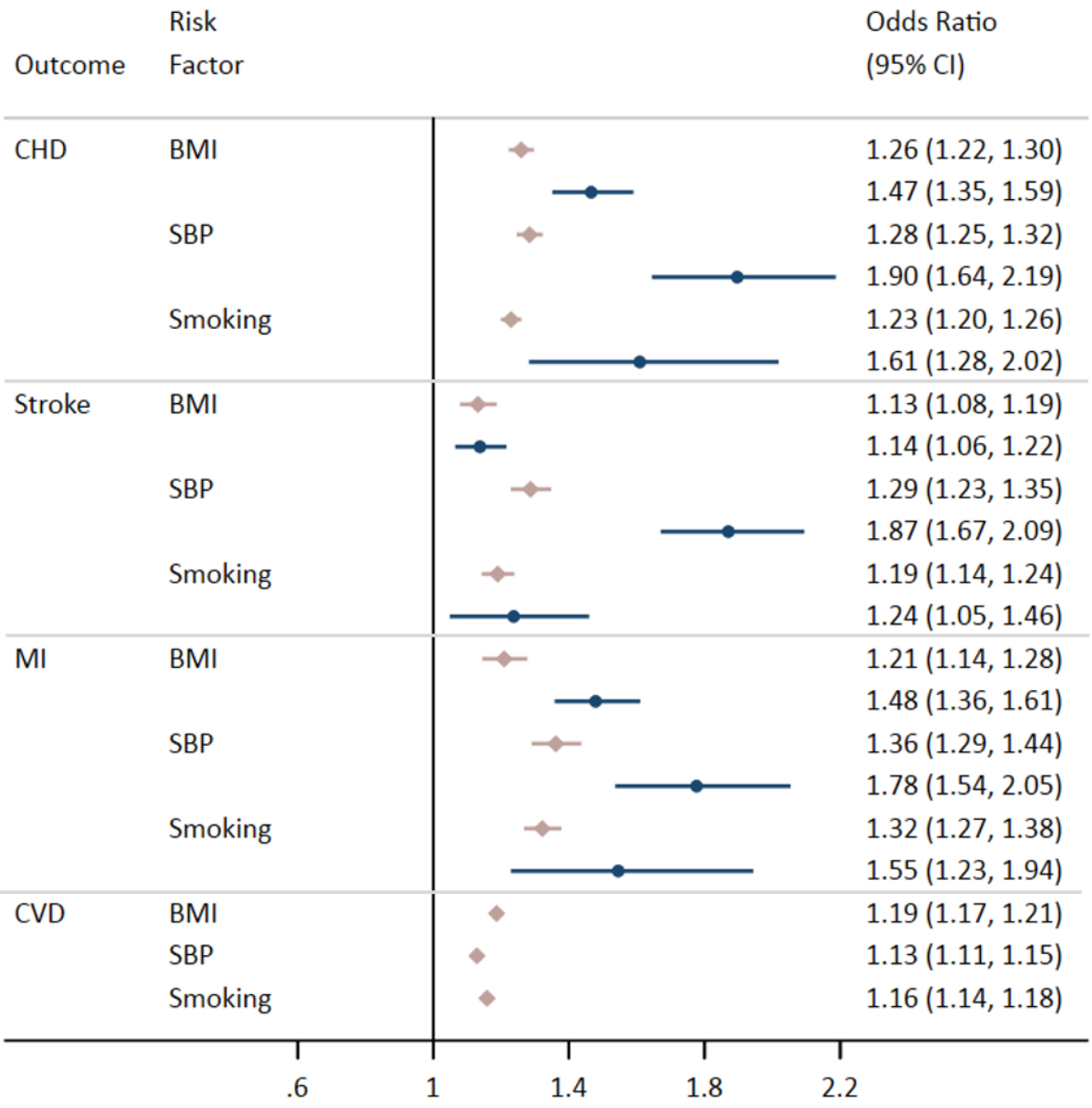
Figure 2: Observational and two-sample MR estimates for the association between 1-SD higher education and BMI, SBP and lifetime smoking respectively. All outcomes are in 1-SD units. Observational multivariable results are plotted in pink, with two sample MR estimates plotted in



navy.

bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3: Observational and two-sample MR associations of 1-SD higher BMI, SBP and lifetime smoking on the risk of cardiovascular disease and its subtypes. Observational multivariable results are plotted in pink, with two sample MR estimates plotted in navy.



bioRxiv preprint doi: https://doi.org/10.1101/488254; this version posted December 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

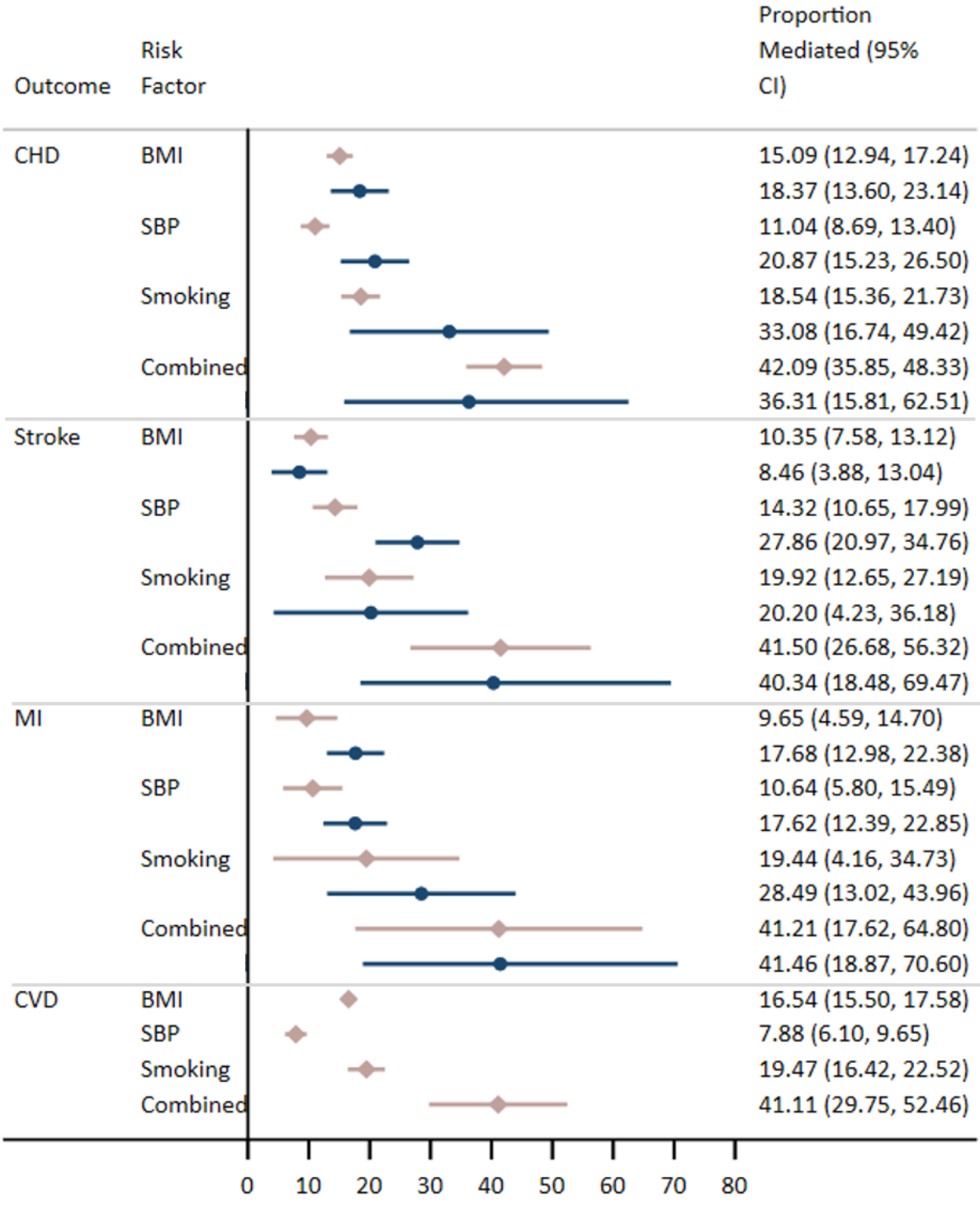



Figure 4: Estimates for the effect of education on CVD and its subtypes explained by BMI, SBP and smoking respectively. Results are provided for the multivariable observational analysis (plotted in pink) and two-sample MR (plotted in navy).