

## **Smoking and subfertility: multivariable regression and Mendelian randomization analyses in the Norwegian Mother, Father and Child Cohort Study**

**Running title:** Smoking and subfertility

Álvaro Hernáez<sup>1,2,3,\*</sup>, Robyn E. Wootton<sup>4,5</sup>, Christian M. Page<sup>1,6</sup>, Karoline H. Skåra<sup>1</sup>, Abigail Fraser<sup>4,7,8</sup>, Tormod Rogne<sup>1,9,10</sup>, Per Magnus<sup>1</sup>, Pål R. Njølstad<sup>11,12</sup>, Ole A. Andreassen<sup>13,14</sup>, Stephen Burgess<sup>15,16</sup>, Deborah A. Lawlor<sup>4,7,8</sup>, Maria Christine Magnus<sup>1,4,7</sup>

1. Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway.
2. Consorcio CIBER, M.P. Fisiopatología de la Obesidad y Nutrición (CIBEROBN), Instituto de Salud Carlos III, Madrid, Spain.
3. Blanquerna School of Health Sciences, Universitat Ramon Llull, Barcelona, Spain.
4. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK.
5. Nic Waals Institute, Lovisenberg Diaconal Hospital, Oslo, Norway.
6. Department of Mathematics, University of Oslo, Oslo, Norway.
7. Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK.
8. NIHR Bristol Biomedical Research Centre, Bristol, UK
9. Department of Chronic Disease Epidemiology, Yale University School of Public Health, New Haven, Connecticut, USA.
10. Gemini Center for Sepsis Research, Department of Circulation and Medical Imaging, NTNU Norwegian University of Science and Technology, Trondheim, Norway.
11. Department of Clinical Science, Center for Diabetes Research, University of Bergen, Bergen, Norway

12. Children and Adolescent Clinic, Haukeland University Hospital, Bergen, Norway
13. Norwegian Centre for Mental Disorders Research, NORMENT, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway
14. Institute of Clinical Medicine, University of Oslo, Oslo, Norway
15. MRC Biostatistics Unit, University of Cambridge, Cambridge, UK.
16. Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.

**Corresponding author:**

Alvaro Hernaez, PharmD, PhD

Centre for Fertility and Health, Norwegian Institute of Public Health

PO 222 Skøyen, 0213 Oslo, Norway

Telephone: (+47) 95523951

e-mail: [alvaro.hernaez@fhi.no](mailto:alvaro.hernaez@fhi.no)

ORCID: 0000-0001-8593-1477

1 **ABSTRACT**

2

3 **Objective.** To investigate the association between smoking-related traits and subfertility.

4 **Design.** Prospective study.

5 **Setting.** Nationwide cohort in Norway.

6 **Patients.** 28,606 women (average age 30) and 27,096 men (average age 33) with  
7 questionnaire and genotype information from the Norwegian Mother, Father and Child  
8 Cohort Study.

9 **Intervention.** Self-reported information on smoking (having ever smoked [both sexes], age  
10 at smoking initiation [women only], smoking cessation [women only], and cigarettes  
11 smoked per week in current smokers [both sexes]) was gathered. Genetically  
12 predetermined levels or likelihood of presenting the mentioned traits were estimated for  
13 Mendelian randomization (MR) analyses.

14 **Main outcome measure.** Subfertility, defined as time-to-pregnancy  $\geq 12$  months.

15 **Results.** A total of 10% of couples were subfertile. In multivariable regression accounting  
16 for age, years of education, body mass index, and number of previous pregnancies,  
17 having ever smoked was not linked to subfertility in women or men. A higher intensity of  
18 tobacco use in women who were current smokers was related to greater odds of  
19 subfertility (+ 1 standard deviation [SD, 48 cigarettes/week]: odds ratio [OR] 1.12, 95%  
20 confidence interval [CI] 1.03 to 1.21), also after adjusting for the partner's tobacco use.  
21 Later smoking initiation (+ 1 SD [3.2 years]: OR 0.89, 95% CI 0.84 to 0.95) and smoking  
22 cessation (relative to not quitting: OR 0.83, 95% CI 0.75 to 0.93) were linked to decreased  
23 subfertility in women who had ever smoked. Nevertheless, MR results were not  
24 directionally consistent for smoking intensity and cessation and were imprecisely  
25 estimated in two-sample MR, with wide confidence intervals that overlapped with the  
26 multivariable regression results. In men, greater smoking intensity was marginally linked to

27 greater odds of subfertility in multivariable analyses, but this association was attenuated  
28 when adjusting for the partner's smoking intensity (+ 1 SD [54 cigarettes/week]: OR 1.05,  
29 95% CI 0.96 to 1.15). MR estimates were directionally consistent but again imprecisely  
30 estimated.

31 **Conclusions.** We did not find robust evidence of an effect of smoking on subfertility. This  
32 may be due to a true lack of effect, weak genetic instruments, or other kinds of  
33 confounding. The relevant limitations across all methods highlights the need for larger  
34 studies with information on subfertility.

## 35 INTRODUCTION

36 Smoking is a well-known source of thousands of pro-oxidative and pro-inflammatory  
37 compounds (1), capable of damaging reproductive tissues which in turn may compromise  
38 fecundity (2-4). Observational studies in women have reported that active smoking was  
39 linked to 14% greater odds of subfertility (trying to conceive for  $\geq 12$  months) (5) and  
40 smoking intensity was dose-dependently associated with greater subfertility risk (6).  
41 Tobacco use has also been related to surrogate indicators of decreased fertility such as  
42 accelerated follicular depletion and earlier menopause (7, 8). Although smoking has been  
43 linked to oligozoospermia and morphological defects of sperm (9), two prospective studies  
44 reported no association between tobacco use and subfertility in men (10, 11). Considering  
45 this evidence, the Practice Committee of the American Society of Reproductive Medicine  
46 has suggested a causal effect of tobacco use on clinical subfertility (12). However, existing  
47 evidence is subject to numerous methodological limitations. The primary concerns raised  
48 were that the relationship between smoking and decreased fertility has not been shown to  
49 be sufficiently strong, residual confounding cannot be ruled out, and most studies were  
50 retrospective and thus unable to reveal any potential exposure-to-effect sequence (12).

51 The use of complementary methodological approaches with different strengths and  
52 sources of bias could help clarify whether there is a causal relationship between tobacco  
53 use and subfertility (13). Mendelian randomization (MR) is based on the use of genetic  
54 variants that are linked to an exposure (e.g. having ever smoked, a greater intensity of  
55 current tobacco use) to assess the unconfounded effect of this exposure on a certain  
56 outcome (e.g. subfertility) (14). It can be performed using data from a single sample (one-  
57 sample MR: exposure, outcome, and genetic variants are measured in the same  
58 population) or from two samples (two-sample MR: the association between genetic  
59 variants and the exposure are assessed in one population and the association between  
60 genetic variants and the outcome are studied in a second population) (15). In this paper,

61 we compare results from multivariable logistic regression, one-sample MR, and two-  
62 sample MR, considering that these are affected by different and unrelated sources of bias  
63 when studying the association between smoking-related traits and subfertility (15-17)  
64 (**Table 1**). Similar results in all of them would allow for more robust conclusions (13).

65 Our aim was to investigate the association between tobacco use and subfertility in  
66 women and men by multivariable logistic regression and one- and two-sample MR.

67

68

## 69 **METHODS**

70

### 71 *Study participants*

72 Our study used data from the Mother, Father and Child Cohort Study (MoBa) (18, 19). The  
73 MoBa Study is a population-based pregnancy cohort study conducted by the Norwegian  
74 Institute of Public Health. Pregnant women and their partners across Norway were  
75 recruited between 1999 and 2008 at the time of the routine ultrasound screening (around  
76 18<sup>th</sup> gestational week). The cohort now includes 114,500 children, 95,200 mothers and  
77 75,200 fathers. The current study is based on version #12 of the quality-assured data.

78 For the current study, we defined a subsample of parents with available genotype  
79 data and pre-pregnancy information on tobacco use. The genotype data in this study came  
80 from blood samples gathered from both parents during pregnancy (20) and followed the  
81 pipeline described by Helgeland et al (genotype calling, imputation, and quality control)  
82 (21). Our work is described according to the STROBE guidelines for reporting MR and  
83 cohort studies (22, 23).

84

### 85 *Tobacco use*

86 Parents responded to questions related to their smoking habits in the questionnaire  
87 completed at recruitment. First, they reported if they had ever smoked. After an affirmative  
88 answer, participants reported their age when they started smoking, if they were smokers  
89 when they conceived, if they had quit smoking (and the date when they quit), and the  
90 amount of cigarettes per week they currently smoked (or used to smoke for former  
91 smokers). Using these data, we defined three exposure variables: ever smoker (yes/no,  
92 available for the whole population), age at smoking initiation (continuous, available in  
93 current/former smokers), and smoking cessation (yes/no, available in current/former  
94 smokers). In addition, we computed the average cigarettes smoked per week at  
95 conception or during the two years prior to conception (continuous, available in current  
96 smokers and in participants who smoked at this time).

97

#### 98 *Selection of genetic variants and generation of instrumental variables*

99 Genetic instruments were extracted from the most recent genome-wide association study  
100 (GWAS) on smoking-related traits (24). It included more than 1.2 million individuals (none  
101 of them participated in the MoBa cohort) and reported 378 conditionally independent single  
102 nucleotide polymorphisms (SNPs) associated with smoking initiation, 10 linked to the age  
103 of smoking initiation, 24 related to smoking cessation, and 55 associated with smoking  
104 heaviness (cigarettes per week) (24). Independent SNPs were defined according to  
105 linkage disequilibrium blocks across the genome using the methods by Li J et al., Gao X et  
106 al., and Chen Z et al. (25-27), presented a minor allele frequency  $\geq 0.1\%$  and were  
107 associated with their respective phenotypes according to the standard genome-wide  
108 significance threshold ( $p < 5 \times 10^{-8}$ ). A total of 355, 10, 23, and 50 independent SNPs were  
109 available for smoking initiation, age of smoking initiation, smoking cessation, and smoking  
110 heaviness, respectively, in the MoBa genotype database.

111 For the one-sample MR analyses, we generated a weighted genetic risk score  
112 (GRS) by multiplying the number of risk alleles by the effect estimate of each variant and  
113 dividing this value by the total number of SNPs (28). SNPs were used individually as the  
114 genetic instruments in the two-sample MR.

115

#### 116 *Subfertility*

117 Women were asked at recruitment if the pregnancy was planned and to provide  
118 information on how many months it had taken them to conceive (19). The answer options  
119 were “<1 month”, “1-2 months” and “3 or more months”. If the mother responded “3 or  
120 more months”, she was asked to further specify how many months they were trying to  
121 conceive. We defined subfertility as trying to conceive for  $\geq 12$  months or having used  
122 assisted reproductive technologies. Those reporting trying for <12 months (defined as  
123 fertile) were included in the reference group. Participants involved in unplanned  
124 pregnancies were excluded.

125

#### 126 *Other variables*

127 Information on the age of the participants (continuous), educational level (years of  
128 education equivalent to the US system (29, 30), continuous), pre-pregnancy body mass  
129 index (BMI, continuous), and previous number of deliveries (0, 1, 2, or  $\geq 3$ ) was gathered in  
130 the baseline visit of the study as potential confounders.

131

#### 132 *Ethical approval*

133 The MoBa cohort is conducted according to the Declaration of Helsinki for Medical  
134 Research involving Human Subjects. The establishment of MoBa and initial data collection  
135 was based on a license from the Norwegian Data Protection Agency. It is now based on  
136 regulations related to the Norwegian Health Registry Act. Participants provided a written



137 informed consent before joining the cohort. This project was approved by the Regional  
138 Committee for Medical and Health Research Ethics of South/East Norway (reference:  
139 2017/1362).

140

#### 141 *Statistical analyses*

142 Normally distributed continuous variables were described by means and standard  
143 deviations (SD), non-normally distributed continuous variables by medians and 25<sup>th</sup>-75<sup>th</sup>  
144 percentiles, and categorical variables by proportions. Differences in baseline  
145 characteristics between subfertile and fertile parents, and between participants with and  
146 without genotype information in the MoBa cohort, were investigated by t-tests in normally  
147 distributed continuous variables, Mann-Whitney U-tests in non-normally distributed  
148 continuous variables, and chi-squared tests in categorical variables.

149

#### 150 Multivariable regression analyses

151 We assessed the association between smoking-related traits and subfertility in women and  
152 men by multivariable logistic regressions (adjusted for predefined subfertility risk factors:  
153 age, years of education, BMI, and number of previous pregnancies (31, 32)). For binary  
154 exposures (smoking initiation and smoking cessation), we investigated the differences in  
155 the odds of subfertility in those exposed compared with those who were not. For  
156 continuous variables, we assessed the relationship between an increase in 1 SD in age at  
157 smoking initiation (in ever smokers) and the number of cigarettes smoked per week (in  
158 current smokers) with odds of subfertility. We also assessed whether a model using  
159 smoothed cubic splines ( $K+4$  degrees of freedom) fitted the data better than a linear  
160 function using a likelihood ratio test. In those traits in which there was information from  
161 both parents (smoking initiation and number of cigarettes smoked per week), we further  
162 adjusted logistic regression models for the partner's trait to minimize bias due to

163 assortative mating. Clustered standard errors were computed to account for non-  
164 independence between pregnancies in participants with data on multiple pregnancies.

165

#### 166 One-sample Mendelian randomization

167 We used logistic regression to estimate the genetically predicted likelihood of smoking  
168 initiation (in both sexes) and smoking cessation (in women), and linear regression to  
169 calculate the genetically predicted values of age at smoking initiation (in women) and the  
170 number of cigarettes smoked per week (in both sexes), using their respective GRSs as  
171 predictors. Next, we assessed the linear relationship between an increase in one SD in the  
172 genetically predicted traits and subfertility using logistic regression.

173

#### 174 Two-sample Mendelian randomization

175 We first performed two GWASs (one for women and one for men) to obtain summary  
176 associations of each SNP, across the genome, with subfertility in the MoBa cohort. Full  
177 details of this procedure are available in the **Supplemental Materials**. Next, in these  
178 subfertility GWAS summary data, we looked for the SNPs related to each of the smoking  
179 traits and extracted the information related to their association with subfertility. We  
180 harmonized both datasets and excluded palindromic SNPs with minor allele frequencies  
181 close to 0.5 (24, 33). Following these exclusions, we had 301, 43, 7 and 16 SNPs for  
182 analyses of smoking initiation, intensity, age at initiation, and cessation, respectively.  
183 Finally, we used inverse variance weighted regression as the main two-sample MR  
184 analysis (15).

185

#### 186 Verification of MR assumptions

187 The key assumptions of MR are: 1) the genetic instrument is robustly related to the  
188 exposure, 2) the genetic instrument is only associated with the outcome through the

189 exposure of interest, and 3) there is no confounding of the genetic instrument-outcome  
190 associations (15).

191         Regarding the first assumption, we checked the strength of the association  
192 between the genetic instruments and their phenotypes. For binary exposures we used  
193 logistic regression, area under the ROC curve, and pseudo- $R^2$  by the Nagelkerke method,  
194 and for continuous exposures we used linear regressions,  $F$ -statistics and  $R^2$ . Since weak  
195 instruments deviate MR causal estimates towards the null in two-sample MR, concordance  
196 between one-sample and two-sample MR reduces the concern that estimates might be  
197 influenced by weak instrument bias (34). We also performed the Robust Adjusted Profile  
198 Score two-sample MR method, which is unbiased by weak instruments (34, 35).

199 The second assumption may be violated when the genetic instruments influence other risk  
200 factors for the outcome independently of the exposure of interest (horizontal pleiotropy)  
201 (16). To evaluate horizontal pleiotropy in one-sample MR, we checked the association of  
202 GRSs with known risk factors for subfertility. We studied the relationship between 1 SD  
203 increase in the GRS and the risk factors (age, years of education, BMI, number of previous  
204 deliveries) (31, 32) using linear regressions. If any of the GRSs was associated with a risk  
205 factor, we considered that a potential pleiotropic effect was present for all the smoking-  
206 related traits. We then performed multivariable MR analyses if GWAS data for the potential  
207 pleiotropic variable was available (17). There was evidence of one or more of the smoking  
208 trait GRS associating with education and BMI and we were able to undertake multivariable  
209 MR for both. For education we used the GWAS by Lee JJ et al. ( $n = 1,271$  independent  
210 SNPs, ~1.1 million participants (36); 1,159 of the SNPs were available in MoBa). For BMI  
211 we used the GWAS by Yengo L et al ( $n = 941$  independent SNPs, ~700,000 participants  
212 (37); 896 of the SNPs were available in MoBa). We generated GRS for education and BMI  
213 using the same method as used for the smoking traits and then included the GRS for  
214 education and BMI in the one-sample MR regression models. (17). The genetic smoking

215 instruments also associated with age, and we performed stratified analyses according to  
216 age (below vs. over the median). In addition, we estimated the association between the  
217 GRSs for smoking traits and subfertility in non-exposed participants (the GRSs for age of  
218 smoking initiation or smoking cessation in never smokers, and the GRS for current  
219 smoking intensity in never + former smokers). As we would expect no association in non-  
220 exposed participants, any evidence of one would indicate the presence of bias.

221 In two-sample MR we explored unbalanced horizontal pleiotropy through sensitivity  
222 analyses comparing the main estimates to those obtained from MR-Egger, the weighted  
223 median and weighted mode methods (38-40). The inverse variance weighted method  
224 assumes no unbalanced horizontal pleiotropy as it forces the regression line through SNP-  
225 exposure and SNP-outcome coordinates to go through zero. MR-Egger does not make  
226 this assumption and it does not force the line through zero. A non-zero intercept is an  
227 indication of unbalanced horizontal pleiotropy and the slope is subsequently corrected.  
228 The weighted median and weighted mode analyses are valid if less than 50% of the weight  
229 comes from SNPs that are not related to other risk factors for the outcome. Concordance  
230 in the estimates across the different approaches reduces the concern regarding  
231 unbalanced pleiotropy (38-40). We also checked for influential outliers in the variant-  
232 specific causal estimates (indicative of horizontal pleiotropy) in scatterplots (38-40).  
233 Finally, we evaluated between SNP heterogeneity, using Cochran's Q and the Rücker's Q'  
234 with the inverse variance weighted and Egger regression methods, respectively.  
235 Heterogeneity indicates a possible violation of the MR assumptions, of which pleiotropy is  
236 probably a major cause (38-40).

237 To reduce the potential for confounding of the genetic instrument-outcome  
238 association due to population stratification (third assumption), we adjusted for the first 10  
239 ancestry-informative principal components in the one-sample MR (41).

240

241 Software

242 Analyses were performed in R Software version 4.0.3 (packages: *compareGroups*,  
243 *estimatr*, *ggplot2*, *miceadds*, and *TwoSampleMR*) and the GWASs to determine which  
244 SNPs were associated with subfertility in the MoBa cohort in Plink v1.9 and GWAMA (42,  
245 43). Code for data management and statistical analysis has been made available in  
246 [https://github.com/alvarohernaez/MR\\_smoking\\_subfertility\\_MoBa/](https://github.com/alvarohernaez/MR_smoking_subfertility_MoBa/) .

247

248

249 **RESULTS**

250

251 *Description of the study population and genetic instruments*

252 Our study population consisted of 28,606 women and 27,096 men with genotype  
253 information (**Figure 1**). A total of 10% of the couples were subfertile. Women and men who  
254 were subfertile were older, had a lower educational attainment, had a higher BMI, and  
255 were more likely to be pregnant for the first time. In relation to smoking habits, there was a  
256 higher proportion of ever smokers in subfertile couples and, among tobacco users,  
257 smoking intensity was higher in subfertile individuals (**Table 2**). MoBa participants without  
258 genotype data were not meaningfully different to those with genetic data in age, years of  
259 education, BMI, or number of previous pregnancies compared to those with genotype  
260 data. However, participants with genotype data were more likely to be subfertile, presented  
261 a different proportion of ever smokers (women were less likely and men were more prone  
262 to have ever smoked), and showed a lower smoking intensity among current smokers  
263 (**Supplemental Table 1**).

264 **Table 3** shows the associations of the GRS for each trait with smoking traits in  
265 MoBa together with statistics related to instrument strengths.

266

267 *Comparison of main multivariable regression, one-sample, and two-sample MR results*

268 **Figure 2A** shows the confounder-adjusted associations, one-sample, and two-sample MR  
269 results for each smoking trait in women. For smoking initiation, there were similar close to  
270 the null results in both multivariable regression (odds ratio [OR] 1.03, 95% confidence  
271 interval [CI] 0.95 to 1.11) and one-sample MR (OR 1.01, 95% CI 0.98 to 1.05), and an  
272 inverse, but also close to the null, relationship in two-sample MR (OR 0.90, 95% CI 0.75 to  
273 1.09). Regarding smoking intensity, it was linearly linked to greater odds of subfertility in  
274 multivariable regression (1 SD increase in the number of cigarettes smoked per week [+48  
275 cigarettes/week]: OR 1.12, 95% CI 1.03 to 1.21;  $p\text{-value}_{\text{non-linearity}} = 0.970$ ), even after  
276 adjusting for the partner's smoking heaviness (OR 1.10, 95% CI 1.01 to 1.19;  
277 **Supplemental Table 2**). However, close to the null results were observed in one-sample  
278 MR (OR 0.96, 95% CI 0.89 to 1.03) and a notable inverse, but imprecisely estimated result  
279 in two-sample MR analyses (OR 0.68, 95% CI 0.40 to 1.15). For age at smoking initiation,  
280 a broadly similar pattern of results across the three measures was seen. We observed a  
281 linear association between later smoking initiation and lower odds of subfertility in  
282 multivariable analyses (1 SD increase in the age of smoking initiation in ever smokers  
283 [+3.2 years]: OR 0.89, 95% CI 0.84 to 0.95;  $p\text{-value}_{\text{non-linearity}} = 0.933$ ), close to the null  
284 results in one-sample MR (OR 0.96, 95% CI 0.91 to 1.02), and a greater, directionally  
285 concordant but imprecise association in two-sample MR (OR 0.47, 95% CI 0.11 to 2.01).  
286 Finally, for cessation amongst those who had ever smoked, there were similar, in terms of  
287 OR, inverse associations in multivariable regression (OR 0.83, 95% CI 0.75 to 0.93) and  
288 two-sample MR (OR 0.80, 95% CI 0.46 to 1.38), the latter with wide confidence intervals,  
289 whereas the result in one-sample MR was close to the null (OR 1.04, 95% CI 0.99 to  
290 1.09).

291 Results from the three methods for smoking initiation in men showed broadly  
292 similar patterns to those seen for women (**Figure 2B**). However, a linear association

293 between higher smoking intensity and greater odds of subfertility was suggested in  
294 multivariable analyses (1 SD increase in the number of cigarettes smoked per week [+54  
295 cigarettes/week]: OR 1.08, 95% CI 0.99 to 1.18;  $p$ -value<sub>non-linearity</sub> = 0.123), attenuated when  
296 adjusting for the partner's smoking intensity (OR 1.05, 95% CI 0.96 to 1.15; **Supplemental**  
297 **Table 2**). A similar but imprecise association was observed in two-sample MR (OR, 1.11,  
298 95% CI 0.78 to 1.59), and close to the null results in one-sample MR (OR 1.02, 95% CI  
299 0.95 to 1.10). It was not possible to complete multivariable regression and one-sample MR  
300 analyses in men, because detailed information on smoking was not obtained from fathers  
301 in MoBa. Nonetheless, in two-sample MR, close to the null, highly imprecise associations  
302 were found for age of smoking initiation (OR 0.95, 95% CI 0.19 to 4.68) and smoking  
303 cessation (OR 0.94, 95% CI 0.57 to 1.54).

304

#### 305 *Sensitivity analyses*

306 We observed associations between some smoking related GRSs and age (in women),  
307 education years (in both sexes) and body mass index (in both sexes) (**Supplemental**  
308 **Table 3**). Results of multivariable one-sample MR accounting for education and BMI and  
309 age-stratified analyses were consistent with the main analyses (**Supplemental Table 4**).  
310 No associations between the GRSs for smoking traits and subfertility were found in non-  
311 smokers (**Supplemental Table 5**).

312 Regarding two-sample MR additional methods, between SNP heterogeneity,  
313 potentially linked to horizontal pleiotropy, was observed for the genetic instrument for  
314 smoking intensity in women, as well as highly imprecise MR estimates (**Supplemental**  
315 **Table 6, Supplemental Figure 1**).

316

317

318 **DISCUSSION**

319 We found close to the null associations between smoking initiation and subfertility, using  
320 confounder adjusted multivariable regression and MR in women and men. Regarding  
321 smoking heaviness, it was associated with greater subfertility odds in both sexes in  
322 multivariable regression. However, we found close to the null relationships in one-sample  
323 MR, a suggested clinically important inverse association in two-sample MR in women, and  
324 a substantial attenuation of the association in multivariable regression in men when  
325 adjusted for the partner's smoking status. A similar pattern of association across the three  
326 methods was observed for age at initiation in women, whereas there were consistent  
327 inverse associations with smoking cessation between multivariable regression and two  
328 sample MR in women. Taken together, our findings do not provide strong evidence for  
329 effects of smoking on subfertility. However, results for two-sample MR were very  
330 imprecise, highlighting the need for large collaborative GWAS of subfertility, one-sample  
331 MR results may have been influenced by weak instrument and, despite numerous  
332 sensitivity analyses, we cannot rule out completely bias due to horizontal pleiotropy in MR  
333 analyses or residual confounding in the multivariable regression analyses.

334 A link between smoking and subfertility has been reported in human studies since  
335 the 1980s, particularly in women (5, 6, 44). Nevertheless, the American Society of  
336 Reproductive Medicine warned in 2018 of some methodological flaws in the available body  
337 of evidence including the risk of residual confounding, the small magnitude of the  
338 association, and the retrospective nature of most studies (14). We address these  
339 limitations here by assessing the association between several smoking-related traits and  
340 subfertility in a large prospective study comparing multivariable regression results to those  
341 from one-sample and two-sample MR, undertaking several sensitivity analyses to explore  
342 possible sources of bias. Our multivariable regression results were similar to previous  
343 studies, including point estimates that were close to the null for several smoking traits (6,  
344 44). The magnitude of these relationships was modest, and they were generally consistent



345 in direction to MR results, though with greater power and hence narrower confidence  
346 intervals. In particular, confidence intervals for estimates for smoking intensity and age at  
347 initiation (which are directly comparable) overlapped, suggesting the three sets of results  
348 were consistent with each other. It is possible that residual confounding due to poorer  
349 health status, worse diet quality, increased alcohol consumption, lower levels of physical  
350 activity, etc. (45-47) could explain the decreased fertility among smokers in multivariable  
351 analyses, with possible weak instruments in some of our one-sample MR analyses being  
352 biased towards these confounded results. Moreover, we could not rule out bias due to  
353 assortative mating completely (individuals who smoke are more likely to select a partner  
354 who also smokes (48), which may distort any potential association in couple-based  
355 outcomes such as fertility). Last, the genetic instruments for age of smoking initiation and  
356 smoking cessation were also weak. Although we used several approaches to minimize the  
357 likelihood of weak instrument bias, this could still have influenced findings, and in the two-  
358 sample MR in particular results were imprecise, suggesting low power to reject the null  
359 hypothesis (33, 38, 39, 49). Regarding men, we did not find any robust relationship  
360 between tobacco-related habits and subfertility across methods. These results are in line  
361 with previous prospective studies (10, 11).

362 Our work has some limitations. First, subfertility is a couple-dependent parameter  
363 that was reported by mothers in the cohort, and we could not determine the cause (female  
364 causes, male, or both). Second, MoBa is a pregnancy cohort, only including couples who  
365 were able to conceive. Further studies considering couples who never conceived  
366 (infertility) are needed. Subfertility is a less severe manifestation of infertility and, therefore,  
367 an association between smoking-related traits and an absolute incapacity to conceive may  
368 be possible. Third, our results may have been affected by selection bias as there was  
369 some differences between participants with and without genetic data. In addition, smokers  
370 are under-represented in the MoBa cohort relative to the whole Norwegian population (50).

371 Nevertheless, associations between smoking and pregnancy outcomes (low birthweight,  
372 placental abruption, and stillbirth) proved unaffected by selection bias in prior studies when  
373 the MoBa cohort and the general Norwegian population were compared (50). Finally, our  
374 study population (adult women and men of a northern European ancestry who were  
375 capable of conceiving) restricts the generalizability. However, our work also presents  
376 several strengths. To our knowledge, this is the first prospective study assessing the  
377 relationship between smoking and subfertility using three complementary approaches  
378 affected by different sources of bias that have been thoroughly investigated and  
379 acknowledged. In addition, our study includes a relatively homogeneous population. This  
380 aspect minimizes the risk of confounding due to population stratification in MR, together  
381 with the adjustment for the first 10 principal components (41).

382

383

## 384 **CONCLUSIONS**

385 We did not find robust evidence of an effect of smoking on subfertility. This may be due to  
386 a true lack of effect, weak genetic instruments, or other kinds of confounding. However,  
387 the comparison of different analytical approaches with complementary sources of bias has  
388 highlighted relevant limitations across all methods, and in particular highlights the needs  
389 for larger studies with information on subfertility.

## 390 **DATA AVAILABILITY STATEMENT**

391 Consent given by the participants does not open for storage of data on an individual level  
392 in repositories or journals. Researchers who want access to datasets for replication should  
393 submit an application to [datatilgang@fhi.no](mailto:datatilgang@fhi.no). Access to datasets requires approval from the  
394 Regional Committee for Medical and Health Research Ethics in Norway and an agreement  
395 with MoBa.

396 Source data of the GWAS on smoking initiation, age of smoking initiation, smoking  
397 cessation, and smoking intensity) are available in the Supplemental Tables of the article by  
398 Liu M et al. (published in Nat Genet in 2019: <https://www.nature.com/articles/s41588-018-0307-5#Sec14>). Source data of the GWAS on education years are available in the  
399 Supplemental Tables of the article by Lee JJ et al. (published in Nat Genet in 2018:  
400 <https://www.nature.com/articles/s41588-018-0147-3#Sec34>). Finally, source data of the  
401 GWAS on BMI (Yengo L et al., Hum Mol Genet, 2018) are available in the GIANT  
402 Consortium website  
403 ([https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT\\_consortium\\_data\\_fil](https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files#GWAS_Anthropometric_2015_BMI_Summary_Statistics)  
404 [es#GWAS\\_Anthropometric\\_2015\\_BMI\\_Summary\\_Statistics](https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files#GWAS_Anthropometric_2015_BMI_Summary_Statistics)).

## 408 **ACKNOWLEDGEMENTS**

409 The MoBa Cohort Study is supported by the Norwegian Ministry of Health and Care  
410 Services and the Ministry of Education and Research. We thank all the participating  
411 families in Norway who take part in this ongoing cohort study, and those who contributed  
412 to the recruitment and the infrastructure of the cohort.

413 We thank the Norwegian Institute of Public Health for generating high-quality  
414 genomic data. This research is part of the HARVEST collaboration, supported by the  
415 Research Council of Norway (project reference: #229624). We also thank the NORMENT

416 Centre for providing genotype data, funded by the Research Council of Norway (project  
417 reference: #223273), South East Norway Health Authority, and Stiftelsen Kristian Gerhard  
418 Jebsen. We further thank the Center for Diabetes Research (University of Bergen) for  
419 providing genotype information and performing quality control and imputation of the data in  
420 research projects funded by the European Research Council Advanced Grant  
421 SELECTIONPREDISPOSED, Stiftelsen Kristian Gerhard Jebsen, the Trond Mohn  
422 Foundation, the Research Council of Norway, the Novo Nordisk Foundation, the University  
423 of Bergen, and the Western Norway Health Authority.

424 This work was performed on the TSD (Tjeneste for Sensitive Data) facilities, owned  
425 by the University of Oslo, operated and developed by the TSD service group at the  
426 University of Oslo, IT-Department ([tsd-drift@usit.uio.no](mailto:tsd-drift@usit.uio.no)).

427 This paper does not necessarily reflect the position or policy of the Norwegian  
428 Research Council.

429

430

#### 431 **FUNDING**

432 The MoBa Cohort Study is supported by the Norwegian Ministry of Health and Care  
433 Services and the Norwegian Ministry of Education and Research. This project received  
434 funding from the European Research Council under the European Union's Horizon 2020  
435 research and innovation program (grant agreement No 947684). This work was also partly  
436 supported by the Research Council of Norway through its Centres of Excellence funding  
437 scheme, project number 262700, and the project "Women's fertility – an essential  
438 component of health and well-being", number 320656. Open Access funding was provided  
439 by the Folkehelseinstituttet/Norwegian Institute of Public Health. D.A.L. is a UK National  
440 Institute for Health Research Senior Investigator (NF-SI-0611–10196) and is supported by  
441 the US National Institutes of Health (R01 DK10324) and a European Research Council

442 Advanced Grant (grant agreement No 669545). The funders had no role in the collection,  
443 analysis, and interpretation of data; in the writing of the report; or in the decision to submit  
444 the article for publication.

445

446

447 **CONFLICT OF INTEREST**

448 D.A.L. receives (or has received in the last 10 years) research support from National and  
449 International government and charitable bodies, Roche Diagnostics and Medtronic for  
450 research unrelated to the current work. The rest of the authors declare that no competing  
451 interests exist.

452    **REFERENCES**

453

- 454    1.     Talhout R, Schulz T, Florek E, van Benthem J, Wester P, Opperhuizen A.  
455    Hazardous compounds in tobacco smoke. *Int J Environ Res Public Health* 2011;8:613-28.
- 456    2.     Amiri M, Ramezani Tehrani F. Potential Adverse Effects of Female and Male  
457    Obesity on Fertility: A Narrative Review. *Int J Endocrinol Metab* 2020;18:e101776.
- 458    3.     Broughton DE, Moley KH. Obesity and female infertility: potential mediators of  
459    obesity's impact. *Fertil Steril* 2017;107:840-7.
- 460    4.     Liu Y, Ding Z. Obesity, a serious etiologic factor for male subfertility in modern  
461    society. *Reproduction* 2017;154:R123-r31.
- 462    5.     Hyland A, Piazza K, Hovey KM, Tindle HA, Manson JE, Messina C *et al.*  
463    Associations between lifetime tobacco exposure with infertility and age at natural  
464    menopause: the Women's Health Initiative Observational Study. *Tob Control* 2016;25:706-  
465    14.
- 466    6.     Baird DD, Wilcox AJ. Cigarette smoking associated with delayed conception. *Jama*  
467    1985;253:2979-83.
- 468    7.     Adena MA, Gallagher HG. Cigarette smoking and the age at menopause. *Ann Hum*  
469    *Biol* 1982;9:121-30.
- 470    8.     Baron JA, La Vecchia C, Levi F. The antiestrogenic effect of cigarette smoking in  
471    women. *Am J Obstet Gynecol* 1990;162:502-14.
- 472    9.     Bundhun PK, Janoo G, Bhurtu A, Teeluck AR, Soogund MZS, Pursun M *et al.*  
473    Tobacco smoking and semen quality in infertile males: a systematic review and meta-  
474    analysis. *BMC Public Health* 2019;19:36.
- 475    10.    de Mouzon J, Spira A, Schwartz D. A prospective study of the relation between  
476    smoking and fertility. *Int J Epidemiol* 1988;17:378-84.

- 477 11. Jensen TK, Henriksen TB, Hjollund NH, Scheike T, Kolstad H, Giwercman A *et al.*  
478 Adult and prenatal exposures to tobacco smoke as risk indicators of fertility among 430  
479 Danish couples. *Am J Epidemiol* 1998;148:992-7.
- 480 12. Penzias A, Bendikson K, Butts S, Coutifaris C, Falcone T, Gitlin S *et al.* Smoking  
481 and infertility: a committee opinion. *Fertil Steril* 2018;110:611-8.
- 482 13. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int*  
483 *J Epidemiol* 2016;45:1866-86.
- 484 14. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian  
485 randomization: using genes as instruments for making causal inferences in epidemiology.  
486 *Stat Med* 2008;27:1133-63.
- 487 15. Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM *et al.*  
488 Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res*  
489 2019;4:186.
- 490 16. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal  
491 inference in epidemiological studies. *Hum Mol Genet* 2014;23:R89-98.
- 492 17. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of  
493 pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol* 2015;181:251-60.
- 494 18. Magnus P, Birke C, Vejrup K, Haugan A, Alsaker E, Daltveit AK *et al.* Cohort  
495 Profile Update: The Norwegian Mother and Child Cohort Study (MoBa). *Int J Epidemiol*  
496 2016;45:382-8.
- 497 19. Magnus P, Irgens LM, Haug K, Nystad W, Skjaerven R, Stoltenberg C. Cohort  
498 profile: the Norwegian Mother and Child Cohort Study (MoBa). *Int J Epidemiol*  
499 2006;35:1146-50.
- 500 20. Paltiel L, Haugan A, Skjerden T, K; H, Bækken S, Stensrud NK *et al.* The biobank  
501 of the Norwegian Mother and Child Cohort Study – present status. *Norsk Epidemiologi*  
502 2014;24:29-35.

- 503 21. Helgeland Ø, Vaudel M, Juliusson PB, Lingaas Holmen O, Juodakis J, Bacelis J *et*  
504 *al.* Genome-wide association study reveals dynamic role of genetic variation in infant and  
505 early childhood growth. *Nat Commun* 2019;10:4448.
- 506 22. Davey Smith G, Davies NM, Dimou N, Egger M, Gallo V, Golub R *et al.* STROBE-  
507 MR: Guidelines for strengthening the reporting of Mendelian randomization studies. *PeerJ*  
508 *Preprints* 2019;7:e27857v1.
- 509 23. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP.  
510 Strengthening the Reporting of Observational Studies in Epidemiology (STROBE)  
511 statement: guidelines for reporting observational studies. *Bmj* 2007;335:806-8.
- 512 24. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F *et al.* Association studies of up  
513 to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol  
514 use. *Nat Genet* 2019;51:237-44.
- 515 25. Chen Z, Liu Q. A new approach to account for the correlations among single  
516 nucleotide polymorphisms in genome-wide association studies. *Hum Hered* 2011;72:1-9.
- 517 26. Gao X, Becker LC, Becker DM, Starmer JD, Province MA. Avoiding the high  
518 Bonferroni penalty in genome-wide association studies. *Genet Epidemiol* 2010;34:100-5.
- 519 27. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of  
520 a correlation matrix. *Heredity (Edinb)* 2005;95:221-7.
- 521 28. Choi SW, Mak TS, O'Reilly PF. Tutorial: a guide to performing polygenic risk score  
522 analyses. *Nat Protoc* 2020;15:2759-72.
- 523 29. Barrabés N, Østli GK. Norsk standard for utdanningsgruppering 2016. In, 2016.
- 524 30. Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW *et al.* GWAS of  
525 126,559 individuals identifies genetic variants associated with educational attainment.  
526 *Science* 2013;340:1467-71.



- 527 31. Dribe M, Hacker JD, Scalone F. The impact of socio-economic status on net fertility  
528 during the historical fertility decline: a comparative analysis of Canada, Iceland, Sweden,  
529 Norway, and the USA. *Popul Stud (Camb)* 2014;68:135-49.
- 530 32. Sharma R, Biedenharn KR, Fedor JM, Agarwal A. Lifestyle factors and  
531 reproductive health: taking control of your fertility. *Reprod Biol Endocrinol* 2013;11:66.
- 532 33. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity analyses for  
533 robust causal inference from Mendelian randomization analyses with multiple genetic  
534 variants. *Epidemiology (Cambridge, Mass)* 2017;28:30.
- 535 34. Barry C, Liu J, Richmond R, Rutter MK, Lawlor DA, Dudbridge F *et al.* Exploiting  
536 collider bias to apply two-sample summary data Mendelian randomization methods to one-  
537 sample individual level data. *medRxiv* 2020.
- 538 35. Zhao Q, Wang J, Hemani G, Bowden J, Small DS. Statistical inference in two-  
539 sample summary-data Mendelian randomization using robust adjusted profile score. *The*  
540 *Annals of Statistics* 2020;48:1742-69, 28.
- 541 36. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M *et al.* Gene discovery  
542 and polygenic prediction from a genome-wide association study of educational attainment  
543 in 1.1 million individuals. *Nat Genet* 2018;50:1112-21.
- 544 37. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN *et al.* Meta-  
545 analysis of genome-wide association studies for height and body mass index in ~700000  
546 individuals of European ancestry. *Hum Mol Genet* 2018;27:3641-9.
- 547 38. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid  
548 instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*  
549 2015;44:512-25.
- 550 39. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in  
551 Mendelian Randomization with Some Invalid Instruments Using a Weighted Median  
552 Estimator. *Genet Epidemiol* 2016;40:304-14.

- 553 40. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in  
554 Mendelian randomization studies. *Hum Mol Genet* 2018;27:R195-r208.
- 555 41. Wang C, Zhan X, Liang L, Abecasis GR, Lin X. Improved ancestry estimation for  
556 both genotyping and sequencing data using projection procrustes analysis and genotype  
557 imputation. *Am J Hum Genet* 2015;96:926-37.
- 558 42. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-  
559 generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*  
560 2015;4:7.
- 561 43. Mägi R, Morris AP. GWAMA: software for genome-wide association meta-analysis.  
562 *BMC Bioinformatics* 2010;11:288.
- 563 44. Howe G, Westhoff C, Vessey M, Yeates D. Effects of age, cigarette smoking, and  
564 other factors on fertility: findings in a large prospective study. *Br Med J (Clin Res Ed)*  
565 1985;290:1697-700.
- 566 45. Dallongeville J, Marécaux N, Fruchart JC, Amouyel P. Cigarette smoking is  
567 associated with unhealthy patterns of nutrient intake: a meta-analysis. *J Nutr*  
568 1998;128:1450-7.
- 569 46. Ng R, Sutradhar R, Yao Z, Wodchis WP, Rosella LC. Smoking, drinking, diet and  
570 physical activity-modifiable lifestyle risk factors and their associations with age to first  
571 chronic disease. *Int J Epidemiol* 2020;49:113-30.
- 572 47. Perkins KA, Rohay J, Meilahn EN, Wing RR, Matthews KA, Kuller LH. Diet,  
573 alcohol, and physical activity as a function of smoking status in middle-aged women.  
574 *Health Psychol* 1993;12:410-5.
- 575 48. Agrawal A, Heath AC, Grant JD, Pergadia ML, Statham DJ, Bucholz KK *et al.*  
576 Assortative mating for cigarette smoking and for alcohol consumption in female Australian  
577 twins and their spouses. *Behav Genet* 2006;36:553-66.

- 578 49. Davies NM, von Hinke Kessler Scholder S, Farbmacher H, Burgess S, Windmeijer  
579 F, Smith GD. The many weak instruments problem and Mendelian randomization. *Stat*  
580 *Med* 2015;34:454-68.
- 581 50. Nilsen RM, Vollset SE, Gjessing HK, Skjaerven R, Melve KK, Schreuder P *et al.*  
582 Self-selection and bias in a large prospective pregnancy cohort in Norway. *Paediatr*  
583 *Perinat Epidemiol* 2009;23:597-608.
- 584

585 **TABLES**

586

587 **Table 1.** Comparison among multivariable logistic regression, one- and two-sample Mendelian randomization regarding sources of

588 bias in the association between smoking and subfertility

	<b>Residual confounding</b> (the effect of smoking on subfertility could be due to behavioral or socioeconomic conditions intimately related to smoking, such as poor diet, poorer health status, etc.)	<b>Weak genetic instruments</b> (the available genetic instruments for smoking traits explain a small proportion of the variability of the exposure)	<b>Horizontal pleiotropy</b> (the genetic instruments for smoking traits influence the risk of subfertility via mechanisms other than smoking)	<b>Confounding of the genetic instrument-outcome association</b> (by population stratification)
Multivariable regression	High risk	Unaffected	Unaffected	Unaffected
One-sample MR	Low risk	Overestimated associations.  Consistence with two-sample MR estimates minimizes the risk of this bias	Possibility to explore the association between genetic instruments and other subfertility risk factors.  Multivariable MR could then be used to correct for such bias	Adjustment for ancestry-informative principal components minimizes the risk of this bias

Two-sample MR	Low risk	Bias towards the null association.  The Robust Adjusted Profile Score approach is immune to this bias	Consistence among MR methods,  lack of between SNP heterogeneity,  and lack of outliers in MR scatterplots  minimizes the risk of this bias	High risk
------------------	----------	---	---	-----------

589

590 **Table 2.** Population characteristics among genotyped participants

	Women				Men			
	All (n = 28,606)	Subfertility reported (n = 3,439)	No subfertility reported (n = 25,167)	p-value	All (n = 27,096)	Subfertility reported (n = 3,275)	No subfertility reported (n = 23,821)	p-value
Age (years), mean ± SD	30.3 ± 4.15	31.5 ± 4.36	30.2 ± 4.09	<0.001	32.7 ± 4.90	34.1 ± 5.36	32.6 ± 4.81	<0.001
Education years, mean ± SD	17.5 ± 3.11	17.0 ± 3.33	17.6 ± 3.08	<0.001	16.6 ± 3.50	16.2 ± 3.54	16.6 ± 3.49	<0.001
BMI (kg/m <sup>2</sup> ), median (25 <sup>th</sup> -75 <sup>th</sup> percentiles)	23.1 (21.2-25.9)	23.7 (21.5-27.2)	23.1 (21.1-25.7)	<0.001	25.5 (23.7-27.7)	25.8 (24.0-28.1)	25.4 (23.7-27.7)	<0.001
Previous pregnancies n (%):				<0.001				<0.001
0	12,888 (45.1%)	2,020 (58.8%)	10,868 (43.2%)		12,282 (45.4%)	1,923 (58.8%)	10,359 (43.5%)	
1 or more	15,680 (54.9%)	1,415 (41.2%)	14,265 (56.8%)		14,784 (54.6%)	1,348 (41.2%)	13,436 (56.5%)	
Ever smokers (all participants), n (%):	13,389 (46.8%)	1,722 (50.1%)	11,667 (46.4%)	<0.001	13,186 (48.7%)	1,668 (50.9%)	11,518 (48.4%)	0.006
Age of smoking initiation (current + former smokers), median (25 <sup>th</sup> -75 <sup>th</sup> percentiles)	17.0 (15.0-19.0)	16.0 (15.0-18.0)	17.0 (15.0-19.0)	0.017	-	-	-	-
Current + former smokers who quit smoking, n (%):	7,627 (57.0%)	910 (52.8%)	6,717 (57.6%)	<0.001	-	-	-	-
Cigarettes/week (current smokers), median (25 <sup>th</sup> -75 <sup>th</sup> percentiles)		42.0 (10.0-70.0)	52.5 (12.0-105)	40.0 (10.0-70.0)	<0.001	70.0 (21.0-105)	70.0 (35.0-105)	70 (21.0

591

592 **Table 3.** Description of genetic instrumental variables<sup>1</sup>

593

<b>Binary exposures</b>					
	SNPs ( <i>n</i> )	GRS value (mean ± SD)	Likelihood of presenting the exposure per each 1-point increase in the GRS OR (95% CI)	Pseudo- <i>R</i> <sup>2</sup>	Area under the ROC curve
GRS for smoking initiation (women)	355	375 ± 11.9	1.02 (1.018 to 1.022)	1.87%	0.57
GRS for smoking initiation (men)	355	376 ± 11.8	1.02 (1.017 to 1.022)	1.69%	0.56
GRS for smoking cessation (women)	23	24.5 ± 2.52	1.03 (1.019 to 1.047)	0.22%	0.52
<b>Continuous exposures</b>					
	SNPs ( <i>n</i> )	GRS value (mean ± SD)	Change in the exposure per each 1-point increase in the GRS Coef. (95% CI)	<i>R</i> <sup>2</sup>	<i>F</i> -statistic
GRS for cigarettes smoked per week (women)	50	51.2 ± 6.90	0.80 (0.62 to 0.98)	1.43%	76
GRS for cigarettes smoked per week (men)	50	51.2 ± 6.96	0.87 (0.66 to 1.09)	1.34%	64
GRS for age of smoking initiation (women)	10	7.78 ± 1.83	0.023 (-0.014 to 0.059)	0.01%	1

594 1. The associations of smoking cessation and age of smoking initiation with their

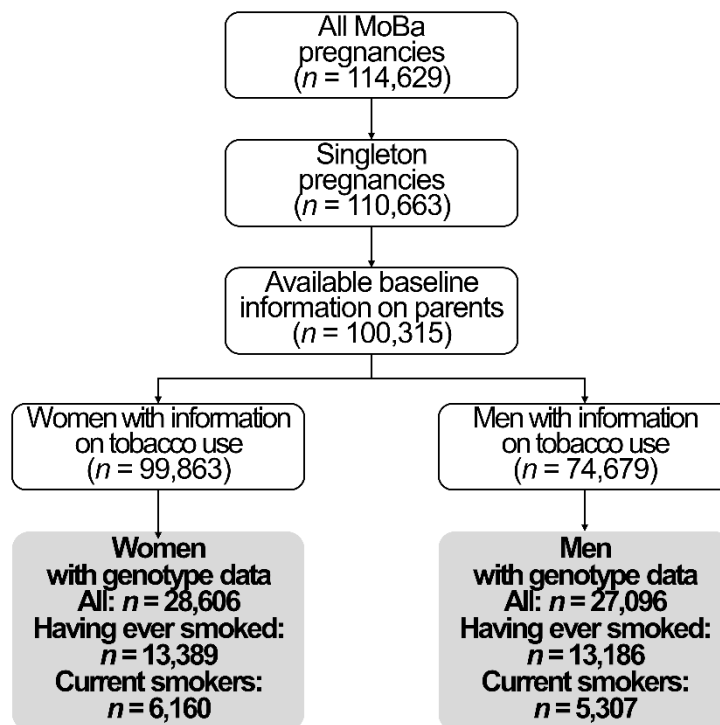
595 respective GRSs in men could not be validated due to the lack of information on this

596 phenotype.

597 **FIGURES**

598

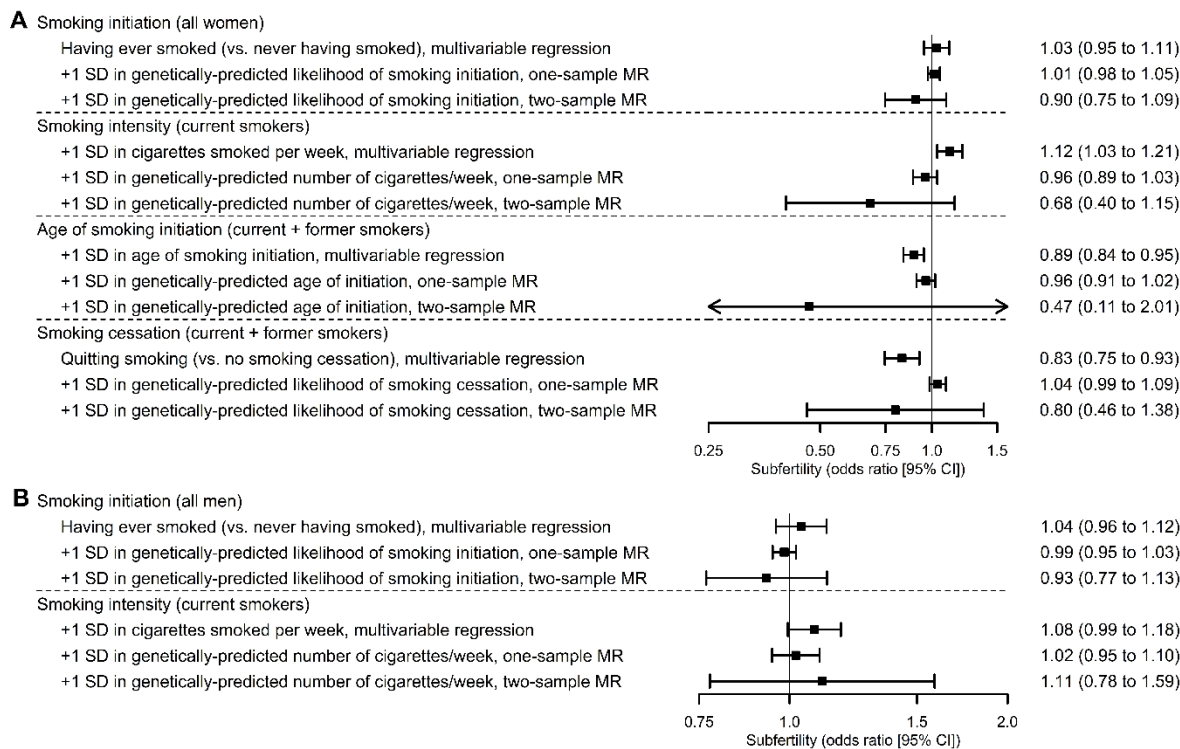
599 **Figure 1.** Study flow chart



600



601 **Figure 2.** Association between smoking-related traits and subfertility in women (A) and  
 602 men (B).



603