bioRxiv preprint doi: https://doi.org/10.1101/299826; this version posted April 16, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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 4.0 International license.

The effect of education and general cognitive ability on smoking:

A Mendelian randomisation study

Eleanor Sanderson [1,2], George Davey Smith [1,2], Marcus R. Munafò [1,3 *]

1. MRC Integrative Epidemiology Unit at the University of Bristol, UK.

2. Population Health Sciences, Bristol Medical School, University of Bristol, UK.

3. UK Centre for Tobacco and Alcohol Studies, School of Experimental Psychology, University of Bristol, UK.

Correspondence to: Marcus R. Munafò, School of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol BS8 1TU, United Kingdom. T: +44.117.9546841; E: marcus.munafo@bristol.ac.uk

Word count: 3,519

bioRxiv preprint doi: https://doi.org/10.1101/299826; this version posted April 16, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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 4.0 International license.

Abstract

Objectives. To determine whether the effect of educational attainment on health outcomes is due in part to an effect on smoking behaviour, and whether this is due to educational attainment or general cognitive ability.

Design. Multivariable Mendelian randomization using genetic variants associated with educational attainment and general cognitive ability.

Setting. Individuals (N = 84,907) in the UK Biobank study.

Main outcome measures. Smoking status, including current smoking, ever smoking and smoking cessation.

Results. One additional year of education leads to a 4.4% decrease in the probability of being a current smoker (95% CI 6.8% decrease to 1.9% decrease), but there is no clear evidence of an effect of general cognitive ability (coefficient 5.5% increase, 95% CI -0.4% decrease to 11.3% increase). Similar results were obtained for smoking initiation and cessation.

Conclusions. More years of education lead to a reduced likelihood of smoking and, among those who do smoke, a greater likelihood of cessation. Given the considerable physical harms associated with smoking, this is likely to account for a substantial proportion of health inequalities associated with differences in educational attainment.

Keywords: education, general cognitive ability, smoking, causal inference, Mendelian randomization

bioRxiv preprint doi: https://doi.org/10.1101/299826; this version posted April 16, 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 4.0 International license.

The effect of education and general cognitive ability on smoking: A Mendelian randomisation study

Introduction

Smoking remains the largest single, modifiable contributor to morbidity and premature mortality in high income countries such as the United Kingdom (1)However, while smoking prevalence has declined steadily since the 1950s, this overall decline masks considerable differences across society. Specifically, while smoking prevalence has declined considerably in higher socioeconomic groups, the decline has been considerably less in lower socioeconomic groups, so that smoking is now strongly socially patterned and an important determinant of health inequalities.

Recent analyses have suggested that educational attainment is an important determinant of health outcomes such as coronary heart disease. Using genetic variants associated with educational attainment in Mendelian randomisation (MR) analyses (2), Tillmann and colleagues provided evidence that lower educational attainment is causally related to increased risk of coronary heart disease (3). Using similar methods, Gage and colleagues showed similar effects with respect to a range of smoking behaviours, with lower educational attainment related to increased risk of smoking initiation and, if a smoker, increased heaviness of smoking and reduced likelihood of smoking cessation (4). It is therefore plausible that some of the effect of educational attainment on risk of coronary heart disease (and other adverse health outcomes) is mediated via effects on smoking behaviour.

However, one limitation of analyses of educational attainment is that they do not allow us to distinguish between the effects of education per se, versus the effects of general cognitive ability (which is associated with higher educational attainment). This distinction has important implications when understanding potential the role of education in reducing health inequalities – if apparent protective effects of educational attainment are in fact due to general cognitive ability, then increasing years in education will not necessarily improve health outcomes or reduce health inequalities. In other words, elucidating the independent

bioRxiv preprint doi: https://doi.org/10.1101/299826; this version posted April 16, 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 4.0 International license.

effects of educational attainment versus general cognitive ability is critical to understanding the nature of health inequalities associated with educational status.

In this study we used genetic variants associated with educational attainment, together with genetic variants associated with general cognitive ability, in a multivariable MR framework using individual level data to determine the unique effects of each on smoking behaviour. This method, recently developed by Sanderson and colleagues (5) allows multiple genetic instruments, capturing distinct exposures, to be investigated simultaneously (even when these exposures are highly correlated, as in the case of educational attainment and general cognitive ability).

Methods

Data

Between 2006 and 2010, UK Biobank recruited 502,641 individuals aged between 37 and 73 years old from across the UK. All participants attended a clinic where they completed a questionnaire and interview about their characteristics and a range of health topics. They additionally provided a range of data including anthropometric measurements and gave samples of blood, urine and saliva. This study is described in more detail elsewhere (6), and data access procedures are described on the UK Biobank website (http://www.ukbiobank.ac.uk/principles-of-access/).

We excluded from our analysis all individuals with a high level of relatedness to other individuals in the study (81,890 exclusions) or who had non-European ancestry (78,647 exclusions). We additionally excluded an additional 849 individuals due to sex chromosomal mismatch with their reported sex or sex-chromosome aneuploidy detection, or who where outliers in heterozygosity or the rate of missing genetic data..

The SNPs that we used to create our instrument for general cognitive ability were discovered using the interim release of UK Biobank, and so we additionally excluded all individuals who were included in the interim release. This gives a resulting potential sample size of 256,142 individuals. A total of 91,016 of the individuals eligible to be included in our

bioRxiv preprint doi: https://doi.org/10.1101/299826; this version posted April 16, 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 4.0 International license.

sample completed the general cognitive ability questions. However, 6,109 of these individuals were missing data on one of the other exposures or controls included in the analysis, giving a final sample size of 84,907 individuals.

Measures

Education. Individuals in UK Biobank were asked to report their highest educational qualification. For each of the levels of education a corresponding age at which the individual would have completed their education has been assigned. These are described in Table 1.

Insert Table 1 about here.

General cognitive ability. General cognitive ability was measured as the number of correct answers recorded in a series of 13 questions designed to measure cognitive ability. This was completed by participants during a computer questionnaire as part of their initial clinic visit or through an online questionnaire completed after the visit. The score is the number of questions correctly completed by the individuals in the time allowed and ranges between 0 and 13 with a mean of 6.0 and a standard deviation of 2.1. For the analyses reported here, the score has been standardised to have a mean of 0 and a standard deviation of 1.

Smoking. All participants where asked if they currently smoked and if they had ever smoked. From this information we constructed three variables: 1) a binary variable for current smokers/current non-smokers; 2) a binary smoking initiation variable for ever/never smokers; and 3) among individuals who have ever smoked, a smoking cessation variable for former vs current smokers.

Genetic Instruments. A recent GWAS of educational attainment by Okbay and colleagues identified 74 SNPs at genome-wide significance that associated with years of education completed (7). A polygenic risk score for education was created for each of the individuals in our sample using the genome-wide significant SNPs from this GWAS. Each

bioRxiv preprint doi: https://doi.org/10.1101/299826; this version posted April 16, 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 4.0 International license.

individual's risk score was calculated as the weighted total number of education increasing alleles they had from these SNPs, weighted by the size of the effect of the SNP in the original GWAS. Three of the SNPs from this GWAS where not available in the UK Biobank data and were substituted; all of the substitutes used where in perfect LD with the original SNP.

For general cognitive ability we created a polygenic risk score for each individual using the results from a recent GWAS of general cognitive ability by Sniekers and colleagues, which identified 18 SNPs at genome-wide significance (8). Again, the risk score was calculated as the weighted number of general cognitive ability increasing alleles each individual had from these SNPs, weighted by the effect of the SNP in the GWAS.

Statistical analysis

We used multivariable MR analysis of individual level data to estimate the direct effects of education and general cognitive ability on our three smoking phenotypes. Multivariable MR includes both education and general cognitive ability as exposure variables in an MR regression with both of these exposures being predicted by a set of SNPs. This analysis controls for any correlation between education and general cognitive ability, and for any pleiotropic effect of the education SNPs on general cognitive ability and the general cognitive ability SNPs on education (i.e., through the education SNPs having a direct effect on general cognitive ability, or the cognitive ability SNPs having a direct effect on education). This analysis therefore provides the direct causal effect of education and general cognitive ability on each of our smoking phenotypes. Multivariable MR is conducted by regressing each of the exposure variables on all of the instruments and all of the control variables to generate predicted values for each exposure that are not correlated with any potential confounders. The outcome of interest is then regressed on these new predicted exposure variables and the control variable to obtain consistent estimates of the direct effect of each of the exposures on the outcome. Multivariable MR is explained in detail elsewhere (5).

bioRxiv preprint doi: https://doi.org/10.1101/299826; this version posted April 16, 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 4.0 International license.

For comparison we also conducted standard MR analysis for each of the exposures on each of our three outcomes using the SNP score for education as the only instrument for educational attainment and the SNP score for general cognitive ability as the only instrument for general cognitive ability. These results give the total causal effect of education and general cognitive ability on each of the smoking phenotypes, including any effect of education on general cognitive ability and any effect of general cognitive ability on education. However, these results are more susceptible to pleiotropy as any direct pleiotropic effect of the education score on general cognitive ability or the general cognitive ability score on education will affect the results obtained from this analysis.

Throughout the analysis we adjusted for age at the time of the clinic, sex, year of birth, and a year of birth × sex interaction term, in order to adjust for any changes in patterns of smoking behaviour over time. We also controlled for the top 10 genetic principal components to account for any residual population stratification. We additionally included an 80% upweighting for individuals who left school with no qualifications.

To test for potential pleiotropy in the effect of the SNPs used as instruments we calculated the multivariable MR Egger effect estimates (9). Multivariable MR Egger analysis extends MR Egger to a two-sample multivariable setting (10). The effect of all of the SNPs included in the analysis on each of the exposures and the outcome are estimated, using separate samples for the exposures and the outcome. The estimated effects for each of the SNPs on the outcome are regressed on the effect of the SNPs on each of the exposures, weighted by the variance of the effect of the SNPs on the outcome. Under a set of assumptions (known as the InSIDE assumptions), the constant from this regression indicates the level of (directional) pleiotropy in the model. A constant of zero suggests that the SNPs used in the analysis do not have a pleiotropic effect on the outcome. The effects of each SNP in the instruments on the exposures was taken from the GWAS of educational attainment (7) and the GWAS of cognitive ability (8) used to calculate the polygenic risk scores in the main analysis. The effect of each of the SNPs on the outcomes was calculated in the UK Biobank data used in the main analysis.

bioRxiv preprint doi: https://doi.org/10.1101/299826; this version posted April 16, 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 4.0 International license.

Results

Observational analysis

Observational analyses, controlling for the full set of controls using ordinary least squares (OLS) regression, indicated a small negative association between educational attainment, general cognitive ability and current smoking. These results also indicated a negative association between increased education and smoking initiation, and a positive association between general cognitive ability and smoking initiation. Increased education and higher general cognitive ability were both positively associated with smoking cessation. These results are shown in Tables 2-4

Insert Tables 2 to 4 about here.

Causal analysis

In all of these analyses we used two instruments in the multivariable MR analysis and one instrument in each of the single variable MR analyses. The F-statistics show that the instruments used strongly predict education and general cognitive ability. In the multivariable MR analysis an additional assumption required for the instruments to be considered strong is that the instruments used can predict both of the exposure variables at the same time. This is measured using the conditional F statistic (11). In all of our multivariable MR estimations the conditional F-statistic is larger than the conventional value of 10 used to indicate a strong instrument, and so the instruments used strongly predict the exposure variables in the multivariable MR analyses. The introduction of a second exposure in the multivariable MR analysis does not introduce collider bias because each exposure is predicted by a set of SNPs that are not associated with smoking. The multivariable MR analysis uses values of education and general cognitive ability that have been predicted from a set of SNPs that are not associated with smoking behaviour; these predicted values will therefore not be affected

bioRxiv preprint doi: https://doi.org/10.1101/299826; this version posted April 16, 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 4.0 International license.

by changes in smoking and so cannot introduce collider bias to the analysis. This is explained further elsewhere (5).

Current smoking. The multivariable MR results indicated a positive effect of education on smoking behaviour, with one additional year of education leading to a 4.4% decrease in the probability of being a current smoker. One standard deviation in age completed education is 2.4 years; therefore, an increase in education of one standard deviation leads to a 10.6% decrease in the probability of being a current smoker. These results provided no clear evidence of an effect of general cognitive ability on smoking behaviour. The single variable MR results provided similar estimates for the effect of education on current smoking, with each year of education leading to a 3.1% decrease in the probability of being a smoker, corresponding to a standard deviation increase in education leading to a 7.4% decrease in the probability of being a smoker. However, the general cognitive ability estimates for multivariable MR differed substantially from the single variable MR result, with a standard deviation increase in general cognitive ability causing an increase in the probability of smoking of 5.5% in the multivariable MR analysis compared to a decrease of 3.1% in the single variable MR analysis. The difference between the multivariable MR result and this result may be due to the high level of correlation between education and general cognitive ability.

Smoking initiation. The results for having ever smoked showed a similar pattern to the results for being a current smoker. An increase in education of one year causes a 9.6% decrease in the probability of having ever smoked, this is equivalent to a 23.0% decrease in the probability of ever smoking for a standard deviation increase in years of education. An increase in general cognitive ability of one standard deviation causes a 14.2% increase in the probability of ever having smoked. Similarly the individual MR results showed that increases in both education and general cognitive ability lead to decreases in the probability of ever having been a smoker.

Smoking cessation. The results for being a former smoker showed that higher education leads to a higher probability of having quit smoking, but did not indicate a causal

bioRxiv preprint doi: https://doi.org/10.1101/299826; this version posted April 16, 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 4.0 International license.

effect of general cognitive ability on quitting smoking. These results reflect the pattern seen in the results for currently or ever smoking showing that increased education makes individuals more likely to quit smoking once they have started as well as being less likely to start smoking in the first place.

Sensitivity analyses

As a sensitivity analysis we repeated the analyses separately with no weighting applied, and with self-reported age at which individual completed education (replaced with 21 for individuals with a degree who were not asked their age when they completed education). Both of these analyses give results that are not substantially different from those presented here (data not shown).

The results for the multivariable MR Egger estimation are given in Table 5. In these analyses, we dropped education SNPs directly associated with cognitive ability since we excluded SNPs in high LD, so that the 89 SNPs were included in total. In all of the analyses the constant is estimated to be close to zero, with narrow 95% confidence intervals around zero. These results suggest that the SNPs used in the main analysis do not have a pleiotropic effect on the outcome, and therefore that the results given above are valid.

For MR Egger analysis it is important to align all of the SNPs so they have a positive effect on the exposure; it is not clear how to do this in multivariable MR Egger and therefore we repeated the analysis with all of the effects aligned in the positive direction for firstly the effect on education and then their effect on cognitive ability. The results from this analysis did not change the conclusion drawn from the results (data not shown).

Discussion

Our multivariable MR results allow us to estimate the direct causal effects of each of education and general cognitive ability on smoking behaviour conditional on general cognitive ability and education respectively. The single variable MR estimates give the total effect of a change of a change in education and general cognitive ability on smoking

bioRxiv preprint doi: https://doi.org/10.1101/299826; this version posted April 16, 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 4.0 International license.

including any effect of general cognitive ability on education and vice versa. Our results suggest that more years of education leads to a reduced likelihood of smoking and, among those who do smoke, a greater likelihood of cessation. Given the considerable physical harms associated with smoking, this is likely to account for a substantial proportion of health inequalities associated with differences in educational attainment. The large difference between the conventional and multivariable MR estimates for the effect of general cognitive ability on smoking suggests that a large part of the total effect of general cognitive ability on smoking behaviour observed in the conventional MR analysis is due to an effect of general cognitive ability on educational attainment. The estimates for the effect of education on smoking are far more similar in each of the estimates suggesting that changes in education will have an effect on smoking behaviour. These results are consistent with recent findings reported by Davies and colleagues, which used the natural experiment of an increase in school leaving age to show that remaining in school causally reduces the risk of adverse health outcomes, including likelihood of smoking (12).

Inferring causality using MR methods does not necessarily clarify the mechanistic pathways that link exposures and outcomes. Our results support a causal pathway from educational attainment to smoking behaviour that appears to be independent of general cognitive ability, but the causal pathway is likely to be complex. More years in education might influence smoking via greater awareness of the harms of smoking, increased self-efficacy, exposure to social groups where smoking is less common, and so on. While years in education might be considered a target for intervention, there is only limited scope to extend this. Identifying other links in the causal pathway may reveal other putative intervention targets that are potentially more tractable. This should be the subject of future research, and in part will depend on the identification of genetic variants associated with, for example, self-efficacy beliefs if MR methods are to be applied.

There are a number of important limitations to this study that should be considered when interpreting these results. First, Mendelian randomization is not without limitations and, critically, relies on key assumptions, in particular regarding the validity of the genetic variants

bioRxiv preprint doi: https://doi.org/10.1101/299826; this version posted April 16, 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 4.0 International license.

used as instruments. We have shown that the genetic variants we used are strongly associated with both education and general cognitive ability. We also assessed the robustness of our analyses to potential pleiotropy using methods that rely on different assumptions (e.g., MR Egger), and these produced similar results. Second, we cannot rule out the possibility of dynastic effects – offspring will share on average 50% of the variegated genotype of their parents, so that the effects we observed may in fact be due to effects of parental education on the offspring environment that have an impact on smoking behaviour. While we cannot rule out this possibility, converging evidence from instrumental variables not subject to potential dynastic effects that educational attainment is causally related to smoking-related health outcomes argues against this possibility. Third, the UK Biobank is highly unrepresentative of the general UK population, and this selection bias may introduce biases from which MR methods are not immune (13). For example, if educational attainment and smoking behaviour are both associated with participation in UK Biobank, then restricting our analyses to this sample may introduce collider bias, and bias associations between these variables (or even introduce spurious associations between them). It will therefore be important to replicate these findings in other samples that are either more representative of the general population, or are subject to different selection biases. Fourth, the interpretation of multivariable MR results is nuanced and potentially complex. Specifically, the estimate for educational attainment is adjusted for general cognitive ability and vice versa. This means that the estimated effect of education is the effect given a constant level of general cognitive ability, and the effect for general cognitive ability is the effect given a constant level of education. While the interpretation of our educational attainment results is relatively straightforward, the interpretation of our conditional general cognitive ability results (which imply increased likelihood of smoking with increasing general cognitive ability) is not. As general cognitive ability affects education what this (weak) effect may reflect is the fact that individuals with high levels of general cognitive ability who do not achieve the educational attainment that might be expected are likely to have done so for specific reasons that are themselves likely to influence smoking. This result indicates that the positive effect of

bioRxiv preprint doi: https://doi.org/10.1101/299826; this version posted April 16, 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 4.0 International license.

general cognitive ability on smoking behaviour observed in the conventional MR analysis is acting via the effect of general cognitive ability on education, and the consequent effect of education on smoking behaviour. This suggests that an increase in general cognitive ability that was not associated with a corresponding increase in educational attainment would have a limited effect on smoking behaviour.

In summary, our results indicate that the observed association between educational attainment and smoking is causal, and unlikely to be due to effects of general cognitive ability on educational attainment. This suggests that a part of the health inequalities associated with differences in educational outcomes may be due to smoking, and that education represents a potential target for intervention to reduce health inequalities. Future research may identify other putative targets on the causal pathway from educational attainment and smoking behaviour, and should also explore whether there are similar relationships to other health behaviours such as alcohol consumption and diet.

bioRxiv preprint doi: https://doi.org/10.1101/299826; this version posted April 16, 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 4.0 International license.

Acknowledgements

This research has been conducted using the UK Biobank Resource as part of application 8786, and was supported by the UK Medical Research Council Integrative Epidemiology Unit at the University of Bristol (MM_UU_12013/1, MM_UU_12013/6). MRM is a member of the UK Centre for Tobacco and Alcohol Studies, a UKCRC Public Health Research: Centre of Excellence. Funding from 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.

Author Contributions

ES, GDS and MRM conceived the study. ES conducted the analyses. ES, GDS and MRM drafted the manuscript. All authors reviewed and approved the final version. The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, a worldwide licence

Competing Interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and 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.

Transparency Declaration

MRM will serve as guarantor for this article. The lead author affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted and that any discrepancies from the study as planned have been explained.

bioRxiv preprint doi: https://doi.org/10.1101/299826; this version posted April 16, 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 4.0 International license.

References

1. Collaborators GBDT. Smoking prevalence and attributable disease burden in 195 countries and territories, 1990-2015: a systematic analysis from the Global Burden of Disease Study 2015. Lancet. 2017 May 13;389(10082):1885-906. PubMed PMID: 28390697. Pubmed Central PMCID: 5439023.

2. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? International journal of epidemiology. 2003 Feb;32(1):1-22. PubMed PMID: 12689998.

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 Aug 30;358:j3542. PubMed PMID: 28855160. Pubmed Central PMCID: 5594424 at http://www.icmje.org/coi_disclosure.pdf and declare: support for the submitted work as detailed above; 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.

4. Gage SH, Bowden J, Davey Smith G, Munafo MR. Investigating causality in associations between education and smoking: A two-sample Mendelian randomization study. bioRxiv. 2017.

5. Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomiation in the single-sample and two-sample summary data settings. in preparation.

6. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS medicine. 2015 Mar;12(3):e1001779. PubMed PMID: 25826379. Pubmed Central PMCID: 4380465.

7. 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.

bioRxiv preprint doi: https://doi.org/10.1101/299826; this version posted April 16, 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 4.0 International license.

Nature. 2016 May 26;533(7604):539-42. PubMed PMID: 27225129. Pubmed Central PMCID: 4883595.

8. Sniekers S, Stringer S, Watanabe K, Jansen PR, Coleman JRI, Krapohl E, et al. Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. Nature genetics. 2017 Jul;49(7):1107-12. PubMed PMID: 28530673. Pubmed Central PMCID: 5665562.

9. Rees JMB, Wood AM, Burgess S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. Statistics in medicine. 2017 Dec 20;36(29):4705-18. PubMed PMID: 28960498. Pubmed Central PMCID: 5725762.

10. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International journal of epidemiology. 2015 Apr;44(2):512-25. PubMed PMID: 26050253. Pubmed Central PMCID: 4469799.

11. Sanderson E, Windmeijer F. A weak instrument [Formula: see text]-test in linear IV models with multiple endogenous variables. Journal of econometrics. 2016 Feb;190(2):212-21. PubMed PMID: 29129953. Pubmed Central PMCID: 5669336.

12. Davies NM, Dickson M, Davey Smith G, van den Berg G, Windmeijer F. The causal effects of education on health outcomes in the UK Biobank. Nat Hum Behav. 2018;2:117-25.

13. Munafo MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: when selection bias can substantially influence observed associations. International journal of epidemiology. 2017 Sep 27. PubMed PMID: 29040562.

bioRxiv preprint doi: https://doi.org/10.1101/299826; this version posted April 16, 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 4.0 International license.

Table 1. Highest educational qualifications.

Highest Educational qualification	Age completed education	% of final sample
None	15	12
CSE / O level / GCSE	16	27
NVQ / HND / HNC	18	6
A level	18	12
Other professional qualification *	20	5
College or University degree	21	37

CSE / O level / GCSE: Academic qualification at 16+; NVQ / HND / HNC: Vocational qualification; A level: Academic qualification at 18+; * e.g. nursing/teaching etc.

Table 2. The effect of education and general cognitive ability on current smoking.

		OLS	Multivariable MR	Single variable MR	
Age completed education	Effect	-0.008	-0.044	-0.031	
	SE	0.0004	0.012	0.006	
	95% CI	-0.009 to -0.007	-0.068 to -0.019	-0.043 to -0.019	
	P-value	<0.001	<0.001	<0.001	
	F	-	536.19	384.08	
	Partial F	-	84.36	-	
Standardised general cognitive ability score	Effect	-0.0090	0.055		-0.031
	SE	0.0009	0.030		0.009
	95% CI	-0.011 to -0.007	-0.004 to 0.113		-0.050 to 0.013
	P-value	<0.001	0.067		0.001
	F	-	557.25		894.27
	Partial F	-	84.57		-

Estimates of the effect of education and general cognitive ability on current smoking from OLS, multivariable MR and single variable MR regressions. All regressions also include a full set of adjustments: age, sex, year of birth, gender interacted with year of birth and 10 genetic principal components. All non-European and related individuals have been excluded from the analysis. Total sample size 119,627.

bioRxiv preprint doi: https://doi.org/10.1101/299826; this version posted April 16, 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 4.0 International license.

Table 3. The effect of education and general cognitive ability on smoking initiation.

		OLS	Multivariable MR	Single variable MR	
Age completed education	Effect	-0.019	-0.096	-0.048	
	SE	0.0006	0.022	0.010	
	95% CI	-0.020 to -0.018	-0.139 to -0.052	-0.069 to -0.027	
	P-value	<0.001	<0.001	<0.001	
	F	-	536.19	384.08	
	Partial F	-	84.36	-	
Standardised general cognitive ability score	Effect	0.001	0.142		-0.047
	SE	0.0016	0.054		0.017
	95% CI	-0.002 to 0.005	0.036 to 0.248		-0.080 to -0.014
	P-value	0.371	0.009		0.005
	F	-	557.25		894.27
	Partial F	-	84.57		-

Estimates of the effect of education and general cognitive ability on current smoking from OLS, multivariable MR and single variable MR regressions. All regressions also include a full set of adjustments: age, sex, year of birth, gender interacted with year of birth and 10 genetic principal components. All non-European and related individuals have been excluded from the analysis. Total sample size 119,627.

bioRxiv preprint doi: https://doi.org/10.1101/299826; this version posted April 16, 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 4.0 International license.

Table 4. The effect of education and general cognitive ability on smoking cessation.

		OLS	Multivariable MR	Single variable MR	
Age completed education	Effect	0.011	0.055	0.047	
	SE	0.0008	0.022	0.012	
	95% CI	0.009 to 0.012	0.011 to 0.098	0.023 to 0.071	
	P-value	<0.001	0.015	<0.001	
	F	-	231.48	188.06	
	Partial F	-	49.27	-	
Standardised general cognitive ability score	Effect	0.020	-0.054		0.045
	SE	0.002	0.052		0.019
	95% CI	0.016 to 0.024	-0.157 to 0.048		0.007 to 0.083
	P-value	<0.001	0.297		0.020
	F	-	252.29		406.62
	Partial F	-	49.63		-

Estimates of the effect of education and general cognitive ability on current smoking from OLS, multivariable MR and single variable MR regressions. All regressions also include a full set of adjustments: age, sex, year of birth, gender interacted with year of birth and 10 genetic principal components. All non-European and related individuals have been excluded from the analysis. Total sample size 52,419.

bioRxiv preprint doi: https://doi.org/10.1101/299826; this version posted April 16, 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 4.0 International license.

Table 5. Results from multivariable MR Egger estimates of the effect of education and general cognitive ability on smoking behaviour.

		Outcome		
		Current Smoking	**Smoking Initiation**	**Smoking Cessation**
Age completed education	Effect	-0.058	-0.102	0.092
	SE	0.015	0.034	0.030
	95% CI	-0.088 to -0.029	-0.170 to -0.035	0.032 to 0.153
	P-value	<0.001	0.003	0.003
Standardised general cognitive ability score	Effect	0.012	0.016	0*0.022
	SE	0.011	0.026	0.023
	95% CI	-0.010 to 0.035	-0.036 to 0.067	-0.068 to 0.024
	P-value	0.279	0.545	0.345
Constant (Estimated pleiotropic effect)	Effect	-0.0002	0.0005	0.0005
	SE	0.0002	0.0004	0.0003
	95% CI	-0.0005 to 0.0002	-0.0003 to 0.0012	-0.0001 to 0.0012
	P-value	0.354	0.198	0.109

Multivariable MR Egger estimates of the effect of educational attainment and cognitive ability on smoking behaviour. Effects of SNPs on each of the exposures are taken from GWAS results (7, 8). Effects of each of the SNPs on the outcomes are estimated from the original analysis sample.

bioRxiv preprint doi: https://doi.org/10.1101/299826; this version posted April 16, 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 4.0 International license.