International Journal of Epidemiology, 2018, 1131–1140
doi: 10.1093/ije/dyy131
Advance Access Publication Date: 28 June 2018
Original article

Tobacco

Investigating causality in associations between education and smoking: a two-sample Mendelian randomization study

Suzanne H Gage,[1]* Jack Bowden,[2] George Davey Smith[2] and Marcus R Munafò[2,3]

[1]Department of Psychological Sciences, University of Liverpool, Liverpool, UK, [2]MRC Integrative Epidemiology Unit and [3]UK Centre for Tobacco and Alcohol Studies, University of Bristol, Bristol, UK

*Corresponding author. Department of Psychological Sciences, University of Liverpool, Eleanor Rathbone Building, Bedford Street South, Liverpool L69 7ZA, UK. E-mail: s.gage@liverpool.ac.uk

Editorial decision 22 May 2018; Accepted 6 June 2018

Abstract

Background: Lower educational attainment is associated with increased rates of smoking, but ascertaining causality is challenging. We used two-sample Mendelian randomization (MR) analyses of summary statistics to examine whether educational attainment is causally related to smoking.

Methods and Findings: We used summary statistics from genome-wide association studies (GWAS) of educational attainment and a range of smoking phenotypes (smoking initiation, cigarettes per day, cotinine levels and smoking cessation). Of 74 single nucleotide polymorphisms (SNPs) that predict educational attainment, 57 (or their highly correlated proxies) were present in the smoking initiation, cigarettes per day and smoking cessation GWAS, and 72 in the cotinine GWAS. Various complementary MR techniques (inverse variance weighted regression, MR Egger, weighted median regression) were used to test the robustness of our results. We found broadly consistent evidence across these techniques that higher educational attainment leads to reduced likelihood of smoking initiation, reduced heaviness of smoking among smokers (as measured via self-report [e.g. inverse variance weighted beta −2.25, 95% confidence interval (CI) −3.81, −0.70, $P = 0.005$] and cotinine levels [e.g. inverse variance weighted beta −0.34, 95% CI −0.67, −0.01, $P = 0.057$]), and greater likelihood of smoking cessation among smokers (inverse variance weighted beta 0.65, 95% CI 0.35, 0.95, $P = 5.54 \times 10^{-5}$). Less consistent across the different techniques were associations between educational attainment and smoking initiation.

Conclusions: Our findings indicate a causal association between low educational attainment and increased risk of smoking, and may explain the observational associations between educational attainment and adverse health outcomes such as risk of coronary heart disease.

Key words: Smoking, education, Mendelian randomization, genetic epidemiology

Key Messages

- We used Mendelian randomization to investigate causal associations between educational attainment and risk of various smoking phenotypes.
- Summary statistics from genome-wide association studies that predict educational attainment and smoking behaviours were used.
- We found broadly consistent evidence of a causal association between low educational attainment and increased risk of smoking related phenotypes.
- Our findings could explain observational associations seen between educational attainment and health outcomes such as coronary heart disease.

Introduction

Smoking prevalence has been consistently declining in most high-income countries over the past 50 years. However, this decline has been seen predominantly in those of higher socioeconomic status, which has resulted in an increase in the social patterning of smoking behaviours and widening health inequality.[1,2] Observational studies consistently show that poor educational outcomes are associated with increased smoking,[3,4] but disentangling causality is challenging due to the higher levels of smoking among those from more disadvantaged backgrounds. A recent study found that lower educational attainment at age 11 predicted later tobacco use, but that the opposite was true for alcohol and cannabis use.[5] Mendelian randomization (MR) is a technique that uses genetic variants as unconfounded proxies for an exposure of interest, as a method of ascertaining better evidence for causality than more traditional observational epidemiological methods.[6–8] Relatedly, a Mendelian randomization study found that genetic variants that predict educational attainment were associated with a decreased risk of coronary artery disease, indicating a causal relationship between education and the disease.[9] One possible pathway through which educational attainment could impact on health is through health behaviours such as smoking.

Here we used two-sample Mendelian randomization[10] to assess the causal relationship between educational attainment and various smoking behaviours, using publicly available summary statistics from genome-wide association studies (GWAS) of educational attainment and smoking behaviours. Specifically, we investigated smoking initiation, heaviness of smoking using self-reported (cigarettes per day) and biomarker (cotinine levels) phenotypes, and smoking cessation. Critical to MR is the assumption that the association of genetic variants acting as a proxy for the exposure operates (directly or indirectly) through the exposure of interest. This assumption can be investigated via the use of a number of complementary MR techniques that rely on orthogonal assumptions.[11] We therefore used a range of MR methods as sensitivity analyses to test the robustness of our conclusions.

Methods

Education variants

Single nucleotide polymorphisms (SNPs) associated with educational attainment were identified from genome-wide significant hits from a GWAS of educational attainment in 305 072 individuals, including 111 349 individuals from the UK Biobank study.[12] The GWAS identified 74 SNPs at genome-wide significance predicting years in education. Beta coefficients and standard errors were extracted from publicly available summary statistics on 305 072 individuals (the discovery and replication sample, minus data from 23andme). Not all the SNPs were present in the outcome (i.e. smoking) GWAS. Where possible, proxy SNPs (correlated $r^2 \leq 0.9$) were identified. The SNPs, proxies, and the analyses they were included in are detailed in Supplementary Table 1, available as Supplementary data at *IJE* online.

Smoking phenotypes

Smoking initiation was assessed by the Tobacco and Genetics (TAG) consortium,[13] as a binary ever/never measure ascertained in 74 053 individuals. Of the 74 possible SNPs associated with years in education, 32 were present in the GWAS of smoking initiation. A further 25 were identified using SNIPA [http://snipa.helmholtz-muenchen.de/snipa/], an online tool to identify SNPs in high linkage disequilibrium that can act as proxy SNPs for those identified, giving a total of 57 SNPs for these analyses.

Heaviness of smoking measured as self-reported cigarettes smoked per day was also assessed by the TAG

consortium, in 38 181 daily smokers. The same 57 SNPs were available for this phenotype as for the initiation phenotype. Heaviness of smoking measured by cotinine levels was assessed by the Cotinine Consortium GWAS[14] in 4548 daily smokers. Of the 74 possible SNPs associated with education, 72 were present in the cotinine GWAS.

Smoking cessation was assessed in the TAG consortium in 41 278 individuals who were either current or former smokers. The same 57 SNP were available for this phenotype as for the initiation and heaviness of smoking (cigarettes per day) phenotypes.

Procedure

Beta coefficients and standard errors of the SNPs associated with years in education were recorded. These SNPs were then identified in the GWASs of the smoking phenotypes, and corresponding log odds ratios or beta coefficients (as appropriate) and standard errors were recoded. The SNP-exposure and SNP-outcome associations for each smoking phenotype were combined in a fixed effect meta-analysis using inverse variance weighting (IVW). We also ran a number of sensitivity analyses that provide causal estimates under less stringent assumptions than the traditional MR approach.

First, we implemented MR Egger regression, which relaxes the assumption that the effects of the variants on the outcome are entirely mediated via the exposure.[12] MR-Egger allows for each variant to exhibit some pleiotropy, but assumes that each gene's association with the exposure is independent in magnitude from its pleiotropic effects (the InSIDE assumption).[15] This is achieved by allowing an intercept term in the weighted regression analysis. The value of the intercept provides an estimate of the degree of pleiotropy affecting the result, and the beta (slope) coefficient represents the causal effect between exposure and outcome adjusted for pleiotropy. We also ran MR Egger analyses using simulation extrapolation (SIMEX) correction, which corrects the standard MR-Egger regression coefficients for regression dilution due to uncertainty in the gene-exposure association estimates. SIMEX creates a series of new datasets under increasing violations of the no measurement error (NOME) assumption, which are then analysed together to infer the effect size if NOME had been perfectly satisfied. This is explained in further detail in Bowden *et al.*[16] Violations of NOME assumption lead to attenuation of the causal effect towards the null.

Second, we conducted weighted median regression analyses, which can provide a consistent estimate for the true causal effect when up to half of the weight in the MR analysis pertains from genetic variants that exert pleiotropic effects on the outcome.[17] We also conducted weighted modal regression analyses, which relax instrumental variable assumptions.[18] Finally, we calculated the Cook's distance and Studentized residual measures for the IVW and MR Egger approaches, in order to ascertain whether any individual SNPs were unduly influential in driving the analysis results. Cook's distance measures the influence of each estimate on the MR regression slopes for these two methods. It is calculated as the sum of all changes in the fitted value of the slope when evaluated at each SNP-exposure association (the x-axis coordinates). SNPs were considered outliers if their Studentized residual value was significant at the 5% level after Bonferroni correction. In this instance, a sensitivity analysis was run with the SNP removed.

Results

Smoking initiation

Using inverse variance weighted regression, more years in education were strongly associated with reduced likelihood of initiating smoking [beta −0.54, 95% confidence interval (CI) −0.71, −0.36, $P = 9.28 \times 10^{-8}$]. Similarly, using both weighted median MR and weighted modal MR, there was strong evidence of an association although confidence intervals were wider (weighted median beta −0.72, 95% CI −1.10, −0.35, $P = 0.0002$). The I^2_{GX} statistic for these data (which quantifies the expected dilution of the MR-Egger estimate[16]) was 0.62, and so a SIMEX correction was deemed necessary. However, using this approach, the point estimate was directionally opposite (SIMEX-corrected beta 0.44, 95% CI −0.54, 1.42, $P = 0.381$). The intercept indicated weak evidence of pleiotropy (beta −0.01, 95% CI −0.03, 0.00, $P = 0.044$). These results are shown in Table 1 and Figure 1.

Heaviness of smoking

There was evidence of an association between more years of education and fewer cigarettes smoked per day (beta

Table 1. Estimates from various Mendelian randomization methods for the association between education and smoking initiation

Method	Beta coefficient	95% CI	*P*-value
IVW MR	−0.538	−0.713, −0.362	9.28×10^{-8}
Weighted median MR	−0.718	−1.094, −0.347	0.0002
Weighted modal MR	−0.648	−0.989, −0.307	0.0002
MR Egger estimate	0.320	−0.344, 0.983	0.338
MR Egger intercept	−0.012	−0.021, −0.003	0.010
Egger SIMEX estimate	0.440	−0.537, 1.417	0.381
Egger SIMEX intercept	−0.014	−0.027, -0.001	0.044

Analysis of 74 053 individuals (SNP $n =$ 57 out of a possible 74).

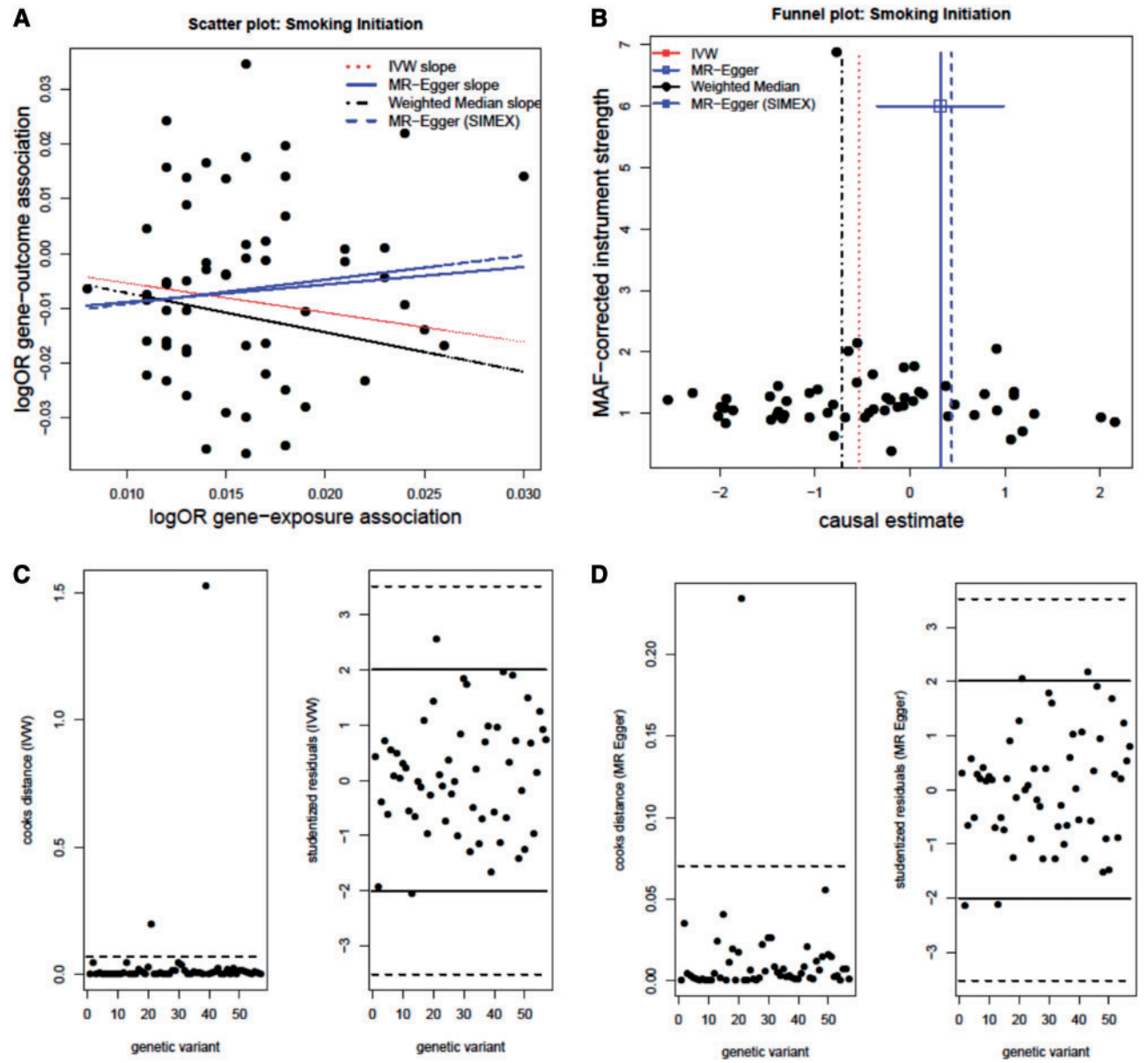

Figure 1. Genetic associations of education with likelihood for smoking initiation. Analysis of 74 053 individuals (SNP $n = 57$ of a possible 74). A) Funnel plot showing minor allele frequency and causal estimate correlations for each SNP. Lines represent the effect sizes of the different regression analyses. B) Scatter plot showing the correlation of genetic associations of education with genetic associations of smoking initiation. Lines represent the slopes of the different regression analyses. C) Cook's distances and Studentized residuals from inverse variance weighted method. D) as C) from MR Egger method, in order to ascertain the presence of outliers. Cook's distance represents the influence of each estimate on the regression slope of the model(s). Dashed lines in Studentized residual graphs represent Bonferroni corrected significance threshold. Any SNPs which fall outside these dashed lines may be considered outliers.

−2.25, 95% CI −3.81, −0.70, $P = 0.005$). I^2_{GX} for these analyses was also 0.62. As with the smoking initiation analysis, results were similar when using weighted median MR, although confidence intervals were again wider (beta −2.17, 95% CI −4.40, 0.06, $P = 0.057$), whereas results using MR Egger were attenuated (SIMEX corrected beta −1.05, 95% CI −9.42, 7.33, $P = 0.804$). The intercept did not indicate any evidence of pleiotropy (SIMEX corrected beta −0.02, 95% CI −0.15, 0.11, $P = 0.759$). The weighted modal estimate was broadly similar to the IVW and weighted median approach (beta −1.70, 95% CI −5.28, 1.89, $P = 0.353$). These results are shown in Table 2 and Figure 2.

There was little evidence of an association between years of education and cotinine levels. However, after assessment of Cook's distances and Studentized residuals, an outlier SNP (rs113520408) was identified in the IVW analysis, although it did not reach significance as an outlier

Table 2. Estimates from various Mendelian randomization methods for the association between education and cigarettes per day

Method	Beta coefficient	95% CI	*P*-value
IVW MR	−2.253	−3.810, −0.696	0.005
Weighted median MR	−2.167	−4.395, 0.061	0.057
Weighted modal MR	−1.698	−5.284, 1.887	0.353
MR Egger estimate	−0.711	−6.711, 5.289	0.813
MR Egger intercept	−0.025	−0.121, 0.070	0.596
Egger SIMEX estimate	−1.045	−9.423, 7.333	0.804
Egger SIMEX intercept	−0.020	−0.149, 0.109	0.759

Analysis of 38 181 daily smokers (SNP $n = 57$ out of a possible 74).

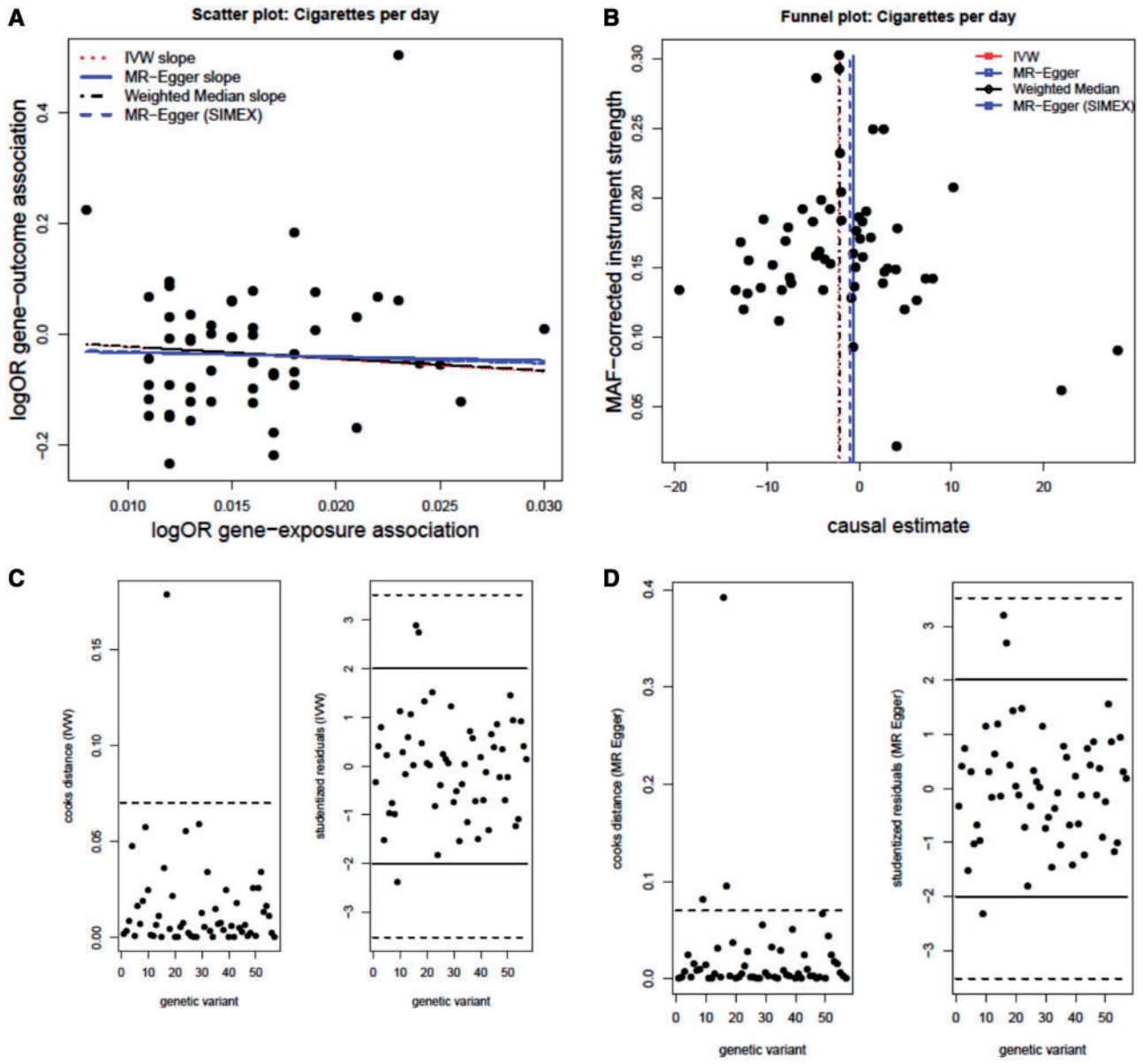

Figure 2. Genetic associations of education with cigarettes per day. Analysis of 38 181 daily smokers (SNP $n = 57$ of a possible 74). A) Funnel plot showing minor allele frequency and causal estimate correlations for each SNP. Lines represent the effect sizes of the different regression analyses. B) Scatter plot showing correlation of genetic associations of education with genetic associations of cigarettes per day. Lines represent the slopes of the different regression analyses. C) Cook's distances and Studentized residuals from inverse variance weighted method. D) as C) from MR Egger method, in order to ascertain the presence of outliers. Cook's distance represents the influence of each estimate on the regression slope of the model(s). Dashed lines in Studentized residual graphs represent Bonferroni corrected significance threshold. Any SNPs which fall outside these dashed lines may be considered outliers.

Table 3. Estimates from various Mendelian randomization methods for the association between education and cotinine levels

Method	Beta coefficient	95% CI	*P*-value
Complete SNP set ($n = 72$)			
IVW MR	0.077	−0.182, 0.335	0.557
Weighted median MR	0.704	0.168, 1.240	0.010
Weighted modal MR	0.577	0.091, 1.064	0.020
MR Egger estimate	−1.462	−2.596, −0.328	0.012
MR Egger intercept	0.024	0.007, 0.040	0.007
Outlier rs113520408 removed (SNP $n = 71$)			
IVW MR	−0.336	−0.667, −0.005	0.057
Weighted median MR	−0.321	−0.822, 0.180	0.209
Weighted modal MR	−0.498	−1.535, 0.539	0.347
MR Egger estimate	−0.635	−1.904, 0.635	0.322
MR Egger intercept	0.005	−0.016, 0.026	0.629
Egger SIMEX estimate	−0.922	−2.721, 0.878	0.311
Egger SIMEX intercept	0.010	−0.019, 0.039	0.512

Analysis of 4548 daily smokers.

in the MR Egger analysis. We removed it in a sensitivity analysis to ascertain whether this SNP was having an undue influence on the association, and the association when this SNP was removed. This SNP was identified in the initial education GWAS as being worthy of further investigation, as it showed sign-discordant effects on height and educational attainment, though we could not identify a specific biological rationale for this SNP being an outlier. Inverse variance weighted MR, weighted median MR and MR Egger all produced beta coefficients in the expected direction in this sensitivity analysis (i.e. with more years of education being associated with lower cotinine levels), although statistical evidence of association was weak for all methods (e.g. inverse variance weighted beta −0.34, 95% CI −0.67, −0.01, $P = 0.057$). The I^2_{GX} value was 0.64. The intercept in the MR Egger analysis did not indicate any evidence of pleiotropy (SIMEX corrected beta 0.01, 95% CI −0.02, 0.04, $P = 0.512$). These results are shown in Table 3 and Figure 3.

Smoking cessation

There was strong evidence of an association between more years of association and greater likelihood of smoking cessation (beta 0.65, 95% CI 0.35, 0.95, $P = 5.54 \times 10^{-5}$). I^2_{GX} was 0.62 for these analyses. Again, results were similar when using weighted median MR (beta 0.60, 95% CI 0.16, 1.04, $P = 0.008$), weighted modal MR (beta 0.39, 95% CI −0.59, 1.37, $P = 0.433$) and MR Egger (SIMEX corrected beta 0.61, 95% CI −1.14, 2.35, $P = 0.500$), although confidence intervals were wider. The intercept did not indicate any evidence of pleiotropy (SIMEX corrected beta 0.00, 95% CI −0.03, 0.03, $P = 0.946$). These results are shown in Table 4 and Figure 4.

Discussion

By triangulating evidence from three complementary MR methods that rely on different underlying assumptions, we find evidence that more years in education leads to reduced likelihood of smoking initiation, reduced heaviness of smoking among smokers and greater likelihood of smoking cessation among smokers. Although statistical evidence was generally weaker when using weighted median and MR Egger methods, results across the methods were broadly consistent in terms of the direction and strength of association observed. Moreover, MR Egger indicated only weak or little evidence of biological pleiotropy. Biological pleiotropy, where one variant has a direct effect on two or more phenotypes, differs from mediated pleiotropy, where a variant impacts on a phenotype via another phenotype. MR assumptions are violated by biological pleiotropy, but are not violated by mediated pleiotropy.

The directionally opposite results obtained using MR Egger for smoking initiation (compared with those obtained using the inverse variance weighted and weighted median methods) is surprising. This may be because smoking initiation might be a less precise phenotype than some of the others. The smoking initiation GWAS could conceivably be measuring another phenotype, such as impulsivity or novelty seeking, rather than a phenotype specifically related to smoking, since the question 'Have you ever smoked a cigarette?' will capture individuals who have experimented only occasionally with smoking (participants were deemed 'ever smokers' if they had smoked 100 cigarettes in their lifetime) as well as those who go on to become daily smokers for many years. MR Egger suggested weak evidence of pleiotropy for this analysis, and the SIMEX confidence intervals are wide. This analysis indicated negative pleiotropy (that is, pleiotropy in the opposite direction to the causal effect), which could mask the true causal estimate. After this is estimated and adjusted for, the causal estimate is inferred to be positive. The variety of analyses conducted for each exposure of interest make different underlying assumptions about the nature of any pleiotropy. The different results obtained using MR Egger compared with other methods could indicate that the InSIDE (instrument strength independent of direct effect) assumption is violated, or it could indicate that the InSIDE assumption is true, with respect to directional pleiotropy. We are unable to distinguish between these possibilities.

The results of the cotinine analysis were initially inconsistent, but this seemed to be largely due to the influence of

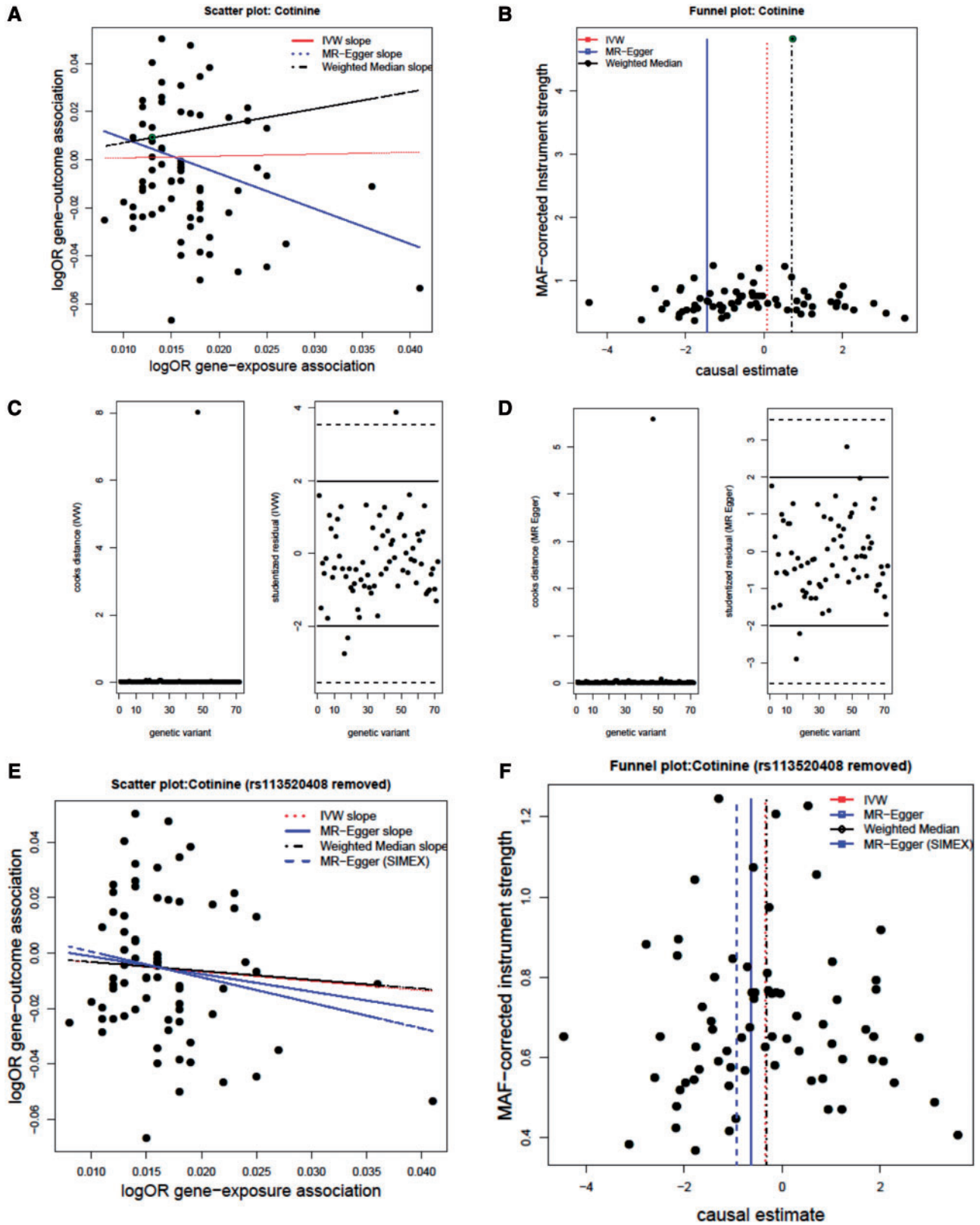

Figure 3. Genetic associations of education with cotinine levels. Analysis of 4548 daily smokers (SNP $n = 72$ of a possible 74). A) Funnel plot showing minor allele frequency and causal estimate correlations for each SNP. Lines represent the effect sizes of the different regression analyses. B) Scatter plot showing correlation of genetic associations of education with genetic associations of cotinine levels. Lines represent the slopes of the different regression analyses. C) Cook's distances and Studentized residuals from inverse variance weighted method. D) as C) from MR Egger method, in order to ascertain the presence of outliers. Cook's distance represents the influence of each estimate on the regression slope of the model(s). Dashed lines in Studentized residual graphs represent Bonferroni corrected significance threshold. Any SNPs which fall outside these dashed lines may be considered outliers. E) and F) as A) and B), after excluding outlier SNP rs113520408 (highlighted in panel A).

Table 4. Estimates from various Mendelian randomization methods for the association between education and smoking cessation

Method	Beta coefficient	95% CI	*P*-value
IVW MR	0.651	0.352, 0.950	5.54×10^{-5}
Weighted median MR	0.600	0.160, 1.040	0.008
Weighted modal MR	0.391	−0.587, 1.369	0.433
MR Egger estimate	0.415	−0.741, 1.571	0.475
MR Egger intercept	0.004	−0.015, 0.022	0.673
Egger SIMEX estimate	0.606	−1.141, 2.353	0.500
Egger SIMEX intercept	0.001	−0.027, 0.029	0.946

Analysis of 41 278 individuals (SNP $n = 57$ out of a possible 74).

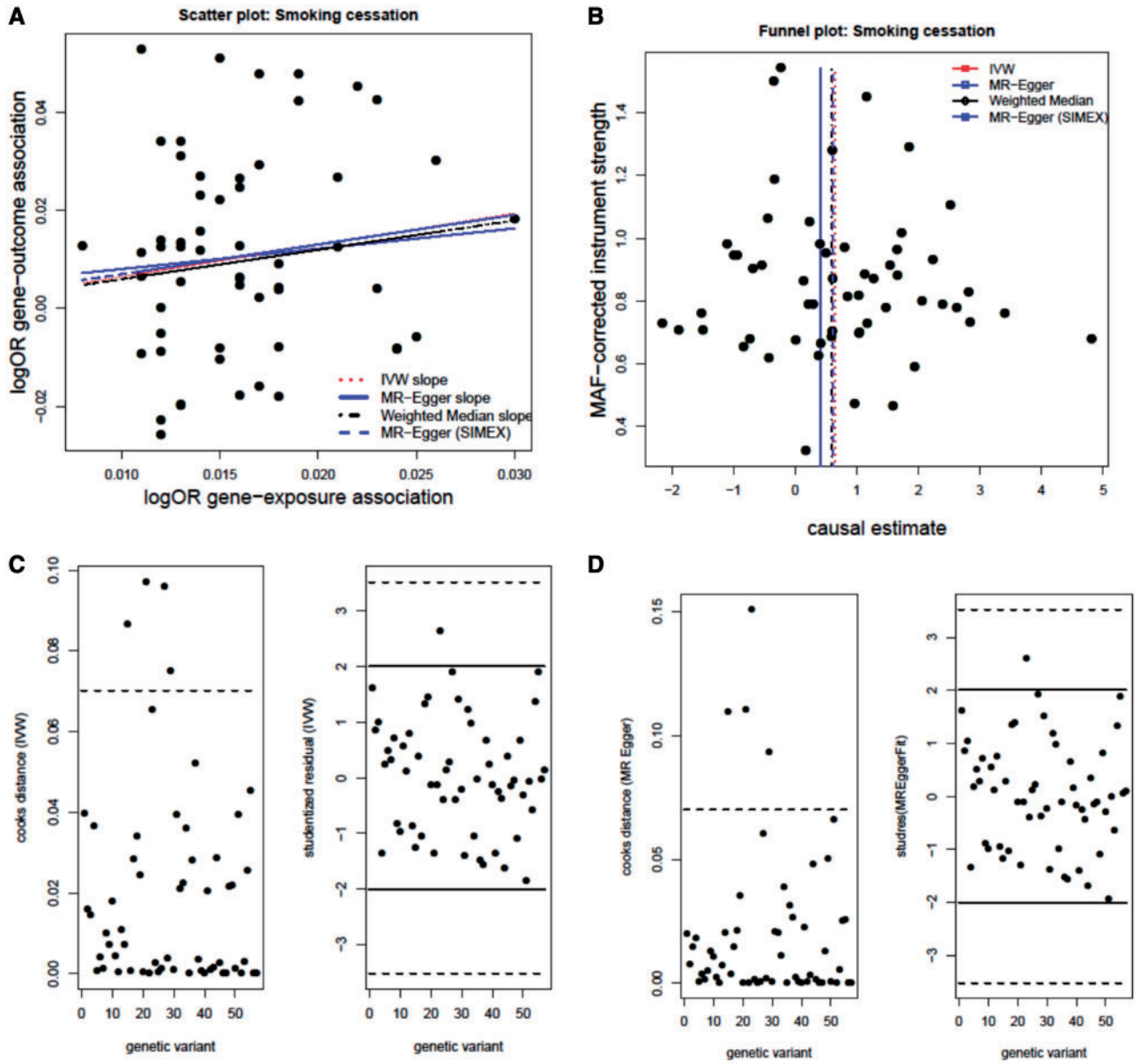

Figure 4. Genetic associations of education with likelihood for smoking cessation. Analysis of 41 278 individuals (SNP $n = 57$ of a possible 74). A) Funnel plot showing minor allele frequency and causal estimate correlations for each SNP. Lines represent the effect sizes of the different regression analyses. B) Scatter plot showing correlation of genetic associations of education with genetic associations of smoking initiation. Lines represent the slopes of the different regression analyses. C) Cook's distances and Studentized residuals from inverse variance weighted method. D) As C) from MR Egger method, in order to ascertain the presence of outliers. Cook's distance represents the influence of each estimate on the regression slope of the model(s). Dashed lines in Studentized residual graphs represent Bonferroni corrected significance threshold. Any SNPs which fall outside these dashed lines may be considered outliers.

one SNP that was identified as an outlier. Once removed, the estimates from the various MR methods were much more similar, and consistent with the results for the analysis of self-reported cigarettes per day (despite being based on a much smaller sample size).

Our use of multiple smoking phenotypes and various different MR techniques is an important strength. However, there are limitations to our results that are important to consider. In particular, not all the genome-wide significant SNPs that predicted educational attainment were available in the outcome GWAS we used, meaning we are not necessarily capturing the full variance with the included variants. Although we were able to identify some proxies, for the analyses that used the TAG consortium we were still missing 17 SNPs. A further limitation relates to the binary nature of our exposure variable. When this is the case, it is difficult to interpret effect sizes as there is the possibility of bias.[19] However, given that the estimate will be unbiased at the causal null, these findings still give us an indication of the likelihood that any association seen is causal in nature, and should be interpreted in this light, rather than giving a precise estimate of the size of any association.

Our findings may explain the observational associations between educational attainment and adverse health outcomes such as risk of coronary heart disease. Indeed, a recent comment piece argues precisely this, that social rank has an impact on health both on lifestyle behaviours (such as smoking) and via other pathways.[20] A recent study concluded that smoking was only a partial mediator of the association between intelligence and mortality, although the measure of smoking used was crude and therefore residual confounding is still possible.[21] It is also possible that psychological or cognitive traits could mediate these associations. Our results indicate that education could represent a worthwhile target for intervention. A recent natural experiment exploiting the raising of school leaving age in the UK changes found evidence of causal associations of increased schooling on a variety of health and socioeconomic factors.[22] Although they found little evidence that the 1-year increase in schooling was associated with smoking, they did find evidence of a causal association between education and decreased risk of various health outcomes including diabetes and stroke. This inconsistency between these findings and the results we report here could be due to the use of UK Biobank data by Davies and colleagues. Higher educational attainment may be associated with greater likelihood of participation in UK Biobank, whereas smoking may be associated with lower likelihood of participation. This could lead to collider bias and a consequent attenuation to the null of associations between educational attainment and smoking.[23] Policy makers should consider the length and quality of education provision, given growing evidence that it can causally impact on health and health-related behaviours.

Supplementary Data

Supplementary data are available at *IJE* online.

Funding

This work was supported by the Medical Research Council and the University of Bristol (MC_UU_12013/1, MC_UU_12013/6). J.B. is supported by an MRC Methodology Research Fellowship (grant MR/N501906/1). M.R.M. and S.H.G. are members of the UK Centre for Tobacco and Alcohol Studies, a UKCRC Public Health Research: Centre of Excellence. Funding from the British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council and the National Institute for Health Research, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged.

Conflict of interest: None declared.

References

1. Jefferis BJ, Power C, Graham H, Manor O. Changing social gradients in cigarette smoking and cessation over two decades of adult follow-up in a British birth cohort. *J Public Health (Oxf)* 2004;**26**:13–18.
2. Peretti-Watel P, Seror V, Constance J, Beck F. Poverty as a smoking trap. *Int J Drug Policy* 2009;**20**:230–36.
3. Fergusson DM, Horwood LJ, Ridder EM. Show me the child at seven II: childhood intelligence and later outcomes in adolescence and young adulthood. *J Child Psychol Psychiatry* 2005;**46**:850–58.
4. Latvala A, Rose RJ, Pulkkinen L, Dick DM, Korhonen T, Kaprio J. Drinking, smoking, and educational achievement: cross-lagged associations from adolescence to adulthood. *Drug Alcohol Depend* 2014;**137**:106–13.
5. Williams J, Hagger-Johnson G. Childhood academic ability in relation to cigarette, alcohol and cannabis use from adolescence into early adulthood: Longitudinal Study of Young People in England (LSYPE). *BMJ Open* 2017;**7**:e012989.
6. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014;**23**:R89–98.
7. Gage SH, Munafo MR, Davey Smith G. Causal inference in developmental origins of health and disease (DOHaD) research. *Annu Rev Psychol* 2016;**67**:567–85.
8. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;**32**:1–22.
9. Tillmann T, Vaucher J, Okbay A *et al*. Education and coronary heart disease: a Mendelian randomization study. *bioRxiv* 2017. doi: https://doi.org/10.1101/106237 [Preprint.]
10. 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.

11. Corbin LJ, Richmond RC, Wade KH *et al*. BMI as a modifiable risk factor for type 2 diabetes: refining and understanding causal estimates using Mendelian randomization. *Diabetes* 2016;**65**: 3002–07.
12. Okbay A, Beauchamp JP, Fontana MA *et al*. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 2016;**533**:539–42.
13. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* 2010;**42**:441–47.
14. Ware JJ, Chen X, Vink J *et al*. Genome-wide meta-analysis of cotinine levels in cigarette smokers identifies locus at 4q13.2. *Sci Rep* 2016;**6**:20092.
15. 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.
16. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. *Int J Epidemiol* 2016;**45**:1961–74.
17. 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.
18. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomisation via the zero modal pleiotropy assumption. *Int J Epidemiol* 2017;**46**:1985–88.
19. Didelez V, Meng S, Sheehan NA. Assumptions of IV methods for observational epidemiology. *Stat Sci* 2010;**25**:22–40.
20. Tobias M. Social rank: a risk factor whose time has come? *Lancet* 2017;**389**:1172–74.
21. Calvin CM, Batty GD, Der G *et al*. Childhood intelligence in relation to major causes of death in 68 year follow-up: prospective population study. *BMJ* 2017;**357**:j2708.
22. 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. *Natur Hum Behav* 2018;**2**:117–25.
23. Munafo M, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: how selection bias can induce spurious associations. *Int J Epidemiol* 2018;**47**:226–35.