Health behaviours prior to pregnancy and fertility outcomes: Triangulation of evidence in the Norwegian Mother, Father and Child Cohort Study (MoBa)

Authors

Robyn E Wootton1,2,3, Rebecca B Lawn4, Maria Magnus5, Jorien Treur6, Elizabeth Corfield1,7, Pål R. Njølstad8,9, Ole A. Andreassen10, Deborah A. Lawlor2,3, Marcus R Munafò2,11, Siri E. Håberg5, George Davey Smith2,3, Ted Reichborn-Kjennerud12,7, Per Magnus5, Alexandra Havdahl1,7

Affiliations

1. Nic Waals Institute, Lovisenberg Diaconal Hospital, Oslo, Norway
2. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK
3. Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK.
4. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, US
5. Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway.
6. Department of Psychiatry, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands
7. Department of Mental Disorders, Norwegian Institute of Public Health, Oslo, Norway
8. Center for Diabetes Research, Department of Clinical Science, University of Bergen, Bergen, Norway
9. Children and Youth Clinic, Haukeland University Hospital, Bergen, Norway
10. NORMENT, Institute of Clinical Medicine, University of Oslo and Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway
11. School of Psychological Science, University of Bristol, Bristol, UK
12. Institute of Clinical Medicine, University of Oslo, Oslo, Norway

Keywords: fertility, alcohol, smoking, BMI, caffeine, Mendelian randomisation, MoBa, MBRN.
Abstract (294 of 300 words)

Introduction: Current advice to improve fertility includes reducing alcohol and caffeine consumption, achieving healthy weight-range, and stopping smoking. Advice is informed by observational evidence, which is often biased by confounding.

Methods: This study uses data from the Norwegian Mother, Father and Child Cohort Study (MoBa) and the Medical Birth Registry of Norway. First, we analysed associations between health behaviours prior to pregnancy (alcohol and caffeine consumption, body-mass index (BMI) and smoking) and multiple indicators of fertility (including number of children, time to conception, and miscarriage) (n=83,128 women, 67,555 men), adjusting for birthyear, education and attention deficit and hyperactive-impulsive (ADHD) traits. Second, we used individual-level Mendelian randomisation (MR) to explore possible causal effects of health behaviours on fertility outcomes (n=27,216 women, 26,131 men). Finally, we performed summary-level MR for available outcomes (n=91,462-1,232,091) and conducted multi-variable MR to control for education and ADHD liability.

Results: In observational analysis, higher BMI and smoking (and to a lesser extent caffeine) were predominantly associated with reduced fertility outcomes. Unexpectedly, higher alcohol consumption was associated with predominantly improved fertility outcomes. There was little evidence from individual-level MR analyses, except smoking and higher BMI were associated with younger age at first birth in women (mean difference in years, per SD increase in genetic score; smoking: -2.65 (95% CI: -3.57, -1.73); BMI: -0.11 (95% CI: -0.16, -0.08)) and men (smoking: -2.82 (95% CI: -4.07, -1.58); BMI: -0.17 (95% CI: -0.23, -0.11)). These results were replicated in the summary-level MR analysis, however effects attenuated after adjusting for education and ADHD liability.

Conclusions: Most observational evidence for associations between health behaviours and fertility was not supported by MR analyses, suggesting possible residual confounding. Evidence from MR analyses supported an effect of smoking and higher BMI on younger age of first birth, but multivariable MR suggested this might be explained by underlying liability to ADHD and low educational attainment.
**Introduction**

Women struggling to conceive are often advised to engage in healthier lifestyle behaviours, for example, reducing their alcohol and caffeine consumption, achieving a healthy weight-range and quitting smoking (1–3). Reviews of observational evidence support an association between these health behaviours and reduced fertility in women (4), with high alcohol consumption and smoking associated with reduced likelihood of conception (5–9) and high caffeine consumption and obesity associated with increased risk of miscarriage (10,11). Guidelines and epidemiological research on fertility often focus on women (2,12,13) however, it is important to also consider the impact of health behaviours in the partner. Meta-analyses suggest that smoking and alcohol consumption can reduce semen quality (14), and paternal obesity has been associated with reduced likelihood of natural conception (15) and an increased time to conception (16).

The majority of evidence to date exploring health behaviours and fertility is observational, and most studies do not adequately control for confounding (17). For example, health behaviours often co-occur with other health behaviours which might instead affect fertility (e.g. diet, physical activity, sleep). Furthermore, health behaviours and fertility outcomes (such as age at first birth and number of children) share common risk factors including low educational attainment and liability to inattention and hyperactive-impulsive behaviour (traits of attention deficit hyperactivity disorder; ADHD)(18–20). In addition to confounding, it is necessary to account for reverse causation because the stress of being unable to conceive might cause couples to engage in unhealthy behaviours. We extended previous epidemiological studies using Mendelian randomisation (MR), which can reduce bias from residual confounding and reverse causation. MR uses genetic variants (single nucleotide polymorphisms; SNPs) to estimate the causal effect of an exposure on an outcome (21). Causal evidence could improve fertility guidance, helping couples successfully conceive as well as removing any unnecessary stress or guilt around unrelated lifestyle factors. The objective of the current study was to explore the associations between health behaviours and fertility outcomes in both men and women using an MR approach. We used the Norwegian Mother, Father and Child (MoBa) pregnancy cohort, containing around 95,000 mothers and 75,000 partners.

**Methods**

**Sample**

MoBa is a prospective population-based pregnancy cohort study conducted by the Norwegian Institute of Public Health. Pregnant women were recruited from all over Norway from 1999-2008 (22,23). Women consented to participation in 41% of the pregnancies. The cohort includes...
114 500 children, 95 200 mothers and 75 200 partners. The current study was based on version 12 of the quality-assured data files released for research on January 2019. Questionnaires were completed across multiple time points during pregnancy and after birth. The present study used measures from the first questionnaires sent to mothers and fathers, received between 13-17 weeks gestation (hereon referred to as the 15-week questionnaire). The questions related to previous pregnancies, medical history and medication use, occupation, exposures in the workplace and home, lifestyle habits and mental health. In addition to questionnaire data, information on maternal characteristics and pregnancy outcomes of the index pregnancies were available through data linkage to the Medical Birth Registry Norway (MBRN). MBRN also provided data on age at first birth and total number of children up to 2018, including (but not limited to) the MoBa index pregnancy. The establishment of MoBa and initial data collection was based on a license from the Norwegian Data Protection Agency and approval from The Regional Committees for Medical and Health Research Ethics. The MoBa cohort is now based on regulations related to the Norwegian Health Registry Act. The current study was approved by The Regional Committees for Medical and Health Research Ethics (2016/1702).

After restricting to those with available exposure and outcome data, a maximum of 83,128 women and 67,555 men were included in multivariable regression analysis. After further restricting to those individuals with quality-controlled genotype data (see details below), a maximum of 27,216 women and 26,131 men were available for individual-level MR analysis.

More details of participant exclusion are given in Supplementary Figure S1.

Health behaviours exposure data

For all health behaviours, we used data from the 15-week questionnaire in pregnancy. Women were asked to report their behaviour 3 months prior to the index pregnancy. Fathers were asked to report their behaviour 6 months prior to the index pregnancy.

Alcohol consumption. Frequency of alcohol consumption was self-reported on a 7-point scale (never, less than once a month, 1-3 times a month, once per week, 2-3 times per week, 4-5 times per week, 6-7 times per week). Women and men were also asked about binge drinking behaviour: “How often did you drink 5 units or more on one occasion?”. Responses were on a 5-point scale (never, less than once a month, 1-3 times a month, once a week, several times per week).

Caffeine consumption. Caffeine consumption in women was calculated from self-reported daily beverage consumption, where one cup contained 125 ml. Caffeine (mg) weights per cup were
taken from a previous study of caffeine consumption in MoBa (24). We excluded outliers if they
consumed more than 3.5 litres of any particular drink in a day (28 cups) or if their total caffeine
consumption per day was greater than 1000mg (25). Values were log transformed to adjust for
skewness. In men, beverage consumption was instead measured categorically over a typical
week. Responses were on a 5-point scale (seldom/never, 1-6 cups a week, 1 cup a day, 2-3 cups
a day, 4+ cups a day). Unlike the questionnaire administered to women, men were not directed
as to the volume of the cup, therefore we followed previous calculations (24) and assumed that a
cup was 125ml for coffee (apart from espresso where we assumed a standard single is 30ml)
and 250ml for tea or fizzy drink. Again consumption was weighted by caffeine (mg) (24) and
divided by 7 to estimate mg per day.

**Smoking behaviour.** Smoking initiation was self-reported: “Have you ever smoked?” where yes
was classed as ever smoking and no was classed as never smoking. Smoking heaviness was self-
reported amongst current smokers as the average number of cigarettes smoked per day prior to
pregnancy. We excluded current smokers who reported smoking no cigarettes.

*Body mass index (BMI).* Height and weight were self-reported. We used pre-pregnancy values for
women and current (15-weeks) values for men. Women were also asked to report their
partner’s height and weight. These reports were highly correlated with partners’ self-report
(r=0.98 for height and r=0.95 for weight). Therefore, we used the woman’s report of their
partner’s height and weight when the partner’s own report was unavailable. We excluded
outliers for women at height <117 or >196 cm and weight <38 or >150 kg and for men at height
<136 or >220 cm and weight <50 or >200 kg as done previously in MoBa (16). From these height
and weight measures we calculated BMI as weight (kg)/height (m²). We note BMI is not a health
behaviour itself but is a biomarker which can crudely proxy for healthy behaviours.

**Fertility outcome data**

All fertility outcomes were self-reported in the 15-week questionnaire by the women apart from
age at first birth and total number of children, which were obtained from the Medical Birth
Registry of Norway (MBRN) (version 12, last update November 2018).

**Age at first birth.** Age at the time of first recorded child’s delivery obtained from MBRN (not
limited to births recorded in MoBa).
Number of children. The total number of children born to women and men up to November 2018 obtained from MBRN (including but not limited to births recorded in MoBa). Totals of more than 6 children were grouped into a 6+ category to adjust for skewness.

Time to conception. Based upon women's self-report from the 15-week questionnaire referring to the index pregnancy. If women reported planning their pregnancy, they were asked: “For how many months did you have regular intercourse without contraception before you became pregnant?” Options were less than one month, 1-2 months or 3+ months. If it took more than 3 months, then they were asked to state the number of months. We combined anyone taking 12 or more months to conceive into one group to reduce skewness and treated as a continuous variable. We have used this woman-reported variable as an outcome in both women and men under the assumption that couples were conceiving together. In our primary analysis, if the couple were not trying to conceive then they were set to missing. However, given differences between couples who planned pregnancy and those who did not, we also conducted a sensitivity analysis where non-planners were included and assigned them the median time to conception from the planning group (2 months) (Supplementary Note 2). We also conducted a sensitivity analysis using dichotomised variables as time to conception was not measured continuously (Supplementary Note 3).

Infertility treatment. Women self-reported in the 15-week questionnaire: “Have you ever been treated for infertility?”. Responses were binary yes or no. We did not use this variable as an outcome in the fathers, as this question did not specify whether the infertility treatment was for the index pregnancy.

Miscarriage. Women self-reported (in the 15-week questionnaire) ever having had a miscarriage, defined as any of their previous pregnancies ending in spontaneous abortion at or before the 20th week of pregnancy.

As a secondary outcome, we also explored the impact of health behaviours on frequency of sexual intercourse (see Supplementary Note 4).

Genotype data
Blood samples were obtained from both parents during pregnancy and from children (umbilical cord) at birth (26). For the current study we used MoBa Genetics genotype data release 1.0 (https://github.com/folkehelseinstituttet/mobagen/wiki/MoBaGenetics1.0). This release has genotype data available for 99,137 individuals (mostly family trios). Details of the genotyping
and QC procedures are available in Supplementary Note 1 (27) and Supplementary Figure S1.
After quality control and relatedness checks, the remaining samples contained 28,929 women and 27,723 men.

**Genetic score construction**

For individual-level MR, our genetic instruments were individual-level genetic scores constructed using PRSice (28) and genome-wide significant variants were possible. SNPs were clumped to ensure independence ($r^2 < 0.01$, clumping window $< 1000$kb) and weighted by effect sizes from discovery GWAS detailed below (selected for being the largest samples using individuals of European ancestry and not containing the MoBa cohort). Prior to analysis, we checked that each genetic score explained significant variance in the exposure. This is presented in Supplementary Table S1 along with number of SNPs passing QC in the MoBa cohort.

**GWAS Summary Statistics for health behaviours**

The following GWAS summary statistics were used to construct genetic scores for individual-level MR, and individual SNP effects sizes were used in summary-level MR.

**Alcohol consumption.** We used two genetic instruments for alcohol consumption: 1) alcohol consumption frequency and 2) binge drinking. Alcohol consumption frequency was measured as average number of drinks per week aggregated across types of alcoholic beverage. The GWAS identified 99 conditionally independent genome-wide significant SNPs in a sample of 941,280 individuals, explaining 2.5% of the variance (29). Binge drinking in the UK Biobank was defined as “How often do you have six or more drinks on one occasion?”, where a drink was defined as a unit of alcohol. The GWAS was conducted by the Neale Lab (http://www.nealelab.is/uk-biobank – round 2, August 2018), and variance explained in an independent sample was not reported. After restricting to independent variants, 4 genome-wide significant SNPs remained. Given the small number of genome-wide significant SNPs, we also used a relaxed p-value threshold of $p<5\times 10^{-6}$, for which 12 SNPs were available.

**Caffeine consumption.** Caffeine consumption was measured as number of cups of coffee per day. The GWAS identified 6 independent genome-wide significant SNPs in a sample of 91,462 coffee drinkers of European ancestry (30). Genome-wide significant SNPs explained 1.3% of the variance in coffee consumption. Given the small number of genome-wide significant SNPs, we also used a relaxed p-value threshold of $p<5\times 10^{-6}$, for which 39 SNPs were available.
**Smoking behaviour.** We used two genetic instruments for smoking behaviour: 1) smoking initiation and 2) smoking heaviness. Smoking initiation was defined as ever v. never regularly smoking (more than 100 cigarettes ever or having ever been a daily smoker). The GWAS of smoking initiation identified 378 conditionally independent genome-wide significant SNPs, in a sample of 1,232,091 individuals, which explained 4% of the variance (29). Smoking heaviness was defined as average number of cigarettes smoked per day. The GWAS of smoking heaviness identified 55 conditionally independent genome-wide significant SNPs in a sample of 337,334 ever smokers, which explained 4% of the variance (29). We also conducted a single-SNP analysis of rs16969968 genotype from the CHRNA5 gene, known to reduce nicotine aversion and consequently increase cigarettes smoked per day (31) (see Supplementary Note 5).

**Body mass index (BMI).** The most recent GWAS of adult BMI identified 941 independent genome-wide significant SNPs in a sample of 681,275 individuals of European ancestry, which explain 6% of the variance in BMI (32).

**Statistical analysis**

All analyses were conducted in R version 4.0.3 (33) and performed separately for women and men. We corrected for multiple testing using a Bonferroni correction of 0.05/48 tests (6 exposures and 8 outcomes) which resulted in an adjusted p-value of p<0.001.

**Stage 1. Multivariable regression analyses.** We first explored the association between each of the fertility outcomes using correlation for continuous traits, chi-squared tests for binary traits and independent t-tests when one trait was continuous and the other binary. Second, we conducted linear regressions for each of the continuous fertility outcomes and logistic regressions for each of the binary fertility outcomes to see if we observed the expected association between health behaviours and fertility. Results are presented unadjusted and adjusted for birth year and educational attainment (at around 15-weeks pregnancy). An additional adjustment for ADHD traits was included as a sensitivity analysis to proxy for ADHD liability. ADHD traits were measured when the index child was age 3 years using the Adult ADHD Self-Report Scale (34).

**Stage 2. Individual-level Mendelian randomisation.** MR can be implemented as an instrumental variable analysis using genetic variants to proxy for an exposure. It can be used to estimate a causal effect of the exposure (health behaviours) on the outcome (fertility) providing certain assumptions are satisfied (21,35). The three core assumptions for valid causal inference are: 1) relevance – the genetic instrument must be robustly associated with the exposure, 2)
independence – there should be no confounding between the genetic instrument and outcome, and 3) exclusion-restriction – the genetic instrument must only be associated with the outcome via the exposure. Additionally, for results to generalise to other populations, there must be no “defiers” – individuals whose exposure is opposite to their genetic predisposition (35). Our individual-level MR analysis used individual-level genetic scores (with weights from external independent GWAS) in an instrumental variable regression controlling for age, genotyping chip and top 10 principal components of population structure. The genetic scores are first regressed on the exposure, and then predicted values are regressed onto the outcome. Analyses were performed using the `ivreg` command from the Applied Econometrics with R (`AER`) package for R. For continuous outcomes, betas are the mean difference per SD increase in the genetic score. When outcomes are binary, estimates approximate the risk difference per SD increase in genetic score.

Sensitivity analyses. We checked for evidence of assortative mating by using the woman’s genetic score to predict the men’s health behaviours and vice versa and by estimating the correlation between the genetic score of women and men. We checked for evidence of possible pleiotropy by testing whether each of the genetic scores predicted any known confounders of the exposure-outcome association (e.g., other health behaviours, income, age) and compared these estimates with the estimated association between observed exposures on confounders. Pleiotropy occurs when one genetic variant influences multiple phenotypes. If these other phenotypes are not on the causal pathway from exposure to outcome (horizontal pleiotropy) then the independence and exclusion-restriction assumptions could be violated, and genetic variants are invalid. Where there was evidence for a causal effect in the individual-level MR analysis, we followed up with additional summary-level MR sensitivity analyses (MR Egger (36), weighted median (37) and weighted mode (38)) which make different assumptions about the nature of pleiotropy. A consistent direction of effect across the different MR sensitivity analyses gives us more confidence that the effects are not due to pleiotropy. We also calculated the MR Egger intercept. If the intercept is significantly different from zero, then this suggests significant directional pleiotropy may be biasing the estimate. To conduct these summary-level sensitivity MR analyses, we generated SNP-outcome association results for the MoBa cohort using the `assoc` command for plink (version 2) (39).

Stage 3. Summary-level Mendelian randomisation. The second MR method used was summary-level MR (i.e. two-sample MR) which uses summary statistics from published GWAS (40). Here we don’t have an effect estimate for each individual but instead an effect size for each SNP from the discovery GWAS and an effect size for that SNP in an outcome GWAS. The ratio of
these two effect sizes can be meta-analysed across multiple SNPs to give us an estimate of the
causal effect. This is our primary analysis known as the inverse-variance weighted estimate. As
sensitivity analyses, again we conducted MR Egger (36), weighted median (37) and weighted
mode (38), which each make orthogonal assumptions about the nature of pleiotropy.

Independent summary GWAS data was available for three of the outcomes: age at first birth (N=
170,498), number of children (N=333,628) and miscarriage (N= 78,700), self-reported in the UK
Biobank (41). We used outcome GWAS from the UK Biobank only to prevent sample overlap
with our health behaviour exposure GWAS. For age at first birth and miscarriage, we used GWAS
summary statistics from the MRC IEU Open GWAS project (42). The GWAS for age at first birth
used the self-reported question "How old were you when you had your first child?" (field 2754)
asked only to women who had previously indicated that they had given birth to at least one
child. The GWAS for miscarriage used the item “How many spontaneous miscarriages have you
had?” (field 3839) which was only asked to women who had previously indicated that they had
ever had a miscarriage, abortion or stillbirth (field 2774). Finally, the GWAS for number of
children combined items "How many children have you fathered?" (field 2405) in men and "How
many children have you given birth to? (Please include live births only)" (field 2734) in women
(43). Age at first birth and miscarriage GWASs are in SD units and units for number of children
are number of children.

For each of the exposures, we used the same health behaviour GWAS as for the individual-level
MR, with the exception of BMI. Here we used an earlier GWAS (44) that did not contain the UK
Biobank to avoid sample overlap which can bias estimates (45). For smoking initiation and
alcohol consumption, we used SNP-exposure estimates from summary statistics with the UK
Biobank and 23andMe removed. Smoking heaviness could not be used as exposure as the
fertility outcome GWASs contain both smokers and non-smokers. Binge drinking could not be
included as this GWAS was also conducted in the UK Biobank.

Sensitivity analyses. The Cochran's Q test of heterogeneity was conducted to estimate possible
pleiotropy and the MR Egger intercept was estimated to test for bias from directional horizontal
pleiotropy. The regression dilution $R^2$ was calculated to assess the suitability of the MR Egger
effect estimate and a SIMEX correction applied where necessary (46). Steiger filtering was
conducted to check for evidence of reverse causation (47). The mean F statistic was calculated as
an indicator of instrument strength, where $F < 10$ is considered to indicate a weak instrument.
Where there was evidence for a causal effect, we conducted multivariable MR analysis (48) to
explore possible pleiotropy via education and ADHD liability (see Supplementary Materials for
details and Table S20).

Results

Average health behaviours and fertility outcomes were relatively consistent between the full
MoBa sample and the genotyped sub-sample (Table 1). Prevalence of smoking, high alcohol
consumption, high caffeine consumption and high BMI were greater in men than women prior to
pregnancy and men were older on average than women. Fertility outcomes did not differ
between women and men, with the exception of age at first birth which was older in men.

Results for frequency of sexual intercourse, and dichotomised time to conception are given in
Supplementary Notes 3 and 4. Associations between the different fertility outcomes found that a
younger age at first birth was associated with having more children in total, shorter time to
conception, being less likely to miscarry and less likely to have infertility treatment
(Supplementary Table S2). In the following section we highlight results that passed Bonferroni
correction.

Stage 1. Multivariable Regression Associations

The results of the observed associations between the health behaviours and fertility outcomes
(adjusted for birth year and education) are given in Figure 1 (Tables S3 and S4).

Alcohol Consumption. Units are per category increase in self-reported alcohol consumption or
binge drinking. In women’s adjusted analyses, greater frequency of alcohol consumption and
binge drinking were both associated with having fewer children (alcohol frequency: -0.090
children, 95% CI: -0.096, -0.085; binge drinking: -0.111 children, 95% CI: -0.118, -0.104) and an
older age at first birth (alcohol frequency: 0.492 years, 95% CI: 0.468, 0.516; binge drinking:
0.333 years, 95% CI: 0.304, 0.362). However, those who consumed more alcohol were less likely
to have had infertility treatment (alcohol frequency: OR = 0.867, 95% CI: 0.847, 0.888; binge
drinking: OR = 0.822, 95% CI: 0.798, 0.847). Binge drinking (but not alcohol frequency) was
associated with decreased odds of miscarriage (OR = 0.959, 95% CI: 0.938, 0.981). Neither
alcohol frequency nor binge drinking were associated with time to conception after correction
for multiple testing. The pattern of results was consistent for outcomes available in men.

Caffeine consumption. Units are per unit increase in log transformed mg of caffeine per day. In
men, higher caffeine consumption was associated with a younger age at first birth (-0.100 years,
95% CI: -0.139, -0.061) and being less likely to have fertility treatment (OR = 0.951, 95% CI:
Caffeine consumption was not associated with any of the other fertility outcomes in men nor with any fertility outcomes in women correction for multiple testing.

**Smoking behaviour.** Ever smoking was associated with having fewer children in women (-0.101 children, 95% CI: -0.114, -0.089), a younger age at first birth in both women and men (women: -0.155 years, 95% CI: -0.208, -0.103; men: -0.218 years, 95% CI: -0.284, -0.151), and an increased time to conception in women (0.095 months, 95% CI: 0.038, 0.151). Smoking heaviness was also associated with having fewer children in women (-0.008 children, 95% CI: -0.010, -0.005) and longer time to conception (women: 0.029 months, 95% CI: 0.016, 0.042; men: 0.021 months, 95% CI: 0.009, 0.034).

**Body mass index.** Units are per kg/m² increase in BMI. Higher BMI was associated with worse outcomes across all indicators for fertility in both sexes, including having fewer children (women: -0.004 children, 95% CI: -0.005, -0.002; men: -0.006 children, 95% CI: -0.008, -0.004) and taking longer to conceive (women: 0.061 months, 95% CI: 0.055, 0.068; men: 0.045 months, 95% CI: 0.035, 0.054). For women, having a higher BMI was also associated with being more likely to have fertility treatment (OR: 1.032, 95% CI: 1.026, 1.037) and being more likely to have a miscarriage (OR: 1.008, 95% CI: 1.004, 1.013). In men, higher BMI was associated with an older age at first birth (0.026 years, 95% CI: 0.016, 0.035). The only exception to this pattern of worsened fertility was an association between higher BMI in women and younger age of first birth (-0.017 years, 95% CI: -0.023, -0.011).

These results were relatively consistent, but less precise to those observed when restricting the analysis to the genotyped sample only and after adjustment for ADHD traits (Supplementary Tables S5 and S6).

**Stage 2. Individual-level Mendelian randomisation**

Genetic scores were associated with the exposures in MoBa with the exception of binge drinking, which was therefore not used in further analyses (Table S1). Results of the individual-level MR are presented in Figure 1 (and Tables S7 for women and S8 for men). All genetic scores were standardised, so that the results are per standard deviation increase in genetic score for the exposure. Genetic liability for higher BMI was associated with a younger age of first birth in both women (-0.113 years, 95% CI: -0.153, -0.073) and men (-0.171 years, 95% CI: -0.232, -0.111). Genetic liability for smoking initiation was also strongly associated with a younger age of first birth in both women (-2.647 years, 95% CI: -3.565, -1.730) and men (-2.824 years, 95% CI: -4.072, -1.575). However, stratified by smoking status, rs16969968 genotype was only associated
with age at first birth in never smokers, suggesting possible pleiotropy (Supplementary Note 5 and Figure S2). We found no robust evidence for association between any of the other exposure and outcomes using individual-level MR.

For both BMI and smoking initiation genetic scores there was evidence of assortative mating (Table S9) and associations between these genetic scores and the other health behaviours and household income (Table S10). Additional sensitivity analyses which are more robust to pleiotropy were consistent and there was no evidence for bias from directional horizontal pleiotropy (Table S11). Sensitivity analyses in women including non-planners with a median time to conception suggested an association between higher alcohol consumption and decreased time to conception (-2.653 months, 95% CI: -4.193, -1.112) and between higher BMI and increased time to conception (0.969 months, 95% CI: 0.811, 1.129), that were not observed when restricting to planners (see Supplementary Note 2 for more details).

Stage 3. Summary-level Mendelian randomisation

There was strong evidence for an effect of smoking initiation on younger age at first birth in women (-0.661, 95% CI: -0.757, -0.566), a greater number of children in both men and women (0.280, 95% CI: 0.205, 0.355) and fewer miscarriages in women (-0.123, 95% CI: -0.182, -0.064) (see Table 2). All pleiotropy robust sensitivity analyses showed consistently strong evidence with the same direction of effect. There was evidence of significant heterogeneity (Table S16), but the MR Egger intercept suggested that these results were not biased by directional pleiotropy (Table S17). Steiger filtering indicated that the majority of SNPs explained more variance in smoking initiation than the outcomes (Table S18), suggesting reverse causation is unlikely.

There was also some evidence for an effect of higher BMI on younger age at first birth in women (-0.108, 95% CI: -0.16, -0.056). All MR sensitivity analyses showed a consistent direction of effect apart from the weighted mode which could indicate possible pleiotropy (Table 2). There was strong evidence of heterogeneity (Table S16), but the MR Egger intercept did not suggest this was due to bias from horizontal pleiotropy (Table S17) and there was no evidence of reverse causation from Steiger filtering (Table S18). There was no evidence for an effect of alcohol consumption or caffeine consumption on reproductive behaviours. All genetic instruments had F-statistic > 10 apart from that for smoking initiation (Table S19).

We conducted multivariable MR analysis (48) to estimate the direct effects of smoking initiation and BMI on age at first birth in women after accounting for education and ADHD liability (see
Supplementary Materials for details and Table S20). Effect estimates for smoking initiation attenuated after adjustment but there was still evidence for an effect (after adjusting for ADHD: -0.435, 95% CI: -0.591, -0.279; after adjusting for education: -0.403, 95% CI: -0.527, -0.279). Effect estimates for BMI were also attenuated, resulting in weak evidence for an effect (ADHD: -0.513, 95% CI: -0.106, 0.003; education: -0.056, 95% CI: -0.113, 0.0008).

Discussion
This study explored the role of multiple health behaviours on fertility outcomes in men and women using the MoBa cohort. We extended previous research by including men as well as women, and strengthening causal inference using MR. We found evidence from multivariable regression for an association between higher BMI and smoking (and to a lesser extent caffeine) prior to pregnancy on reduced fertility outcomes. Our MR results were not in line with the observational associations, finding only support for an effect of higher BMI and smoking initiation on a younger age at first birth, with evidence of possible horizontal pleiotropy from ADHD liability and education.

Frequency of alcohol consumption was associated with being less likely to have infertility treatment in multivariable regression analyses for both women and men. In contrast, previous meta-analyses have shown a dose-response relationship between alcohol consumption and reduced likelihood of conception in women (5) and reduced semen quality in men (14). However, it is important to note that our analyses were conducted in a pregnancy cohort, so do not capture those who failed to conceive, introducing possible selection bias. Furthermore, levels of alcohol consumption were low in our sample, and are possibly below the threshold which affects fertility (49). To capture more harmful drinking behaviours, we used a measure of binge drinking, which showed weaker associations, supporting our interpretation. It might be feasible that low levels of alcohol consumption are not detrimental for fertility, but it seems unlikely that alcohol could improve fertility, as observed here. Highly confounded multivariable regression associations of alcohol consumption are a common phenomenon, with low levels of consumption often being associated with positive outcomes due to confounding by socio-economic position, or due to never drinkers being a selected group (50). Bias from confounding is supported by our MR analyses, finding no evidence for effects of alcohol consumption on any fertility outcomes. Alternatively, our observational results might be affected by reporting bias, as underreporting of alcohol consumption might be more pronounced amongst those struggling to conceive. Unfortunately, we were unable to follow up binge drinking associations with MR, as the genetic instrument did not predict binge drinking in the MoBa cohort. This is not altogether surprising, as MoBa represents a very different sample to that of the UK Biobank, where the
binge drinking GWAS was conducted: 1) the UK Biobank were aged between 40 and 69 years at recruitment (51) compared with an average age of 31 years in MoBa mothers and 33 years in MoBa fathers, 2) response rate was higher for the MoBa cohort (41%) compared to the UK Biobank (5%) (22, 52), 3) drinking behaviours differ greatly between the UK and Norway (53), with only 6.31% of MoBa mums reporting binge drinking at least weekly in the 3 months prior to pregnancy, and 4) MoBa is a pregnancy cohort rather than a population cohort, capturing only fertile couples.

There was no evidence for an association between caffeine consumption and miscarriage risk, in contrast to what has been reported in several meta-analyses (54). This could be because our study explored reported caffeine consumption levels prior to pregnancy rather than during pregnancy, which has been the primary focus of most previous meta-analyses (54).

Alternatively, it could be due to social patterning of caffeine consumption in Norway, with higher consumption associated with lower levels of education (24) and consequently a younger age at first birth. Older age is a strong predictor of miscarriage risk (55), so education could be masking the association. MR results (which are more robust to bias from confounding) did not find evidence for a causal effect, so it could be possible that previous associations were due to confounding from other lifestyle factors (56). However, it is important to note that the caffeine genetic instrument was the weakest, explaining only 0.1-0.2% of the variance. Due to this weak instrument bias, causal effects cannot be ruled out (57) and replication is warranted.

We found evidence for an association between smoking heaviness and increased time to conception in both women and men. This is supported by a previous meta-analysis finding increased odds of infertility in smokers compared to non-smokers (6). Smoking is hypothesised to negatively affect sperm production and quality in males (58) and to affect the follicles and hormone levels in females (49). However, the effect sizes observed in the current study were quite small and were not supported by the MR results.

Finally, we saw evidence from multivariable regressions for associations between higher BMI and reduced fertility, including taking longer to conceive and an increased risk of miscarriage. These associations are thought to be due to obesity leading to hormone imbalances and ovulatory dysfunction (similar to very low body weights) (59). However, we did not see evidence that this effect was causal in the MR analyses. This could be due to confounding or due to the MR analysis being unable to detect non-linear effects. Another MR analysis of BMI on subfertility in the MoBa cohort did find evidence for non-linear effects (60). Intervention evidence is mixed, with a meta-analysis including randomised control trials finding evidence
that interventions for weight loss increased chance of pregnancy but did not affect risk of miscarriage (61).

A review of studies into health behaviours and fertility concluded that the evidence was robust for smoking and higher weight reducing fertility (49). However, in the current study, our MR results did not find evidence to support this effect. In both the individual-level and summary-level MR we saw consistent evidence for an effect of smoking initiation and higher BMI on having a younger age at first birth. There is also the possibility that these results are due to selection bias: both the GWAS for age at first birth and being a participant in the MoBa sample are conditioned on having had at least one child. If BMI and smoking are associated with reduced fertility, perhaps only those who had children younger were able to conceive and consequently be in the sample. However, if this were the case, we might expect to also see effects of smoking and BMI on other indicators of fertility. An alternative explanation could be that our estimates are biased by horizontal pleiotropy. We used a range of sensitivity analyses to explore potential assumption violation by horizontal pleiotropy. Methods which are agnostic to the specific sources of pleiotropy, (for example, MR-Egger), suggested that the results were not importantly biased by unbalanced horizontal pleiotropy. However, our exploratory multivariable MR analyses did show strong attenuation of the effects of BMI and smoking initiation on age at first birth after accounting for ADHD liability (proxy for ADHD traits of inattentive and hyperactive-impulsive behaviour) and educational attainment (known risk factors for age at first birth). This is likely because ADHD traits affect smoking behaviour and some of the smoking genetic variants relate to smoking via their relationship to ADHD traits. Previous studies have shown a strong association between the smoking initiation instrument and risk-taking behaviours including number of sexual partners (62), which could increase the likelihood of having children younger. Previous Mendelian randomisation studies have also found evidence for bi-directional causal effects between smoking and education (63,64), smoking and ADHD (65), BMI and education (66,67) and BMI and ADHD (68). Bi-directional effects between the exposure and the non-exposure traits can make it difficult to disentangle horizontal from vertical pleiotropy (69). However, several of these previous studies did find evidence of horizontal pleiotropy, especially for the smoking initiation instrument (64,65,68). Therefore, we conclude that horizontal pleiotropy is the most plausible pathway. If there is indeed horizontal pleiotropy from ADHD liability and education, then direct effects from MVMR accounting for these traits will be closer to the true causal effect. Finally, there was also strong evidence for assortative mating for both BMI and smoking initiation instruments which can bias MR results, even for methods which are robust to horizontal pleiotropy (70).
Strengths and Limitations

The current study has several strengths. The majority of epidemiological research to date has focused on women (13), but we also included fathers in our analysis. We observed a similar pattern of results for women and men in our multivariable regression analyses, despite different hypothesised mechanisms. This supports our interpretation of widespread confounding, because pathways through different mechanisms are unlikely to have the same magnitude of effect (71). Second, this was a very large sample of genotyped individuals with detailed measures of fertility. Third, we combined multivariable regression and MR methods which each rely on different assumptions and therefore triangulating across them can strengthen causal inference. Finally, we included a wide range of health behaviours and different indicators of fertility.

This study does have several limitations. First, all MoBa participants were recruited during pregnancy (12-18 weeks gestation). This means that we are unable to capture the full range of fertility in the population. Those who never managed to conceive were not observed, and this could induce selection bias. Second, multivariable regression analyses were cross-sectional and it is therefore difficult to assess temporality for these associations. Specifically, health behaviours were retrospectively reported about behaviours 3 or 6 months prior to the index pregnancy however some couples had been trying to conceive for longer than 6 months. Furthermore, variables from the Medical Birth Registry of Norway (age at first birth and total number of children) are across all births, and therefore health behaviours may have differed compared to before the index pregnancy. Relatedly, there was also a difference between those in the sample who were planning to conceive compared with those who were not. We hypothesise that planners are more likely to be cautious about their health behaviours, especially if they have been having trouble conceiving and have been advised to quit smoking, stop drinking and lose weight. This could lead to reverse or weakened patterns of association in the multivariable regression analyses. However, MR would be robust to this type of bias, given that genetic propensities to health behaviours are fixed at birth. This might explain the different pattern of results between the multivariable regression and MR analyses. Third, for this paper we have assumed that partners in MoBa were men. In our genetic analyses this is the case, because individuals who were not chromosomally XY were removed from the father analysis. However, in the observational analysis, some partners might have been female partners of the mother, and these individuals could not be identified. Finally, this study conducted triangulation within a single sample, and an approach for stronger causal inference would be to triangulate across different samples in future studies.
Conclusions

For accurate fertility guidance it is extremely important to establish causality. The majority of associations between health behaviours and fertility observed here, were not replicated using MR methods. This could suggest that previous observational associations were due to confounding or other sources of bias, and consequently improving these health behaviours may not increase the likelihood of conception. We found evidence of potential horizontal pleiotropy, as our genetic instruments for smoking initiation and BMI were also capturing educational attainment and ADHD liability. Therefore, triangulation across a broader range of methods, including those not susceptible to pleiotropy, are required to establish causality.
Funding and Acknowledgements

This work was funded by a Gro Harlem Bruntland Scholarship from the Centre of Fertility and Health NIPH to Dr Wootton. Dr Wootton is currently funded by a postdoctoral fellowship from the South-Eastern Regional Health Authority (2020024). AH was supported by a career grant from the South-Eastern Regional Health Authority (2020022), and by the Research Council of Norway (300668, 288083, 274611). TRK was supported by the Research Council of Norway (274611). PM and SEH were partly funded by the Research Council of Norway (262700 and 320656). GDS and MR M is part of the Medical Research Council Integrative Epidemiology Unit (MRC IEU) at the University of Bristol (MM_UU_00011/1, MM_UU_00011/7). JLT was supported by a Veni grant from the Netherlands Organization for Scientific Research (NWO: grant number 016.Veni.195.016) and by the Foundation Volksbond Rotterdam. RBL is supported by the National Institute of Mental Health grant (R01 MH101269).

This study includes data from the Norwegian Mother, Father and Child Cohort Study (MoBa) conducted by the Norwegian Institute of Public Health. The Norwegian Mother, Father and Child Cohort Study is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research. We are grateful to all the participating families in Norway who take part in this on-going cohort study. We thank the Norwegian Institute of Public Health (NIPH) for generating high-quality genomic data. This research is part of the HARVEST collaboration, supported by the Research Council of Norway (#229624). For providing genotype data we thank deCODE Genetics, and the NORMENT Centre for providing genotype data, funded by the Research Council of Norway (#223273), South East Norway Health Authority and KG Jebsen Stiftelsen. We further thank the Center for Diabetes Research, the University of Bergen for providing genotype data and performing quality control and imputation of the data funded by the ERC AdG project SELECTionPREDISPOSED, Stiftelsen Kristian Gerhard Jebsen, Trond Mohn Foundation, the Research Council of Norway, the Novo Nordisk Foundation, the University of Bergen, and the Western Norway health Authorities (Helse Vest). This research was also supported by the Research Council or Norway through its Centres of Excellence funding scheme (project No 262700). MCM has funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement No 947684).

Data availability

The consent and ethical approvals for MoBa does not allow storage of data in repositories or journals, but it is possible to apply for access to summary statistics datasets for replication or reproduction of studies by sending an application to datatilgang@fhi.no. Data access requires
approval from The Regional Committee for Medical and Health Research Ethics in Norway and an agreement with MoBa.

**Ethical statement**

The establishment and data collection in MoBa was previously based on a license from the Norwegian Data protection agency and approval from The Regional Committee for Medical Research Ethics, and it is now based on regulations related to the Norwegian Health Registry Act. The current study was approved by The Regional Committees for Medical and Health Research Ethics (2016/217). All GWAS summary statistics are publicly available.
References


Table 1. Descriptive statistics comparing exposure and outcome data across mothers and fathers.

<table>
<thead>
<tr>
<th></th>
<th>Mothers Full Sample</th>
<th>Genotyped Sample</th>
<th>Fathers Full Sample</th>
<th>Genotyped Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>N = 94,418</td>
<td>Mean (SD)/%</td>
<td>N = 28,849</td>
<td>Mean (SD)/%</td>
</tr>
<tr>
<td>BMI (kg/m2)</td>
<td>83,191</td>
<td>30.61 (20.66)</td>
<td>27,272</td>
<td>24.05 (4.30)</td>
</tr>
<tr>
<td>Alcohol Consumption</td>
<td>79,562</td>
<td></td>
<td>26,819</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>5,490</td>
<td>7.15%</td>
<td>1,885</td>
<td>7.03%</td>
</tr>
<tr>
<td>Less than once a month</td>
<td>23,618</td>
<td>29.69%</td>
<td>7,848</td>
<td>29.26%</td>
</tr>
<tr>
<td>1-3 times a month</td>
<td>28,575</td>
<td>35.92%</td>
<td>9,004</td>
<td>36.56%</td>
</tr>
<tr>
<td>Once per week</td>
<td>14,263</td>
<td>17.93%</td>
<td>4,811</td>
<td>17.94%</td>
</tr>
<tr>
<td>2-3 times per week</td>
<td>6,527</td>
<td>8.20%</td>
<td>2,188</td>
<td>8.16%</td>
</tr>
<tr>
<td>4-5 times per week</td>
<td>724</td>
<td>0.91%</td>
<td>221</td>
<td>0.82%</td>
</tr>
<tr>
<td>6-7 times per week</td>
<td>165</td>
<td>0.21%</td>
<td>62</td>
<td>0.23%</td>
</tr>
<tr>
<td>Binge Drinking</td>
<td>Total</td>
<td>N = 78,615</td>
<td>Mean (SD)/%</td>
<td>N = 26,546</td>
</tr>
<tr>
<td></td>
<td>No report</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Never</td>
<td>26,229</td>
<td>33.36%</td>
<td>8,707</td>
<td>32.80%</td>
</tr>
<tr>
<td>Less than once a month</td>
<td>31,002</td>
<td>39.44%</td>
<td>10,647</td>
<td>40.11%</td>
</tr>
<tr>
<td>1-3 times a month</td>
<td>16,428</td>
<td>20.90%</td>
<td>5,606</td>
<td>21.12%</td>
</tr>
<tr>
<td>Once a week</td>
<td>4,313</td>
<td>5.49%</td>
<td>1,411</td>
<td>5.32%</td>
</tr>
<tr>
<td>Several times per week</td>
<td>175</td>
<td>0.82%</td>
<td>706</td>
<td>2.46%</td>
</tr>
<tr>
<td>Smoking initiation</td>
<td>Total</td>
<td>N = 82,824</td>
<td>Mean (SD)/%</td>
<td>N = 27,601</td>
</tr>
<tr>
<td></td>
<td>No report</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Never smokers</td>
<td>41,326</td>
<td>49.90%</td>
<td>13,897</td>
<td>50.35%</td>
</tr>
<tr>
<td>Ever smokers</td>
<td>41,498</td>
<td>50.10%</td>
<td>13,704</td>
<td>49.65%</td>
</tr>
<tr>
<td>Smoking Heaviness</td>
<td>(cigarettes per day)</td>
<td>Total</td>
<td>N = 16,074</td>
<td>Mean (SD)/%</td>
</tr>
<tr>
<td></td>
<td>No report</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Never smokers</td>
<td>10,744</td>
<td>61.41%</td>
<td>4,961</td>
<td>61.18%</td>
</tr>
<tr>
<td>Ever smokers</td>
<td>5,330</td>
<td>38.59%</td>
<td>1,413</td>
<td>38.82%</td>
</tr>
<tr>
<td>Caffeine Consumption (mg per day)</td>
<td>Total</td>
<td>N = 75,172</td>
<td>Mean (SD)/%</td>
<td>N = 24,512</td>
</tr>
<tr>
<td></td>
<td>No report</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Never smokers</td>
<td>14,160</td>
<td>141.60 (138.36)</td>
<td>25,142</td>
<td>139.40 (136.63)</td>
</tr>
<tr>
<td>Ever smokers</td>
<td>60,012</td>
<td>217.15 (465.65)</td>
<td>2,369</td>
<td>273.22 (442.42)</td>
</tr>
<tr>
<td>Age at First Birth (years)</td>
<td>Total</td>
<td>N = 94,740</td>
<td>Mean (SD)/%</td>
<td>N = 28,871</td>
</tr>
<tr>
<td></td>
<td>No report</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Never smokers</td>
<td>27,154</td>
<td>217.15 (465.65)</td>
<td>2,369</td>
<td>273.22 (442.42)</td>
</tr>
<tr>
<td>Ever smokers</td>
<td>67,586</td>
<td>217.15 (465.65)</td>
<td>2,369</td>
<td>273.22 (442.42)</td>
</tr>
<tr>
<td>Number of children N children)</td>
<td>Total</td>
<td>N = 94,740</td>
<td>Mean (SD)/%</td>
<td>N = 28,871</td>
</tr>
<tr>
<td></td>
<td>No report</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Never smokers</td>
<td>27,154</td>
<td>217.15 (465.65)</td>
<td>2,369</td>
<td>273.22 (442.42)</td>
</tr>
<tr>
<td>Ever smokers</td>
<td>67,586</td>
<td>217.15 (465.65)</td>
<td>2,369</td>
<td>273.22 (442.42)</td>
</tr>
<tr>
<td>Time to Conception (months)</td>
<td>Total</td>
<td>N = 63,068</td>
<td>Mean (SD)/%</td>
<td>N = 21,749</td>
</tr>
<tr>
<td></td>
<td>No report</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Never smokers</td>
<td>27,154</td>
<td>4.63 (7.04)</td>
<td>2,369</td>
<td>4.63 (7.04)</td>
</tr>
<tr>
<td>Ever smokers</td>
<td>65,914</td>
<td>4.63 (7.04)</td>
<td>2,369</td>
<td>4.63 (7.04)</td>
</tr>
<tr>
<td>Infertility treatment</td>
<td>Total</td>
<td>N = 83,876</td>
<td>Mean (SD)/%</td>
<td>N = 27,616</td>
</tr>
<tr>
<td></td>
<td>No report</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Never smokers</td>
<td>76,195</td>
<td>9.08%</td>
<td>25,253</td>
<td>9.14%</td>
</tr>
<tr>
<td>Ever smokers</td>
<td>7,681</td>
<td>9.16%</td>
<td>2,363</td>
<td>8.56%</td>
</tr>
<tr>
<td>Miscarriage</td>
<td>Total</td>
<td>N = 55,696</td>
<td>Mean (SD)/%</td>
<td>N = 17,896</td>
</tr>
<tr>
<td></td>
<td>No report</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Never smokers</td>
<td>39,114</td>
<td>70.23%</td>
<td>12,617</td>
<td>70.50%</td>
</tr>
<tr>
<td>Ever smokers</td>
<td>16,582</td>
<td>29.77%</td>
<td>5,279</td>
<td>29.50%</td>
</tr>
</tbody>
</table>

Note. * = supplemented with mother’s report when father’s report was unavailable. ∇ = Father variable reported by the mother.
Figure 1. The association between health behaviours and fertility outcomes comparing mothers and fathers, and comparing multivariable regression associations with estimates of causal effects from Mendelian randomisation. For continuous outcomes, units are betas. For binary outcomes, regression units are on the log odds scale and MR units are risk differences.
Table 2. Summary level Mendelian randomisation results for health behaviours on age at first birth, number of children and number of miscarriages

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Outcome</th>
<th>Method</th>
<th>N SNP</th>
<th>Beta (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinks per week</td>
<td>Age at First Birth</td>
<td>IVW</td>
<td>94</td>
<td>-0.055 (-0.145, 0.034)</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>MR Egger</td>
<td>94</td>
<td>0.04 (-0.095, 0.175)</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weighted median</td>
<td>94</td>
<td>-0.048 (-0.139, 0.042)</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weighted mode</td>
<td>94</td>
<td>-0.046 (-0.129, 0.038)</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of children</td>
<td>IVW</td>
<td>83</td>
<td>0.027 (-0.057, 0.110)</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>MR Egger (SIMEX)</td>
<td>83</td>
<td>0.014 (-0.029, 0.056)</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weighted median</td>
<td>83</td>
<td>-0.010 (-0.094, 0.074)</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weighted mode</td>
<td>83</td>
<td>0.009 (-0.068, 0.086)</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of miscarriages</td>
<td>IVW</td>
<td>94</td>
<td>-0.029 (-0.097, 0.038)</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>MR Egger</td>
<td>94</td>
<td>-0.018 (-0.121, 0.084)</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weighted median</td>
<td>94</td>
<td>0.001 (-0.099, 0.102)</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weighted mode</td>
<td>94</td>
<td>0.011 (-0.084, 0.103)</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>Caffeine</td>
<td>Age at First Birth</td>
<td>IVW</td>
<td>6</td>
<td>-0.036 (-0.13, 0.057)</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>Weighted median</td>
<td>6</td>
<td>0.013 (-0.073, 0.099)</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weighted mode</td>
<td>6</td>
<td>0.033 (-0.067, 0.134)</td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of children</td>
<td>IVW</td>
<td>6</td>
<td>0.021 (-0.031, 0.073)</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>Weighted median</td>
<td>6</td>
<td>0.021 (-0.047, 0.089)</td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of miscarriages</td>
<td>IVW</td>
<td>6</td>
<td>-0.042 (-0.106, 0.022)</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>Weighted median</td>
<td>6</td>
<td>-0.052 (-0.128, 0.025)</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weighted mode</td>
<td>6</td>
<td>-0.076 (-0.178, 0.025)</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>Smoking initiation</td>
<td>Age at First Birth</td>
<td>IVW</td>
<td>322</td>
<td>-0.661 (-0.757, -0.566)</td>
<td>3.57 x 10^{-6}</td>
</tr>
<tr>
<td></td>
<td>Weighted median</td>
<td>322</td>
<td>-0.504 (-0.603, -0.405)</td>
<td>2.71 x 10^{-2}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weighted mode</td>
<td>322</td>
<td>-0.555 (-0.785, -0.326)</td>
<td>3.30 x 10^{-6}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of children</td>
<td>IVW</td>
<td>323</td>
<td>0.280 (0.205, 0.355)</td>
<td>3.17 x 10^{-3}</td>
</tr>
<tr>
<td></td>
<td>Weighted median</td>
<td>323</td>
<td>0.207 (0.125, 0.289)</td>
<td>7.11 x 10^{-6}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weighted mode</td>
<td>323</td>
<td>0.177 (-0.009, 0.363)</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of miscarriages</td>
<td>IVW</td>
<td>322</td>
<td>-0.123 (-0.182, -0.064)</td>
<td>4.68 x 10^{-5}</td>
</tr>
<tr>
<td></td>
<td>Weighted median</td>
<td>322</td>
<td>-0.106 (-0.196, -0.017)</td>
<td>0.019</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weighted mode</td>
<td>322</td>
<td>-0.176 (-0.399, 0.047)</td>
<td>0.124</td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>Age at First Birth</td>
<td>IVW</td>
<td>95</td>
<td>-0.108 (-0.16, -0.056)</td>
<td>4.32 x 10^{-6}</td>
</tr>
<tr>
<td></td>
<td>MR Egger</td>
<td>95</td>
<td>-0.076 (-0.225, 0.072)</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weighted median</td>
<td>95</td>
<td>-0.043 (-0.101, 0.015)</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weighted mode</td>
<td>95</td>
<td>0.006 (-0.114, 0.126)</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of children</td>
<td>IVW</td>
<td>91</td>
<td>-0.014 (-0.054, 0.027)</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>MR Egger</td>
<td>91</td>
<td>-0.002 (-0.117, 0.113)</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weighted median</td>
<td>91</td>
<td>-0.008 (-0.054, 0.030)</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weighted mode</td>
<td>91</td>
<td>-0.012 (-0.094, 0.069)</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of miscarriages</td>
<td>IVW</td>
<td>95</td>
<td>0.004 (-0.033, 0.041)</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>MR Egger</td>
<td>95</td>
<td>0.021 (-0.086, 0.128)</td>
<td>0.70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weighted median</td>
<td>95</td>
<td>0.019 (-0.033, 0.072)</td>
<td>0.47</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weighted mode</td>
<td>95</td>
<td>0.013 (-0.091, 0.116)</td>
<td>0.81</td>
<td></td>
</tr>
</tbody>
</table>

Note. Caffeine genetic instrument using a relaxed p-value threshold of p<5x10^{-6}. IVW = inverse-variance weighted; BMI = body mass index. The 12Gx suggested that the smoking initiation and caffeine genetic instruments were not suitable for conducting MR Egger (Supplementary Table S14). Unweighted SIMEX corrections were conducted for smoking initiation.