Causal effects of lifetime smoking on risk for depression and schizophrenia: Evidence from a Mendelian randomisation study

Authors. Robyn E Wootton*[1,3,4], Rebecca C Richmond[2,3], Bobby G Stuijfzand[6], Rebecca B Lawn[1,3], Hannah M Sallis[1,2,3], Gemma M. J. Taylor[1,3], Hannah J. Jones[2,3], Stanley Zammit[2,7], George Davey Smith[2,3], Marcus R Munafò[1,3,5]

Author Affiliations.

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

[2]Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, BS8 2PR

[3]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, BS8 2PR

[4]NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol, (BS8 2BN).

[5]UK Centre for Tobacco and Alcohol Studies, University of Bristol, Bristol, BS8 1TU

[6]Jean Golding Institute, Royal Fort House, University of Bristol, Bristol, BS8 1UH

[7]MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff CF24 4HQ

Abstract (150 words). Smoking prevalence is higher amongst individuals with schizophrenia and depression, compared to the general population. Mendelian randomisation (MR) can examine whether this association is causal using genetic variants identified in genome-wide association studies (GWAS). We conducted a GWAS of lifetime smoking behaviour (capturing smoking duration, heaviness and cessation) in a sample of 463,003 individuals from the UK Biobank, and validated the findings via MR analyses of positive control outcomes (e.g., lung cancer). Further MR analyses provided evidence that smoking is a causal risk factor for both schizophrenia and depression. We also found some evidence that genetic liability for both depression and schizophrenia cause increased lifetime smoking. These findings suggest that the association between smoking, schizophrenia and depression is due, at least in part, to a causal effect of smoking. The genetic variants we identify for lifetime smoking have the potential to be used in further MR studies.

Introduction

Smoking is a major risk factor for lung cancer, cardiovascular and respiratory diseases making it the leading cause of preventable death worldwide[1]. In developed nations, smoking is more common amongst individuals with mental health conditions[2–5], in particular schizophrenia[6] and depression[7–9]. In the UK, estimates suggest that up to 45% of individuals with schizophrenia, and 31% of individuals with depression smoke[6], compared to around 15% of the general population[10]. Individuals with mental health conditions smoke more heavily[2] and experience up to 25 years reduced life expectancy compared with the general population[11]. Much of this reduction can be explained by high smoking rates[11], making it important to understand the relationship between smoking and mental health.

It is often assumed that the association between mental health and smoking can be explained by a self-medication model – that is, symptoms of mental illness, or side effects of psychiatric medications, are alleviated by the chemical properties of tobacco[12]. For example, cigarettes have been suggested to reduce the side effects of antipsychotic medications[13,14], and smokers with depression report a reduction in negative affect in response to cigarette smoking[15]. However, observational evidence cannot determine whether the association between smoking and mental health is causal or the result of confounding from correlates of both smoking and mental health[16]. Furthermore, traditional observational evidence cannot robustly identify the direction of causation[16], and there is growing evidence to suggest that smoking may be a causal risk factor for poor mental health. The genome-wide association study (GWAS) of schizophrenia conducted by the Psychiatric Genomics Consortium (PGC) found that variants in the gene cluster *CHRNA5-A3-B4* were associated with increased schizophrenia risk[17]. These variants are known to be strongly associated with heaviness of smoking[18–21]. Therefore, one interpretation of the PGC results is that they reflect the causal effect of smoking on schizophrenia[22]. Prospective observational studies using related individuals to control for genetic and environmental confounding have also suggested a dose-response effect of smoking on schizophrenia[23] and depression[24]. Meta-analyses show further

evidence for an increased relative risk of schizophrenia in daily smokers over non-smokers using unadjusted estimates[25] and a reduction in depressive symptoms following smoking cessation[26]. Although these studies suggest a potential causal effect, more robust methods are required to triangulate evidence and allow for stronger causal inference[27].

One way to overcome the limitations of residual confounding and reverse causation is Mendelian randomisation (MR)[28,29]. This method uses genetic variants as a proxy for an exposure in an instrumental variable analysis to estimate the causal effect on an outcome[16]. Previous MR studies have failed to show any clear evidence for a causal effect of smoking on depression[30–32] and show suggestive but inconclusive evidence for an effect of smoking on schizophrenia[32,33]. However, the genetic instruments for smoking used in these MR studies are limited, only capturing individual aspects of smoking behaviour such as smoking initiation or smoking heaviness, rather than total lifetime smoking exposure[30–33]. Furthermore, any instrument for smoking heaviness requires stratifying samples into smokers and non-smokers. Stratification is not possible using the most common two-sample MR method because it only requires GWAS summary data where individual smoking status is unknown. Stratification might not even be possible at the individual level in a one-sample MR analysis if smoking measures have not been collected in all contributing samples.

Here we report the development of a novel genetic instrument for lifetime smoking exposure that takes into account smoking status (i.e., both smokers and non-smokers) and among ever smokers takes into account smoking duration, heaviness and cessation. This instrument can therefore be used in unstratified samples of smokers and non-smokers. We validate this instrument in an independent sample and against positive control disease outcomes known to be caused by smoking (lung cancer, coronary artery disease and DNA methylation). We next apply the instrument in MR analyses to determine whether the observational associations between smoking and schizophrenia and depression are causal, and the directionality of these relationships.

Methods

Instrument discovery for lifetime smoking index

Sample. We generated the lifetime smoking measure using data from the UK Biobank, a national health research resource of 502,647 participants aged 40-69 years. Participants were recruited from assessment centres across the United Kingdom between 2006 and 2010 (http://www.ukbiobank.ac.uk). Our sample consisted of 463,003 individuals of European ancestry who had phenotype data and passed genotype inclusion criteria (54% female; mean age = 56.7 years; SD = 8.0 years). Overall, 30% of the sample had ever smoked (8% current smokers and 22% former smokers).

Measures of smoking. Phenotypic measures of smoking available in the UK Biobank were self-reported and collected at initial assessment. They included: smoking status (i.e., current, former, never), age at initiation in years, age at cessation in years and number of cigarettes smoked per day. Anyone self-reporting to smoke more than 100 cigarettes per day was contacted to confirm their reported cigarettes per day was accurate. Hand-rolled cigarette smokers were told one gram of tobacco equates to one cigarette. We calculated duration of smoking and time since cessation (calculated from current age and age at cessation). Individuals reporting smoking fewer than 1 cigarette a day or more than 150 cigarettes a day were excluded from instrument generation analysis.

Construction of the lifetime smoking index. Following the method outlined by Leffondré and colleagues[34], we combined the smoking measures into a lifetime smoking index along with a simulated half-life (τ) constant. Half-life captures the exponentially decreasing effect of smoking at a given time on health outcomes. The value of half-life was determined by simulating the effects of lifetime smoking on lung cancer and overall mortality in the UK Biobank. Both suggested the best fitting value was 18. For full details on construction of the lifetime smoking index see Supplementary Note.

Genotyping. UK Biobank participants provided blood samples at initial assessment centre. Genotyping was performed using the Affymetrix UK BiLEVE Axiom array for 49,979 participants and using the Affymetrix UK Biobank Axiom® array for 438,398 participants. The two arrays share 95% coverage, but chip is adjusted for in all analyses because the UK BiLEVE sample is over represented for smokers. Imputation and initial quality control steps were performed by the Wellcome Trust Centre for Human Genetics resulting in over 90 million single nucleotide polymorphisms (SNPs) and indels[35].

Individuals were excluded if there were sex-mismatches between reported and chromosomal sex or aneuploidy (N=814). Individuals were restricted to European ancestry based on the first four principal components of population structure and related individuals were removed following MRC Integrative Epidemiology Unit filtering steps[36]. After excluding individuals who had withdrawn consent, 463,033 of the participants remained[36]. We restricted our analysis to autosomes only and used filtering thresholds for SNPs of minor allele frequency (MAF) >0.01 and info score (measure of imputation uncertainty) >0.8.

Genome-wide association study of lifetime smoking index. GWAS was conducted using the UK Biobank GWAS pipeline set up for the MRC Integrative Epidemiology Unit[37]. BOLT-LMM was used to conduct the analysis[38], which accounts for population stratification and relatedness using linear mixed modelling. Genotyping chip and sex were included as covariates. Genome-wide significant SNPs were selected at $P < 5 \times 10^{-8}$ and were clumped to ensure independence at linkage disequilibrium (LD) $r^2 = 0.001$ and a distance of 10000kb.

Prediction in an independent sample. To validate our instrument, we used an independent sample of 2,712 mothers from the Avon Longitudinal Study of Parents and Children (ALSPAC), a longitudinal birth cohort, which recruited 14,541 pregnant women between 1991 and 1992 with detailed descriptions reported elsewhere[39,40]. Self-reported measures of smoking status, age at initiation, age at cessation and cigarettes per day were collected from these women when their offspring were 18 years of age (mean mothers age = 48 years, SD = 4.3). The genome-wide SNPs for

lifetime smoking were extracted and PLINK was used to generate a polygenic risk score[41]. Linear regression was used to estimate the percentage variance of lifetime smoking explained by the polygenic score.

Instrument validation

Data sources. We tested our genetic instrument using the positive controls of lung cancer, coronary artery disease (CAD) and hypomethylation at the aryl-hydrocarbon receptor repressor (*AHRR*) site cg05575921 because smoking has been robustly shown to predict these outcomes[1,42]. We conducted these analyses using GWAS summary data in a two-sample MR framework using our exposure instrument for lifetime smoking from our GWAS in the UK Biobank. For lung cancer, we used the summary data from the ILCCO consortium GWAS, which comprises 11,348 cases and 15,861 controls of European ancestry[43]. For CAD, we used the GWAS summary data from CARDIoGRAMplusC4D which comprises 60,801 cases and 123,504 controls of mixed ancestry[44]. We conducted a GWAS of *AHRR* (cg05575921) methylation in the Accessible Resource for Integrated Epigenomic Studies (ARIES) subset of ALSPAC[45]. Genome-wide DNA methylation profiling in ARIES was performed using the Illumina Infinium HumanMethylation450 BeadChip (450K) array for ~1000 mother-offspring pairs[45]. For this analysis, we used methylation data derived from whole blood which was collected from mothers (846 smokers and non-smokers) when the index offspring were around 18 years of age. Full details of the GWAS methods used for *AHRR* locus methylation are described elsewhere[46].

Two-sample Mendelian randomisation of positive controls. Analyses were conducted using MR Base, an R[47] package for two-sample MR using summary data[48]. We compared results across five different MR methods: inverse-variance weighted, MR Egger[49], weighted median[50], weighted mode[51] and MR RAPS[52]. Each method makes different assumptions and therefore a consistent effect across multiple methods strengthens causal evidence[53]. If a SNP was unavailable in the outcome GWAS

summary statistics, then proxy SNPs were searched for with a minimum LD $r^2 = 0.8$. We aligned palindromic SNPs with MAF<0.3.

Bidirectional two-sample Mendelian randomisation of mental health

Data sources. For lifetime smoking, we used the summary data from our GWAS of the lifetime smoking index in the UK Biobank. For schizophrenia, we used summary data from the PGC consortium GWAS, which comprises 36,989 cases and 113,075 controls of mixed ancestry[17]. For depression, we used the most recent GWAS summary data from the PGC for major depressive disorder (MDD), which comprises 130,664 MDD cases and 330,470 controls of European ancestry[54]. 23andMe data (75,607 cases and 231,747 controls) were excluded when MDD was the outcome because genome-wide summary statistics are not available with 23andMe data included.

Genetic instruments. For lifetime smoking, we identified 124 independent loci at genome-wide significance. The schizophrenia PGC identified 114 independent SNPs at 108 loci explaining around 3.4% of the variance in schizophrenia liability[17]. The genetic instrument for MDD from the PGC GWAS was 40 genome-wide significant SNPs which explain 1.9% of the variance in liability[54]. Weighting for MDD instrument comes from summary data of the full sample including 23andMe.

Two-sample Mendelian randomisation of mental health outcomes. Analyses were run using the MR Base R Package[47,48] and compared across the five different methods. This time, analyses of GWAS summary data were run bi-directionally, with lifetime smoking first as the exposure and then as the outcome. Steiger filtering was used to provide evidence of the direction of effect[55]. All three genetic instruments were clumped at LD $r^2 = 0.001$ and a distance of 10000kb. Again, if a SNP from the instrument was unavailable in the outcome, an attempt to find proxies was made with a minimum LD $r^2 = 0.8$ and palindromic SNPs were aligned with MAF<0.3.

Results

Instrument discovery for lifetime smoking index

The results of our GWAS of lifetime smoking (N = 463,033) are presented in Figure 1 and Supplementary Figure S1. The most strongly associated regions on chromosome 15 and chromosome 9 have been previously shown to be associated with smoking heaviness and cessation respectively[56]. We identified 10,398 SNPs at the genome-wide level of significance ($P < 5 \times 10^{-8}$). After clumping and filtering, 124 independent SNPs remained. A full list of these SNPs and their z-scored effect sizes can be found in Supplementary Table S1. In the ALSPAC independent sample, the 124 SNPs explained 0.31% of the variance in lifetime smoking (P = 0.003).

Instrument validation

Two-sample Mendelian randomisation of positive controls. We validated the genetic instrument for lifetime smoking exposure using two-sample MR of smoking on positive control outcomes: lung cancer, CAD and hypomethylation at the *AHRR* locus. All five MR methods indicated the expected direction of effect (see Table 1) increasing risk of both disease outcomes and decreasing *AHRR* methylation, with the exception of the MR Egger and MR Egger SIMEX adjusted estimates for coronary artery disease. However, these should be interpreted with caution given the low I^2_{GX} (see Supplementary Table S2)[57]. There was strong evidence of an effect of lifetime smoking on increased odds of lung cancer and CAD. There was weaker evidence of an effect on AHRR methylation, possibly due to lower power. Sensitivity analyses are presented in Supplementary Figures S2-S7. Tests of MR Egger intercepts generally indicated weak evidence of directional pleiotropy (see Supplementary Table S3), with the exception of CAD.

Bidirectional two-sample Mendelian randomisation of mental health

Bi-directional MR analyses provided strong evidence that increased lifetime smoking increases risk of both schizophrenia (OR = 2.33, 95% CI = 1.71 - 3.18, $P = 8.71 \times 10^{-08}$) and depression (OR = 2.03, 95% CI = 1.74 - 2.37, $P = 3.77 \times 10^{-19}$) (see Table 2), with consistent direction of effect across all five MR methods. Again, MR Egger results are the least reliable due to low I^2_{GX} (see

Supplementary Table S2). There was also evidence of a consistent but smaller causal effect of higher genetic risk for schizophrenia on lifetime smoking (β = 0.021, 95% CI = 0.005 - 0.037, P = 0.009) and of genetic risk for depression on lifetime smoking (β= 0.098, 95% CI = 0.037 - 0.160, P = 0.002) (see Table 2), although the evidence was weaker than in the reverse direction. The I^2_{GX} value for schizophrenia and depression was less than 0.6, therefore MR Egger was not conducted (see Supplementary Table S2). MR Egger intercepts suggest directional pleiotropy is not biasing the estimates (see Supplementary Table S3) and Steiger filtering supported the conclusion that the effects operate in both directions (see Supplementary Table S3). A total of 14,260 cases and 15,480 controls from the UK Biobank were included in the latest GWAS for MDD[54] which could lead to bias from sample overlap[58]. Therefore, this analysis was repeated using summary data from an earlier GWAS of MDD[59], which showed a consistent direction of effect with slightly weaker statistical evidence, possibly due to reduced sample size (N = 18,759) (see Supplementary Table S4). Further sensitivity tests were conducted and are presented in Supplementary Figures S9-S20.

Discussion

We conducted a GWAS of lifetime smoking exposure which provides a novel genetic instrument that can be used in two-sample MR of summary data without the need to stratify on smoking status. Our validation analyses confirm that the lifetime smoking instrument predicts increased risk of lung cancer, CAD and hypo-methylation at the *AHRR* locus. Of the 124 independent loci identified, many associations are in novel regions, but we also replicated associations with the *CHRNA5* gene on chromosome 15 previously associated with smoking heaviness and a region near the *DBH* gene on chromosome 9 previously associated with smoking cessation[56]. Notably, the weighting of these regions is different when looking at overall lifetime smoking behaviour rather than individual smoking behaviour phenotypes.

We used this novel genetic instrument to explore causal pathways between smoking and both schizophrenia and depression. The two-sample MR results provide evidence that smoking is a

causal risk factor for both schizophrenia and depression. This supports prospective observational evidence controlling for genetic confounding[23,24], as well as meta-analyses of observational studies[25,26] (although it should be noted that these meta-analyses include unadjusted estimates). Some studies which adjust for potential confounders find the effect of smoking attenuates to the null[60] or even becomes protective[61], demonstrating that there are likely to be substantial confounding effects in observational studies. Previous MR studies have not found clear evidence to support smoking as a causal risk factor for neither schizophrenia nor depression[30–33], but our approach offers greater power and better captures lifetime smoking rather than simply individual components of smoking behaviour. Previous studies of smoking have either used an initiation phenotype (not taking into account subsequent heaviness of use or cessation), or a heaviness of smoking phenotype (which only examines the dose-response relationship between smoking and the outcome of interest, and requires stratification on smoking status). Our approach overcomes these limitations, allowing two-sample MR using summary data in unstratified samples. However, it is not possible to precisely estimate from our results what proportion of the observational association between smoking, schizophrenia and depression is causal, or the population attributional fraction of these disorders due to smoking.

In support of the self-medication hypothesis[12], we found evidence to suggest that genetic risk for schizophrenia and depression increases lifetime smoking. This supports previous observational evidence[13–15] and might explain why smoking rates remain so high amongst individuals with schizophrenia and depression compared with the general population[62]. However, the evidence for liability to mental illness causing lifetime smoking is weaker than the evidence for smoking as a causal risk factor for mental illness. This could be due to self-medication only affecting one component of smoking behaviour (e.g., heaviness of smoking in current smokers), rather than overall lifetime smoking. Additionally, the effects might be weaker because MR methods typically capture the long-term effects of exposures. Clinical observations suggest that smoking increases when symptoms are exacerbated among patients with schizophrenia, which is typically interpreted

as evidence of self-medication[63]. However, the mechanisms underlying acute effects of psychiatric symptoms on short-term smoking behaviour may differ from the chronic effects of smoking on mental health, with cigarettes having positive effects on symptoms only in the short-term. This highlights the need to differentiate between acute and chronic exposures, and short- and long-term effects. Different methods may be required to examine these, with MR generally being better suited to interrogating chronic exposures and long-term effects.

Bi-directional effects of lifetime smoking with schizophrenia and depression might be explained by the misattribution hypothesis. This proposes that smokers misattribute the ability of cigarettes to relieve withdrawal, to their ability to relieve symptoms of psychological distress[64,65]. For example, withdrawal symptoms include depressed mood, anxiety and irritability, and smoking a cigarette alleviates those symptoms[66]. Since many withdrawal symptoms are similar to the negative symptoms of schizophrenia and mood symptoms in depression, their alleviation by smoking could give rise to the strong belief that smoking helps to alleviate mental health symptoms. This, in turn, might serve to maintain smoking among individuals with schizophrenia and depression[67]. Given that the symptoms of withdrawal are most similar to the symptoms of depression, this might explain why we see stronger effects of depression risk on smoking than schizophrenia risk on smoking.

A potential biological mechanism for the bi-directional causal effects of smoking, schizophrenia and depression could be neuroadaptations in the dopaminergic and serotonin systems. Nicotine acts on nicotinic cholinergic receptors in the brain stimulating the release of neurotransmitters including dopamine and serotonin[67]. Dopamine and serotonin dysfunction have been implicated in the aetiology of schizophrenia and depression respectively[68,69]. It is plausible, therefore, that disruption of these pathways has a causal effect on these disorders. Alternatively, cannabis use could be a mediating mechanism for the effects of smoking on schizophrenia and depression. In prospective studies, cigarette smoking has been shown to increase risk of cannabis dependence even after adjusting for cannabis use at baseline[70]. There is strong evidence suggesting that cannabis use increases the risk of psychosis and affective disorders, with the strongest evidence

for an effect on psychosis[71]. This vertical pleiotropy does not violate the assumptions of MR, but simply means there are intermediate mechanisms operating between the exposure and the outcome. However, the strength of the effects of lifetime smoking that we observe suggest that any mediating influence of cannabis use is likely to only partially account for these effects, given the relatively low prevalence of cannabis use (e.g., annual prevalence in the UK of ~7% in 2010[72]). Multivariable MR analysis of tobacco and cannabis use would help resolve this question.

Future work should attempt to elucidate the underlying mechanisms with a hope to intervene, inform public health messages or further advance our knowledge on the aetiology of mental illness. In particular it will be important to consider other constituents of tobacco smoke to determine whether it is exposure to nicotine or some other constituent that increases risk of schizophrenia and depression. This is particularly important in the context of the recent growth in electronic cigarette use.

Strengths and Limitations

Our study is the first to generate a genetic instrument for lifetime smoking behaviour in a large sample, which allows the use of two-sample MR with summary data from unstratified samples to answer questions about the association between smoking and other health outcomes. However, there are some limitations which should be noted. First, there is evidence to suggest that even after seemingly controlling for population structure in GWAS of samples as large as the UK Biobank, coincident apparent structure remains[73]. This might confound the association between smoking and mental health, increasing the risk of false positives. As independent samples with adequate sample size become available, the influence of structure should be further explored. However, it is reassuring that our instrument predicted lifetime smoking in the ALSPAC replication sample, where such issues would not arise in the same manner. Second, sample overlap in two-sample MR can bias results towards the observational estimate[58]. There was some sample overlap between the MDD GWAS[54] and the UK Biobank (used to derive the lifetime smoking instrument) meaning that the

effects could be inflated. Therefore, a sensitivity analysis was conducted using a previous GWAS of MDD[59] which showed the same direction of effect despite lower power. This gives us confidence in the bi-directional effects of smoking and depression, despite sample overlap. Third, including multiple aspects of smoking behaviour could have introduced more potential for horizontal pleiotropy. The more diffuse the definition of smoking, the more lifestyle factors might be correlated, making it especially important to check for horizontal pleiotropy. We conducted multiple MR sensitivity analyses (which all make different and largely uncorrelated assumptions), and all demonstrated the same direction of effect. This increases our confidence that the results are not biased by pleiotropy. Furthermore, MR Egger intercepts did not show evidence of directional pleiotropy for schizophrenia or depression. However, further work should still attempt to understand the biological mechanisms underpinning the association in order to reduce the likelihood of pleiotropic effects. Finally, there is known selection bias in the UK Biobank sample, with participants being more highly educated, less likely to be a smoker and overall healthier than the general UK population[74]. Of the 9 million individuals contacted, only ~5% consented to take part[74]. Due to the lack of representativeness in the UK Biobank sample, disease prevalence and incidence rates will not reflect underlying population levels and there is potential for collider bias. If both smoking and risk for schizophrenia and depression reduce the likelihood of participating in the UK Biobank, then this would induce a negative correlation between schizophrenia or depression and smoking. That is the opposite of the effects we observe, suggesting our estimates may, if anything, be conservative. Relatedly, we adjusted for genotyping chip in our analyses, but the UK BiLEVE sub-sample (for which a different genotyping chip was used) were selected on the basis of smoking status, so this might induce collider bias. However, this bias would have to be substantial to generate spurious associations of the magnitude we observe.

Conclusion

In conclusion, we have developed a novel genetic instrument of lifetime smoking behaviour which can be used in two-sample MR of summary data without stratifying on smoking status or reducing power. In two-sample MR analysis, there was evidence of an association between smoking and mental health in both directions; however, the evidence was stronger for a causal effect of smoking on risk of schizophrenia and depression. This emphasises the detrimental public health consequences of smoking, especially for mental health, and the need to reduce smoking prevalence not only to reduce the burden of physical illness, but also to reduce the burden of mental illness.

Sources of Support

REW, RCR, RBL, HMS, SZ, GDS and MRM are all members of the MRC Integrative Epidemiology Unit at the University of Bristol funded by the MRC: http://www.mrc.ac.uk [MC_UU_00011/1, MC_UU_00011/5, MC_UU_00011/7]. This study was supported by the NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health and Social Care. GMJT is funded by Cancer Research UK Population Researcher Postdoctoral Fellowship award (reference: C56067/A21330). This work was also supported by CRUK (grant number C18281/A19169) and the ESRC (grant number ES/N000498/1). The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. The Accessible Resource for Integrated Epigenomics Studies (ARIES) was funded by the UK Biotechnology and Biological Sciences Research Council (BB/I025751/1 and BB/I025263/1). This research has been conducted using the UK Biobank Resource under Application Number 9142.

Conflict of Interest declaration.

REW declares no conflicts of interest on behalf of all co-authors.

Acknowledgements

We are grateful to the participants of the UK Biobank, the Avon Longitudinal Study of Parents and Children and the individuals who contributed to each of the previous GWAS analyses conducted as well as all the research staff who worked on the data collection.

Data and code availability

GWAS summary data for lifetime smoking will be made available following publication.

Author contribution statement

Contributors MRM conceived the study. REW conducted the analysis with statistical assistance from BGS and theoretical assistance from HMS, RBL and MRM. RCR conducted the two-sample MR analysis of lifetime smoking on methylation and performed sensitivity analyses. REW and MRM drafted the initial manuscript. AET provided access to the UK Biobank data. All authors guided the analysis, assisted with interpretation, commented on drafts of the manuscript and approved the final version.

References

1. World Health Organization. *WHO report on the global tobacco epidemic*. (World Health Organization, 2011).
2. Coulthard, M., Farrell, M., Singleton, N. & Meltzer, H. Tobacco, alcohol and drug use and mental health. *Lond. Station. Off.* (2002).
3. Lasser, K. *et al.* Smoking and mental illness: A population-based prevalence study. *JAMA* **284,** 2606–2610 (2000).
4. Lawrence, D., Mitrou, F. & Zubrick, S. R. Smoking and mental illness: results from population surveys in Australia and the United States. *BMC Public Health* **9,** 285 (2009).

5. McClave, A. K., McKnight-Eily, L. R., Davis, S. P. & Dube, S. R. Smoking characteristics of adults with selected lifetime mental illnesses: results from the 2007 National Health Interview Survey. *Am. J. Public Health* **100,** 2464–2472 (2010).
6. Royal College of Physicians. *Royal College of Physicians, Royal College of Psychiatrists. Smoking and mental health.* (2013).
7. Byers, A. L. *et al.* Twenty-year depressive trajectories among older women. *Arch. Gen. Psychiatry* **69,** 1073–1079 (2012).
8. Leung, J., Gartner, C., Dobson, A., Lucke, J. & Hall, W. Psychological distress is associated with tobacco smoking and quitting behaviour in the Australian population: evidence from national cross-sectional surveys. *Aust. N. Z. J. Psychiatry* **45,** 170–178 (2011).
9. Tjora, T. *et al.* The association between smoking and depression from adolescence to adulthood. *Addiction* **109,** 1022–1030 (2014).
10. Adult smoking habits in the UK - Office for National Statistics. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/bulletins/adultsmokinghabitsingreatbritain/2016#main-points. (Accessed: 9th July 2018)
11. Chang, C.-K. *et al.* Life Expectancy at Birth for People with Serious Mental Illness and Other Major Disorders from a Secondary Mental Health Care Case Register in London. *PLOS ONE* **6,** e19590 (2011).
12. Khantzian, E. The self-medication hypothesis of substance use disorders: a reconsideration and recent applications. *Harv. Rev. Psychiatry* **4,** 231–244 (1997).
13. Desai, H. D., Seabolt, J. & Jann, M. W. Smoking in patients receiving psychotropic medications. *CNS Drugs* **15,** 469–494 (2001).

14. Levin, E. D., Wilson, W., Rose, J. E. & McEvoy, J. Nicotine–haloperidol interactions and cognitive performance in schizophrenics. *Neuropsychopharmacology* **15,** 429 (1996).
15. Lerman, C. *et al.* Investigation of mechanisms linking depressed mood to nicotine dependence. *Addict. Behav.* **21,** 9–19 (1996).
16. Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N. & Davey Smith, G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27,** 1133–1163 (2008).
17. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,** 421–427 (2014).
18. Munafò, M. R. *et al.* Association Between Genetic Variants on Chromosome 15q25 Locus and Objective Measures of Tobacco Exposure. *JNCI J. Natl. Cancer Inst.* **104,** 740–748 (2012).
19. Thorgeirsson, T. E. *et al.* A Variant Associated with Nicotine Dependence, Lung Cancer and Peripheral Arterial Disease. *Nature* **452,** 638–642 (2008).
20. Tobacco Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* **42,** 441–447 (2010).
21. Ware, J. J., van den Bree, M. B. & Munafò, M. R. Association of the CHRNA5-A3-B4 gene cluster with heaviness of smoking: a meta-analysis. *Nicotine Tob. Res.* **13,** 1167–1175 (2011).
22. Gage, S. H. & Munafò, M. R. Rethinking the association between smoking and schizophrenia. *Lancet Psychiatry* **2,** 118 (2015).
23. Kendler, K. S., Lönn, S. L., Sundquist, J. & Sundquist, K. Smoking and schizophrenia in population cohorts of Swedish women and men: a prospective co-relative control study. *Am. J. Psychiatry* **172,** 1092–1100 (2015).

24. Kendler, K. S. *et al.* Smoking and major depression: a causal analysis. *Arch. Gen. Psychiatry* **50,** 36–43 (1993).

25. Gurillo, P., Jauhar, S., Murray, R. M. & MacCabe, J. H. Does tobacco use cause psychosis? Systematic review and meta-analysis. *Lancet Psychiatry* **2,** 718–725 (2015).

26. Taylor, G. *et al.* Change in mental health after smoking cessation: systematic review and meta-analysis. *BMJ* **348,** g1151 (2014).

27. Munafò, M. R. & Smith, G. D. Robust research needs many lines of evidence. *Nature* (2018). doi:10.1038/d41586-018-01023-3

28. Davey Smith, G. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **32,** 1–22 (2003).

29. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23,** R89–R98 (2014).

30. Bjørngaard, J. H. *et al.* The causal role of smoking in anxiety and depression: a Mendelian randomization analysis of the HUNT study. *Psychol. Med.* **43,** 711–719 (2013).

31. Taylor, A. E. *et al.* Investigating the possible causal association of smoking with depression and anxiety using Mendelian randomisation meta-analysis: the CARTA consortium. *BMJ Open* **4,** e006141 (2014).

32. Wium-Andersen, M. K., Ørsted, D. D. & Nordestgaard, B. G. Tobacco smoking is causally associated with antipsychotic medication use and schizophrenia, but not with antidepressant medication use or depression. *Int. J. Epidemiol.* **44,** 566–577 (2015).

33. Gage, S. H. *et al.* Investigating causality in associations between smoking initiation and schizophrenia using Mendelian randomization. *Sci. Rep.* **7,** 40653 (2017).

34. Leffondré, K., Abrahamowicz, M., Xiao, Y. & Siemiatycki, J. Modelling smoking history using a comprehensive smoking index: application to lung cancer. *Stat. Med.* **25,** 4132–4146 (2006).

35. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* 166298 (2017). doi:10.1101/166298

36. Mitchell, R., Hemani, G., Dudding, T. & Paternoster, L. UK Biobank Genetic Data: MRC-IEU Quality Control, Version 1. *data.bris* (2017). doi:10.5523/bris.3074krb6t2frj29yh2b03x3wxj

37. Elsworth, B. L. *et al.* MRC IEU UK Biobank GWAS pipeline version 1. *data.bris* (2017). doi:10.5523/bris.2fahpksont1zi26xosyamqo8rr

38. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47,** 284–290 (2015).

39. Boyd, A. *et al.* Cohort profile: the 'children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int. J. Epidemiol.* **42,** 111–127 (2013).

40. Fraser, A. *et al.* Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int. J. Epidemiol.* **42,** 97–110 (2012).

41. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).

42. Zeilinger, S. *et al.* Tobacco Smoking Leads to Extensive Genome-Wide Changes in DNA Methylation. *PLOS ONE* **8,** e63812 (2013).

43. Wang, Y. *et al.* Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. *Nat. Genet.* **46,** 736 (2014).

44. CARDIoGRAMplusC4D Consortium & others. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47,** 1121–1149 (2015).

45. Relton, C. L. *et al.* Data resource profile: accessible resource for integrated epigenomic studies (ARIES). *Int. J. Epidemiol.* **44,** 1181–1190 (2015).

46. Taylor, A. *et al.* The effect of body mass index on smoking behaviour and nicotine metabolism: a Mendelian randomization study. *bioRxiv* 299834 (2018).

47. R. Core Team. *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013*. (ISBN 3-900051-07-0, 2014).

48. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7,** e34408 (2018).

49. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44,** 512–525 (2015).

50. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40,** 304–314 (2016).

51. Hartwig, F. P., Smith, G. D. & Bowden, J. Robust inference in two-sample Mendelian randomisation via the zero modal pleiotropy assumption. *bioRxiv* 126102 (2017).

52. Zhao, Q., Wang, J., Hemani, G., Bowden, J. & Small, D. S. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *arXiv* (2018).

53. Lawlor, D. A., Tilling, K. & Davey Smith, G. Triangulation in aetiological epidemiology. *Int. J. Epidemiol.* **45,** 1866–1886 (2016).

54. Wray, N. R., Sullivan, P. F. & others. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50,** 668–681 (2018).

55. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLOS Genet.* **13,** e1007081 (2017).

56. Furberg, H. *et al.* Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* **42,** 441 (2010).

57. Bowden, J. *et al.* Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I 2 statistic. *Int. J. Epidemiol.* **45,** 1961–1974 (2016).

58. Burgess, S. *et al.* Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur. J. Epidemiol.* **30,** 543–552 (2015).

59. Ripke, S. *et al.* A mega-analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry* **18,** 497–511 (2013).

60. Jones, H. J. *et al.* Association of Combined Patterns of Tobacco and Cannabis Use in Adolescence With Psychotic Experiences. *JAMA Psychiatry* **75,** 240–246 (2018).

61. Zammit, S. *et al.* Investigating the association between cigarette smoking and schizophrenia in a cohort study. *Am. J. Psychiatry* **160,** 2216–2221 (2003).

62. Cook, B. L. *et al.* Trends in Smoking Among Adults With Mental Illness and Association Between Mental Health Treatment and Smoking Cessation. *JAMA* **311,** 172–182 (2014).

63. Jarvik, M. E. *et al.* Nicotine blood levels and subjective craving for cigarettes. *Pharmacol. Biochem. Behav.* **66,** 553–558 (2000).

64. Parrott, A. C. Does cigarette smoking cause stress? *Am. Psychol.* **54,** 817 (1999).

65. Parrott, A. C. Cigarette-derived nicotine is not a medicine. *World J. Biol. Psychiatry* **4,** 49–55 (2003).

66. Hughes, J. R. Effects of abstinence from tobacco: valid symptoms and time course. *Nicotine Tob. Res.* **9,** 315–327 (2007).

67. Benowitz, N. L. Nicotine Addiction. *N. Engl. J. Med.* **362,** 2295–2303 (2010).

68. Howes, O., McCutcheon, R. & Stone, J. Glutamate and dopamine in schizophrenia: An update for the 21st century. *J. Psychopharmacol. (Oxf.)* **29,** 97–115 (2015).

69. Jakubovski, E., Varigonda, A. L., Freemantle, N., Taylor, M. J. & Bloch, M. H. Systematic Review and Meta-Analysis: Dose-Response Relationship of Selective Serotonin Reuptake Inhibitors in Major Depressive Disorder. *Am. J. Psychiatry* **173,** 174–183 (2015).

70. Hindocha, C., Shaban, N. D., Freeman, T. P., Das, R. K., Gale, G., Schafer, G., ... & Curran, H. V. Associations between cigarette smoking and cannabis dependence: a longitudinal study of young cannabis users in the United Kingdom. *Drug and alcohol dependence,* **148,** 165-171 (2015).

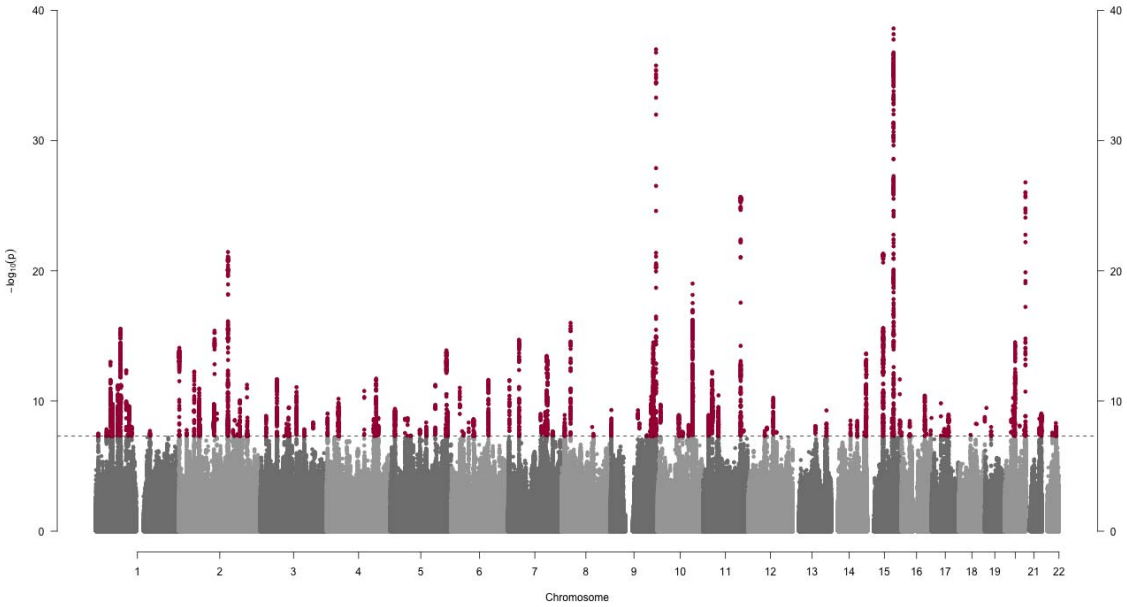
71. Moore, T. H., Zammit, S., Lingford-Hughes, A., Barnes, T. R., Jones, P. B., Burke, M., & Lewis, G. Cannabis use and risk of psychotic or affective mental health outcomes: a systematic review. *The Lancet,* **370(9584),** 319-328 (2007).

72. UNODC. World Drug Report 2011. Vienna: United Nations Office on Drugs and Crime, (2011).

73. Haworth, S. *et al.* Common genetic variants and health outcomes appear geographically structured in the UK Biobank sample: Old concerns returning and their implications. *bioRxiv* 294876 (2018). doi:10.1101/294876

74. Munafò, M. R., Tilling, K., Taylor, A. E., Evans, D. M. & Davey Smith, G. Collider scope: when selection bias can substantially influence observed associations. *Int. J. Epidemiol.* **47,** 226–235 (2018).

Figure 1. Manhattan plot of genome-wide association study lifetime smoking index (N = 337,106).



The x-axis represents chromosomal position and the y-axis represents $-\log_{10}$ P-value for the association of each single nucleotide polymorphism (SNP) with the lifetime smoking index using an additive model and linear regression. The dashed line indicates the genome-wide level of significance ($P < 5 \times 10^{-8}$) and genome-wide significant SNPs are indicated in red.

Table 1. Two-sample Mendelian randomisation analyses of the effect of lifetime smoking on coronary artery disease, lung cancer and *AHRR* Methylation.

Outcome	Method	OR (95% CI)	P-value
Coronary Artery Disease	Inverse-Variance Weighted	1.57 (1.35, 1.83)	7.42×10^{-09}
	MR Egger	0.69 (0.37, 1.29)	0.25
	MR Egger (SIMEX)	0.76 (0.44, 1.31)	0.33
	Weighted Median	1.64 (1.35, 2.00)	6.04×10^{-07}
	Weighted Mode	1.76 (1.03, 3.02)	0.04
	MR RAPS	1.64 (1.40, 1.91)	4.07×10^{-10}
Lung Cancer	Inverse-Variance Weighted	4.16 (2.94, 5.90)	9.44×10^{-16}
	MR Egger	18.05 (4.23, 77.07)	1.58×10^{-04}
	MR Egger (SIMEX)	17.60 (4.43, 69.98)	8.54×10^{-05}
	Weighted Median	2.79 (1.93, 4.04)	4.97×10^{-08}
	Weighted Mode	6.42 (2.14, 19.23)	0.001
	MR RAPS	3.63 (2.69, 4.90)	2.84×10^{-17}
		Beta (95% CI)	**P-value**
***AHRR* Methylation**	Inverse-Variance Weighted	-0.085 (-0.154, -0.016)	0.02
	MR Egger	-0.024 (-0.302, 0.254)	0.87
	MR Egger (SIMEX)	-0.194 (-0.456, 0.068)	0.15
	Weighted Median	-0.122 (-0.225, -0.018)	0.02
	Weighted Mode	-0.207 (-0.503, 0.088)	0.17
	MR RAPS	-0.084 (-0.159, -0.010)	0.03

Of the 124 genome-wide significant SNPs associated with the lifetime smoking index, 124 were available from the GWAS of coronary artery disease (Nikpay et al., 2015), 118 from the GWAS of lung cancer (Wang et al., 2014) and 118 from our GWAS of *AHRR* locus methylation. SIMEX-corrected estimates are unweighted. Due to low regression dilution I^2_{GX} (see Supplementary Table S2), MR Egger and MR Egger SIMEX estimates should be interpreted with caution. SIMEX = simulation extrapolation, MR RAPS = robust adjusted profile score.

Table 2. Bidirectional two-sample Mendelian randomisation analyses of the effect of lifetime smoking on schizophrenia and major depression.

Exposure	Outcome	Method	OR (95% CI)	P-value
Smoking	**Schizophrenia**	Inverse-Variance Weighted	2.33 (1.71, 3.18)	8.71×10^{-08}
		MR Egger	3.46 (0.97, 12.34)	0.06
		MR Egger (SIMEX)	5.33 (1.80, 15.83)	0.003
		Weighted median	2.08 (1.59, 2.72)	1.12×10^{-07}
		Weighted mode	1.77 (0.73, 4.27)	0.21
		MR RAPS	2.57 (1.92, 3.43)	2.30×10^{-10}
Smoking	**Depression**	Inverse-Variance Weighted	2.03 (1.74, 2.37)	3.77×10^{-19}
		MR Egger	1.26 (0.68, 2.33)	0.47
		MR Egger (SIMEX)	14.25 (0.02, 12554.04)	0.45
		Weighted median	2.02 (1.69, 2.40)	2.34×10^{-15}
		Weighted mode	1.96 (1.24, 3.09)	0.005
		MR RAPS	2.02 (1.73, 2.36)	8.25×10^{-19}
			Beta (95% CI)	**P-value**
Schizophrenia	**Smoking**	Inverse-Variance Weighted	0.021 (0.005, 0.037)	0.009
		MR Egger	-	-
		MR Egger (SIMEX)	-	-
		Weighted median	0.015 (0.004, 0.026)	0.007
		Weighted Mode	0.016 (-0.011, 0.044)	0.24
		MR RAPS	0.017 (0.003, 0.031)	0.003
Depression	**Smoking**	Inverse-Variance Weighted	0.098 (0.037, 0.160)	0.002
		MR Egger	-	-
		MR Egger (SIMEX)	-	-
		Weighted median	0.115 (0.073, 0.158)	9.62×10^{-08}
		Weighted mode	0.110 (0.041, 0.180)	0.004
		MR RAPS	0.088 (0.027, 0.149)	0.005

Of the 124 genome-wide significant SNPs associated with lifetime smoking index, 123 were available in the GWAS of schizophrenia (Ripke et al, 2014) and 124 were available in the GWAS of major depression (Wray et al., 2018). SIMEX-corrected estimates are unweighted. Of the 114 SNPs associated with schizophrenia, 102 were available in the lifetime smoking index GWAS. Of the 40 SNPs associated with major depression, 36 were available in the GWAS of lifetime smoking. MR Egger regression was not conducted for schizophrenia or major depression as exposures because regression dilution I^2_{GX} was below

0.3 (see Supplementary Table S4). Due to low regression dilution I^2_{GX} for lifetime smoking as the exposure (see Supplementary Table S4), MR Egger and MR Egger SIMEX estimates should be interpreted with caution. SIMEX = simulation extrapolation, MR RAPS = robust adjusted profile score.