Investigating the possible causal association of smoking with depression and anxiety using Mendelian randomisation meta-analysis: the CARTA consortium


ABSTRACT

Objectives: To investigate whether associations of smoking with depression and anxiety are likely to be causal, using a Mendelian randomisation approach.

Design: Mendelian randomisation meta-analyses using a genetic variant (rs16969968/rs1051730) as a proxy for smoking heaviness, and observational meta-analyses of the associations of smoking status and smoking heaviness with depression, anxiety and psychological distress.

Participants: Current, former and never smokers of European ancestry aged ≥16 years from 25 studies in the Consortium for Causal Analysis Research in Tobacco and Alcohol (CARTA).

Primary outcome measures: Binary definitions of depression, anxiety and psychological distress assessed by clinical interview, symptom scales or self-reported recall of clinician diagnosis.

Results: The analytic sample included up to 58,176 never smokers, 37,428 former smokers and 32,028 current smokers (total N=127,632). In observational analyses, current smokers had 1.85 times greater odds of depression (95% CI 1.65 to 2.07), 1.71 times greater odds of anxiety (95% CI 1.54 to 1.90) and 1.69 times greater odds of psychological distress (95% CI 1.56 to 1.83) than never smokers. Former smokers also had greater odds of depression, anxiety and psychological distress than never smokers. There was evidence for

Strengths and limitations of this study

- This is the largest Mendelian randomisation study of the relationship between smoking and depression and anxiety conducted to date.
- By using a genetic variant associated with smoking heaviness as a proxy for smoking heaviness, bias from confounding is minimised and findings not affected by reverse causality.
- Measurement of depression and anxiety differed across studies so we were unable to use a consistent definition.
- While results are consistent with no causal association between smoking heaviness and depression or anxiety, we cannot rule out the possibility of a small effect.

positive associations of smoking heaviness with depression, anxiety and psychological distress (ORs per cigarette per day: 1.03 (95% CI 1.02 to 1.04), 1.03 (95% CI 1.02 to 1.04) and 1.02 (95% CI 1.02 to 1.03) respectively). In Mendelian randomisation analyses, there was no strong evidence that the minor allele of rs1696968/rs1051730 was associated with depression (OR=1.00, 95% CI 0.95 to 1.05), anxiety (OR=1.02, 95% CI 0.97 to 1.07) or psychological distress (OR=1.02, 95% CI 0.98 to 1.06) in current smokers. Results were similar for former smokers.

**Conclusions:** Findings from Mendelian randomisation analyses do not support a causal role of smoking heaviness in the development of depression and anxiety.

---

**INTRODUCTION**

Smoking is highly comorbid with both depression and anxiety across many different populations.1-9 Furthermore, there is evidence to suggest that tobacco control interventions may not be as effective in populations with mental health conditions; for example, recent trends in the USA suggest that, since 2004, smoking rates have declined less rapidly in individuals with anxiety than in the general population.10 Given the profound public health burden of both tobacco-related disease,11 and depression and anxiety,12 understanding this relationship is of great importance.

Unfortunately, it is difficult to infer causal links between smoking and depression and anxiety from observational data, due to confounding. There may be factors associated with both smoking and depression and anxiety, such as other substance use (eg, alcohol), socioeconomic adversity and education which cannot be fully accounted for in observational studies.13 In addition, even if a causal association does exist, the direction of the relationship between smoking and depression and anxiety is unclear.14 Prospective studies have provided evidence that depressive symptoms are associated with increased likelihood of smoking initiation,2 7 15-17 while smoking cessation appears to be associated with a short-term increase in depressive symptoms during their quit attempt among a subgroup of smokers, and these individuals have poor smoking cessation outcomes.18 This evidence is consistent with the popular belief that cigarette smoking can reduce anxiety and improve mood, particularly among those experiencing anxiety or low mood (the self-medicating hypothesis). However, there is also a growing body of evidence suggesting that smoking may contribute to the development of these conditions19-21 and that smoking cessation is associated with improvements in mental health, including depression and anxiety, compared to continued smoking.22

Plausible biological mechanisms through which constituents of tobacco smoke may cause depression and anxiety have been described. In animal studies, for example, there is evidence that nicotine administration produces dysregulation in the hypothalamic-pituitary-adrenal system, which leads to hypersecretion of cortisol and changes in the activity of associated monoamine neurotransmitters.23 These systems function to regulate the biological and psychological reactions to stressors. Similarly, human data have demonstrated elevated cortisol levels in smokers compared to non-smokers.24 Constituents of tobacco smoke inhibit the activity of monoamine oxidase, enzymes that are involved in the breakdown of monoamines (including dopamine, serotonin and norepinephrine); this effect appears to normalise following smoking cessation.25 Animal studies also indicate that both drugs of abuse (including nicotine) and environmental stressors appear to trigger changes in midbrain dopaminergic function.26 Consequently, prolonged smoking may act to sensitise stress response systems, weakening adaptive coping responses and making smokers more susceptible to emotional distress in response to environmental stressors.

Mendelian randomisation methods allow us to investigate causal relationships in humans by using genetic variants as proxies for exposures of interest. The principle of Mendelian randomisation relies on the basic (but approximate) laws of Mendelian genetics (segregation and independent assortment). When these principles hold, genetic variants, at a population level, will not be associated with the confounding factors that generally distort conventional observational studies.27 28 In addition, as an outcome measure cannot alter the genotype that an individual is born with, these analyses should not be biased by reverse causality. A genetic variant, single nucleotide polymorphism (SNP) number rs1696968, in the CHRNA5-CHRHA3-CHRNB4 (CHRNA5-A3-B4) nicotinic receptor subunit gene cluster on chromosome 15 has demonstrated robust association with smoking heaviness within smokers.29-32 The rs16969698 variant is functional and leads to an amino acid change (D398N) in the nicotinic receptor α5 subunit protein.33 The minor (risk) allele of this variant is associated with an average increase in smoking amount of one cigarette per day in smokers, and even more strongly associated with increases in cotinine (a metabolite of nicotine) levels.34 However, given the known role of the variant in altering receptor function,33 it is likely that the greater variance explained for cotinine levels is due to this measure better capturing total tobacco exposure, and not because the variant directly affects nicotine metabolism.31 There is also good evidence that the rs1696968 variant, unlike smoking heaviness, does not associate with confounding factors that may distort associations with health outcomes, for example, socioeconomic status and education level.36 37

The rs16969698 variant (or its proxy rs1051730) has been used as an instrument for smoking heaviness in Mendelian randomisation studies to demonstrate that smoking causally lowers body mass index38 and that maternal smoking during pregnancy lowers offspring birth weight39 (see figure 1 for an illustration of the Mendelian randomisation approach). Using the rs1051730 variant, two recent studies have applied Mendelian randomisation to examine the causal...
Figure 1  Diagram of Mendelian randomisation analysis of smoking and depression/anxiety. The genetic variant rs16969968/rs1051730 is associated with smoking heaviness but should not be associated with the confounders of the association between smoking heaviness and depression/anxiety. In addition, there is no pathway from depression and anxiety to the genetic variant (reverse causality).

METHODS

Study populations

We used data on individuals aged ≥16 years and of self-reported European ancestry from 25 studies from the CARTA consortium: the 1958 Birth Cohort (1958BC), the Avon Longitudinal Study of Parents and Children (ALSPAC, including both mothers and children), the British Regional Heart Study (BRHS), the British Women’s Heart and Health Study (BWHS), the Caerphilly Prospective Study (CaPS), the Christchurch Health and Development Study (CHDS), Cohorte Lausannoise (CoLaus), the English Longitudinal Study of Ageing (ELSA), the National FINRISK Study (FINRISK), Generation Scotland: the Scottish Family Health Study (GS:SFHS), Genomics of Overweight Young Adults (GOYA) females, the Helsinki Birth Cohort Study (HBCS), Health2006, Health2008, the second wave of the Nord-Trondelag health study (HUNT 2), Inter99, the Northern Finland Birth Cohorts (NFBC1966 and NFBC1986), the National Health and Nutrition Examination Survey (NHANES), the MRC National Survey of Health and Development (NSHD), the Netherlands Twin Registry (NTR), Patch 2, the Rotterdam Study, the Saguenay Youth Study-Parents (SYS-P) and the Whitehall II study. Further details of these studies are provided in online supplementary material.

Genotype

Within each study, individuals were genotyped for one of two SNPs in the CHRNA5-A3-B4 nicotinic receptor subunit gene cluster, rs16969968 or rs1051730. These SNPs are in perfect linkage disequilibrium with each other in Europeans ($R^2=1.00$ in HapMap 3, http://hapmap.ncbi.nlm.nih.gov/) and therefore represent the same genetic signal. Where studies had data available for both SNPs, we used the SNP that was genotyped in the largest number of individuals. Details of genotyping methods within each study are provided in online supplementary material.

Measures of depression, anxiety and psychological distress

Depression and anxiety were assessed by clinical interview, symptom scales or self-reported recall of clinician diagnosis (see table 1). As some of the scales do not distinguish between symptoms of depression and anxiety, we used the term ‘psychological distress’ to refer to a composite phenotype.

To compare measures across studies, we created two case definitions for each of depression, anxiety and psychological distress (see table 2). According to case definition 1, individuals were classified as depressed or anxious if they self-reported clinician diagnosis of depression or anxiety, met clinical criteria for depression (excluding bereavement where known) or anxiety, or were above previously published cut points for depression or anxiety on symptom scales. Individuals were classified as having psychological distress if they met case definition 1 for depression or anxiety, or if they were above a cut point on a general scale for psychiatric symptoms. As not all scale measures used for assessing mental health have published cut-offs for defining cases, we created a second definition. According to case definition 2, individuals were classified as depressed or anxious if they were above the 90th centile for the specific depression or anxiety scales, and psychologically distressed if they were above the 90th centile on either the depression or anxiety scales or above the 90th centile on the general scales of psychiatric symptoms. Where both case definitions 1 and 2 were available within a study, case definition 1 was used. Full details of the measures and cut points used are provided in online supplementary table S1.

For the majority of studies (k=17), diagnoses were based on current depression and anxiety (at the time of measurement). Where current diagnoses were not available, diagnoses of depression or anxiety in the previous 12 months or lifetime diagnoses were used. For lifetime diagnoses, if information on age at first diagnosis was collected, individuals reporting diagnoses prior to 16 years of age were excluded.
Symptom scales were also used as continuous measures of depression, anxiety and psychological distress (see online supplementary table S2). To compare across studies, these were converted to z-scores within each study. Most measures of depression, anxiety and psychological distress were strongly right skewed. However, standard transformations (eg, log, square root) did not greatly improve distributions in several of the samples. Therefore, z-scores were constructed using the untransformed data in all samples (z-score=(individual score−sample mean)/sample SD).

Table 1  Measures of depression, anxiety and psychological distress in the CARTA studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Psychological Distress</th>
<th>Depression</th>
<th>Anxiety</th>
</tr>
</thead>
<tbody>
<tr>
<td>1958BC</td>
<td>CIS-R</td>
<td>CIS-R</td>
<td>CIS-R</td>
</tr>
<tr>
<td>ALSPAC children</td>
<td>CIS-R</td>
<td>CIS-R</td>
<td>CIS-R</td>
</tr>
<tr>
<td>ALSPAC mothers</td>
<td>EPDS or CCEI</td>
<td>EPDS</td>
<td>CCEI</td>
</tr>
<tr>
<td>BRHS</td>
<td></td>
<td>Clinician diagnosis (lifetime)</td>
<td></td>
</tr>
<tr>
<td>BWHHS</td>
<td></td>
<td>Clinician diagnosis (lifetime)</td>
<td></td>
</tr>
<tr>
<td>CapPS</td>
<td>GHQ-30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHDS</td>
<td>CIDI (previous 12 months)</td>
<td>CIDI (previous 12 months)</td>
<td>STAI</td>
</tr>
<tr>
<td>CoLaus</td>
<td>DIGS</td>
<td>DIGS</td>
<td>DIGS</td>
</tr>
<tr>
<td>ELSA</td>
<td>CES-D or clinician diagnosis of anxiety</td>
<td>CES-D (8-item)</td>
<td>Clinician diagnosis (lifetime)</td>
</tr>
<tr>
<td>FINRISK</td>
<td></td>
<td>Clinician diagnosis (previous 12 months)</td>
<td></td>
</tr>
<tr>
<td>Generation Scotland</td>
<td>GHQ-28</td>
<td>SCIDI/NP (Lifetime diagnosis)</td>
<td></td>
</tr>
<tr>
<td>GOYA females</td>
<td>Clinician diagnosis (since giving birth)</td>
<td>Clinician diagnosis (since giving birth)</td>
<td>Clinician diagnosis (since giving birth)</td>
</tr>
<tr>
<td>HBCS</td>
<td>CES-D or STAI</td>
<td>CES-D (20 items)</td>
<td>STAI</td>
</tr>
<tr>
<td>Health2006</td>
<td>SCL-90-R</td>
<td>SCL-90-R</td>
<td>SCL-90-R</td>
</tr>
<tr>
<td>Health2008</td>
<td>SCL-90-R</td>
<td>SCL-90-R</td>
<td>SCL-90-R</td>
</tr>
<tr>
<td>HUNT</td>
<td>HADS</td>
<td>HADS</td>
<td>HADS</td>
</tr>
<tr>
<td>Inter99</td>
<td>SCL-90-R</td>
<td>SCL-90-R</td>
<td>SCL-90-R</td>
</tr>
<tr>
<td>NFBC1966</td>
<td>SCL-25</td>
<td>SCL-25</td>
<td>SCL-25</td>
</tr>
<tr>
<td>NFBC1986</td>
<td>YSR</td>
<td>YSR</td>
<td>YSR</td>
</tr>
<tr>
<td>NHANES</td>
<td></td>
<td>DIS (lifetime diagnosis)</td>
<td></td>
</tr>
<tr>
<td>NSHD</td>
<td>GHQ-28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NTR</td>
<td>ASR</td>
<td>ASR</td>
<td>ASR</td>
</tr>
<tr>
<td>Patch 2</td>
<td></td>
<td>SCID (lifetime diagnosis)</td>
<td></td>
</tr>
<tr>
<td>Rotterdam</td>
<td>CES-D or M-CIDI</td>
<td>CES-D (20 items)</td>
<td>M-CIDI</td>
</tr>
<tr>
<td>SYS-P</td>
<td>CES-D or DSM instrument</td>
<td>CES-D (12 items)</td>
<td>10 question DSM-based instrument</td>
</tr>
<tr>
<td>Whitehall II</td>
<td>GHQ-30</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All scales measure current depression and anxiety unless otherwise stated. Clinician diagnosis was assessed by self-reported recall in all studies.

ALSPAC, Avon Longitudinal Study of Parents and Children; ASR, Adult Self Report; BC, Birth Cohort; BRHS, British Regional Heart Study; BWHHS, British Women’s Heart and Health Study; CARTA, Causal Analysis Research in Tobacco and Alcohol; CCEI, Crown–Crisp Experiential Index; CES-D, Centre for Epidemiologic Studies Depression; CHDS, Christchurch Health and Development Study; CIDI, Composite International Diagnostic Interview; CIS-R, Computerised interview schedule-revised; CoLaus, Cohorte Lausannoise; DIGS, Diagnostic Interview for Genetic Studies; DIS, Diagnostic Interview Schedule; ELSA, English Longitudinal Study of Ageing; EPDS, Edinburgh Postnatal Depression Scale; GHQ, General Health Questionnaire; GOYA, Genomics of Overweight Young Adults; HADS, Hospital Anxiety and Depression Scale; HBCS, Helsinki Birth Cohort Study M-CIDI, Munich version of CIDI; NHANES, National Health and Nutrition Examination Survey; NSHD, National Survey of Health and Development; NTR, the Netherlands Twin Registry; SCID, Structured Clinical Interview for DSM-III-R diagnosis; SCIDI/NP, Structured Clinical Interview for DSM-IV Axis disorders non-patient edition; SCL, symptoms checklist; STAI, State Trait Anxiety Inventory; YSR, Youth Self Report.

Smoking status
Smoking status was self-reported (either by questionnaire or interview) at the same time as mental health assessment for all studies, with the exception of 1958BC, CoLaus and HBCS which used smoking status and depression/anxiety data collected up to 3 years apart (see online supplementary material). Individuals were classified as never, former, current or ever (ie, current and former combined) cigarette smokers. Where information on smoking frequency was available, current smokers were restricted to individuals smoking at least one cigarette per day. Where information on pipe and cigar smoking was available, individuals reporting being current or former smokers of pipes or cigars but not cigarettes were excluded from all analyses.

For studies with adolescent populations (ALSPAC children and NFBC1986), analyses were restricted to current daily smokers who reported smoking at least one
cigarette per day (current smokers) and individuals who had never tried smoking (never smokers).

Smoking heaviness in current smokers, measured as cigarettes smoked per day, was collected in some studies as a continuous variable and in some studies as a categorical variable. Further details of the smoking measures collected within each study are provided in the online supplementary material.

**Statistical analysis**

Analyses were conducted within each contributing study using Stata or R software, following the same analysis plan. Analyses were restricted to individuals with full data on depression and anxiety outcomes, smoking status and rs16969968/rs1051730 genotype.

Sex-adjusted and age-adjusted associations of smoking status (never, former, current, ever) and smoking heaviness with binary measures of depression, anxiety and psychological distress were assessed using logistic regression. For the smoking status analysis, never smokers were used as the reference group. The smoking heaviness analysis was restricted to current daily smokers, and ORs represent differences in odds of the outcome measure per additional cigarette consumed per day. These analyses were restricted to studies with continuous measures of cigarettes per day.

Within each study, genotype frequencies were tested for deviation from Hardy-Weinberg equilibrium (HWE) using a $\chi^2$ exact test. Mendelian randomisation analyses of the association between rs16969968/rs1051730 and binary measures of depression, anxiety and psychological distress using linear regression stratified by smoking status. These analyses were adjusted for age and sex and additionally for use of depression or anxiety medication where available. For studies with a survey (NHANES) or family-based design (SYS-P), appropriate methods were used to adjust SEs (see online supplementary material for further information).

**Table 2** Case definitions for depression, anxiety and psychological distress

<table>
<thead>
<tr>
<th>Case definition 1</th>
<th>Case definition 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Depression</strong></td>
<td></td>
</tr>
<tr>
<td>Self-report of clinical diagnosis OR Meeting clinical criteria for depression OR Scoring above published cut point on specific depression scale</td>
<td>Scoring &gt;90th centile on specific depression scale</td>
</tr>
<tr>
<td><strong>Anxiety</strong></td>
<td></td>
</tr>
<tr>
<td>Self-report of clinical diagnosis OR Meeting clinical criteria for anxiety disorders OR Scoring above published cut point on specific anxiety scale</td>
<td>Scoring &gt;90th centile on specific anxiety scale</td>
</tr>
<tr>
<td><strong>Psychological distress</strong></td>
<td></td>
</tr>
<tr>
<td>Depression or anxiety as defined above OR Scoring above published cut point on general scale for psychiatric symptoms</td>
<td>Depression or anxiety as defined above OR &gt;90th centile on general scale for psychiatric symptoms</td>
</tr>
</tbody>
</table>

Results from individual studies were meta-analysed in Stata (V.11) using the ‘metan’ command. Where there was evidence of heterogeneity between studies ($I^2>50\%$), both fixed and random effects analyses were performed. Within the Mendelian randomisation analyses, the Cochran Q statistic was used to test for interactions between genotype and smoking status on the outcome measures. Analyses were also performed stratified by sex because there is some evidence from observational studies that the association between smoking and mental health outcomes may differ by sex. Sex differences in the association between genotype and outcomes measures were tested for using meta-regression after taking into account potential differences by smoking status.
RESULTS

Descriptive statistics

In total, data on up to 127,632 individuals were available for analysis, including 58,176 never smokers, 37,428 former smokers and 32,028 current smokers. Overall, 45% of the combined study population was male. The median age within the contributing studies ranged between 16 and 68 years. The mean prevalence of depression, anxiety and psychological distress (using case definition 1) was 12.5% (range 6.1–37.5%), 10.2% (range 2.6–19.9%) and 17.4% (range 8.9–27%), respectively. Descriptive statistics for each of the study populations are provided in the online supplementary table S3.

The minor allele frequency for rs16969968/rs1051730 ranged between 0.31 and 0.39 (see online supplementary table S4). There was no strong evidence for devi- ation from HWE in any of the studies (p values all ≥0.06).

Observational analysis

Levels of depression, anxiety and psychological distress differed by smoking status (see figure 2). In age-adjusted and sex-adjusted analyses, current smokers had 1.85 times (95% CI 1.65 to 2.07, p<0.001) greater odds of depression, 1.71 times (95% CI 1.54 to 1.90, p<0.001) greater odds of anxiety and 1.69 times (95% CI 1.56 to 1.83, p<0.001) greater odds of psychological distress than never smokers. Former smokers had 1.22 times (95% CI 1.14 to 1.30, p<0.001) greater odds of depression, 1.23 times (95% CI 1.12 to 1.36, p<0.001) greater odds of anxiety and 1.17 times (95% CI 1.11 to 1.25, p<0.001) greater odds of psychological distress than never smokers.

Among smokers, smoking heaviness was positively associated with levels of depression, anxiety and psychological distress (figure 3). In age-adjusted and sex-adjusted analyses, a one cigarette per day increase in smoking heaviness was associated with a 1.03-fold (95% CI 1.02 to 1.04, p<0.001) increase in the odds of having depression, a 1.03-fold (95% CI 1.02 to 1.04, p<0.001) increase in the odds of having anxiety and a 1.02-fold (95% CI 1.02 to 1.03, p<0.001) increase in the odds of having psychological distress.

As there was evidence of between-study heterogeneity for analyses of both smoking status and smoking heaviness, random effects meta-analyses are presented. However, results from fixed effects meta-analyses were similar (data not shown). Individual study estimates for observational analyses are provided in online supplementary figures S1 and S2.

Mendelian randomisation analysis

There was no clear evidence that rs16969968/rs1051730 was associated with binary measures of depression in never (OR per minor allele 1.02, 95% CI 0.97 to 1.06, p=0.47), former (OR 1.00, 95% CI 0.95 to 1.05, p=0.99), current (OR 1.00, 95% CI 0.95 to 1.05, p=0.99) or ever (OR 1.01, 95% CI 0.98 to 1.05, p=0.58) smokers. However, in never smokers there was some evidence of a positive association between the minor allele of rs16969968/rs1051730 and anxiety (OR 1.05, 95% CI 1.01 to 1.10, p=0.03). For psychological distress there was a similar pattern, with no strong evidence for an association between rs16969968/rs1051730 in smokers, but some evidence of a positive association in never smokers.
(OR 1.03, 95% CI 1.00 to 1.07, p=0.09). For all outcomes, there was no clear statistical evidence for a difference in the effect of rs16969968/rs1051730 between never, former and current smokers (p values for heterogeneity between never, former and current smokers from Cochran Q test all >0.57). Individual study estimates for observational analyses are provided in online supplementary figure S3.

Results were similar for continuous measures of symptoms of depression, anxiety and psychological distress (see online supplementary figures S4 and S5). There was no clear evidence for associations of rs16969968/rs1051730 with continuous outcomes in any of the smoking categories.

Finally, there was no clear evidence for sex differences in either observational or Mendelian randomisation analyses for associations between smoking or smoking-related genotype and depression or anxiety (data available on request).

**DISCUSSION**

In the largest Mendelian randomisation study on the association of smoking with depression and anxiety conducted to date, we found no evidence to suggest that smoking causes either depression or anxiety. Despite higher levels of depression, anxiety and psychological distress in current and former smokers compared to never smokers, and a positive association between the number of cigarettes smoked per day and depression and anxiety, there was no clear evidence for associations between the CHRNA5-A3-B4 variant and these outcomes in smokers. If heavier smoking were to cause depression or anxiety, we would expect to see an increased risk of depression or anxiety per copy of the minor allele of rs16969968/rs1051730, which increases smoking heaviness, in current smokers and potentially also in former smokers, but no difference in risk for never smokers. In our meta-analyses, ORs for the effect of rs16969968/rs1051730 in current, former and ever smokers were all close to the null, with the CIs for these estimates all overlapping the null. In addition, we found no evidence to suggest that the variant was differentially associated with depression or anxiety according to smoking status.

Our results are consistent with those of the two previous Mendelian randomisation studies, which did not find evidence that smoking increases depression or anxiety or ante-natal depression. Both of these studies were included in this meta-analysis and the HUNT study made up more than half of the study sample in some analyses. However, exclusion of either of these samples did not make a substantial difference to effect estimates (see online supplementary figures S6 and S7). These findings suggest that previous findings linking smoking to higher levels of depression and anxiety may be due to residual confounding, a shared vulnerability to both mental disorders and...
smoking behaviour, or reverse causality (eg, if smokers smoke in an attempt to alleviate the symptoms of depression or anxiety or depressed smokers find it more difficult to quit). Numerous longitudinal studies have reported that depressive symptoms in childhood and adolescence are associated with increased risk of smoking initiation or progression to tobacco dependence. At the same time, smoking cessation appears to be associated with an acute increase in depressive symptoms among a subgroup of smokers, and these individuals have poor smoking cessation outcomes. Taken together, this suggests that people experiencing depressive symptoms may smoke (or relapse to smoking) in an attempt to self-medicate these symptoms. A Mendelian randomisation study of the association of genetic variants for depression and anxiety with smoking behaviour would be required to investigate the self-medication hypothesis. However, genetic variants robustly associated with depression and anxiety have yet to be identified.

Some caution should be taken in completely ruling out an effect of this variant on depression and anxiety within this analysis. The CIs for the associations of rs16969968/rs1051730 with depression and anxiety in current smokers overlap the estimates for the per cigarette per day increase in ORs of depression and anxiety, so we cannot conclusively say that the Mendelian randomisation analysis results differ from the observational results (figure 3). This is the most direct comparison that we can make with observational estimates in our data, given that the minor allele of rs16969968/rs1051730 is associated with an average of one cigarette per day increase in smoking heaviness. However, this comparison may be problematic because cigarettes per day, a self-reported measure of tobacco exposure, does not take into account variation in smoking topography, such as the amount of a cigarette an individual smokes or the depth of inhalation. The CHRNA5-A3-B4 variant is an instrument for lifetime tobacco exposure within current smokers, and this is not fully captured by cigarettes per day. It has been shown that rs16969968/rs1051730 explains more of the variance in an objective measure of tobacco exposure, cotinine (4%), than in self-reported cigarettes per day (1%). This appears to be why the variant shows a much stronger association with lung cancer than predicted from the observed associations with self-reported cigarettes per day. Therefore if higher levels of smoking did cause depression or anxiety, we might expect the effects of rs16969968/rs1051730 to be considerably larger than those seen observationally per cigarette per day. For the same reason, we did not perform instrumental variable analysis to estimate the magnitude of the causal effect of smoking heaviness on depression or anxiety. It has been demonstrated that using cigarettes per day as an intermediate variable in Mendelian randomisation analyses using rs16969968/rs1051730 can lead to large biases in causal effect size estimates.

The rs16969968/rs1051730 variant associates with smoking heaviness within smokers but is not an instrument for smoking status (ever smoking vs never smoking). Therefore we cannot rule out the possibility that being a smoker, rather than smoking heaviness, could influence likelihood of depression or anxiety. However, we do see an observational association between smoking heaviness and depression and anxiety (figure 3), and a dose-dependent relationship between an exposure and an outcome strengthens support for a causal association. We might therefore expect to observe an association between rs16969968/rs1051730 and depression or anxiety if smoking were to cause these conditions. Furthermore, while rs16969968/rs1051730 is the strongest genetic contributor to smoking behaviour identified to date, this variant only explains a fraction of the estimated 50% of total variation in smoking behaviour within a population at any one time that is due to genetic factors. Further signals for smoking heaviness have been identified in the same gene cluster, in other nicotinic receptor units and in other genes, such as those related to nicotine metabolism like CYP2A6. Combining these variants in genetic risk scores for smoking behaviour in Mendelian randomisation studies will be an important future direction for validation of these results.

We would not expect to see an effect of rs16969968/rs1051730 on depression or anxiety in never smokers, because the variant is not associated with smoking heaviness within these individuals. Thus, this group can be used to test potential bias due to pleiotropy (that the gene affects more than one exposure) in Mendelian randomisation analyses. We did observe some evidence for an association between the variant and anxiety in never smokers, a finding previously reported by the HUNT study. Removal of HUNT from this analysis did not affect the point estimate, suggesting that this association is not driven solely by the data from this study (see online supplementary figure S6). However, using case definition 2 (where available) for anxiety in preference to case definition 1 slightly attenuated this association in never smokers (see online supplementary figure S8). While this may be a chance finding, it is possible that rs16969968/rs1051730 or a variant in linkage disequilibrium with this variant, may affect anxiety directly, not through tobacco consumption. There is some suggestion from animal studies that nicotinic acetylcholine receptors may play a role in anxiety (eg, mice lacking the α4 subunit show increases in anxiety-related behaviour). However, there is currently little evidence for this association in humans, and rs16969968/rs1051730 has not been identified in genome-wide association studies of depression or anxiety to date.

It is important to note that stratifying by the measured exposure variable in Mendelian randomisation studies can lead to collider bias. In this specific analysis, if both the genetic variant and anxiety cause individuals to smoke, then stratifying on smoking could, in theory,
induce an association between the variant and anxiety.\textsuperscript{13} We do not think that collider bias is likely to be a major issue in these analyses because rs16069968/rs1051730 does not appear to be associated with smoking initiation in this sample \cite{29} or in previous studies.\textsuperscript{29, 30} There is, however, as reported previously in a few specific populations,\textsuperscript{36, 62, 63} some evidence that the minor allele of rs16069968/rs1051730 is associated with smoking cessation; ORs of being a current compared to a former smoker were 1.08-fold higher per copy of the minor allele (95% CI 1.06 to 1.11) in this sample.

**Strengths and limitations**

The key strength of this study is the large sample size, using data on over 125,000 individuals from 25 different populations. Despite this, we did not have the power to rule out the possibility of a causal effect. A substantial increase in sample size would be required to be confident that what we observe is a true null association in smokers. We hope that our estimates may be combined with those of further studies addressing the same question in future meta-analyses, to provide more definitive answers.

One of the main limitations is the use of broad definitions of depression and anxiety rather than clinical definitions, which were not available in all studies. It is possible that use of more precise phenotypic measures of depression and anxiety based solely on clinical criteria could yield stronger results because non-differential misclassification of a binary outcome is likely to attenuate associations towards the null.\textsuperscript{2, 64} However, we showed the expected observational associations between smoking and depression, anxiety and psychological distress. In addition, we used two case definitions and performed a sensitivity analysis using case definition 2 in preference (where both were available) which produced similar results (see online supplementary figure S8).

Sensitivity analyses performed excluding lifetime definitions of depression or anxiety also produced similar results (see online supplementary figure S10). Finally, restricting our analyses to those studies with questionnaire based on clinical criteria or self-report of doctor diagnosis produced consistent results, although these analyses were underpowered (see online supplementary figure S11).

Although we analysed depression and anxiety separately, these conditions are highly comorbid in the general population.\textsuperscript{65, 66} and symptom scale questionnaires are not adequate to distinguish between them.\textsuperscript{67, 68} In addition, the definition of anxiety we used encompassed general anxiety disorder, panic, and phobias. It is possible that these conditions have different aetiologies.\textsuperscript{69} Therefore, we cannot make inferences about specific anxiety disorders from these results. Furthermore, the sample encompasses a wide age range, so it is unlikely that this analysis would be able to capture any age-specific effects of smoking on depression and anxiety.

**CONCLUSION**

In conclusion, our Mendelian randomisation analyses do not support a causal role of smoking heaviness among smokers in the development of depression and anxiety. While we cannot directly address the question of whether smoking initiation (ie, starting smoking) plays a causal role in relation to these outcomes, we expect that if it did we would also see a dose-dependent relationship between smoking heaviness and depression and anxiety. We see such an association in our observational analyses, but no strong evidence for this in our Mendelian randomisation analyses. Future research should focus on the possible role of depression and anxiety in increasing susceptibility to smoking. As larger genome-wide association studies of depression and anxiety emerge, it is likely that genetic variants will be identified that can be utilised in Mendelian randomisation analyses for this purpose.

**Author affiliations**

\begin{itemize}
  \item 1MRC Integrative Epidemiology Unit (IEU) at the University of Bristol, UK
  \item 2UK Centre for Tobacco and Alcohol Studies, School of Experimental Psychology, University of Bristol, Bristol, UK
  \item 3Department of Public Health, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway
  \item 4Forensic Department, Research Centre Brøset St. Olav’s University Hospital Trondheim, Trondheim, Norway
  \item 5Faculty of Medicine, Department of Laboratory Medicine, Children’s and Women’s Health, Norwegian University of Science and Technology, Trondheim, Norway
  \item 6Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK
  \item 7Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
  \item 8Queensland Brain Institute, The University of Queensland, Brisbane, QLD, Australia
  \item 9Department of Epidemiology and Public Health, University College London, London, UK
  \item 10Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands
  \item 11University of Helsinki, Hjelt institute, Helsinki, Finland
  \item 12University of Eastern Finland, Institute of Public Health & Clinical Nutrition, Kuopio, Finland
  \item 13Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland
  \item 14Hospital District of North Karelia, Joensuu, Finland
  \item 15Department of Mental Health and Substance Abuse Services, National Institute for Health and Welfare, Helsinki, Finland
  \item 16Institute of Health Sciences, FI-90014 University of Oulu, Finland
  \item 17Biocenter Oulu, FI-90014 University of Oulu, Finland
  \item 18South West London and St George’s Mental Health Trust, London, UK
  \item 19Population, Policy and Practice, UCL Institute of Child Health, University College London, UK
  \item 20Research Centre for Prevention and Health, The Capital Region of Denmark, Denmark
  \item 21Metabolic Genetics Section, Faculty of Health and Medical Sciences, Novo Nordisk Foundation Centre for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark
  \item 22Copenhagen Prospective Studies on Asthma in Childhood, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
  \item 23Danish Pediatric Asthma Center, Gentofte Hospital, The Capital Region, Copenhagen, Denmark
  \item 24Institute of Public Health and Center for Healthy Aging, University of Copenhagen, Denmark
\end{itemize}
Acknowledgements

ALSPAC: The authors are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. BRHS: The British Regional Heart Study has been made possible through grant funding from the UK government (MRCUnitforLifelongHealth, Ageing at UCL, UK). BRHS has been made possible through grant funding from UK government (MRC Unit for Lifelong Health, Ageing at UCL, UK) and supported by U01 HL64781 and R01HL65988 which were sponsored by the National Heart, Lung, and Blood Institute (US Department of Health and Human Services). The rotation and coordination of the data analysis were conducted by the Wellcome Trust Sanger Institute, Cambridge, UK. The Wellcome Trust Sanger Institute was supported by grants from the Wellcome Trust and the National Institutes of Health. The authors would like to acknowledge the contribution of the late Academician of Science Leena Peltonen. NSHD: The authors are very grateful to the members of the NSHD Collaborating Centers, the general practitioners and their staff who have supported data collection since the study inception. CaPS: The Caerphilly Prospective Study was conducted by the former MRC Epidemiology Unit (South Wales). The Caerphilly archive is now maintained by the School of Social and Community Medicine in Bristol University. The authors thank the Health and Social Care Information Centre (HCISIC) for helping us maintain long term follow-up with the cohort. They also thank all the men who have given their time to be participants in CaPS. HUNT: Nord-Trndelag Health Study (The HUNT Study) is a collaboration between HUNT Research Centre (Faculty of Medicine, Norwegian University of Science and Technology NTNU), Nord-Trndelag County Council and the Norwegian Institute of Public Health. NBFC: The authors thank the late Professor Paula Rantakallio (launch of NBFSs), and Ms Outi Tornwall and Ms Minttu Jussila (DNA biobanking). The authors would like to acknowledge the contribution of the late Academician of Science Leena Peltonen. NSHD: The authors are very grateful to the members of this birth cohort for their continuing interest and participation in the study.

Collaborators

ALSPAC (AET, MEF, MRG, DGS), HUNT (JHB, PRR, MEG, FS), Generation Scotland (REM, AC, DJP, CH), Whitehall/ELSA (Mek, JE, Mirk), Rotterdam (HT, SM), FINRISK (AL, TL, TP, TK), NFBC (M-JF, MAF, JF, FD), 1958BC (AC, CP, EH), Health2006/Health2008/Inter99 (LLNH, RJK, TS, JFE, Tsa, AL, ELM), NTR (CCM, GW, Dib, JMV), COLAUS/PsyCoLaus (PMIV, MP), BWHS (CED, AA, JPC), BRHS (RWM, LTL), HBSLS (JC, AP, JGE, KR), NSHD (AW, DK), Goya females (EAN, LP), CaPs (YS-B, SY-S, AP-YW, ZP, TP), CHDS (LJH, MAK), Patch 2 (EJC, MM).

Contributors

MRM and GDS conceived the study. AET and MEF drafted the analysis protocol and co-ordinated the data analysis. AET conducted the final meta-analyses. AET and MEF performed analyses within individual studies. MM, EJCA, MAK, ZP, TP, YBS, DJP, CH and PRR all substantially contributed to data acquisition in contributing studies. All authors commented on a draft of the manuscript and approved the final version.

Funding

1958BC: Statistical analyses were funded by the Academy of Finland (Project 24300796 and SALVE/PREVMEDESYN). DNA collection was funded by MRC grant G0000834 and cell-line creation by Welcome Trust grant 068545/Z/02. This research used resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases, National Human Genome Research Institute, National Institute of Child Health and Human Development, and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418. Funding for the project was provided by the Wellcome Trust under the award 076113, the British Heart Foundation, and unrestricted grants from GlaxoSmithKline as well as by the Faculty of Biology and Medicine of the Lausanne University. Wellcome Trust Sanger Institute was supported by the Wellcome Trust (grant number 090049) and the Medical Research Council (grant numbers MR/J01351X/1, G0800612, G0802736, MC_UU_120131/1, MC_UU_120136/3). BRHS: The British Women’s Heart and Health Study has been supported by funding from the British Heart Foundation (BHF) (grant PG/09/022) and the UK Department of Health Policy Research Programme (England) (grant 090094). The BWHS HumanCVD data were funded by the BHF (PG/07/17213/24254). CHDS: The Christchurch Heart and Development Study has been supported by funding from the Health Research Council of New Zealand, the National Child Health Research Foundation (Cure Kids), the Canterbury Medical Research Foundation, the New Zealand Lottery Grants Board, the University of Otago, the Carney Centre for Pharmacogenomics, the James Hume Bequest Fund, US NIH grant MH077874, and NIDA grant “A developmental model of gene-environment interplay in SUDs” (R01DA024413) 2007–2012. Colaus/PsyCoLaus: The CoLaus/PsyCoLaus study was supported by four grants of the Swiss National Science Foundation (#105993, 113830, 139468 and 122661), two unrestricted grants from GlaxoSmithKline as well as by the Faculty of Biology and Medicine of the Lausanne University. ELSA: ELSA is funded by the National Institute on Aging in the US (RO1 AG017644, R01AG176440613) and by a consortium of UK Government departments (including: Department for Open Access
Communities and Local Government, Department for Transport, Department for Work and Pensions, Department of Health, HM Revenue and Customs and Office for National Statistics). FINRISK: This study was supported by the Academy of Finland Center of Excellence in Complex Disease Genetics (grant numbers 213506, 129860), the Academy of Finland (grant numbers 139635, 129494, 136895, 263836 and 141054), the Sigrid Juselius Foundation, and ENGAGE – European Network for Genetic and Genomic Epidemiology, FP7-HEALTH-F4-2007, grant agreement number 201413, Generation Scotland: Generation Scotland has received core funding from the Chief Scientist Office of the Scottish Government Health Directorates CZD1/6 and the Scottish Funding Council HR03006. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Funding Council HR03006. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland and was funded by the UK's Medical Research Council. GOYA females: The GOYA study was conducted as part of the activities of the Danish Obesity Research Centre (DanORC, www.danorc.dk) and The MRC centre for Causal Analyses in Translational Epidemiology (MRC CAITE). The genotyping for GOYA was funded by the Wellcome Trust (WT 084762). GOYA is a nested study within The Danish National Birth Cohort which was established with major funding from the Danish National Research Foundation. Additional support for this cohort has been obtained from the Pharmacy Foundation, the Egmont Foundation, The March of Dimes Birth Defects Foundation, the Augustinus Foundation, and the Health Foundation. HBCS: Helsinki Birth Cohort Study has been supported by grants from the Academy of Finland, the Finnish Diabetes Research Society, Novo Nordisk Foundation, Finska Läkareassistentenkapet, Signe and Ane Gyllenberg Foundation, University of Helsinki. Ministry of Education, Akhaos Foundation, Emil Aaltonen Foundation. NFBC: NFBC1966 and NFBC1986 received financial support from the Academy of Finland (project grants 104781, 120315, 129269, 1114194, 24300796, 141042 Center of Excellence in Complex Disease Genetics and SALVE), University Hospital Oulu, Biocenter, University of Oulu, Finland (75617), NHLBI grant 5R01HL087679-02 through the STAPPEED program (1R1L1MH082688-01), NIH/NIH (5R01HL3706/02), the European Commission (EURO-BLCS, Framework 5 award QLG1-CT-2000-01643), ENGAGE project and grant agreement HEALTH-F4-2007-201413, EU FP7 EuroHEALTHAging -277849, the Medical Research Council, UK (G0500539, G0600705, G1002319, PrevMetSyn/SALVE) and the MRC, Centenary Early Career Award. The DNA extractions, sample quality controls, biobank up-keeping and aliquotting was performed in the National Public Health Institute, Biomedicum Helsinki, Finland and supported financially by the Academy of Finland and Biocentre Helsinki. NSHD: This work was funded by the Medical Research Council [MC_UU_12019/1]. NTR: This study was supported by the European Research Council (ERC Starting Grant 284167 PI Vink), Netherlands Organization for Scientific Research (NWO) (MagW/ZonMW grants 904-61-099, 985-10-002, 904-61-193, 480-04-004, 400-05-717, Addiction-31160008 Middelgroot-911-09-032, Spinozepremie 56-464-14192), BBRMI-NL (Biobanking and Biomedical Resources Research Infrastructure), VU University’s Institutes for Health and Care Research and Neuroscience Campus Amsterdam. Patch 2: The Patch 2 study was funded by the Imperial Cancer Research Fund (ICAF) and Cancer Research UK (CRUK) Programme Grants to the General Practice Research Group at the University of Oxford. Rotterdam: The Rotterdam Study was supported by the Erasmus Medical Center and the Erasmus University Rotterdam, the Netherlands Organisation for Scientific Research (NWO), the Netherlands Organisation for Health Research and Development (ZonMw), the Ministry of Education, Culture, and Science, the Ministry of Health, Welfare, and Sports, the European Commission (DG-XII). SYS-P: The SYS-P is funded by the Canadian Institutes of Health Research, the Canadian Foundation for Innovation, and the Heart and Stroke Foundation of Ontario. Whitehall: The Whitehall II study has been supported by grants from the Medical Research Council; British Heart Foundation; Health and Safety Executive; Department of Health; National Heart Lung and Blood Institute (NHLBI: HL36310) and National Institute on Aging (AG13196); US, NIH; Agency for Health Care Policy Research (HS06516); and the John D and Catherine T MacArthur Foundation Research Networks on Successful Midlife Development and Socio-economic Status and Health.

Competing interests MRM: Medical Research Council, LTL: British Heart Foundation, LJJH: New Zealand Health Research Council, MK: Jim and Mary Carney Charitable Trust, New Zealand Health Research Council, CCM: European Research Council, DIB, JMV: NWO, European Research Council, National Institutes of Health, HT: Dutch Medical Research Council, TP: Canadian Institutes of Health Research, MK: Medical Research Council, NIH Heart, Lung and Blood Institute and Economic and Social Research Council); financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years (LJJH: Australian National Health and Medical Research council, PMV: Swiss National Science Foundation, MP: Eli Lilly and Lumentuck advisory board membership, Swiss National Science Foundation, TK: Pfizer consultancy on tobacco dependence, Juho Vainio Foundation, The Finnish Medical Society; TSA: Gene Diet Interactions in Obesity project, HT: KOC University, Pfizer); YBS: grants from NIH, MRC, POLUK, BHF, AET, MRM and MF are members 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. LTL is supported by BHF programme grant RG/08/013/25942. LP is funded by an MRC Population Health Scientist Fellowship (MR/J012165/1). TSA was supported by the Gene Diet Interactions in Obesity (GENDINOB, www.gendinob.dk) postdoctoral fellowship grant. LLNH was supported by the Health Insurance Foundation (grant no. 2010 B 131). The work of SSM was funded by the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) (grant numbers 050-060-810 NCHA); the work of HT was supported by the NHLBI (grant 5R01HL087679-02) through the STAMPEED program (1RL1MH083288-01), the National Heart, Lung and Blood Institute and the Medical Research Council, PMV: Swiss National Science Foundation, MPV: Swiss National Science Foundation, HT: supported financially by the Economic and Social Research Council International Centre for Life Course Studies in Society and Health (RES-596-28-0001).

Patient consent Obtained.

Ethics approval 1958BC: Ethics approval for the study was obtained from the South-East Multi-Centre Research Ethics Committee (Ref: 01/144) and the Joint UCL/UCLH Committees on the Ethics of Human Research (Committee A) Ref: 08/H0714/40. ALSPAC: Ethics approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committee. BRHS: The BRHS has local (from each of the districts in which the study was based) and multicentre ethical committee approvals. Ethical approval was granted for the BWHHS from the London Multi-Centre Research Ethics Committee and 23 Local Research Ethics Committees. BWHHS: Ethical approval was granted for the BWHHS from the London Multi-Centre Research Ethics Committee and 23 Local Research Ethics Committees. CaPcs: Ethics approval was obtained from the South Glamorgan Area Health Authority, the Gwent REC, and the South Wales Research Ethics Committee D. CHDS: All phases of the study have received ethical approval from the regional and local ethics committees and a national ethics committee (Medical Research Council, SALVE), University of Helsinki, Ministry of Education, Culture, Science, the Ministry of Health, Welfare, and Sports, the European Commission (ZonMw), the Ministry of Education, Culture, and Science, the Ministry of Health, Welfare, and Sports, the European Commission (DG-XII). SYS-P: The SYS-P is funded by the Canadian Institutes of Health Research, the Canadian Foundation for Innovation, and the Heart and Stroke Foundation of Ontario. Whitehall: The Whitehall II study has been supported by grants from the Medical Research Council; British Heart Foundation; Health and Safety Executive; Department of Health; National Heart Lung and Blood Institute (NHLBI: HL36310) and National Institute on Aging (AG13196); US, NIH; Agency for Health Care Policy Research (HS06516); and the John D and Catherine T MacArthur Foundation Research Networks on Successful Midlife Development and Socio-economic Status and Health.

Open Access
residence of the patients. Rotterdam: The Medical Ethics Committee of Erasmus Medical Center Rotterdam approved the study. SYS-P: The Research Ethics Committee of the Chichoum Hospital in Quebec, Canada approved the study protocol. Whitelab: Ethical approval for the Whitelab II study was obtained from the University College London Medical School committee on the ethics of human research.

**Provenance and peer review** Not commissioned; externally peer-reviewed.

**Data sharing statement** 1958BC: This study makes use of data generated by the Wellcome Trust Case-Control Consortium. A full list of investigators who contributed to generation of the data is available from the Wellcome Trust Case-Control Consortium website. The 1958 birth cohort data can be accessed via the UK Data Service (http://ukdataservice.ac.uk/). ALSPAC: Data used for this submission will be made available on request to the ALSPAC executive committee (alspac-exec@bristol.ac.uk). The ALSPAC data management plan (available here http://www.bristol.ac.uk/alspac/researchers/data-access/) describes in detail the policy regarding data sharing, which is through a system of managed open access, BRHS: We welcome proposals for collaborative projects and data sharing (http://www.ucl.ac.uk/pcph/research-groups-themes/brhs-pub). For general data sharing enquiries, please contact Lucy Lennon (l.lennon@ucl.ac.uk). BWHHS: All BWHHS data collected is held by the research team based at London School of Hygiene and Tropical Medicine, for ongoing analysis. If you would like to collaborate with the BWHHS team, contact the study coordinator, AA (antoinette.amuzu@lshtm.ac.uk). Data and biological samples provided to the collaborators can only be used for the purposes originally stated and must not be used in any other way without re-application to the BWHHS team. No data should be passed on to any third party unless they were specified in the original application. CaPs: Data used for the Caerphilly Prospective study (CaPs) was made available by the CaPs access committee (Chair: Professor Kay Tew Khaw). More information about its managed access procedure is available on the study website (http://www.bris.ac.uk/social-community-medicine/people/project/1392). CHDS: Data contributed for this submission are available on request from the CHDS (john.horwood@otago.ac.nz). CoLus/PsyCoLaus: Data from the CoLus/PsyCoLaus study can be requested according to the procedure described on the CoLus website (http://www.colus.ch/en/cls_home/cls_pro_home/cls-research-3.htm). ELSA: ELSA data are made available through the EDS data website (http://www.elsa-project.ac.uk/availableData). FINRISK: Data used for this submission will be made available on request to the FINRISK Management Group, according to the given ethical guidelines and Finnish legislation. Generation Scotland: Data is available on request (access@generationscotland.org). GOYA females: An anonymized copy of the data used for this submission will be made available on request to the GOYA analysts after permission have been given by the DNBC executive committee (www.dnbc.dk). HBCS: Data used for this submission will be made available on request to the HBCS executive committee (johan.eriksson@helsinki.fi). Health2006/Health2008/Inter99: Data used for this submission can be made available on request to the Research Centre for Prevention and Health (http://www.region.dk/rcfs/Menu/). Please contact LLNH (lise.lotte.nystrup@regionh.dk) or AL (allan.linneberg@regionh.dk). HUNT: Data used from the HUNT Study for this submission will be made available on request to the HUNT Data Access Committee (hunt@medisin.ntnu.no). The HUNT data access information (http://www.ntnu.edu/hunt/data) describes in detail the policy regarding data availability. NFBC: Data used for this submission can be made available on request to Tuula Ylitalo (tuula.ylitalo@oulu.fi), Minna Mannikko (minna.mannikko@oulu.fi) or M-RJ (mi.jarvelin@imperial.ac.uk). NHANES: NHANES data can be accessed here: (http://www.cdc.gov/nchs/nhanes.htm). The genotype used in this analysis is a restricted variable. Applications for access to these data must be made through the Research Data Center: (http://www.cdc.gov/rdc/). NSHD: The NSHD data are made available to researchers who submit data requests (tomrcba.switinfo@ucl.ac.uk). More information is available in the full policy documents (http://www.nshd.mrc.ac.uk/data.aspx). Managed access is in place for this study in order to ensure that use of the data are within the bounds of consent given previously by participants, and to safeguard any potential threat to anonymity since the participants are all born in the same week. NTR: Data used for this submission will be made available on request to the NTR committee (ntr@psych.vu.nl). SYS-P: Data used for this submission will be made available on request to the SYS principal investigators (http://www.saguenay-youth-study.org) (scientific community/collaborative interests). Whitelab: Data from the Whitelab II study are made publicly available as described in the Whitelab II data sharing policy (http://www.ucl.ac.uk/whitelabll/datasharing).

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


