

1 Evaluating the causal impact of reproductive factors on 2 breast cancer risk: a multivariable mendelian 3 randomization approach

4 Claire Prince^{1,2}, Laura D Howe^{1,2}, Eleanor Sanderson^{1,2}, Gemma C Sharp^{1,2,3}, Abigail Fraser^{1,2},
5 Bethan Lloyd-Lewis⁴ and Rebecca C Richmond^{1,2}

6 ¹ MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK, ² Population Health Sciences,
7 Bristol Medical School, University of Bristol, Bristol, UK, ³ School of Psychology, University of Exeter,
8 Exeter, UK, ⁴ School of Cellular and Molecular Medicine, University of Bristol, Bristol, UK

9 Abstract

10 **Background:** Observational evidence proposes a protective effect of having children and an
11 early age at first birth on the development of breast cancer, however the causality of this
12 association remains uncertain. In this study we assess whether these reproductive factors
13 impact breast cancer risk independently of age at menarche, age at menopause, adiposity
14 measures and other reproductive factors that have been identified as being causally related
15 to or genetically correlated with the reproductive factors of interest.

16 **Methods:** We used genetic data from UK Biobank (273,238 women) for reproductive
17 factors, age at menarche and menopause, and adiposity measures, and the Breast Cancer
18 Association Consortium for risk of overall, estrogen receptor (ER) positive and negative
19 breast cancer as well as breast cancer subtypes. We applied univariable and multivariable
20 Mendelian randomization (MR) to estimate direct effects of ever parous status, ages at first
21 birth and last birth, and number of births on breast cancer risk.

22 **Results:** We found limited evidence of an effect of age at first birth on overall or ER positive
23 breast cancer risk in either the univariable or multivariable analyses. While the univariable
24 analysis revealed an effect of later age at first birth decreasing ER negative breast cancer
25 risk (Odds ratio (OR): 0.76, 95% confidence interval:0.61-0.95 per standard deviation (SD)
26 increase in age at first birth), this effect attenuated with separate adjustment for age at
27 menarche and menopause (e.g., OR 0.83, 0.62-1.06 per SD increase in age at first birth,
28 adjusted for age at menarche). In addition, we found evidence for an effect of later age at

29 first birth on decreased human epidermal growth factor receptor 2 enriched breast cancer
30 risk but only with adjustment for number of births (OR 0.28 (0.11-0.57) per SD increase in
31 age at first birth).

32 We found little evidence for direct effects of ever parous status, age at last birth or number of
33 births on breast cancer risk, however, analyses of ever parous status and age at last birth
34 were limited by weak instruments in the multivariable analysis.

35 **Conclusions:** This study found minimal evidence of a protective effect of earlier age at first
36 birth on breast cancer risk, while identifying some evidence for an adverse effect on ER
37 negative breast cancer risk. However, multivariable MR of ever parous status and age at last
38 birth is limited by weak instruments which might be improved in future studies with larger
39 sample sizes and when additional genetic variants related to reproductive factors are
40 identified.

41 Introduction

42 Reproductive factors relating to pregnancy, including ever having children, the number of
43 children a woman has and age at first and last birth, have been found to be related to the
44 development of breast cancer.⁽¹⁻³⁾ However, the relationships between these factors and
45 breast cancer risk are complex, and the underlying mechanisms remain unclear.

46 Experiencing pregnancy earlier in life has been shown to be associated with a reduced long-
47 term breast cancer risk in previous studies,^(2,4-8) with one study identifying a 7% increase in
48 risk (odds ratio (OR): 1.07; 95% confidence interval (CI): 1.01-1.13) per 5-year increase in
49 age at first full-term pregnancy.⁽⁵⁾

50 There are several proposed mechanisms which may underlie the protective role of an early
51 pregnancy on breast cancer development. Parity stimulates the differentiation of mammary
52 epithelial lobules to its mature state, and cells in a differentiated state have reduced
53 susceptibility to mutation and tumorigenesis.^(9,10) Additionally parous, compared to

54 nulliparous, women have lower responsiveness to estrogen, via a downregulation of
55 estrogen receptor (ER) α , and upregulation of ER β .⁽¹¹⁾ ER α have higher affinity to estrogen
56 compared to ER β , and heterodimerising of ER α and ER β reduces the affinity of ER α to
57 estrogen.⁽¹²⁾ Since estrogen is known to be proliferative and genotoxic, there may be an
58 increased risk of mutation and subsequent development of breast cancer among nulliparous
59 women.^(9,13)

60 However, a short-term adverse effect of pregnancy on breast cancer risk has also been
61 observed (for women aged 35 years at first birth, 5 years post-delivery OR of 1.26; 1.10-
62 1.44, compared to nulliparous women).⁽¹⁴⁾ This short-term elevated risk could occur mainly
63 through two mechanisms.⁽⁵⁾ Firstly, levels of ovarian hormones, which have proliferative
64 effects, increase during pregnancy and subsequently there is a higher chance of a mutation
65 occurring.⁽⁹⁾ Secondly, involution, that occurs following lactation, involves programmed cell
66 death and tissue remodelling of the breast which has similar properties to wound healing and
67 inflammation,^(13,15,16) which are known to be pro-oncogenic.^(13,15)

68 The long-term protective benefits of pregnancy may therefore not emerge until at least 10
69 years after delivery,^(5,13-15) with a proposed cross-over point where the group at higher risk
70 changes from parous women to nulliparous women.^(14,15,17-19)

71 Considering the proposed mechanism linking age at first pregnancy and breast cancer
72 development, it is thought that a shorter interval between menarche and first pregnancy
73 reduces breast cancer risk given the prevalence of undifferentiated cells that are more
74 susceptible to mutation-causing events during this time.⁽¹³⁾ The first pregnancy is therefore
75 mainly protective against later breast cancer risk when experienced earlier in life. In addition,
76 a late age at first pregnancy may increase risk because the proliferative effects of pregnancy
77 cause underlying mutations, that arise during the high-risk window, to be exacerbated and
78 result in cancer formation.^(4,20) There appears to be little difference in risk if the first
79 pregnancy occurs between 30 and mid-30s compared to nulliparous women, while having

80 the first pregnancy after the age of 35 may increase risk.^(2,4) However, evidence for this
81 increased risk is not consistent across all studies.^(2,21,22)

82 Considering the role of multiparity, there is evidence to suggest that each additional
83 pregnancy conveys further protection beyond an early age at first pregnancy.^(16,23) While
84 other studies show that a higher age at last birth is associated with higher breast cancer
85 risk,^(7,8,24-26) with one study reporting an increased breast cancer risk per every 5-year rise in
86 age at last birth (OR: 1.09; 1.04-1.13).⁽²⁵⁾ However, the mechanisms by which number of
87 births and age at last birth lead to breast cancer are unclear.^(7,27)

88 Observational studies are known to be limited by confounding bias as it is difficult to capture
89 all confounders accurately. Mendelian randomisation (MR), a method less likely to be
90 affected by confounding and reverse causation,^(28,29) has been used to assess the causal
91 relationship between reproductive factors and breast cancer risk by using genetic variants
92 robustly associated with the exposure as instruments. MR studies have consistently found
93 little evidence of an effect of age at first birth,^(30,31) (OR 0.92; 0.79-1.07⁽³⁰⁾),⁽³⁰⁾ but
94 interestingly, some evidence for a protective effect of a later age at last birth on breast
95 cancer risk (OR 0.69; 0.54-0.88), in contrast to observational data.⁽³¹⁾ An MR identified an
96 estimate reflecting a higher number of births reduces breast cancer risk although confidence
97 intervals span the null (OR 0.70; 0.44-1.11).⁽³⁰⁾

98 It is currently unclear how each reproductive event affects risk in isolation since these traits
99 are highly correlated with, and/or causally linked to other reproductive factors as well as age
100 at menarche and menopause, and adiposity measures,^(32,33) which are established breast
101 cancer risk factors.^(34,35) Multivariable MR (MVMR) is an extension of MR which can estimate
102 the direct effect of two or more risk factors while accounting for genetic correlation between
103 them,^(36,37) which may be valuable for investigating the impact of correlated reproductive
104 factors.

105 The effects of reproductive factors on risk of different subtypes of breast cancer are not well
106 understood.^(1,30,38-44) There is some consensus that a lower age at first birth and higher parity
107 decreases risk of hormone-receptor positive breast cancer subtypes, however there is less
108 consistency for the effects of these traits on risk of hormone-receptor negative subtypes.^{(1,41-}
109 ⁴⁴⁾ Hormone-receptor negative subtypes are more common in women under 40,⁽⁴⁵⁾ therefore,
110 pregnancy may increase short-term risk for these subtypes more so than those classified as
111 hormone-receptor positive. Additionally, as more women are choosing not to have
112 children,⁽⁴⁶⁾ or having children later in life,⁽⁴⁷⁾ it is important to understand how these traits
113 might convey different risk depending on subtype.

114 **Aims**

115 We aimed to use univariable and multivariable MR methods to untangle the effects of ever
116 having children, age at first birth, age at last birth and number of children birthed on breast
117 cancer risk. We further aimed to investigate whether effects are independent of age at
118 menarche and menopause, and adiposity measures as well as other reproductive factors,
119 and whether effects differ for ER positive compared to ER negative breast cancer.
120 Furthermore, based on these findings we aimed to assess the effects of reproductive factors
121 on risk of intrinsic breast cancer subtypes.

122 **Methods**

123 **UK Biobank**

124 The UK Biobank study is a large population-based cohort of 502,682 individuals who were
125 recruited at ages 37–73 years across the UK between 2006 and 2010. The study includes
126 extensive health and lifestyle questionnaire data, physical measures, and biological samples
127 from which genetic data has been generated. The study protocol is available online, and
128 more details have been published elsewhere.⁽⁴⁸⁾ At recruitment the participants gave
129 informed consent to participate and be followed up. Unless stated otherwise, in UK Biobank,

130 the factors investigated in the current study were derived from questionnaire responses at
131 the baseline assessment.

132 **Reproductive factors**

133 The reproductive factors investigated in this study were: age at first live birth, age at last live
134 birth, number of live births and parous status (ever/never given birth at the time of
135 assessment). Hereafter these will be referred to as age at first birth, age at last birth, number
136 of births and ever parous status. In UK Biobank these reproductive factors were derived from
137 questionnaire responses at the baseline assessment, further details are in **Supplementary**
138 **File 1**.

139 To identify genetic variants robustly related to each of the reproductive factors, we
140 performed genome-wide association study (GWAS) for each reproductive factor among
141 women in the UK Biobank. Each GWAS was performed using the Medical Research Council
142 (MRC) Integrative Epidemiology Unit (IEU) UK Biobank GWAS pipeline,^(49,50) and further
143 details can be found in **Supplementary File 1**.

144 **Additional variables**

145 We additionally included age at menarche and menopause, and adiposity measures in
146 childhood and adulthood. Further details on how age at menarche and menopause were
147 defined can be found in **Supplementary File 1**. We explored adiposity in childhood using
148 the comparative body size measure obtained from the baseline questionnaire in UK Biobank.
149 Participants were asked, "When you were 10 years old, compared to average would you
150 describe yourself as:", and were given the options: "Thinner", "Plumper" and "About
151 average". We investigated adiposity in adulthood using body size based on BMI. BMI was
152 derived from height and weight measured during the initial UK Biobank Assessment Centre
153 visit. The categorical body size measure was composed on three groups based on the same
154 proportions as the childhood body size variable.

155 We performed GWAS for age at menarche and menopause among women in the UK
156 Biobank similarly to the reproductive factors stated above. We obtained female only GWAS
157 summary statistics for childhood and adulthood body size from Richardson et al. 2020,
158 where they performed GWAS using a similar approach.⁽⁵¹⁾

159 **Breast cancer association consortium**

160 **Overall breast cancer**

161 The overall breast cancer risk GWAS summary statistics were obtained from the Breast
162 Cancer Association Consortium (BCAC) which used iCOGS, OncoArray and other GWAS
163 data for 133,384 breast cancer cases and 113,789 controls of European ancestry from 82
164 studies.⁽⁵²⁾

165 **Estrogen receptor status**

166 GWAS summary statistics for breast cancer risk stratified by ER status were obtained from
167 BCAC.⁽⁵³⁾ The data contained 69,501 ER positive cases, 21,468 ER negative cases and
168 105,974 controls of European ancestry.

169 **Estrogen receptor negative subtypes**

170 We additionally investigated the birth-related reproductive factors, which showed an effect
171 on risk of ER positive or negative breast cancer, in relation to intrinsic breast cancer
172 subtypes that are characterised by being hormone-receptor positive or negative. GWAS
173 summary statistics used in this study were for risk of two breast cancer subtypes; human
174 epidermal growth factor receptor 2 (HER2) enriched (718 cases) and triple negative (2,006
175 cases), were obtained from the BCAC. The data involved individuals of European ancestry
176 and contained 20,815 controls.⁽⁵²⁾

177 **Genetic correlation**

178 Genetic correlations between the reproductive factors, age at menarche and menopause,
179 adiposity measures and breast cancer risk outcomes were calculated using linkage
180 disequilibrium score regression (LDSC) and the UK Biobank, and BCAC GWAS summary
181 statistics.^(54,55) Further details can be found in **Supplementary File 1**.

182 **STROBE-MR**

183 The Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian
184 Randomization (STROBE-MR) guidelines have been followed in the analysis of this work
185 (**Supplementary File 1**).^(56,57)

186 **Univariable analysis**

187 We performed univariable MR (UVMR) of ever parous status, age at first birth, age at last
188 birth and number of births on overall, ER positive and ER negative breast cancer risk. We
189 additionally performed UVMR on reproductive factors that appear to have an effect on risk of
190 ER negative breast cancer, on ER negative breast cancer subtypes (HER2 enriched and
191 triple negative). We used the function “mr()” from TwoSampleMR R package.⁽⁵⁰⁾ The primary
192 analysis focused on the inverse variance weighted (IVW) MR method.⁽⁵⁸⁾ GWAS estimates
193 for age at first birth, age at last birth, number of births were standardized based on the
194 standard deviation of the phenotypic exposure (mean = 0 and standard deviation (SD) = 1)
195 prior to performing MR.

196 **Evaluating univariable mendelian randomization assumptions**

197 Mendelian randomization has three main assumptions: the relevance assumption states the
198 genetic variants instrumented are associated with the exposure of interest; the
199 independence assumption states the genetic variants and outcome under investigation do
200 not share any common causes; the exclusion restriction assumption states that the

201 instrumented genetic variants do not affect the outcomes via any pathway other than the
202 exposure being examined.⁽⁵⁹⁾

203 To evaluate the strength of the genetic instruments, we determined the mean F statistic for
204 each trait,⁽⁵⁹⁾ and considered a F statistic of 10 or above as indicative of a strong instrument.

205 To evaluate whether the genetic instruments are pleiotropic, we performed MR using
206 additional methods: Weighted mode,⁽⁶⁰⁾ Weighted median,⁽⁶¹⁾ and MR Egger.^(62,63) The
207 intercept from MR Egger regression was used to determine directional pleiotropy.⁽⁶²⁾

208 We also applied MR-PRESSO (Mendelian Randomisation Pleiotropy RESidual Sum and
209 Outlier) which detects presence of horizontal pleiotropy and accounts for horizontal
210 pleiotropy by correcting for single nucleotide polymorphisms (SNPs) identified as outliers.⁽⁶⁴⁾

211 Detail on the functions used in this analysis can be found in **Supplementary File 1**.

212 **Multivariable analysis**

213 As previous outlined, MVMR is an extension of MR that allows the estimation of effects of
214 two or more exposures that may be genetically correlated, on an outcome.^(36,37) It is
215 advantageous as it enables adjustment for possible pleiotropy that may occur if the
216 exposures have a shared genetic component,^(36,37) which may violate the exclusion
217 restriction assumption of UVMR. In addition, MVMR allows adjustment for factors that may
218 be confounders of the association between the genetic variants being used as instruments
219 for the exposure of interest and the outcome, which would violate the independence
220 assumption. Finally, if the exposures are on the same causal pathway, MVMR can be used
221 to estimate a direct effect independently of a potential mediator.⁽⁶⁵⁾

222 We performed MVMR to evaluate the direct effect, using the IVW method, of ever parous
223 status, age at first birth, age at last birth and number of births on overall, ER positive and ER
224 negative breast cancer risk. We additionally performed MVMR of reproductive factors, that
225 appeared to have an effect on risk of ER negative breast cancer, on breast cancer subtypes
226 (HER2 enriched and triple negative). We used the “ivw_mvmmr()” from the MVMR R

227 package.⁽⁶⁶⁾ GWAS estimates for age at first birth, age at last birth and number of births were
228 standardized based on the standard deviation of the phenotypic exposure (mean = 0 and SD
229 = 1) prior to performing MVMR.

230 Factors that were considered for MVMR adjustment were 1) traits we identified have an
231 causal effect on the exposure and therefore could violate the independence assumption 2)
232 traits that we have shown have a genetic correlation with the exposure but for which there is
233 little evidence for a causal relationship with the exposure 3) traits that we have identified as
234 potential mediators.^(32,33) We adjusted for reproductive factors, age at menarche and
235 menopause, and finally adiposity measures, we adjusted for each trait in turn. (**Figure 1**)

236 While it may seem implausible to adjust for ever parous status in the MVMR analysis of age
237 at first and last birth, since the age at first and last birth GWAS are conditioned on ever
238 parous status by being performed only on parous women, while the breast cancer risk
239 GWAS is not, it is important to adjust for ever parous status to reduce bias and allow
240 estimation of direct effects.⁽⁶⁷⁾

241 **Evaluating multivariable mendelian randomization assumptions**

242 We evaluated the instrument strength for the two exposures in the MVMR setting using a
243 conditional F-statistic,⁽⁶⁸⁾ and used a modified form of Cochran's Q statistic to evaluate
244 evidence of horizontal pleiotropy.⁽⁶⁶⁾ Where we identify weak instruments and/or evidence of
245 pleiotropy we additionally performed MVMR estimation using Q-statistic minimisation
246 allowing for weak instruments and balanced heterogeneity.⁽⁶⁶⁾ We did not use this method
247 when the F statistic falls below 4, since the method does not perform well in this
248 circumstance.⁽⁶⁶⁾ We additionally performed the MVMR analysis using the MR Egger method
249 to evaluate the presence of directional pleiotropy.^(69,70) This analysis was performed similarly
250 to the initial MVMR analysis in relation to overall, ER positive and ER negative breast cancer
251 risk.

252 Further details on the functions used to perform these analyses can be found in

253 **Supplementary File 1.**⁽⁶⁶⁾

254 Results

255 UK Biobank

256 273,238 women from UK Biobank were included. The mean age at assessment was 56

257 years (SD=8), further sample characteristics are shown in **Table 1**.

258

259 **Table 1 UK Biobank study characteristics included in this study**

260 N=Sample size, SD= Standard deviation. IQR = Interquartile range. UKBB: UK Biobank

UKBB trait	N	Mean (SD)
Age at menarche (years)	243 898	13.0 (1.6)
Age at first live birth (years)	203 606	25.9 (5.1)
Age at last live birth (years)	203 356	30.1 (5.2)
Age at menopause (years)	143 791	49.7 (5.1)
Number of live births	250 746	1.8 (1.2)
Adulthood BMI	250 746	27.1 (5.2)
UKBB trait	% (N)	
Never parous	18.69 (49 358)	
Childhood body size	About average	50.47 (135 399)
	Thinner	31.80 (85 316)
	Plumper	17.74 (47 585)

261

262 Genome wide association studies

263 **Table 2** displays the number of SNPs associated with each reproductive factor at genome-

264 wide significance ($p \text{ value} < 5 \times 10^{-8}$) after LD clumping and harmonising with the outcome.

265 **Table 2 Univariable instrument strength of each trait of interest. N: sample size,**
 266 **nSNPs: number of SNPs.**

Trait	Trait N	Breast cancer outcome	nSNPs	F statistic
Age at first live birth	203 606	Overall	34	39.18
		ER status	41	38.28
		Subtypes	34	39.18
Age at last live birth	203 356	Overall	8	34.11
		ER status	9	33.98
		Subtypes	8	34.11
Number of live births	250 746	Overall	8	43.64
		ER status	9	42.79
		Subtypes	8	43.64
Ever parous status	250 746	Overall	4	39.43
		ER status	4	39.43
		Subtypes	4	39.43

267

268 **Genetic correlation**

269 Many of the reproductive factors, adiposity measures, as well as age at menarche and
 270 menopause were strongly genetically correlated. Exceptions for the four reproductive factors
 271 were: childhood body size and age at first birth, age at last birth and number of births.
 272 **(Figure 2, Supplementary File 2: Table S1)** In addition, age at first and last birth were
 273 inversely genetically correlated with overall and ER negative breast cancer risk but not ER
 274 positive. Ever parous status and number of births were not genetically correlated with any of
 275 the breast cancer risk outcomes. **(Figure 2, Supplementary File 2: Table S1)**

276 **Estimated causal effects between reproductive factors and** 277 **adiposity measures**

278 Previous research has shown interrelationships between the four reproductive factors of
279 interest, as well as causal effects between the reproductive factors, adiposity in childhood
280 and adulthood, and age at menarche and menopause (**Figure 3**). Further details on these
281 findings are provided elsewhere.^(32,33)

282 **Mendelian randomization**

283 In the univariable MR analysis, all traits had an F statistic over the standard threshold of 10
284 (**Table 2**). However, in the multivariable analysis the conditional F statistic was reduced for
285 all traits and was below 10 for all reproductive factors investigated; age at first birth, age at
286 last birth, number of births and ever parous status, with adjustment for at least one factors in
287 the MVMR analysis (**Supplementary File 2: Table S1**). We therefore present MVMR
288 estimated using Q-statistic minimisation, which is robust to weak instruments and balanced
289 heterogeneity, as the primary analysis. However, this was not performed where the
290 instruments were deemed too weak, with an F statistic of less than 4. This was the case for
291 age at last birth adjusted for age at menarche, age at menopause, and adulthood body size
292 (each separately); number of births adjusted for age at menarche; and, ever parous status
293 adjusted for age at menarche, age at menopause, age at first birth, childhood and adulthood
294 body size (each separately) (**Supplementary File 2: Table S2**). In these cases, MVMR
295 using the IVW method are presented but these results should be interpreted with caution
296 due to the risk of weak instrument bias. In the MVMR analysis, we adjusted for each factor in
297 turn, thus there were only two exposures: the exposure of interest and the adjustment, and
298 the outcome in each MVMR model.

299 ORs for age at first birth, age at last birth and number of births are shown as per SD in the
300 phenotypic exposure.

301 **Ever parous status**

302 Ever parous status is a binary exposure and therefore, to facilitate interpretation, the effects
303 in **Supplementary File 2: Table S3** and **Figure 4A** have been converted from "ORs per log
304 odds" in the main text here to per doubling in the genetic liability to ever being parous.

305 In the univariable analysis we identified a protective effect of being parous on overall
306 (UVMR: 0.17, CI: 0.05, 0.58), ER positive breast cancer risk (UVMR OR: 0.17, CI: 0.05,
307 0.56) and ER negative breast cancer risk (UVMR OR: 0.20, CI: 0.04, 1.02).

308 Multivariable adjustments revealed wide confidence intervals spanning the null for the effect
309 of ever parous status on overall breast cancer risk when adjusted for age at menarche
310 (MVMR OR: 0.74, CI: 0.38, 1.47), childhood body size (MVMR OR: 0.99, CI: 0.43, 2.28) and
311 adulthood body size (MVMR OR: 1.52, CI: 0.80, 2.89); ER positive when adjusted for age at
312 first birth (MVMR OR: 0.42, CI: 0.16, 1.14), age at menarche (MVMR OR: 0.94, CI: 0.47,
313 1.89), childhood body size (MVMR OR: 1.01, CI: 0.41, 2.47) and adulthood body size
314 (MVMR OR: 1.52, CI: 0.75, 3.05); and ER negative breast cancer risk when adjusted for age
315 at first birth (MVMR OR: 0.44, CI: 0.11, 1.81), age at last birth (MVMR estimated using Q-
316 statistic minimisation OR: 0.16, CI: 0.02, 1.47), age at menarche (MVMR OR: 0.78, CI: 0.31,
317 1.92), age at menopause (MVMR OR: 0.46, CI: 0.15, 1.40), childhood body size (MVMR OR:
318 1.27, CI: 0.42, 3.89) and adulthood body size (MVMR OR: 1.76, CI: 0.71, 4.38) (**Figure 4A,**
319 **Supplementary File 2: Table S3**). It is worth noting that conditional F statistics for ever
320 parous status fell below 4 for all adjustment except age at last birth, and the number of SNPs
321 used as instruments were 4 or less (**Supplementary File 2: Table S2**).

322 **Age at first birth**

323 In the univariable analysis we found little evidence for an effect of age at first birth on overall
324 (UVMR OR: 1.00, 95% CI 0.85, 1.18) or ER positive breast cancer risk (UVMR OR: 1.04, CI:
325 0.89, 1.20) but an inverse effect on ER negative breast cancer risk (UVMR OR: 0.76, CI:
326 0.61, 0.95). (**Figure 4B, Supplementary File 2: Table S4**)

327 We identified a protective effect of later age at first birth on overall breast cancer risk after
328 adjusting for ever parous status in the multivariable analysis (MVMR estimated using Q-
329 statistic minimisation OR: 0.74, CI: 0.57, 0.93). Conversely, similarly to the univariable
330 analysis we find minimal evidence of an effect of age at first birth on ER positive breast
331 cancer risk. Although it's worth noting that adjusting for age at menarche reveals the largest
332 positive effect although confidence intervals are wide, overlapping the univariable estimate
333 and the null (MVMR estimated using Q-statistic minimisation OR: 1.15, CI: 0.86, 1.50). The
334 inverse effect on ER negative breast cancer risk we identified in the univariable analysis
335 somewhat attenuated with adjustment for age at menarche with confidence intervals just
336 crossing the null (MVMR estimated using Q-statistic minimisation OR: 0.83, CI: 0.62, 1.06)
337 and age at menopause (MVMR estimated using Q-statistic minimisation OR: 0.80, CI: 0.66,
338 1.01). (**Figure 4B, Supplementary File 2: Table S4**)

339 **Age at last birth**

340 We found little evidence for an effect of age at last birth on overall (UVMR OR: 0.98, CI:
341 0.75, 1.29), ER positive (UVMR OR: 0.97, CI: 0.73, 1.29) and ER negative (UVMR OR: 0.83,
342 CI: 0.48, 1.45) breast cancer risk in the univariable analysis (**Figure 4C, Supplementary**
343 **File 2: Table S5**).

344 However, we found evidence of an inverse effect of age at last birth adjusting for adulthood
345 body size on overall (MVMR OR:0.70, CI: 0.49, 0.99), and ER negative breast cancer risk
346 (MVMR OR: 0.54, CI: 0.34, 0.87) (**Figure 4C, Supplementary File 2: Table S5**).

347 Adjusting for age at first birth revealed an inverse effect of larger magnitude on overall
348 (MVMR OR 0.56, CI: 0.22, 1.38) and ER positive breast cancer risk (MVMR OR 0.50, CI:
349 0.21, 1.19) compared to the univariable analysis. Adjusting for number of births revealed
350 little evidence for an effect of age at last birth on overall, ER positive and ER negative breast
351 cancer risk, similar to the UVMR analysis (**Figure 4C, Supplementary File 2: Table S5**).

352 **Number of births**

353 We found limited evidence for an effect of number of births on overall (UVMR OR: 0.78, CI:
354 0.43, 1.41), ER positive (UVMR OR: 0.78, CI: 0.46, 1.34) or ER negative breast cancer risk
355 (UVMR OR: 0.73, CI: 0.37, 1.43) in the univariable analysis, with a similar result in the
356 multivariable analyses. (**Figure 4D, Supplementary File 2: Table S6**)

357 **Estrogen receptor negative subtypes**

358 Since we found some evidence for an effect of age at first birth on ER negative breast
359 cancer risk with relatively strong instruments, we investigated the effects of age at first birth
360 on HER2 enriched and triple negative breast cancer risk. We used UVMR and MVMR in a
361 similar manner as in relation to overall, ER positive and negative breast cancer risk, detail on
362 evaluating the MR assumptions can be found in **Supplementary File 1**.

363 The univariable analysis revealed limited evidence of an effect of age at first birth on HER2
364 enriched breast cancer risk (UVMR OR: 0.78, CI:0.47, 1.30), while adjusting for number of
365 births (MVMR estimated using Q-statistic minimisation OR: 0.28, CI: 0.11, 0.57) and ever
366 parous status (MVMR estimated using Q-statistic minimisation OR: 0.38, CI: 0.17, 0.80)
367 revealed inverse effects (**Supplementary File 2: Table S4, Supplementary File 3: Figure**
368 **S5**). We found minimal evidence for an effect of age at first birth on triple negative breast
369 cancer risk in the UVMR analysis (UVMR OR: 0.88, CI: 0.63, 1.24) which was similar in the
370 MVMR analysis (**Supplementary File 2: Table S4, Supplementary File 3: Figure S5**).

371 **Evaluating univariable mendelian randomization assumptions**

372 There was consistency between the univariable analysis using the IVW method and the
373 additional MR methods (MR Egger, weighted median and weighted mode) for the effects of
374 age at last birth and number of births. However, there were inconsistencies with ever parous
375 status and age at first birth. While the effect estimates for ever parous status on overall and
376 ER negative breast cancer risk were largely similar to the IVW method across the additional
377 methods, confidence intervals for the MR Egger and weighted mode spanned the null, with

378 the MR Egger estimate being largely attenuated (**Supplementary File 2: Table S7**). In
379 addition, the effect of age at first birth on ER negative breast cancer risk was similar when
380 using the weighted median and mode methods but not MR Egger, where the estimate was in
381 the positive direction. However, across all additional methods confidence intervals spanned
382 the null (**Supplementary File 2: Table S7**).

383 Using the MR-PRESSO method, we identified potential outliers for all relationships
384 assessed, except for the effect of age at last birth on overall and ER positive breast cancer
385 risk, and the effect of age at first birth on HER2 enriched and Triple negative breast cancer
386 risk. Correcting for identified outliers didn't appear to change the evidence for effects
387 identified in the initial analysis using the IVW method. However, it is worth noting that of the
388 4 SNPs used as instruments for ever parous status, 2 were identified as outliers in relation to
389 overall and ER positive breast cancer risk and 1 in relation to ER negative, meaning this
390 analysis was based on a limited number of instruments (**Supplementary File 2: Table S8**)
391 The Egger intercept test didn't identified evidence for pleiotropy across any relationships
392 investigated. (**Supplementary File 2: Table S9**)

393 **Evaluating multivariable mendelian randomization assumptions**

394 In the multivariable analysis we identified evidence of heterogeneity for all models except for
395 those that included: age at first birth and age at last birth; and, age at last birth and ever
396 parous status in relation to ER positive breast cancer risk, those that included age at first
397 birth and age at last birth in relation to ER negative breast cancer risk, and in the analysis of
398 age at first birth for all adjustments except childhood body size on HER2 enriched breast
399 cancer risk. (**Supplementary File 2: Table S10**). Where there was evidence for
400 heterogeneity and the F statistic was 4 or higher, we performed MVMR estimation using Q-
401 statistic minimisation which is robust to weak instruments and balanced heterogeneity
402 (**Supplementary File 2: Table S3-6**). As previously stated, these results were presented as
403 the primary analysis.

404 *MR Egger*

405 We performed the multivariable analysis using the MR Egger method to assess for
406 pleiotropy and the genetic associations with the first exposure were set to be positive. The
407 estimates were mostly consistent with the main analysis, although in many cases, the
408 confidence intervals spanned the null. However, there were inconsistencies. Of note, we
409 identified some evidence for an effect of ever parous status on ER positive breast cancer
410 risk adjusted for age at first birth (OR: 0.30, CI: 0.09, 0.97) which we do not find strong
411 evidence for in the main analysis. (**Supplementary File 2: Table S11, Supplementary File**
412 **3: Figure S1**) In addition, the main analysis identified strong evidence for an effect of age at
413 first birth on overall breast cancer risk adjusted for ever parous status which was not seen
414 using the MR Egger method. (**Supplementary File 2: Table S11, Supplementary File 3:**
415 **Figure S2**) The MR Egger method in the multivariable analysis revealed an inverse effect of
416 number of births, adjusted for age at first birth, on overall (OR: 0.49, CI: 0.26, 0.91) and ER
417 negative breast cancer risk (OR: 0.42, CI: 0.19, 0.91) which was not identified in the main
418 analysis. (**Supplementary File 2: Table S11, Supplementary File 3: Figure S4**)

419 Where we identified an effect using the IVW method of age at first birth on HER2 enriched
420 breast cancer risk in the multivariable analysis, wide confidence intervals included the null
421 using the MR Egger method (adjusted for ever parous status: OR: 1.08, CI: 0.07, 15.71,
422 adjusted for number of births: OR: 0.27, CI: 0.05, 1.62). (**Supplementary File 2: Table S11**)
423 Although we did identify some evidence for an inverse effect of age at first birth on HER2
424 enriched breast cancer risk adjusting for age at menopause (OR: 0.46, CI: 0.21, 1.00).
425 (**Supplementary File 2: Table S11**) We additionally identified minimal evidence for an effect
426 of age at first birth on triple negative breast cancer risk in the MVMR analysis using the MR
427 Egger method, similarly to the IVW method. (**Supplementary File 2: Table S11**)

428 Discussion

429 In this study we used univariable, and multivariable Mendelian randomization methods to
430 investigate the role of parity-related reproductive factors on breast cancer development.

431 Using these approaches, while we found evidence to support that having children has a
432 protective effect on breast cancer risk in the univariable analysis, we could not determine the
433 direct effect independently of age at menarche, age at menopause, reproductive factors and
434 adiposity measures due to the analysis of ever parous status being subject to weak
435 instruments across all the multivariable models. It may be that a larger GWAS of ever parous
436 status is required to use MR more effectively, resulting in a higher number of SNPs identified
437 as robustly associated with ever parous status.

438 In addition, we found little evidence to support a protective effect of an earlier age at first
439 birth on breast cancer risk.^(2,4-8) This may be due to biases such as residual pleiotropy that
440 have driven our MR estimates towards the null. We discuss this limitation in more detail
441 below. However, we do find evidence for an adverse effect of earlier age at first birth on ER
442 negative breast cancer risk in the univariable analysis, although this attenuated somewhat
443 with adjustment for age at menarche and menopause, with confidence intervals just crossing
444 the null. We further evaluated the effect of age at first birth on ER negative breast cancer risk
445 by looking at the HER2 enriched and triple negative subtypes and we find little evidence of
446 an effect in the univariable analysis. However, after adjustment for number of births and ever
447 parous status, we found that earlier age at first birth increases HER2 enriched breast cancer
448 risk.

449 Since ER negative breast cancer is typically seen in younger, premenopausal women, the
450 increased risk immediately after an earlier first pregnancy is more likely to relate to ER
451 negative breast cancer. This risk is likely to be due to the proliferative effects of pregnancy
452 due to high levels of ovarian hormones which up-regulate proliferation related genes,
453 increasing the risk of mutation.^(9,13) Since the effect of age at first birth on breast cancer risk

454 we find is isolated to HER2 enriched breast cancer, it could be that an increased risk of
455 HER2 mutations during pregnancy leads to an overexpression of HER2, driving the
456 development of breast cancer.⁽⁷¹⁾ Additionally, pregnancy is thought to increase breast
457 cancer risk in the shorter-term because of the inflammatory and wound healing properties of
458 mammary involution which occurs after lactation.^(13,15,16)

459 Previous observational research has identified that having an age at first birth at or younger
460 than 25 is associated with a lower level of estrogen post menopause compared to
461 nulliparous women.⁽⁷²⁾ Therefore, it would be valuable for future studies to evaluate whether
462 levels of ovarian hormones such as estrogen and progesterone play a role in the relationship
463 between birth-related factors and breast cancer risk, for example, in a mediation MVMR
464 framework.

465 Point estimates for the effect of age at last birth across the univariable and multivariable
466 models suggest an adverse effect of a younger age at last birth on ER negative breast
467 cancer risk, which is consistent with a previous MR study,⁽³¹⁾ but conflicts with the
468 observational literature.^(7,8,24-26)

469 It may seem implausible to adjust for ever parous status in the analysis of age at first and
470 last birth on breast cancer risk outcomes since these events only occur in women who have
471 given birth and therefore ever being parous could not observationally confound or mediate
472 this relationship. However, a recent study has shown that performing MR without adjusting
473 for a factor that was conditioned on in one of the exposure or outcome GWAS can bias the
474 resulting MR estimate, and adjusting for that factor can allow estimation of direct effects.⁽⁶⁷⁾

475 While age at first and last birth are conditioned on ever parous status since the GWAS was
476 for these traits only being performed among parous women, the GWAS of breast cancer risk
477 was not performed solely on parous women or adjusted for ever parous status, and so it is
478 important to adjust for ever parous status in an MVMR model. Moreover, in the analysis of
479 age at first birth on overall breast cancer risk, adjusting for ever parous status did modify the
480 estimate to reveal an inverse effect, compared to the limited evidence identified in the

481 univariable analyses. It would have been informative to perform the analysis of age at first
482 and last birth using GWAS from the BCAC consortium performed solely in parous women,
483 however this data, to our knowledge, is not publicly available.

484 We find limited evidence that number of births has an impact on breast cancer risk in either
485 the univariable or multivariable analysis, which is consistent with a previous MR study.⁽³⁰⁾

486 Due to weak instruments, we cannot make any definitive conclusions regarding age at last
487 birth or ever parous status for most MVMR models. Where instruments were at least slightly
488 stronger, with a F statistic of 4 or higher, in the multivariable analyses, we performed MVMR
489 estimation using Q-statistic minimisation, and the direction and strength of evidence was
490 largely similar to the UVMR model. While weak instruments prevent in depth speculation, in
491 the case of age at last birth, it may be that we are capturing the effect of age at first birth
492 since these factors are highly genetically correlated. Indeed, in the multivariable analysis of
493 age at last birth on breast cancer risk outcomes adjusted for age at first birth confidence
494 intervals were particularly wide, compared to the other models with age at last birth as the
495 exposure.

496 We used the MR Egger method to assess for evidence of pleiotropy, for the most part, this
497 method revealed similar evidence to the main analysis. However, we did not identify
498 evidence for the inverse effect of age at first birth on overall breast cancer risk, adjusting for
499 ever parous status in the MVMR analysis using the MR Egger method which we did find in
500 the main analysis. In addition, using the MR Egger method, adjusting for age at first birth
501 uncovered evidence that a higher number of children leads to reduced overall and ER
502 negative breast cancer risk. While this suggests the main analysis may be biased by
503 pleiotropy, the instrument strength was deemed weak in the multivariable analyses, with an
504 F statistic below 10, which limited confidence in these results and prevented further
505 assessment of pleiotropy using robust methods.⁽⁷³⁾

506 **Strengths**

507 The main strength of this study is the use of a multivariable MR approach allowing the
508 evaluation of direct effects of reproductive factors on breast cancer risk outcomes. There has
509 not previously been an informed, systematic MVMR study, using information from previous
510 work on genetic correlation and causal relationships between the reproductive factors of
511 interest and age at menarche and menopause, adiposity measures and other reproductive
512 factors.^(32,33)

513 Additionally, this study uses the UK Biobank study which has phenotypic and genetic data
514 on a large number of individuals which increases the likelihood of identifying strong genetic
515 instruments for each exposure of interest. Another strength is the use of GWAS data from
516 the BCAC consortium, allowing the investigation of risk of breast cancer by ER status, and
517 risk of breast cancer subtypes.

518 **Limitations**

519 This study has a number of limitations that should be considered when interpreting results.
520 Firstly, weak instruments were a key limitation when investigating the effects of age at last
521 birth and ever parous status on breast cancer risk outcomes, especially in the multivariable
522 models. This can be attributed to the genetic correlation that is present between these traits
523 and the other factors included in the multivariable analyses,⁽³²⁾ as well as the low number of
524 SNPs that arise as genome-wide significant in relation to these traits (8 and 4 for age at last
525 birth and ever parous status, respectively). Where the instrument strength was not deemed
526 too weak, with an F statistic of 4 or higher, we additionally performed MVMR estimation
527 using Q-statistic minimisation that is developed to be robust to weak instruments. However,
528 this was not possible for number of births adjusted for age at menarche, and all MVMR
529 models for ever parous status, except with adjustment for age at last birth, and additionally
530 for age at last birth, except with adjustment for number of births, ever parous status and
531 childhood body size. Furthermore, it would have been useful to include multiple adjustments

532 within the MVMR models, however instruments were too weak to facilitate this. Larger
533 GWAS and stronger instruments are required to extend our MVMR approach, however it
534 may be that, due to the aforementioned genetic correlation, an exceptionally large sample
535 size would be required.

536 Secondly, as previously mentioned, we may not find a protective effect of an early age at
537 first birth on breast cancer risk because of biases due to violations of the MR assumptions. It
538 is plausible that we may be capturing residual pleiotropic effects as a result of age at
539 menarche confounding the relationship between age at first birth and breast cancer risk. The
540 univariable analysis reveals little evidence for an effect of age at first birth on ER positive
541 breast cancer risk (OR: 1.04, CI: 0.89, 1.20 per SD increase in age at first birth), while the
542 effect in the multivariable analysis, adjusted for age at menarche, is larger in the protective
543 direction, although confidence intervals crossed the null (OR: 1.18, CI: 0.93, 1.51 per SD
544 increase in age at first birth). Given the conditional F statistics in this MVMR model (age at
545 first birth: 5.3, age at menarche: 25.0) and that age at menarche and first birth are strongly
546 genetically correlated, the effect in the multivariable analysis may be biased by weak
547 instruments or from pleiotropy via another trait. Stronger instruments for age at first birth in
548 an MVMR model adjusted for age at menarche might reveal the protective effect seen in the
549 observational literature.

550 Thirdly, data on the reproductive factors investigated have been derived from self-report
551 which can be unreliable. However, we can be relatively confident that reports of age at first
552 and last birth and number of births are accurate since these are significant life events that
553 are likely to be reliably recalled. Additionally, there has typically been good replication of the
554 genetic scores for the reproductive traits in other cohorts.⁽⁷⁴⁻⁷⁶⁾

555 In addition, since the reproductive factors investigated in this study are bio-social, findings
556 may not be generalisable to other non-UK populations that have different social norms.
557 Additionally, the findings may not be generalisable to younger populations. It would have
558 been valuable to compare the findings in the study, where the population was predominately

559 European, to populations with other ancestries such as Asian and African. However, while
560 Asian and African ancestry GWAS summary statistics for breast cancer risk are available,
561 those for the four reproductive factors investigated here are not, to the best of our
562 knowledge.

563 Finally, the MR analysis performed assumes a constant effect of age at first and last birth,
564 and number of births, on breast cancer risk. Future studies could use MR to evaluate the
565 effects of an early age at first birth compared to an average and late age, and vice versa.
566 This approach would also enable the evaluation of whether there is a cross-over point where
567 breast cancer risk changes from parous women to nulliparous women at a certain age at first
568 birth as previously proposed.^(14,15) Again, further analysis would require the identification of
569 additional and stronger genetic instruments.

570 **Conclusion**

571 Using a series of univariable and multivariable MR analyses, we found minimal evidence for
572 a protective effect of early age at first birth on breast cancer risk that has been identified in
573 observational studies, which may be due to unmeasured biases. On the other hand, we
574 identify some evidence of an adverse effect on ER negative breast cancer risk, which
575 supports the notion that there is an immediate increase in breast cancer risk post pregnancy.
576 Understanding of mechanism behind ER negative breast cancers is important since it has
577 worse prognosis compared with those that are ER positive.

578 In addition, we show the number of births that women have has little effect on overall, ER
579 positive or ER negative breast cancer risk. Future studies considering non-linear
580 relationships between age at first birth and breast cancer risk may provide additional
581 insights.

582 While our findings for ever parous status and age at last birth do not all concur with the
583 existing literature from non-genetic studies, this may partly be due to limitations of the

584 MVMR approach, specifically relating to instrument strength, which may be overcome with
585 larger GWAS of these traits.

586 **Conflicts of interests**

587 The authors have declared no competing interests.

588 **Funding**

589 All authors work in a unit that receives funding from the University of Bristol and the UK
590 Medical Research Council (MRC) (MC_UU_00011/1, MC_UU_00011/5, MC_UU_00011/6).
591 C.P., G.C.S., L.D.H., A.F., and R.C.R. are members of the Menarche, Menstruation,
592 Menopause and Mental Health (4M) consortium, which was established with support from
593 the GW4 Alliance. C.P. is supported by a Wellcome Trust PhD studentship in Molecular,
594 Genetic and Lifecourse Epidemiology (108902/B/15/Z). L.D.H. is supported by Career
595 Development Awards from the UK MRC (MR/M020894/1). R.C.R. is supported by the
596 CRUK-funded Integrative Cancer Epidemiology Programme (C18281/A19169). B.L.-L. is
597 funded by a Vice-Chancellor's Research Fellowship from the University of Bristol and
598 acknowledges support from the Academy of Medical Sciences/Wellcome Trust/the
599 Government Department of Business, Energy and Industrial Strategy/British Heart
600 Foundation/Diabetes UK Springboard Award (SBF003/1170), Elizabeth Blackwell Institute
601 for Health Research (University of Bristol), and Wellcome Trust Institutional Strategic
602 Support Fund (204813/Z/16/Z).

603 **Authors' Contributions**

604 C.P. was responsible for Analysis, Investigation and Writing - Original Draft. E.S., G.C.S.,
605 L.D.H., A.F., B.L.-L. and R.C.R. were responsible for Conceptualization, Writing - Review &
606 Editing and Supervision. R.C.R. and B.L.-L. were additionally responsible for Investigation.
607 All authors read and approved the final manuscript.

608 **Acknowledgements**

609 This research has been conducted using the UK Biobank Resource under Application
610 Number 6326. We thank the participants and researchers from the UK Biobank who
611 contributed or collected data.

612 The breast cancer genome-wide association analyses for BCAC and CIMBA were supported
613 by Cancer Research UK (PPRPGM-Nov20\100002, C1287/A10118, C1287/A16563,
614 C1287/A10710, C12292/A20861, C12292/A11174, C1281/A12014, C5047/A8384,
615 C5047/A15007, C5047/A10692, C8197/A16565) and the Gray Foundation, The National
616 Institutes of Health (CA128978, X01HG007492- the DRIVE consortium), the PERSPECTIVE
617 project supported by the Government of Canada through Genome Canada and the
618 Canadian Institutes of Health Research (grant GPH-129344) and the Ministère de
619 l'Économie, Science et Innovation du Québec through Genome Québec and the PSRSIIRI-
620 701 grant, the Quebec Breast Cancer Foundation, the European Community's Seventh
621 Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175)
622 (COGS), the European Union's Horizon 2020 Research and Innovation Programme (634935
623 and 633784), the Post-Cancer GWAS initiative (U19 CA148537, CA148065 and CA148112 -
624 the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian
625 Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer
626 (CRN-87521), the Komen Foundation for the Cure, the Breast Cancer Research Foundation
627 and the Ovarian Cancer Research Fund.

628 **Ethics**

629 UK Biobank received ethical approval from the North West Multi-Centre Research Ethics
630 Committee (REC reference: 16/NW/0274) and was conducted in accordance with the
631 principles of the Declaration of Helsinki.

632 **Availability of data and materials**

633 The availability of all data analysed in this study has been referenced throughout the
634 manuscript and supplementary materials.

635 GWAS summary statistics for breast cancer risk outcomes were obtained from

636 <https://bcac.ccge.medschl.cam.ac.uk/bcacdata/>

637 **Figure legends**

638 Figure 1 **Variables included in multivariable mendelian randomization analysis.**

639 Figure 2 **Genetic correlation between reproductive factors, age at menarche and age at**
640 **menopause, and adiposity measures, and breast cancer risk outcomes.** * p-value<0.05,
641 ** p-value<0.01, *** p-value<0.001.

642 Figure 3 **Causal relationships estimated using mendelian randomization in previous**
643 **work (*Prince et al. 2022, Prince et al. 2023*).** OR: Odds ratio, CI: 95% confidence interval.
644 The width of the line/arrowhead indicates the magnitude of the estimate.

645 Figure 4 **Univariable and multivariable mendelian randomization assessing the effects**
646 **of A) ever parous status, B) age at first birth, C) age at last birth and D) number of**
647 **births on overall, ER positive and ER negative breast cancer risk.** No adjustment
648 indicated findings from the univariable analysis. Effect of ever parous status is presented per
649 log odds of ever parous status and effects of age at first birth, age at last birth and number of
650 births are presented per standard deviation increase in the phenotypic exposure. * initial
651 MVMR result are presented due to weak instruments preventing the use of MVMR estimated
652 using Q-statistic minimisation.

653 **Supplementary files**

654 Supplementary File 1: Supplementary Note

655 Supplementary File 2: Supplementary Tables

656 Supplementary File 3: Supplementary Figures

657 References

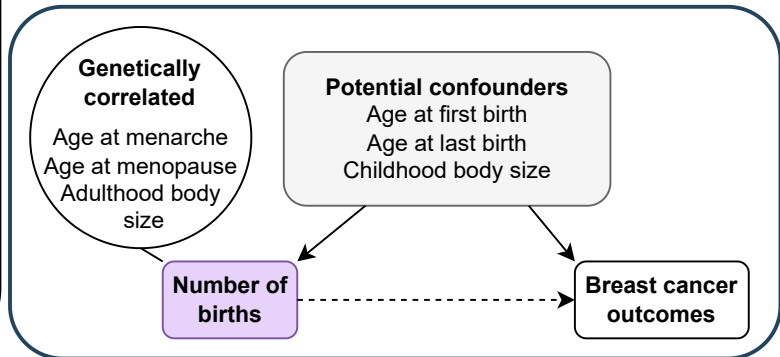
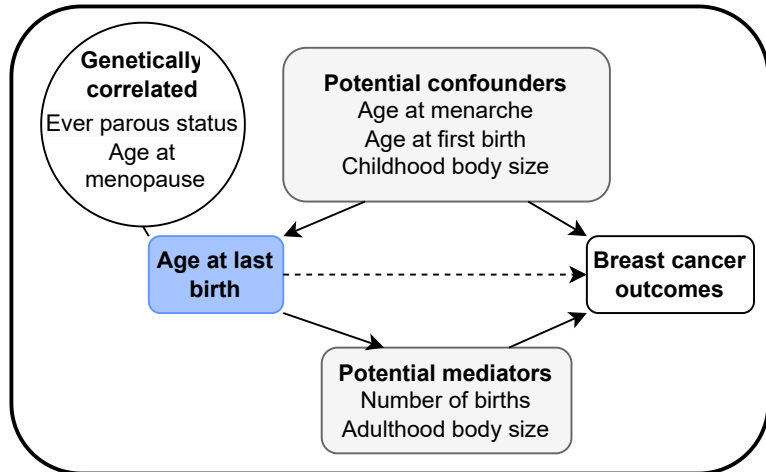
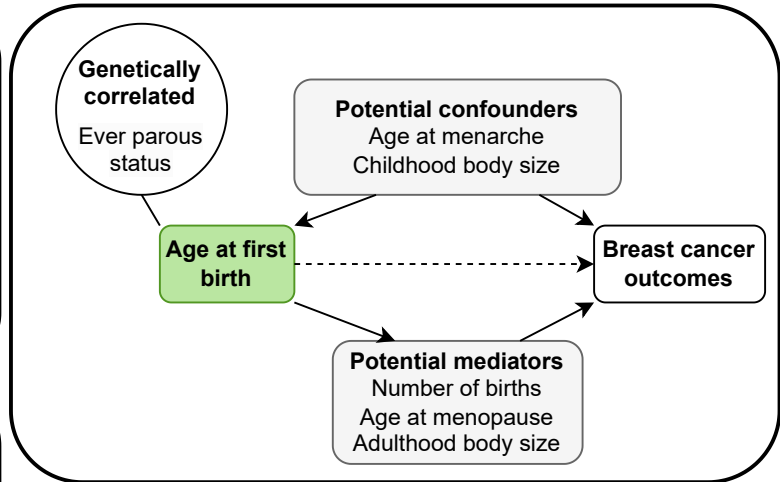
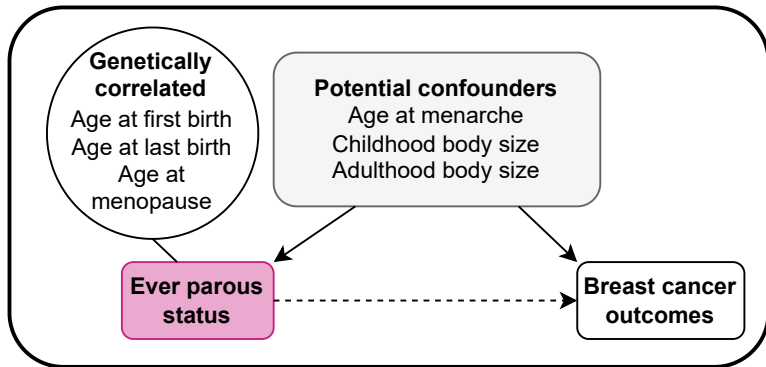
- 658 1. Ma H, Bernstein L, Pike MC, Ursin G. Reproductive factors and breast cancer risk
659 according to joint estrogen and progesterone receptor status: a meta-analysis of
660 epidemiological studies. *Breast Cancer Res*. 2006;8(4):R43.
- 661 2. Ewertz M, Duffy SW, Adami HO, Kvale G, Lund E, Meirik O, et al. Age at first birth,
662 parity and risk of breast cancer: a meta-analysis of 8 studies from the Nordic countries. *Int J*
663 *Cancer*. 1990;46(4):597-603.
- 664 3. Wu Y, Wang M, Sun W, Li S, Wang W, Zhang D. Age at last birth and risk of
665 developing breast cancer: a meta-analysis. *Eur J Cancer Prev*. 2020;29(5):424-32.
- 666 4. MacMahon B, Cole P, Lin TM, Lowe CR, Mirra AP, Ravnihar B, et al. Age at first birth
667 and breast cancer risk. *Bull World Health Organ*. 1970;43(2):209-21.
- 668 5. Chie WC, Hsieh C, Newcomb PA, Longnecker MP, Mittendorf R, Greenberg ER, et
669 al. Age at any full-term pregnancy and breast cancer risk. *Am J Epidemiol*. 2000;151(7):715-
670 22.
- 671 6. Dupont WD, Page DL. Breast cancer risk associated with proliferative disease, age
672 at first birth, and a family history of breast cancer. *Am J Epidemiol*. 1987;125(5):769-79.
- 673 7. Lambe M, Hsieh CC, Chan HW, Ekblom A, Trichopoulos D, Adami HO. Parity, age at
674 first and last birth, and risk of breast cancer: a population-based study in Sweden. *Breast*
675 *Cancer Res Treat*. 1996;38(3):305-11.
- 676 8. Trichopoulos D, Hsieh CC, MacMahon B, Lin TM, Lowe CR, Mirra AP, et al. Age at
677 any birth and breast cancer risk. *Int J Cancer*. 1983;31(6):701-4.
- 678 9. Russo IH, Russo J. Role of hormones in mammary cancer initiation and progression.
679 *J Mammary Gland Biol Neoplasia*. 1998;3(1):49-61.
- 680 10. Dall GV, Britt KL. Estrogen Effects on the Mammary Gland in Early and Late Life and
681 Breast Cancer Risk. *Front Oncol*. 2017;7:110.
- 682 11. Asztalos S, Gann PH, Hayes MK, Nonn L, Beam CA, Dai Y, et al. Gene expression
683 patterns in the human breast after pregnancy. *Cancer Prev Res (Phila)*. 2010;3(3):301-11.
- 684 12. Okoh V, Deoraj A, Roy D. Estrogen-induced reactive oxygen species-mediated
685 signalings contribute to breast cancer. *Biochim Biophys Acta*. 2011;1815(1):115-33.
- 686 13. Ruiz R, Herrero C, Strasser-Weippl K, Touya D, St Louis J, Bukowski A, et al.
687 Epidemiology and pathophysiology of pregnancy-associated breast cancer: A review.
688 *Breast*. 2017;35:136-41.
- 689 14. Lambe M, Hsieh C, Trichopoulos D, Ekblom A, Pavia M, Adami HO. Transient
690 increase in the risk of breast cancer after giving birth. *N Engl J Med*. 1994;331(1):5-9.
- 691 15. Schedin P. Pregnancy-associated breast cancer and metastasis. *Nat Rev Cancer*.
692 2006;6(4):281-91.
- 693 16. Slepicka PF, Cyrill SL, Dos Santos CO. Pregnancy and Breast Cancer: Pathways to
694 Understand Risk and Prevention. *Trends Mol Med*. 2019;25(10):866-81.
- 695 17. Pike MC, Spicer DV, Dahmouch L, Press MF. Estrogens, progestogens, normal
696 breast cell proliferation, and breast cancer risk. *Epidemiol Rev*. 1993;15(1):17-35.
- 697 18. MacMahon B, Cole P, Brown J. Etiology of human breast cancer: a review. *J Natl*
698 *Cancer Inst*. 1973;50(1):21-42.
- 699 19. Janerich DT, Hoff MB. Evidence for a crossover in breast cancer risk factors. *Am J*
700 *Epidemiol*. 1982;116(5):737-42.
- 701 20. Russo J, Tay LK, Russo IH. Differentiation of the mammary gland and susceptibility
702 to carcinogenesis. *Breast Cancer Res Treat*. 1982;2(1):5-73.
- 703 21. Leon DA. A prospective study of the independent effects of parity and age at first
704 birth on breast cancer incidence in England and Wales. *Int J Cancer*. 1989;43(6):986-91.
- 705 22. Wohlfahrt J, Melbye M. Age at any birth is associated with breast cancer risk.
706 *Epidemiology*. 2001;12(1):68-73.

- 707 23. Russo J, Moral R, Balogh GA, Mailo D, Russo IH. The protective role of pregnancy in
708 breast cancer. *Breast Cancer Res.* 2005;7(3):131-42.
- 709 24. Kalache A, Maguire A, Thompson SG. Age at last full-term pregnancy and risk of
710 breast cancer. *Lancet.* 1993;341(8836):33-6.
- 711 25. Hsieh CC, Chan HW, Lambe M, Ekblom A, Adami HO, Trichopoulos D. Does age at
712 the last birth affect breast cancer risk? *Eur J Cancer.* 1996;32A(1):118-21.
- 713 26. Albrektsen G, Heuch I, Tretli S, Kvale G. Breast cancer incidence before age 55 in
714 relation to parity and age at first and last births: a prospective study of one million Norwegian
715 women. *Epidemiology.* 1994;5(6):604-11.
- 716 27. Dall G, Risbridger G, Britt K. Mammary stem cells and parity-induced breast cancer
717 protection- new insights. *J Steroid Biochem Mol Biol.* 2017;170:54-60.
- 718 28. Richmond RC, Davey Smith G. Mendelian Randomization: Concepts and Scope.
719 *Cold Spring Harb Perspect Med.* 2022;12(1).
- 720 29. Sanderson E, Glymour MM, Holmes MV, Kang H, Morrison J, Munafo MR, et al.
721 Mendelian randomization. *Nat Rev Methods Primers.* 2022;2.
- 722 30. Vabistsevits M, Davey Smith G, Sanderson E, Richardson TG, Lloyd-Lewis B,
723 Richmond RC. Deciphering how early life adiposity influences breast cancer risk using
724 Mendelian randomization. *Commun Biol.* 2022;5(1):337.
- 725 31. Jia L, Lv W, Liang L, Ma Y, Ma X, Zhang S, et al. The Causal Effect of Reproductive
726 Factors on Breast Cancer: A Two-Sample Mendelian Randomization Study. *J Clin Med.*
727 2023;12(1).
- 728 32. Prince C, Sharp GC, Howe LD, Fraser A, Richmond RC. The relationships between
729 women's reproductive factors: a Mendelian randomisation analysis. *BMC Med.*
730 2022;20(1):103.
- 731 33. Prince C, Howe LD, Sharp GC, Fraser A, Richmond RC. Establishing the
732 relationships between adiposity and reproductive factors: a multivariable Mendelian
733 randomization analysis. *BMC Med.* 2023;21(1):350.
- 734 34. Collaborative Group on Hormonal Factors in Breast C. Menarche, menopause, and
735 breast cancer risk: individual participant meta-analysis, including 118 964 women with breast
736 cancer from 117 epidemiological studies. *Lancet Oncol.* 2012;13(11):1141-51.
- 737 35. Guo Y, Warren Andersen S, Shu XO, Michailidou K, Bolla MK, Wang Q, et al.
738 Genetically Predicted Body Mass Index and Breast Cancer Risk: Mendelian Randomization
739 Analyses of Data from 145,000 Women of European Descent. *PLoS Med.*
740 2016;13(8):e1002105.
- 741 36. Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of
742 multivariable Mendelian randomization in the single-sample and two-sample summary data
743 settings. *Int J Epidemiol.* 2019;48(3):713-27.
- 744 37. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of
745 pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol.* 2015;181(4):251-60.
- 746 38. Holm J, Eriksson L, Ploner A, Eriksson M, Rantalainen M, Li J, et al. Assessment of
747 Breast Cancer Risk Factors Reveals Subtype Heterogeneity. *Cancer Res.*
748 2017;77(13):3708-17.
- 749 39. De Mulder H, Laenen A, Wildiers H, Punie K, Poppe A, Remmerie C, et al. Breast
750 cancer subtype and survival by parity and time since last birth. *Breast Cancer Res Treat.*
751 2018;169(3):481-7.
- 752 40. McCarthy AM, Friebel-Klingner T, Ehsan S, He W, Welch M, Chen J, et al.
753 Relationship of established risk factors with breast cancer subtypes. *Cancer Med.*
754 2021;10(18):6456-67.
- 755 41. Ma H, Ursin G, Xu X, Lee E, Togawa K, Duan L, et al. Reproductive factors and the
756 risk of triple-negative breast cancer in white women and African-American women: a pooled
757 analysis. *Breast Cancer Res.* 2017;19(1):6.
- 758 42. Barnard ME, Boeke CE, Tamimi RM. Established breast cancer risk factors and risk
759 of intrinsic tumor subtypes. *Biochim Biophys Acta.* 2015;1856(1):73-85.

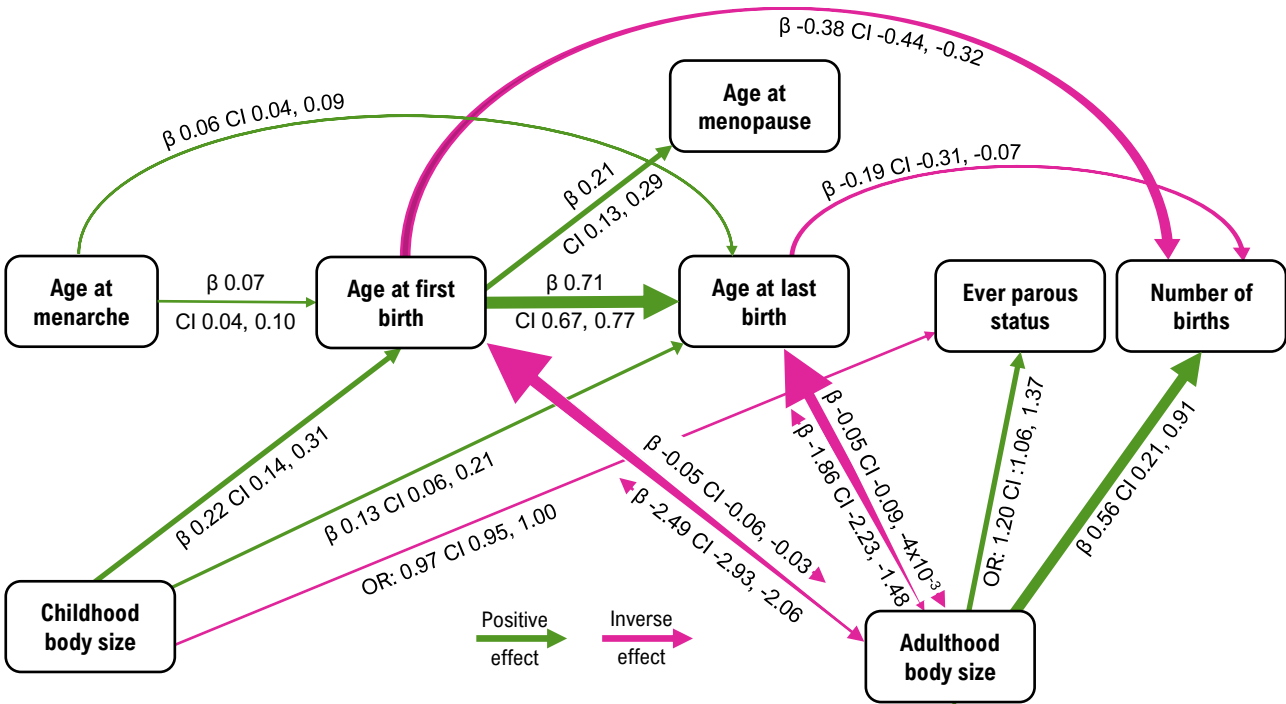
- 760 43. Li C, Fan Z, Lin X, Cao M, Song F, Song F. Parity and risk of developing breast
761 cancer according to tumor subtype: A systematic review and meta-analysis. *Cancer*
762 *Epidemiol.* 2021;75:102050.
- 763 44. Anderson KN, Schwab RB, Martinez ME. Reproductive risk factors and breast cancer
764 subtypes: a review of the literature. *Breast Cancer Res Treat.* 2014;144(1):1-10.
- 765 45. Kim NH, Bang HW, Eom YH, Choi SH. The different prognostic impact of age
766 according to individual molecular subtypes in breast cancer. *Ann Surg Treat Res.*
767 2022;103(3):129-44.
- 768 46. Berrington A. Childlessness in the UK. In: Kreyenfeld M, Konietzka D, editors.
769 *Childlessness in Europe: Contexts, Causes, and Consequences.* Cham: Springer
770 International Publishing; 2017. p. 57-76.
- 771 47. Births by Parents' Characteristics, England and Wales. Office for National Statistics:
772 www.ons.gov.uk; 2019.
- 773 48. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an
774 open access resource for identifying the causes of a wide range of complex diseases of
775 middle and old age. *PLoS Med.* 2015;12(3):e1001779.
- 776 49. Mitchell R, Elsworth BL, Mitchell R, Raistrick CA, Paternoster L, Hemani G, et al.
777 MRC IEU UK Biobank GWAS pipeline version 2. . 2019.
- 778 50. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-
779 Base platform supports systematic causal inference across the human phenome. *Elife.*
780 2018;7.
- 781 51. Richardson TG, Sanderson E, Elsworth B, Tilling K, Davey Smith G. Use of genetic
782 variation to separate the effects of early and later life adiposity on disease risk: mendelian
783 randomisation study. *BMJ.* 2020;369:m1203.
- 784 52. Zhang H, Ahearn TU, Lecarpentier J, Barnes D, Beesley J, Qi G, et al. Genome-wide
785 association study identifies 32 novel breast cancer susceptibility loci from overall and
786 subtype-specific analyses. *Nat Genet.* 2020;52(6):572-81.
- 787 53. Michailidou K, Lindstrom S, Dennis J, Beesley J, Hui S, Kar S, et al. Association
788 analysis identifies 65 new breast cancer risk loci. *Nature.* 2017;551(7678):92-4.
- 789 54. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working
790 Group of the Psychiatric Genomics C, et al. LD Score regression distinguishes confounding
791 from polygenicity in genome-wide association studies. *Nat Genet.* 2015;47(3):291-5.
- 792 55. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of
793 genetic correlations across human diseases and traits. *Nat Genet.* 2015;47(11):1236-41.
- 794 56. Skrivankova VW, Richmond RC, Woolf BAR, Yarmolinsky J, Davies NM, Swanson
795 SA, et al. Strengthening the Reporting of Observational Studies in Epidemiology Using
796 Mendelian Randomization: The STROBE-MR Statement. *JAMA.* 2021;326(16):1614-21.
- 797 57. Skrivankova VW, Richmond RC, Woolf BAR, Davies NM, Swanson SA,
798 VanderWeele TJ, et al. Strengthening the reporting of observational studies in epidemiology
799 using mendelian randomisation (STROBE-MR): explanation and elaboration. *BMJ.*
800 2021;375:n2233.
- 801 58. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with
802 multiple genetic variants using summarized data. *Genet Epidemiol.* 2013;37(7):658-65.
- 803 59. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies:
804 a guide, glossary, and checklist for clinicians. *BMJ.* 2018;362:k601.
- 805 60. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data
806 Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol.*
807 2017;46(6):1985-98.
- 808 61. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in
809 Mendelian Randomization with Some Invalid Instruments Using a Weighted Median
810 Estimator. *Genet Epidemiol.* 2016;40(4):304-14.
- 811 62. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid
812 instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol.*
813 2015;44(2):512-25.

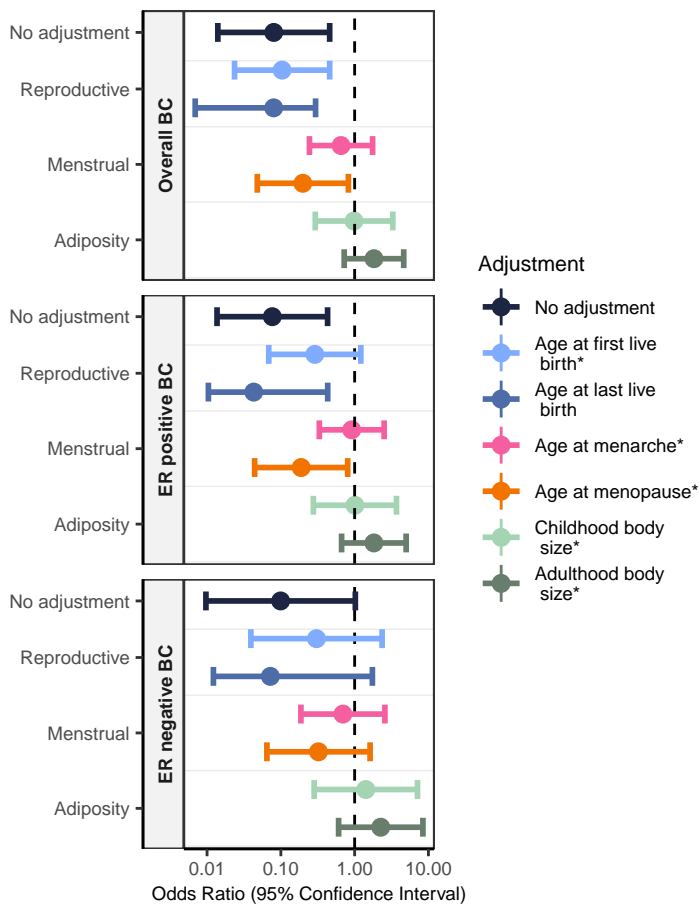
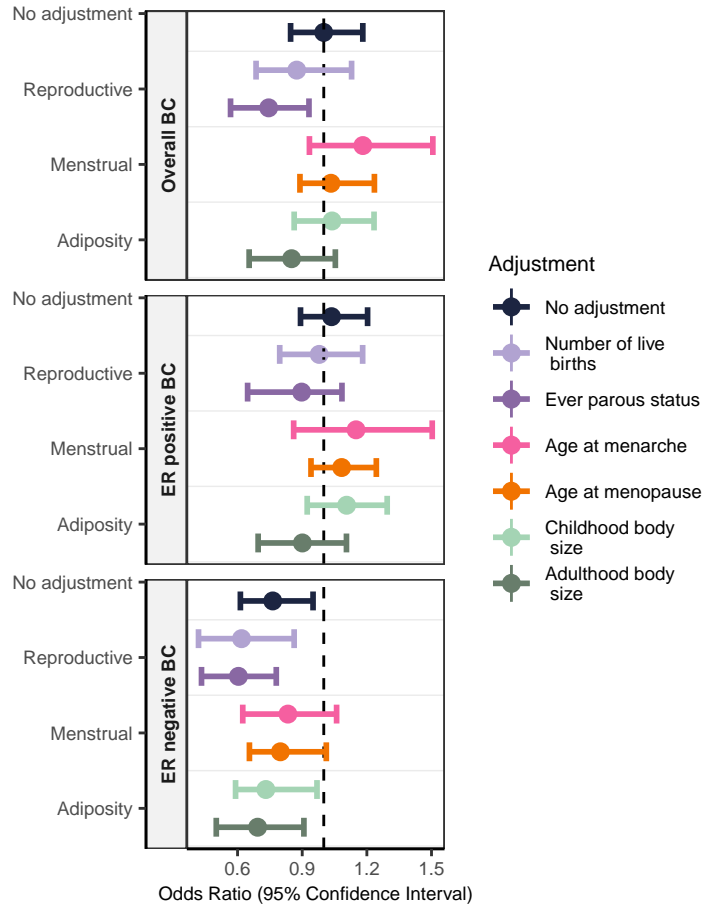
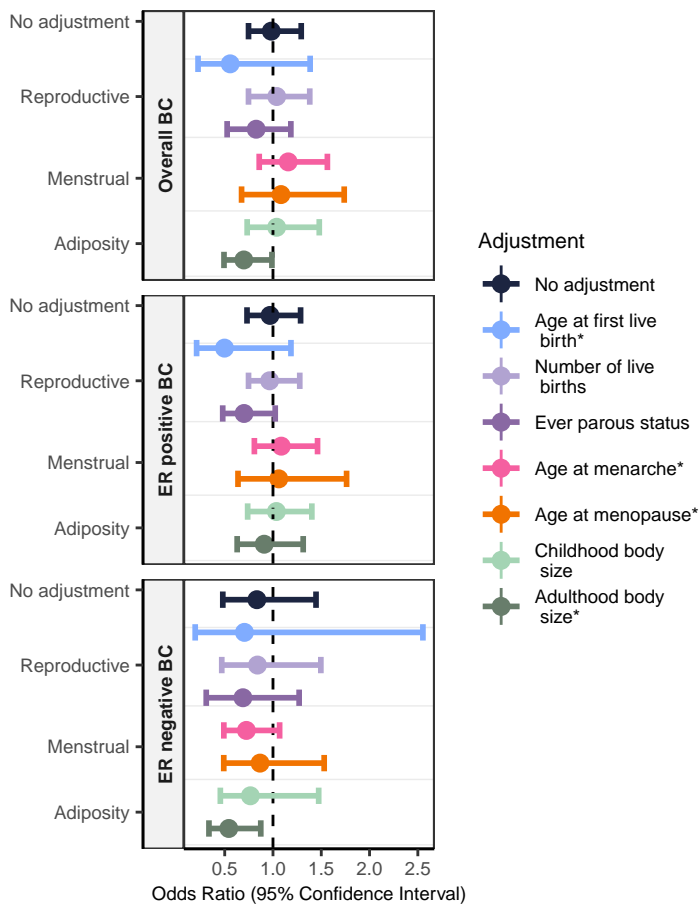
- 814 63. Lawlor DA, Wade K, Borges MC, Palmer TM, Hartwig FP, Hemani G. A Mendelian
815 Randomization dictionary: Useful definitions and descriptions for undertaking, understanding
816 and interpreting Mendelian Randomization studies [Internet]. OSF Preprints. 2019.
817 64. Verbanck M, Chen CY, Neale B, Do R. Publisher Correction: Detection of
818 widespread horizontal pleiotropy in causal relationships inferred from Mendelian
819 randomization between complex traits and diseases. *Nat Genet.* 2018;50(8):1196.
820 65. Carter AR, Sanderson E, Hammerton G, Richmond RC, Davey Smith G, Heron J, et
821 al. Mendelian randomisation for mediation analysis: current methods and challenges for
822 implementation. *Eur J Epidemiol.* 2021;36(5):465-78.
823 66. Sanderson E, Spiller W, Bowden J. Testing and correcting for weak and pleiotropic
824 instruments in two-sample multivariable Mendelian randomization. *Stat Med.*
825 2021;40(25):5434-52.
826 67. Gilbody J, Borges MC, Smith GD, Sanderson E. Multivariable MR can mitigate bias
827 in two-sample MR using covariable-adjusted summary associations. medRxiv.
828 2022:2022.07.19.22277803.
829 68. Sanderson E, Windmeijer F. A weak instrument [Formula: see text]-test in linear IV
830 models with multiple endogenous variables. *J Econom.* 2016;190(2):212-21.
831 69. Rees JMB, Wood AM, Burgess S. Extending the MR-Egger method for multivariable
832 Mendelian randomization to correct for both measured and unmeasured pleiotropy. *Stat*
833 *Med.* 2017;36(29):4705-18.
834 70. Yavorska O, Staley J, Burgess S. MendelianRandomization: Mendelian
835 Randomization Package. <https://cran.r-project.org/web/packages/MendelianRandomization/>.
836 2023.
837 71. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast
838 cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene.
839 *Science.* 1987;235(4785):177-82.
840 72. Yaghjian L, Darville LNF, Cline J, Martinez YC, Rich S, Austin-Datta RJ, et al.
841 Associations of established breast cancer risk factors with urinary estrogens in
842 postmenopausal women. *Cancer Causes Control.* 2022;33(2):279-91.
843 73. Grant AJ, Burgess S. Pleiotropy robust methods for multivariable Mendelian
844 randomization. *Statistics in medicine.* 2021;40(26):5813-30.
845 74. Perry JR, Day F, Elks CE, Sulem P, Thompson DJ, Ferreira T, et al. Parent-of-origin-
846 specific allelic associations among 106 genomic loci for age at menarche. *Nature.*
847 2014;514(7520):92-7.
848 75. Day FR, Ruth KS, Thompson DJ, Lunetta KL, Pervjakova N, Chasman DI, et al.
849 Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast
850 cancer susceptibility and BRCA1-mediated DNA repair. *Nat Genet.* 2015;47(11):1294-303.
851 76. Barban N, Jansen R, de Vlaming R, Vaez A, Mandemakers JJ, Tropf FC, et al.
852 Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat*
853 *Genet.* 2016;48(12):1462-72.

854







A) Effect of ever parous status**B) Effect of age at first live birth****C) Effect of age at last live birth****D) Effect of number of live births**