

1 **Age at Menarche and Coronary Artery Disease Risk: Divergent Associations with Different**
2 **Sources of Variation**

3

4 Ambreen Sonawalla, MD,^{a,b} Daniel I. Chasman, PhD,^{b,c} Yee-Ming Chan, MD, PhD^{a,b}.

5 ^aDivision of Endocrinology, Department of Pediatrics, Boston Children's Hospital, ^bDepartments
6 of Pediatrics (A.S., Y.M.C.) and Medicine (D.I.C.), Harvard Medical School, ^cDivision of
7 Preventive Medicine, Department of Medicine, Brigham and Women's Hospital

8

9 **Short title:** Age at Menarche and Coronary Artery Disease

10

11 **Address for correspondence:** Ambreen Sonawalla, MD;
12 ambreen.sonawalla@childrens.harvard.edu; 300 Longwood Avenue, Boston, MA 02115. Phone:
13 617-355-0796, fax: 617-730-0194.

14

15 Total word count: 7097

16 **Abstract**

17

18 **Background:** Both earlier and later age at menarche (AAM) are associated with increased risk
19 of coronary artery disease (CAD) in women. This study sought to determine if the relationship of
20 AAM with CAD and CAD risk factors differs for underlying sources of variation in AAM –
21 specifically, variation attributable to common genetic variants as represented by a polygenic
22 score (PGS) vs. variation independent of the PGS.

23

24 **Methods:** Primary analyses were conducted on data from 201,037 women in the UK Biobank
25 and validation studies on data from 23,268 women in the Women’s Genome Health Study
26 (WGHS). For each individual, a PGS for AAM was calculated, then two variables were
27 estimated from linear regression models: the PGS-associated change in AAM and the PGS-
28 independent change in AAM. Logistic regression and linear splines were then used to study the
29 relationships of these variables with CAD and CAD risk factors: hemoglobin A1c, triglycerides,
30 high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C),
31 systolic and diastolic blood pressure, body mass index, and waist-hip ratio.

32

33 **Results:** PGS-associated change in AAM demonstrated a linear relationship with CAD and
34 linear or roughly linear relationships with CAD risk factors. In contrast, PGS-independent
35 change in AAM demonstrated a U-shaped relationship with CAD and with hemoglobin A1c,
36 triglycerides, HDL-C, and waist-hip ratio. Validation studies using WGHS data produced similar
37 results.

38

39 **Conclusions:** These results suggest that later AAM itself does not cause increased risk of CAD;
40 rather, upstream sources of variation other than common genetic variants can cause both later
41 AAM and increased risk of CAD. Dysglycemia, dyslipidemia, and central adiposity are
42 candidate mediators of the association of later AAM with increased risk of CAD.

43

44 Key words: pubertal timing, polygenic score, women, UK Biobank.

- 45 Non-standard Abbreviations and Acronyms:
- 46 AAM: age at menarche
- 47 BMI: body-mass index
- 48 CAD: coronary artery disease
- 49 DBP: diastolic blood pressure
- 50 HbA1c: hemoglobin A1c
- 51 HDL-C: high-density lipoprotein cholesterol
- 52 LDL-C: low-density lipoprotein cholesterol
- 53 PGS: polygenic score
- 54 SBP: systolic blood pressure
- 55 WGHS: Women’s Genome Health Study

56 INTRODUCTION

57 Many pathologies manifesting in adulthood have antecedents in childhood. There is growing
58 evidence that coronary artery disease (CAD) in women, a leading cause of morbidity and
59 mortality in the world, is associated with both earlier and later age at menarche (AAM), a
60 hallmark of pubertal timing (1–4). *Earlier* puberty is associated with increased risk of CAD in
61 both men and women; however, the association between *later* puberty and increased risk of CAD
62 appears to be unique to women; in men, later puberty is associated with a decreased risk of CAD
63 (3). A deeper exploration of these childhood antecedents would allow a better understanding of
64 the pathogenesis of CAD in adulthood, specifically identify factors that uniquely affect women
65 and facilitate the development of targeted interventions, potentially as early as childhood.

66

67 Multiple studies have associated earlier AAM with a higher risk of developing components of
68 the metabolic syndrome, namely obesity, type 2 diabetes mellitus, hypertension, and
69 dyslipidemia (3,5–9). Studies have further suggested that the association between earlier AAM
70 and risk of CAD is mediated by adiposity (10,11). In contrast, the associations of later AAM
71 reported to date do not fit neatly into the paradigm of metabolic syndrome. Later AAM is
72 associated with lower rather than higher body-mass index (BMI) (3,12,13), and studies on other
73 components of the metabolic syndrome have produced conflicting results, with some studies
74 showing association of later AAM with higher risk of hypertension (1,14), others showing a
75 lower risk of hypertension (3) or type 2 diabetes (15), and yet others showing no association with
76 hypertension, type 2 diabetes, or hypercholesterolemia (3,16). Hence, while earlier AAM has
77 been associated with several CAD risk factors, there may be distinct mechanisms underlying the
78 association of later AAM with increased risk of CAD.

79

80 Variation in AAM can stem from several upstream sources, including genetics (both common
81 genetic variants and rare genetic variants), acquired factors such as chronic illness, chronic
82 stress, underweight, and undernutrition, and environmental factors such as family composition
83 (17). It is possible that some of these sources of variation may influence risk of CAD only
84 through their influence on AAM, while others may directly influence CAD risk and features of
85 the metabolic syndrome (Figure 1). Thus, dissecting variation in AAM based on underlying
86 sources of variation could provide a clearer understanding of the relationship of AAM with
87 CAD.

88

89 Genetics is a major source of variation in AAM, with half to three-quarters of variation
90 attributable to genetics (18,19). The largest published genome-wide association study (GWAS)
91 on AAM identified 389 independent single-nucleotide polymorphisms (SNPs) associated with
92 AAM at genome-wide significance. The results from this GWAS allow the calculation of a
93 polygenic score (PGS) for a given individual to reflect the cumulative contribution of common
94 genetic variants to AAM.

95

96 Previous studies have used PGSs to dissect the influence of genetics vs. environmental factors
97 (or other factors not captured by the PGS) that contribute to traits such as BMI and LDL-C
98 (20,21). These studies have found that associations of these traits with health outcomes differ
99 between genetically and environmentally influenced traits. For example, obesity driven by
100 environmental factors was associated with more harmful cardiovascular outcomes than

101 genetically predicted obesity (20), suggesting that dissecting effects based on underlying source
102 of variation can allow deeper insights into pathogenic mechanisms.

103

104 The aims of this study were two-fold: first, to determine if the association between later AAM
105 and increased risk of CAD depends on the underlying source of variation in AAM – specifically,
106 common genetic variation vs. other sources of variation; second, to study the relationships of
107 these different sources of variation in AAM with CAD risk factors.

108

109 **METHODS**

110 To study how different sources of variation in the timing of menarche affect risk of CAD,
111 variation in AAM was subdivided into variation attributable to the effects of common genetic
112 influences, as estimated by a PGS for AAM, and variation due to other sources, i.e., independent
113 of the PGS.

114

115 *Study cohorts*

116 This study used data from two cohorts: the UK Biobank for primary analyses and the Women's
117 Genome Health Study (WGHS) for validation analyses. The UK Biobank is a population-based
118 cohort of over 500,000 men and women in the UK 40 years and older at the time of recruitment,
119 with extensive health-related phenotypic and laboratory data as well as individual-level genetic
120 data (22). This study analyzed data from 201,037 unrelated women in the UK Biobank of non-
121 Finnish European ancestry (as determined through principal component analysis) (23) who had

122 genetic and self-reported AAM data. During data collection in the UK Biobank, any value of
123 AAM <5 years or >25 years was rejected, and any AAM entered as <6 years or >20 years
124 required confirmation from participants (24). Women with missing AAM data were excluded
125 from analysis. The UK Biobank obtained the multiple ethical and regulatory approvals required
126 for recruitment and research procedures, and participants provided written consent (22). The
127 WGHS is a cohort of initially healthy American women aged 45 years and older at enrollment
128 with genetic and phenotypic data, followed over 26-28 years for cardiovascular and other
129 outcomes (25). This analysis studied data from 23,268 women in the WGHS who had genetic
130 and self-reported AAM data. Self-reported AAM ≤ 9 years or ≥ 17 years were entered as 9 years
131 or 17 years, respectively. The WGHS was approved by the institutional review board of Brigham
132 and Women's Hospital, and participants consented to ongoing analyses (25).

133

134 *Polygenic score calculation*

135 A PGS was calculated for each woman in the above cohorts in two steps. The first step used the
136 PRS-CS algorithm, which allows the inclusion of all available SNPs from the GWAS on AAM
137 (26), not just those that meet a given p -value threshold, and weights the SNP effect sizes based
138 on their significance and adjusts for linkage disequilibrium (27). The use of this algorithm has
139 the potential to explain more variability in AAM than algorithms that use only the SNPs that
140 meet a given significance threshold. The second step calculated the PGS using PRSice-2
141 (without clumping or thresholding) to sum the weighted effect sizes for all SNPs in each
142 individual with the ability to incorporate the probabilistic genotype dosages generated by
143 imputation (28).

144

145 *Subdividing variation in age at menarche*

146 Using each full cohort, regression of self-reported AAM was performed against the PGS for
147 AAM, with the first 10 genetic principal components, assessment center and technical variables
148 such as array number as covariates. The following two variables were then calculated for each
149 individual:

- 150 1. PGS-associated change in AAM (PGS-associated Δ AAM): This represents the change in
151 AAM attributable to the contributions of common genetic variants as estimated by the
152 PGS. This was calculated as the difference between the individual's AAM predicted by
153 the regression and the AAM corresponding to the mean PGS for the cohort
154 (Supplemental Figure 1).
- 155 2. PGS-independent change in AAM (PGS-independent Δ AAM): This represents the
156 change in AAM attributable to sources of variation not represented by the PGS. This was
157 the residual for the individual of the regression of AAM against the PGS (Supplemental
158 Figure 1).

159

160 *Outcomes and analytical methods*

161 The study's primary outcome variable was CAD risk. In the UK Biobank, prevalent CAD at
162 baseline was determined as previously described using a combination of self-report, ICD-9/10
163 codes, and procedure codes (23). The WGHS recruited women with no history of CAD at
164 baseline and identified validated incident CAD during 26-28 years of follow-up as described
165 previously (25). Secondary outcome variables were CAD risk factors at baseline for both

166 cohorts: hemoglobin A1c (HbA1c), triglycerides, high-density lipoprotein cholesterol (HDL-C),
167 low-density lipoprotein cholesterol (LDL-C), systolic blood pressure (SBP), diastolic blood
168 pressure (DBP), and BMI; data on waist-hip ratio was available at baseline in the UK Biobank
169 and 6 years after recruitment in the WGHS.

170

171 The relationships of PGS-associated Δ AAM and PGS-independent Δ AAM with CAD prevalence
172 were studied using logistic regression, with incident CAD using Cox proportional hazards
173 models, and with each continuous variable (HbA1c, triglycerides, HDL-C, LDL-C, SBP, DBP,
174 BMI, waist-hip ratio) using linear splines, with a knot at Δ AAM of 0 years. For analyses with
175 LDL-C, results were corrected for self-reported use of cholesterol-lowering medications – these
176 were specific LDL-lowering medications in the UK Biobank, (statins, ezetimibe, and bile-acid
177 sequestrants) and collective cholesterol-lowering medications in the WGHS; additional analyses
178 included only women not taking these medications.

179 Covariates of age and age² were used for all analyses. Analyses were conducted using R v.4.3.1.
180 A significance threshold of 0.05 was used.

181

182 To determine if results in the UK Biobank were biased by overfitting as a result of the UK
183 Biobank having contributed to the GWAS for AAM, we also repeated analyses of CAD using a
184 PGS calculated using an earlier GWAS that did not include the UK Biobank (29). Furthermore,
185 to determine if outlier values of AAM were disproportionately affecting results, analyses were
186 repeated after excluding women with extreme values of AAM such that up to 0.1% of women
187 were excluded at each extreme.

188

189 **RESULTS**

190 To study the relationship between AAM and CAD risk, our study divided variation in AAM into
191 1) variation attributable to the effects of common genetic variants as estimated by a PGS for
192 AAM, which we term PGS-associated Δ AAM, and 2) variation due to factors not captured by the
193 PGS, which we term PGS-independent Δ AAM.

194

195 To examine the differential effects of these two sources of variation on CAD risk, we analyzed
196 data from 201,037 unrelated, non-Finnish European women in the UK Biobank. We first
197 calculated a PGS for AAM for each woman, then regressed self-reported AAM against the AAM
198 PGS to calculate PGS-associated Δ AAM and PGS-independent Δ AAM (Supplemental Figure 1).
199 For women in the UK Biobank, the regression of self-reported AAM against the AAM PGS
200 demonstrated that the PGS accounted for 15.8% of the variation in AAM. PGS-associated
201 Δ AAM had a mean of 0.016 years, with a standard deviation of 0.639 years, and the residual,
202 PGS-independent Δ AAM, had a mean of 0 years with a standard deviation of 1.47 years
203 (Supplemental Figure 2). We then studied the associations of PGS-associated Δ AAM and PGS-
204 independent Δ AAM with risk of CAD and with CAD risk factors.

205

206 *Risk of coronary artery disease*

207 Risk of CAD demonstrated a linear relationship with PGS-associated Δ AAM but a non-linear,
208 U-shaped relationship with PGS-independent Δ AAM (Figure 2). In the linear relationship of

209 PGS-associated Δ AAM with risk of CAD; each 1-year increase in PGS-associated Δ AAM was
210 associated with an odds ratio (OR) for CAD of 0.91. In other words, for every 1 year that AAM
211 was later due to the effects of common genetic variants, as estimated by the PGS, the odds of
212 CAD were lower by 9% (Table 1). Linear spline analyses showed no difference in slopes when
213 PGS-associated Δ AAM was negative vs. positive, i.e., driving earlier vs. later AAM (Table 1).

214
215 In contrast, the association between PGS-independent Δ AAM and risk of CAD was U-shaped,
216 with both increasingly negative and increasingly positive values (driving earlier and later AAM,
217 respectively) associated with increased CAD risk (Figure 2). For negative values of PGS-
218 independent Δ AAM, each 1-year increase (causing AAM to be less early) was associated with an
219 OR for CAD of 0.91 (Table 1). In other words, for every 1 year that AAM was less early due to
220 factors independent of the PGS, the odds of CAD were lower by 9%. In contrast, for positive
221 values of PGS-independent Δ AAM, each 1-year increase (causing AAM to be even later) was
222 associated with an OR for CAD of 1.11, i.e., the odds of CAD were higher by 11% (Table 1).

223
224 Because the 2018 AAM GWAS meta-analysis included data from the UK Biobank, which could
225 introduce bias in the above analyses, we conducted sensitivity analyses using the results from an
226 AAM GWAS meta-analysis published in 2014 that did not include the UK Biobank. These
227 analyses showed similar results for CAD (Supplemental Figure 3, Supplemental Table 1). We
228 also conducted sensitivity analyses excluding women with extreme values of AAM and obtained
229 similar results.

230

231 *Hemoglobin A1c*

232 We then examined the associations of PGS-associated Δ AAM and PGS-independent Δ AAM
233 with CAD risk factors. Just as the relationship between PGS-associated Δ AAM and CAD was
234 linear, we observed linear (or roughly linear) relationships between PGS-independent Δ AAM
235 and most CAD risk factors. In contrast, for PGS-independent Δ AAM we observed mostly non-
236 linear relationships with CAD risk factors.

237

238 For hemoglobin A1c, which correlates with average blood glucose over the preceding 3 months,
239 PGS-associated Δ AAM demonstrated a roughly negative linear relationship, with higher PGS-
240 associated Δ AAM associated with lower HbA1c (Figure 3). The slope of the association with
241 HbA1c for negative values of PGS-associated Δ AAM was slightly steeper than for positive
242 values ((-0.034 and -0.022, respectively; p for difference between slopes: 7×10^{-5} ; Table 1). In
243 contrast, PGS-independent Δ AAM demonstrated a U-shaped relationship with HbA1c: both
244 increasingly negative and increasingly positive values of PGS-independent Δ AAM were
245 associated with an increase in HbA1c (for negative values: slope = -0.022, for positive values:
246 slope = 0.007; p for difference between slopes $< 2 \times 10^{-16}$; Figure 3, Table 1). This was similar
247 to the U-shaped relationship of PGS-independent Δ AAM with CAD risk.

248

249 Comparing PGS-associated Δ AAM and PGS-independent Δ AAM, for negative values (driving
250 earlier AAM) the slopes for the associations with HbA1c were comparable. However, for
251 positive values (driving later AAM), different slopes were seen for PGS-associated vs. PGS-

252 independent Δ AAM; as noted above, for every 1 year that AAM was delayed, HbA1c *decreased*
253 by 0.01% for PGS-associated Δ AAM but *increased* by 0.007% for PGS-independent Δ AAM.

254

255 *Lipids*

256 PGS-associated Δ AAM demonstrated linear relationships with triglycerides and HDL-C (Figure
257 3), with higher values of PGS-associated Δ AAM associated with a decrease in triglycerides and
258 an increase in HDL-C (Figure 3) and no significant difference in slopes between negative and
259 positive values of PGS-associated Δ AAM (p for difference in slopes for triglycerides = 0.83, for
260 HDL-C = 0.38; Table 1). In contrast, PGS-independent Δ AAM demonstrated non-linear
261 relationships with HDL-C and triglycerides – U-shaped for triglycerides and inverted-U-shaped
262 for HDL-C (Figure 3), with both increasingly negative and increasingly positive values of PGS-
263 independent Δ AAM associated with increases in triglycerides and decreases in HDL-C (Figure 3,
264 Table 1).

265

266 For LDL-C, PGS-associated Δ AAM showed no association for negative values of PGS-
267 associated Δ AAM and a negative linear relationship for positive values of PGS-associated
268 Δ AAM (Table 1). For PGS-independent Δ AAM, there was an association for negative values –
269 increasing PGS-independent Δ AAM was associated with lower LDL-C – and no significant
270 association with LDL-C for positive values of PGS-independent Δ AAM (Figure 3, Table 1).
271 Similar results were seen when analyses excluded those taking LDL-lowering medications
272 (Table 1).

273

274 *Blood pressure*

275 For PGS-associated Δ AAM, increasing PGS-associated Δ AAM was associated with decreases in
276 both SBP and DBP (Figure 3), with no significant difference between the slopes for negative vs.
277 positive values of PGS-associated Δ AAM (Table 1). For PGS-independent Δ AAM, increasing
278 PGS-independent Δ AAM was also associated with decreased SBP and DBP (Figure 3), but the
279 slopes of the associations were steeper for negative values of PGS-independent Δ AAM
280 compared to the slopes for positive values of PGS-independent Δ AAM (Table 1).

281

282 *Adiposity*

283 We studied two estimates of adiposity: BMI and waist-hip ratio. For BMI, we found a roughly
284 linear negative relationship between PGS-associated Δ AAM and BMI, with higher PGS-
285 associated Δ AAM associated with lower BMI (Figure 3, Table 1). The slope of the association
286 with BMI was slightly steeper for negative values of PGS-associated Δ AAM than for positive
287 values (Table 1). This pattern was similar to the relationship seen between PGS-associated
288 Δ AAM and HbA1c.

289

290 For PGS-independent Δ AAM, a reverse-J-shaped relationship was seen with BMI (Figure 3).
291 While higher values of PGS-independent Δ AAM were consistently associated with lower BMI,
292 the slope of the association was ten-fold steeper for negative than for positive values of PGS-
293 independent Δ AAM (-0.80 vs. -0.079 respectively; p for difference between slopes 2×10^{-16} ;
294 Table 1). This resembled the patterns seen with SBP and DBP.

295

296 For waist-hip ratio, PGS-associated Δ AAM showed a negative linear association, with no
297 significant difference in slopes for the association when PGS-associated Δ AAM was negative vs.
298 positive (Figure 3, Table 1). In contrast, the relationship of PGS-independent Δ AAM with waist-
299 hip ratio was U-shaped, with both increasingly negative and increasingly positive values of PGS-
300 associated Δ AAM associated with an increase in waist-hip ratio (Figure 3, Table 1), similar to
301 the associations seen with CAD, HbA1c, triglycerides, and HDL-C.

302

303 *Validation in the Women's Genome Health Study*

304 Validation studies using data from the WGHS produced results similar to the above results for
305 the UK Biobank. In the WGHS, PGS-associated Δ AAM showed linear relationships with
306 HbA1c, triglycerides, HDL-C, SBP, DBP, BMI and waist-hip ratio and no clear relationship with
307 LDL-C (Figure 4). Just as in the UK Biobank, PGS-independent Δ AAM showed U-shaped
308 relationships with triglycerides and HbA1c, an inverted-U-shaped relationship with HDL-C, and
309 a reverse-J-shaped relationship with BMI in the WGHS (Figure 4).

310

311 There were two differences between results from the WGHS and the UK Biobank: 1) for SBP
312 and DBP, PGS-independent Δ AAM showed roughly U-shaped relationships in the WGHS
313 (Figure 4) rather than the reverse-J-shaped relationships seen in the UK Biobank (Figure 3), and
314 2) for waist-hip ratio, PGS-independent Δ AAM showed a negative linear relationship for
315 negative values of PGS-independent Δ AAM but no significant relationship for positive values in
316 the WGHS (Figure 4) compared to the clear U-shaped relationship seen in the UK Biobank
317 (Figure 2).

318

319 For risk of CAD itself, the relationship of PGS-associated Δ AAM with CAD in the WGHS was
320 similar to that in the UK Biobank (Figure 5, Supplemental Table 1), with a negative linear
321 relationship in both cohorts. For the relationship of PGS-independent Δ AAM with CAD,
322 analyses in the WGHS demonstrated a reverse-J shaped relationship (Figure 5, Supplemental
323 Table 1), slightly different from the U-shaped relationship seen in the UK Biobank.

324

325 **DISCUSSION**

326 In this study, we dissected variation in AAM into variation attributable to common genetic
327 variants, as estimated by a PGS, and variation independent of the PGS, and we found different
328 relationships with these two sources of variation in AAM with risk of CAD and with CAD risk
329 factors, particularly when causing AAM to occur later.

330

331 In general, later AAM showed favorable associations when attributable to PGS-associated
332 variation and harmful or neutral associations when attributable to PGS-independent variation. If
333 all sources of variation were affecting CAD and CAD risk factors wholly through AAM itself,
334 the associations would be similar regardless of the source of variation studied. Our finding that
335 these associations varied based on the underlying source of variation driving later AAM
336 therefore indicates that it is not later AAM itself that causes increased risk of CAD. Rather, there
337 appear to be PGS-independent factors that cause both later AAM and increased risk of CAD and
338 unfavorable cardiometabolic risk profiles. Such factors could include environmental or acquired
339 influences (such as chronic illnesses, chronic stress, undernutrition) as well as genetic influences

340 not captured by the PGS, and future studies will identify these factors and determine how they
341 contribute to CAD risk.

342

343 Earlier AAM, whether attributable to PGS-associated or PGS-independent sources of variation,
344 was consistently associated with greater cardiometabolic risk. Thus, it seems that earlier AAM
345 itself is intrinsically associated with risk for CAD and worsening CAD risk factors; indeed, some
346 Mendelian randomization studies suggest that earlier AAM is causative of these negative
347 outcomes (10,11,30). However, other Mendelian randomization studies have suggested that the
348 association of earlier AAM and increased risk of CAD is due to pleiotropic effects of genetic
349 factors influencing both AAM and BMI, such that AAM itself may have only a small influence
350 (31).

351

352 The U-shaped relationship of PGS-independent Δ AAM with CAD was mirrored by the
353 relationships with HbA1c, triglycerides, HDL-C, and waist-hip ratio. This raises the possibility
354 that dysglycemia, dyslipidemia, and central adiposity contribute to the relationship between later
355 AAM and increased risk of CAD; further studies will be required to formally evaluate these
356 CAD risk factors as potential mediators of this relationship. As noted above, previous studies
357 evaluating associations of later AAM with these CAD risk factors have had conflicting results.
358 While differences between study cohorts may have accounted for some of the differing results,
359 our findings raise the additional possibility that the relationships between later AAM and these
360 outcomes may have been obscured by opposing effects of PGS-associated and PGS-independent
361 variation in AAM. Of note, the associations with waist-hip ratio differed from those with BMI,

362 suggesting that central obesity, reflected by waist-hip ratio, is more relevant for CAD, as has
363 been suggested by prior studies (32,33).

364

365 Our analyses in the Women's Genome Health Study (WGHS) largely validated our results from
366 the UK Biobank. PGS-associated change in AAM demonstrated linear relationships with CAD
367 and CAD risk factors, and PGS-independent change in AAM demonstrated mostly non-linear
368 relationships with these outcomes, supporting the conclusions described earlier.

369

370 However, the two cohorts also demonstrated some differences, most notably in the association of
371 PGS-independent Δ AAM with the risk of CAD and waist-hip ratio. In the UK Biobank, PGS-
372 independent variation driving AAM *later* was associated with an increased risk of CAD, but
373 there was no association in the WGHS. For waist-hip ratio, PGS-independent variation
374 demonstrated a clear U-shaped association in the UK Biobank but a reverse J-shaped relationship
375 in the WGHS. There are several potential reasons for this difference. First, it is possible that the
376 smaller sample size and lower power in the WGHS affected the ability to find significant
377 associations with PGS-independent variation in AAM and risk of CAD and waist-hip ratio.
378 Second, environmental influences affecting AAM (which would contribute to PGS-independent
379 variation in AAM) could differ between the two cohorts. The WGHS recruited women in the
380 United States of America born in 1950 or earlier, while the UK Biobank recruited women in the
381 UK who were born between 1932 and 1969. The different impact of global events such as World
382 War II on the two countries could contribute to differences in the PGS-independent factors
383 (which includes environmental factors) and, in turn, to different associations with risk of CAD.

384 Third, the WGHS excluded women with a history of CAD at the time of enrollment whereas the
385 UK Biobank did not, and this may also have led to differences in PGS-independent factors
386 between the cohorts. Fourth, the participants in the WGHS were health professionals while the
387 UK Biobank drew from the general UK population, and this may have led to further differences
388 in PGS-independent factors, such as higher socioeconomic status, greater knowledge of CAD
389 and its risk factors, healthier diets and lifestyles, and use of preventative interventions, as well as
390 potentially less variation in these factors.. While the analyses also differed in the CAD measure
391 used from each cohort – prevalence of CAD was analyzed in the UK Biobank compared to
392 incidence of CAD in the WGHS – it is unlikely to account for the difference in results, as an
393 analysis of incident CAD in the Million Women Study in the UK also showed a U-shaped
394 relationship between AAM and CAD (1). Despite these differences in the results between the
395 two cohorts, results from both cohorts consistently demonstrated differences between the
396 associations of PGS-independent variation driving AAM *earlier* (negative values) vs. driving
397 AAM *later* (positive values) with these outcomes.

398

399 In 2021, Liang et al. used a PGS to represent genetically predicted AAM and examined
400 associations with all-cause mortality, also using data from the UK Biobank. Interestingly, they
401 found a U-shaped association with mortality, with both earlier and later genetically predicted
402 AAM associated with higher risk; this finding contrasts with the linear relationship we found
403 between PGS-associated Δ AAM and CAD. This difference suggests that later genetically
404 predicted AAM increases the risk of causes of mortality other than CAD, and future studies of
405 genetically predicted AAM are needed to identify these causes.

406

407 Prior studies that have also dissected influences on human traits into genetic vs. environmental
408 influences have found differences in the magnitude of effect of genetics vs. environmental
409 influences on health outcomes (20,21). Interestingly, our results show not only different
410 magnitudes of influence of PGS-associated and PGS-independent variation in AAM on risk of
411 CAD, but also opposite directionality. This further underscores the value of separating the effects
412 of genetics vs. other influences while studying human traits as they can have starkly different
413 effects.

414

415 Sensitivity analyses using a GWAS that did not include the primary cohort, the UK Biobank,
416 produced similar results and suggested that these results were not biased by overfitting.
417 Additionally, sensitivity analyses excluding women with extreme values of AAM also produced
418 similar results, suggesting that results were not heavily influenced by outliers.

419

420 One limitation of this study is that while the PGS represents genetic influences on AAM, it does
421 not represent all genetic factors that influence AAM. Our analyses demonstrated that the PGS
422 explains 15.8% of the variation in the observed AAM, but prior studies suggest that 49-73% of
423 variation in AAM is inherited (18,19); hence, a large amount of the variation in AAM due to
424 genetic factors remains unexplained. Another limitation of this study is that the PGS represents
425 just one method of capturing the effects of common genetic variants that affect AAM. There may
426 be several pathways causing later AAM represented within these common genetic variants, and
427 using a single PGS to represent all those effects may obscure relationships with each individual
428 pathway. Future studies may identify these different pathways by methods such as clustering

429 analyses (34,35), which would then allow an estimation of multiple polygenic scores, each
430 representing a different pathway, to study their associations with risk of CAD and CAD risk
431 factors.

432

433 Distinguishing between sources of variation in AAM has provided a novel lens through which to
434 study associations of AAM with CAD and has allowed us to uncover differences in the
435 associations of PGS-associated vs. PGS-independent variation in AAM with CAD and CAD risk
436 factors. Because later puberty in women, but not men, has been associated with an increased risk
437 of CAD, studying these differences further may provide unique insights into mechanisms that
438 affect CAD risk specifically in women.

439 ACKNOWLEDGEMENTS

440 The authors thank Evan Schafer and Jia Zhu for their support with statistical analyses. This
441 research has been conducted using data from UK Biobank, a major biomedical database:
442 www.ukbiobank.ac.uk

443 SOURCES OF FUNDING

444 A.S. was supported by National Institutes of Health (NIH) grant T32DK007699. The WGHS and
445 its parent cohort, the Women’s Health Study (WHS), have been supported by the National Heart,
446 Lung, and Blood Institute (HL043851 and HL080467) and the National Cancer Institute
447 (CA047988 and UM1CA182913), with support and funding for genotyping provided by Amgen.
448 Some cardiovascular endpoints in the WHS were funded by HL099355.

449 DISCLOSURES

450 A.S. has no disclosures or relationships with industry. D.I.C has no disclosures or relationships
451 with industry. Y.-M.C. receives royalties from UpToDate on topics related to puberty.

452

453

454 SUPPLEMENTAL MATERIALS

455 Supplemental Figures 1-3

456 Supplemental Table 1

457 **REFERENCES**

- 458 1. Canoy D, Beral V, Balkwill A, Wright FL, Kroll ME, Reeves GK, et al. Age at menarche and risks of
459 coronary heart and other vascular diseases in a large UK cohort. *Circulation* 2015;131(3):237–44.
460
- 461 2. Zhu J, Chan YM. Adult Consequences of Self-Limited Delayed Puberty. *Pediatrics* [Internet]. 2017
462 Jun 1;139(6).
463
- 464 3. Day FR, Elks CE, Murray A, Ong KK, Perry JRB. Puberty timing associated with diabetes,
465 cardiovascular disease and also diverse health outcomes in men and women: the UK Biobank
466 study. *Sci Rep*. 2015 Jun 18;5.
467
- 468 4. Lee JJ, Cook-Wiens G, Johnson BD, Braunstein GD, Berga SL, Stanczyk FZ, et al. Age at Menarche
469 and Risk of Cardiovascular Disease Outcomes: Findings From the National Heart Lung and Blood
470 Institute-Sponsored Women’s Ischemia Syndrome Evaluation. *J Am Heart Assoc*. 2019 Jun
471 18;8(12).
472
- 473 5. Santos MP, Li Y, Bazzano LA, He J, Rexrode KM, Ley SH. Age at menarche, type 2 diabetes and
474 cardiovascular disease complications in US women aged under 65 years: NHANES 1999-2018.
475 *BMJ Nutr Prev Health*. 2023 Dec 1;6(2):293–300.
476
- 477 6. Janghorbani M, Mansourian M, Hosseini E. Systematic review and meta-analysis of age at
478 menarche and risk of type 2 diabetes. *Acta Diabetol*. 2014;51(4):519–28.
479
- 480 7. Stöckl D, Meisinger C, Peters A, Thorand B, Huth C, Heier M, et al. Age at menarche and its
481 association with the metabolic syndrome and its components: results from the KORA F4 study.
482 *PLoS One*. 2011 Oct 18;6(10).
483
- 484 8. Liu D, Qin P, Liu Y, Sun X, Li H, Wu X, et al. Association of age at menarche with hypertension in
485 rural Chinese women. *J Hypertens*. 2021 Mar 1;39(3):476–83.
486
- 487 9. Kheradmand M, Hamzehgardeshi Z, Shahhosseini Z, Mirjalili R, Moosazadeh M. The association
488 between early menarche and higher-risk cardiometabolic profile: a dose-response analysis of the
489 Tabari cohort at enrollment phase. *Front Cardiovasc Med*. 2023;10.
490

- 491 10. Cao M, Cui B. Negative Effects of Age at Menarche on Risk of Cardiometabolic Diseases in
492 Adulthood: A Mendelian Randomization Study. *J Clin Endocrinol Metab.* 2020 Jan 23;105(2).
493
- 494 11. Gill D, Brewer CF, Del Greco M F, Sivakumaran P, Bowden J, Sheehan NA, et al. Age at menarche
495 and adult body mass index: a Mendelian randomization study. *International Journal of Obesity*
496 2018 42:9. 2018 Feb 26;42(9):1574–81.
497
- 498 12. Prentice P, Viner RM. Pubertal timing and adult obesity and cardiometabolic risk in women and
499 men: a systematic review and meta-analysis. *Int J Obes (Lond).* 2013 Aug;37(8):1036–43.
500
- 501 13. Yang L, Li L, Millwood IY, Lewington S, Guo Y, Sherliker P, et al. Adiposity in relation to age at
502 menarche and other reproductive factors among 300 000 Chinese women: findings from China
503 Kadoorie Biobank study. *Int J Epidemiol.* 2017 Jan 1;46(2):502–12.
504
- 505 14. Chen L, Zhang L, Chen Z, Wang X, Zheng C, Kang Y, et al. Age at menarche and risk of
506 hypertension in Chinese adult women: Results from a large representative nationwide
507 population. *J Clin Hypertens (Greenwich).* 2021 Aug 1;23(8):1615–21.
508
- 509 15. Liu D, Zhao Y, Liu Y, Sun X, Li H, Yin Z, et al. Adiposity and insulin resistance as mediators between
510 age at menarche and type 2 diabetes mellitus. *Menopause.* 2020 May 1;27(5):579–85.
511
- 512 16. Elks CE, Ong KK, Scott RA, Van Der Schouw YT, Brand JS, Wark PA, et al. Age at menarche and
513 type 2 diabetes risk: the EPIC-InterAct study. *Diabetes Care.* 2013 Nov;36(11):3526–34.
514
- 515 17. Schlomer GL, Marceau K. Father absence, age at menarche, and genetic confounding: A
516 replication and extension using a polygenic score. *Dev Psychopathol.* 2022 Feb 27;34(1):355–66.
517
- 518 18. Morris DH, Jones ME, Schoemaker MJ, Ashworth A, Swerdlow AJ. Familial concordance for age at
519 menarche: Analyses from the breakthrough generations study. *Paediatr Perinat Epidemiol.* 2011
520 May;25(3):306–11.
521
- 522 19. Towne B, Czerwinski SA, Demerath EW, Blangero J, Roche AF, Siervogel RM. Heritability of age at
523 menarche in girls from the Fels Longitudinal Study. *Am J Phys Anthropol.* 2005 Sep;128(1):210–9.
524

- 525 20. Ojalehto E, Zhan Y, Jylhävä J, Reynolds CA, Dahl Aslan AK, Karlsson IK. Genetically and
526 environmentally predicted obesity in relation to cardiovascular disease: a nationwide cohort
527 study. *EClinicalMedicine*. 2023 Apr 1;58.
528
- 529 21. Hawkes G, Yengo L, Vedantam S, Marouli E, Beaumont RN, Tyrrell J, et al. Identification and
530 analysis of individuals who deviate from their genetically-predicted phenotype. *PLoS Genet*. 2023
531 Sep 21;19(9).
532
- 533 22. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access
534 resource for identifying the causes of a wide range of complex diseases of middle and old age.
535 *PLoS Med*. 2015 Mar 1;12(3).
536
- 537 23. Zhu J, Pujol-Gualdo N, Wittemans LBL, Lindgren CM, Laisk T, Hirschhorn JN, et al. Evidence From
538 Men for Ovary-independent Effects of Genetic Risk Factors for Polycystic Ovary Syndrome. *J Clin*
539 *Endocrinol Metab*. 2022 Mar 24;107(4):e1577–87.
540
- 541 24. UK Biobank age at menarche data collection:
542 <https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2714>.
543
- 544 25. Ridker PM, Chasman DI, Zee RYL, Parker A, Rose L, Cook NR, et al. Rationale, design, and
545 methodology of the Women’s Genome Health Study: a genome-wide association study of more
546 than 25,000 initially healthy american women. *Clin Chem*. 2008;54(2):249–55.
547
- 548 26. Day FR, Thompson DJ, Helgason H, Chasman DI, Finucane H, Sulem P, et al. Genomic analyses
549 identify hundreds of variants associated with age at menarche and support a role for puberty
550 timing in cancer risk. *Nat Genet*. 2017 Jun 1;49(6):834.
551
- 552 27. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and
553 continuous shrinkage priors. *Nat Commun*. 2019 Dec 1;10(1).
554
- 555 28. Choi SW, O’Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience*.
556 2019 Jul 1;8(7).
557
- 558 29. Perry JRB, Day F, Elks CE, Sulem P, Thompson DJ, Ferreira T, et al. Parent-of-origin specific allelic
559 associations among 106 genomic loci for age at menarche. *Nature*. 2014 Oct 10;514(7520):97.
560

- 561 30. Zheng J, Chen K, Huang T, Shao C, Li P, Wang J, et al. Genetically Determined Lifestyle and
562 Cardiometabolic Risk Factors Mediate the Association of Genetically Predicted Age at Menarche
563 With Genetic Predisposition to Myocardial Infarction: A Two-Step, Two-Sample Mendelian
564 Randomization Study. *Front Cardiovasc Med.* 2022 Apr 25;9.
- 565
- 566 31. Bell JA, Carslake D, Wade KH, Richmond RC, Langdon RJ, Vincent EE, et al. Influence of puberty
567 timing on adiposity and cardiometabolic traits: A Mendelian randomisation study. *PLoS Med.*
568 2018 Aug 1;15(8).
- 569
- 570 32. Huxley R, Mendis S, Zheleznyakov E, Reddy S, Chan J. Body mass index, waist circumference and
571 waist:hip ratio as predictors of cardiovascular risk--a review of the literature. *Eur J Clin Nutr.* 2010
572 Jan;64(1):16–22.
- 573
- 574 33. Sahakyan KR, Somers VK, Rodriguez-Escudero JP, Hodge DO, Carter RE, Sochor O, et al. Normal-
575 Weight Central Obesity: Implications for Total and Cardiovascular Mortality. *Ann Intern Med.*
576 2015 Dec 1;163(11):827–35.
- 577
- 578 34. Chasman DI, Giulianini F, Demler O V., Udler MS. Pleiotropy-Based Decomposition of Genetic Risk
579 Scores: Association and Interaction Analysis for Type 2 Diabetes and CAD. *Am J Hum Genet.* 2020
580 May 7;106(5):646–58.
- 581
- 582 35. Udler MS, Kim J, von Grothuss M, Bonàs-Guarch S, Cole JB, Chiou J, et al. Type 2 diabetes genetic
583 loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft
584 clustering analysis. *PLoS Med.* 2018 Sep 1;15(9).
- 585
- 586

587 **FIGURE LEGENDS**

588

589 **Figure 1:** Relationships of different sources of variation in age at menarche (AAM) with
590 coronary artery disease (CAD) and CAD risk factors. Variation in AAM was dissected into
591 variation associated with common genetic variants (as estimated by a polygenic score for AAM)
592 and variation independent of the polygenic score, then associations of these sources of variation
593 with CAD and CAD risk factors were studied. When associated with common genetic variants,
594 later AAM was associated with a lower risk of CAD and favorable changes in CAD risk factors
595 (blue arrows). However, when occurring independently of the polygenic score, later AAM was
596 associated with higher risk of CAD and unfavorable changes in CAD risk factors (red arrows).
597 The “other factors” causing later AAM independent of the PGS have yet to be identified and
598 could potentially include chronic illness, underweight, and rare genetic variants.

599

600 **Figure 2: Association of variation in AAM with odds of CAD in the UK Biobank.** Variation
601 in AAM was dissected into variation associated with common genetic variants as estimated by a
602 PGS for AAM (PGS-associated Δ AAM) and variation independent of the PGS (PGS-
603 independent Δ AAM), then associations of these sources of variation with risk of CAD were
604 studied. PGS-associated Δ AAM showed linear or roughly linear associations with CAD whereas
605 PGS-independent Δ AAM showed a U-shaped association with risk of CAD. The different
606 associations of positive values of PGS-associated Δ AAM with risk of CAD demonstrates that it
607 is not later AAM itself that causes the increased risk of CAD and that the increased risk of CAD
608 observed in later AAM is being driven by factors other than common genetic variants and that.
609 To achieve bin sizes of ≥ 1000 individuals, the first and last bins for each variable represent the

610 group of individuals with $\Delta\text{AAM} \leq$ or \geq the value listed on the x-axis, respectively. Dots
611 represent estimates; bars represent standard errors.

612 AAM: age at menarche; ΔAAM : change in AAM; CAD: coronary artery disease; PGS:
613 polygenic score.

614

615 **Figure 3: Associations of variation in AAM with risk factors for CAD in the UK Biobank.**

616 We observed linear (or roughly linear) relationships between PGS-associated ΔAAM and most
617 CAD risk factors. In contrast, for PGS-independent ΔAAM we observed mostly non-linear
618 relationships with CAD risk factors. The U-shaped relationship of PGS-independent ΔAAM with
619 CAD was mirrored by the relationships with HbA1c, triglycerides, HDL-C, and waist-hip ratio.
620 This raises the possibility that dysglycemia, dyslipidemia and central adiposity may mediate the
621 relationship between later AAM and increased risk of CAD. For LDL-C, $n = 177,542$ after
622 excluding those taking LDL-C lowering medications. Dots represent estimates, bars represent
623 standard errors.

624 AAM: age at menarche; ΔAAM : change in AAM; BMI: body mass index; CAD: coronary artery
625 disease; HDL-C: high-density lipoprotein cholesterol; LDL-C: low-density lipoprotein
626 cholesterol; PGS: polygenic score

627

628 **Figure 4: Associations of variation in AAM with risk factors for CAD in the WGHS.** PGS-

629 associated ΔAAM showed linear or roughly linear associations with most CAD risk factors and
630 PGS-independent ΔAAM showed mostly non-linear associations with most CAD risk factors.
631 Similar to results in the UK Biobank, hemoglobin A1c, triglycerides and HDL-C showed U-
632 shaped associations with PGS-independent ΔAAM which support conclusions from the primary

633 analyses that dysglycemia, dyslipidemia and central adiposity may mediate the relationship
634 between later AAM and increased risk of CAD. For LDL-C, $n = 22,495$ after excluding those
635 taking cholesterol-lowering medications. Dots represent estimates, bars represent standard errors.
636 AAM: age at menarche; Δ AAM: change in AAM; BMI: body mass index; CAD: coronary artery
637 disease; HDL-C: high density lipoprotein cholesterol; LDL-C: low density lipoprotein
638 cholesterol; PGS: polygenic score; WGHS: Women's Genome Health Study.

639

640 **Figure 5: Associations of variation in AAM with hazard ratio for risk of CAD in the**
641 **WGHS.** PGS-associated Δ AAM demonstrated a negative linear association with risk of CAD,
642 similar to the results in the UKBB. PGS-independent Δ AAM demonstrated a reverse-J shaped
643 relationship, slightly different from the U-shaped relationship seen in the UK Biobank. Analyses
644 in the WGHS largely validated the results from the UK Biobank. The different associations of
645 PGS-independent Δ AAM driving AAM later with risk of CAD may have been due to differences
646 in environmental factors between the two study cohorts.

647 AAM: age at menarche; Δ AAM: change in AAM; CAD: coronary artery disease; PGS:
648 polygenic score. Dots represent estimates, bars represent standard errors.

649 TABLES

Coronary artery disease					
	Change per year increase in AAM				
	Negative values of ΔAAM		Positive values of ΔAAM		<i>p</i> for difference between slopes
	Odds ratio (95% CI)	<i>p</i>	Odds ratio (95% CI)	<i>p</i>	
PGS-associated ΔAAM	0.91 (0.83 to 0.99)	3×10^{-2}	0.92 (0.84 to 1)	0.063	0.87
PGS-independent ΔAAM	0.92 (0.88 to 0.95)	3×10^{-6}	1.11 (1.08 to 1.15)	2×10^{-9}	7×10^{-10}
Coronary artery disease risk factors					
	Change per year increase in AAM				
	Negative values of ΔAAM		Positive values of ΔAAM		<i>p</i> for difference between slopes
	Slope (95% CI)	<i>p</i>	Slope (95% CI)	<i>p</i>	
Hemoglobin A1c (%)					
PGS-associated ΔAAM	-0.034 (-0.041 to -0.027)	$< 2 \times 10^{-16}$	-0.01 (-0.02 to -0.006)	3×10^{-3}	6×10^{-5}
PGS-independent ΔAAM	-0.022 (-0.025 to -0.019)	$< 2 \times 10^{-16}$	0.007 (0.004 to 0.01)	1.3×10^{-6}	$< 2 \times 10^{-16}$
Triglycerides (mg/dL)					
PGS-associated ΔAAM	-2.82 (-3.84 to -1.79)	7×10^{-8}	-2.63 (-3.62 to -1.64)	2×10^{-7}	0.83
PGS-independent ΔAAM	-4.25 (-4.7 to -3.8)	$< 2 \times 10^{-16}$	1.2 (0.78 to 1.62)	2×10^{-8}	$< 2 \times 10^{-16}$
High-density lipoprotein (HDL) cholesterol (mg/dL)					

PGS-associated Δ AAM	0.90 (0.69-1.11)	$<2 \times 10^{-16}$	0.74 (0.54-0.95)	8×10^{-13}	0.38
PGS-independent Δ AAM	0.85 (0.76-0.95)	$<2 \times 10^{-16}$	-0.28 (-0.37 to -0.20)	2×10^{-10}	$<2 \times 10^{-16}$
Low-density lipoprotein (LDL) cholesterol (mg/dL)					
Adjusted for LDL-lowering medications					
PGS-associated Δ AAM	0.27 (-0.15 to 0.69)	2×10^{-1}	-0.53 (-0.93 to -0.12)	1×10^{-2}	0.026
PGS-independent Δ AAM	-0.46 (-0.65 to -0.28)	1×10^{-6}	0.067 (-0.11 to 0.24)	0.45	7×10^{-4}
Including only those not taking LDL-lowering medications (n = 177,542)					
PGS-associated Δ AAM	0.14 (-0.31 to 0.59)	0.054	-0.59 (-1.02 to -0.16)	8×10^{-3}	0.059
PGS-independent Δ AAM	-0.53 (-0.72 to -0.33)	3×10^{-7}	0.0065 (-0.18 to 0.19)	0.9	2×10^{-3}
Systolic blood pressure (mmHg)					
PGS-associated Δ AAM	-0.73 (-0.99 to -0.47)	4×10^{-8}	-0.71 (-0.96 to -0.45)	5×10^{-8}	0.91
PGS-independent Δ AAM	-0.69 (-0.81 to -0.58)	$<2 \times 10^{-16}$	-0.21 (-0.31 to -0.10)	2×10^{-4}	5×10^{-7}
Diastolic blood pressure (mmHg)					
PGS-associated Δ AAM	-0.49 (-0.63 to -0.34)	5×10^{-11}	-0.49 (-0.63 to -0.35)	9×10^{-12}	0.98
PGS-independent Δ AAM	-0.52 (-0.59 to -0.46)	$<2 \times 10^{-16}$	-0.14 (-0.20 to -0.08)	5×10^{-6}	2×10^{-12}
Body-mass index (kg/m²)					
PGS-associated Δ AAM	-0.98 (-1.05 to -0.91)	$<2 \times 10^{-16}$	-0.77 (-0.84 to -0.70)	$<2 \times 10^{-16}$	4×10^{-4}
PGS-independent	-0.80 (-0.83 to -0.77)	$<2 \times 10^{-16}$	-0.079 (-0.11 to -0.051)	4×10^{-8}	$<2 \times 10^{-16}$

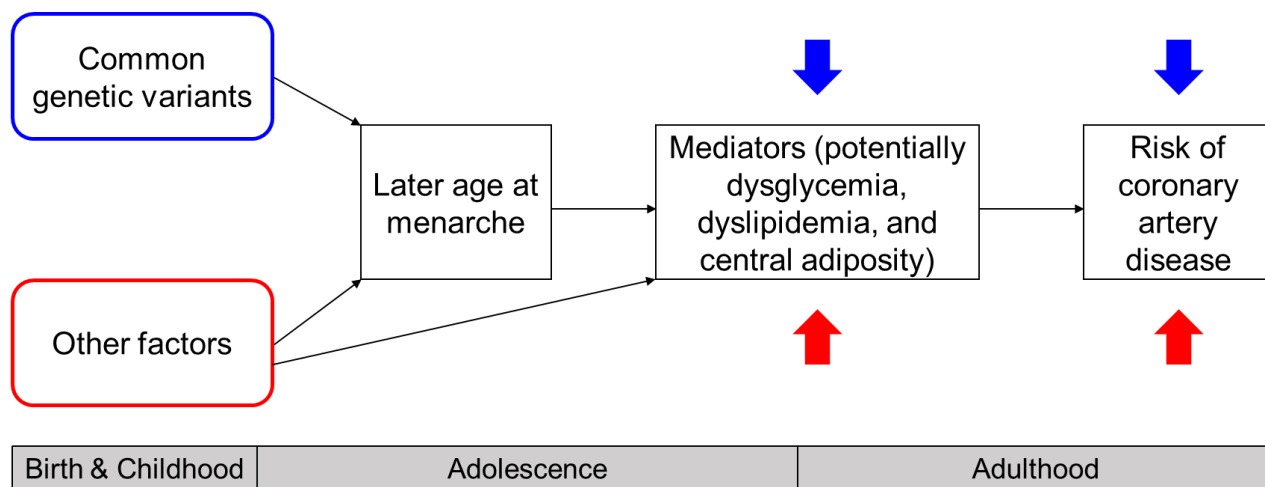
Δ AAM					
Waist-hip ratio					
PGS-associated Δ AAM	-0.0038 (-0.0047 to -0.0029)	1.1×10^{-15}	-0.0029 (-0.0038 to -0.0020)	2×10^{-10}	0.28
PGS-independent Δ AAM	-0.0046 (-0.0050 to -0.0042)	$< 2 \times 10^{-16}$	0.0016 (0.0012 to 0.0020)	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$

650

651 **Table 1:** Associations of variation in age at menarche with coronary artery disease and its risk
 652 factors.

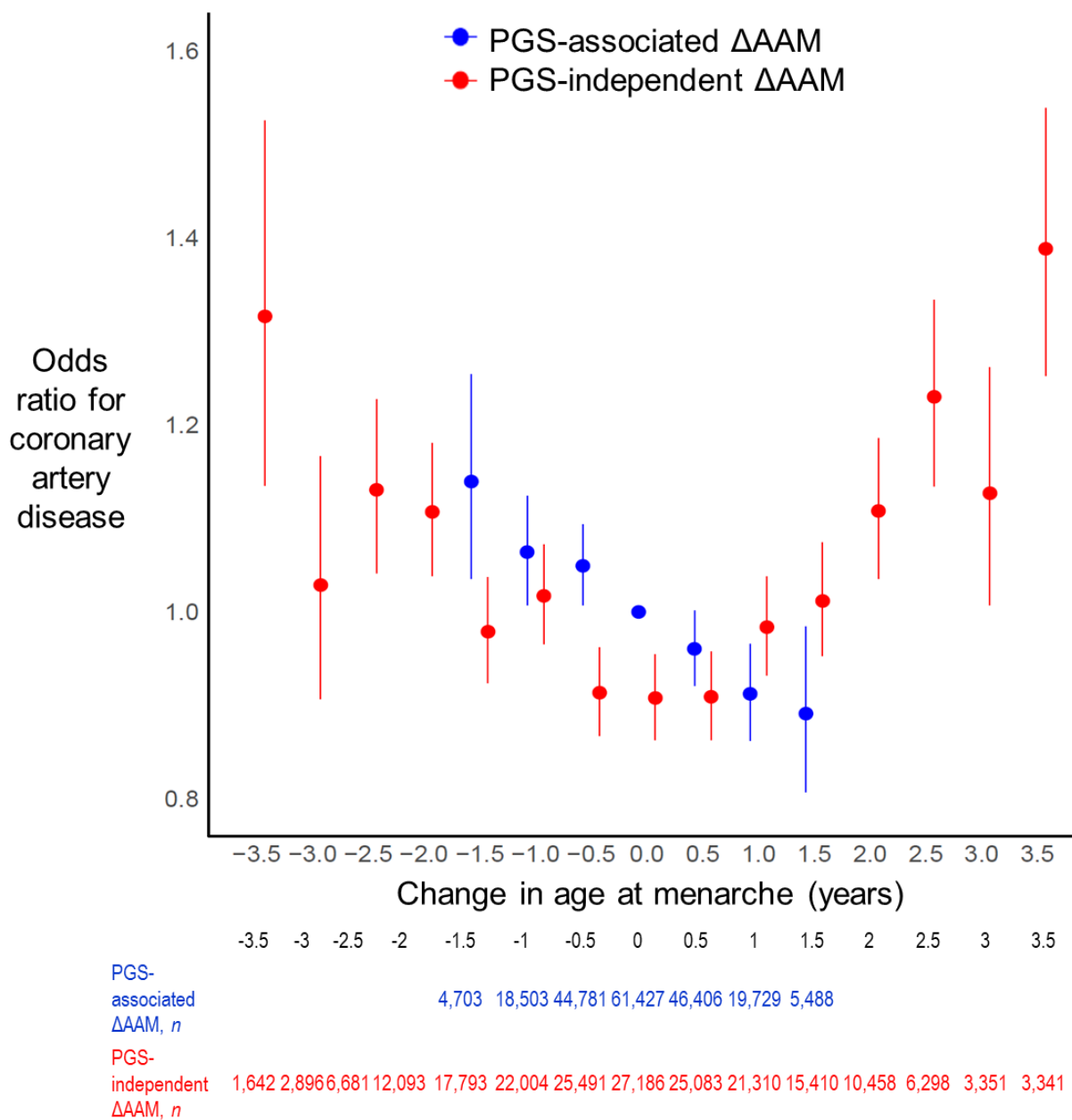
653 AAM: age at menarche; Δ AAM: change in AAM; PGS: polygenic score

654 **FIGURES**



655

656 **Figure 1:** Relationships of different sources of variation in age at menarche (AAM) with
657 coronary artery disease (CAD) and CAD risk factors. Variation in AAM was dissected into
658 variation associated with common genetic variants (as estimated by a polygenic score for AAM)
659 and variation independent of the polygenic score, then associations of these sources of variation
660 with CAD and CAD risk factors were studied. When associated with common genetic variants,
661 later AAM was associated with a lower risk of CAD and favorable changes in CAD risk factors
662 (blue arrows). However, when occurring independently of the polygenic score, later AAM was
663 associated with higher risk of CAD and unfavorable changes in CAD risk factors (red arrows).
664 The “other factors” causing later AAM independent of the PGS have yet to be identified and
665 could potentially include chronic illness, underweight, and rare genetic variants.

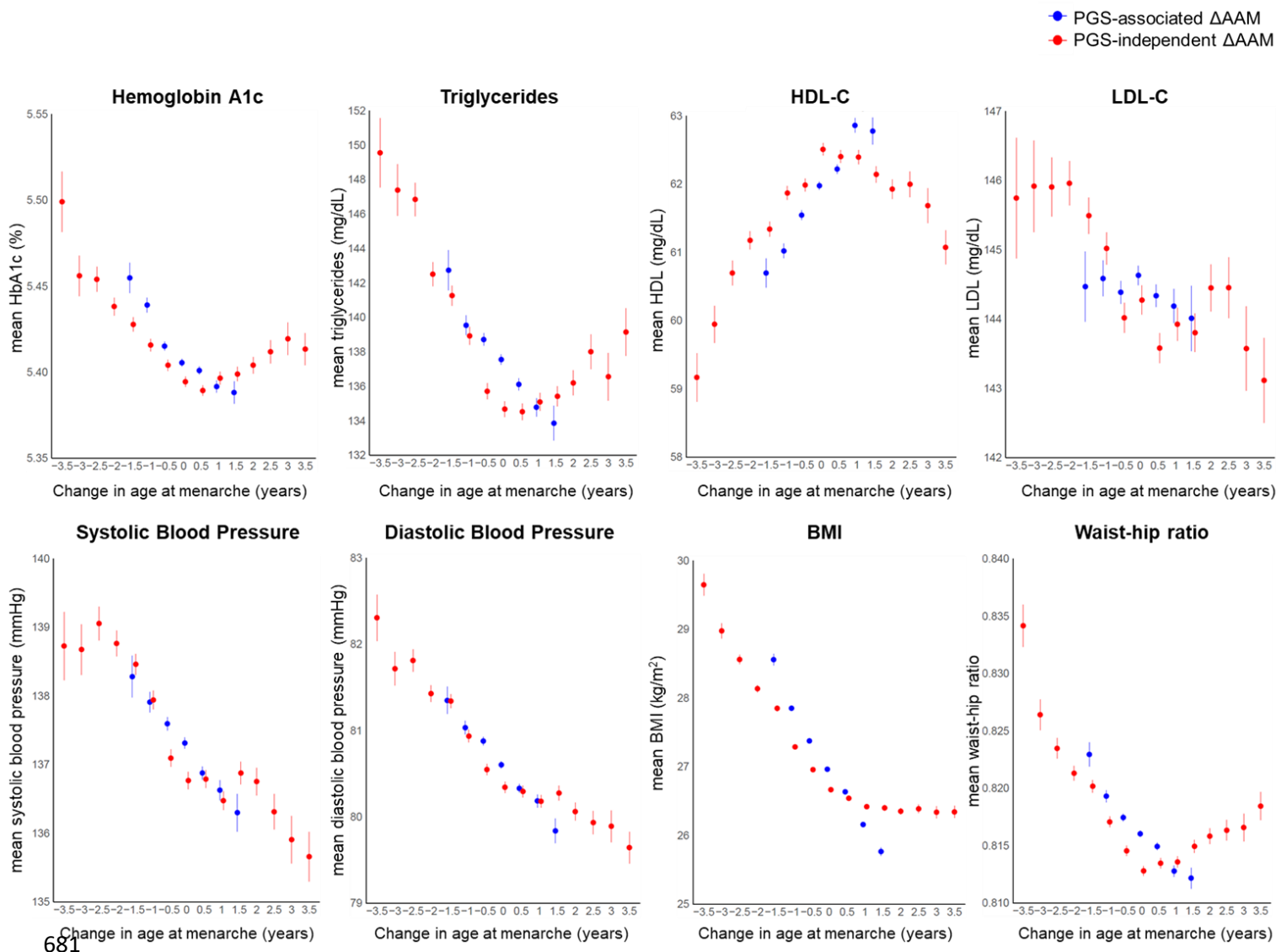


666

667 **Figure 2: Association of variation in AAM with odds of CAD in the UK Biobank.** Variation
 668 in AAM was dissected into variation associated with common genetic variants as estimated by a
 669 PGS for AAM (PGS-associated Δ AAM) and variation independent of the PGS (PGS-
 670 independent Δ AAM), then associations of these sources of variation with risk of CAD were
 671 studied. PGS-associated Δ AAM showed linear or roughly linear associations with CAD whereas
 672 PGS-independent Δ AAM showed a U-shaped association with risk of CAD. The different

673 associations of positive values of PGS-associated Δ AAM with risk of CAD demonstrates that it
674 is not later AAM itself that causes the increased risk of CAD and that the increased risk of CAD
675 observed in later AAM is being driven by factors other than common genetic variants and that.
676 To achieve bin sizes of ≥ 1000 individuals, the first and last bins for each variable represent the
677 group of individuals with Δ AAM \leq or \geq the value listed on the x-axis, respectively. Dots
678 represent estimates; bars represent standard errors.

679 AAM: age at menarche; Δ AAM: change in AAM; CAD: coronary artery disease; PGS:
680 polygenic score.



682 **Figure 3: Associations of variation in AAM with risk factors for CAD in the UK Biobank.**

683 We observed linear (or roughly linear) relationships between PGS-associated Δ AAM and most
 684 CAD risk factors. In contrast, for PGS-independent Δ AAM we observed mostly non-linear
 685 relationships with CAD risk factors. The U-shaped relationship of PGS-independent Δ AAM with
 686 CAD was mirrored by the relationships with HbA1c, triglycerides, HDL-C, and waist-hip ratio.
 687 This raises the possibility that dysglycemia, dyslipidemia and central adiposity may mediate the
 688 relationship between later AAM and increased risk of CAD. For LDL-C, n = 177,542 after

689 excluding those taking LDL-C lowering medications. Dots represent estimates, bars represent
690 standard errors.

691 AAM: age at menarche; Δ AAM: change in AAM; BMI: body mass index; CAD: coronary artery
692 disease; HDL-C: high-density lipoprotein cholesterol; LDL-C: low-density lipoprotein
693 cholesterol; PGS: polygenic score

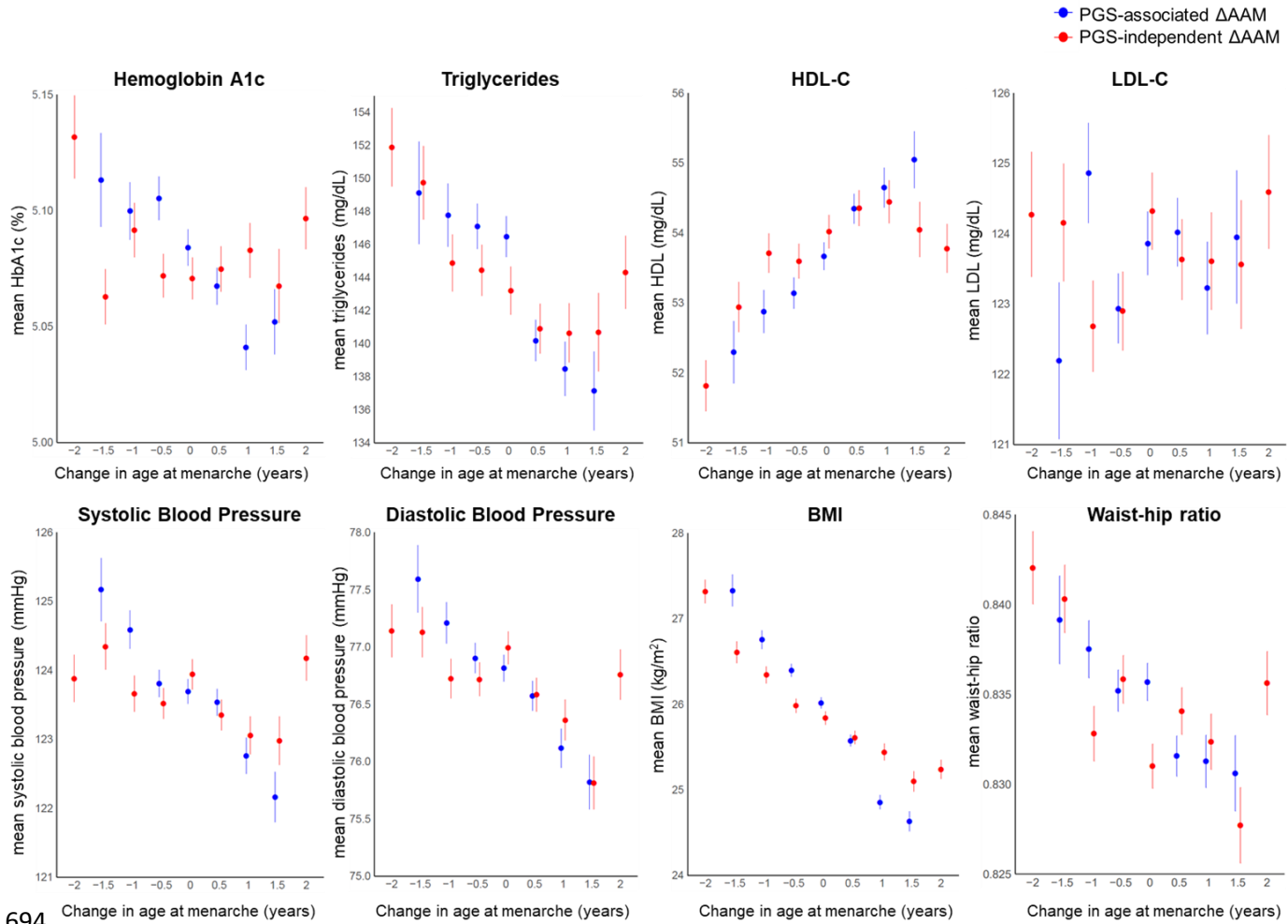
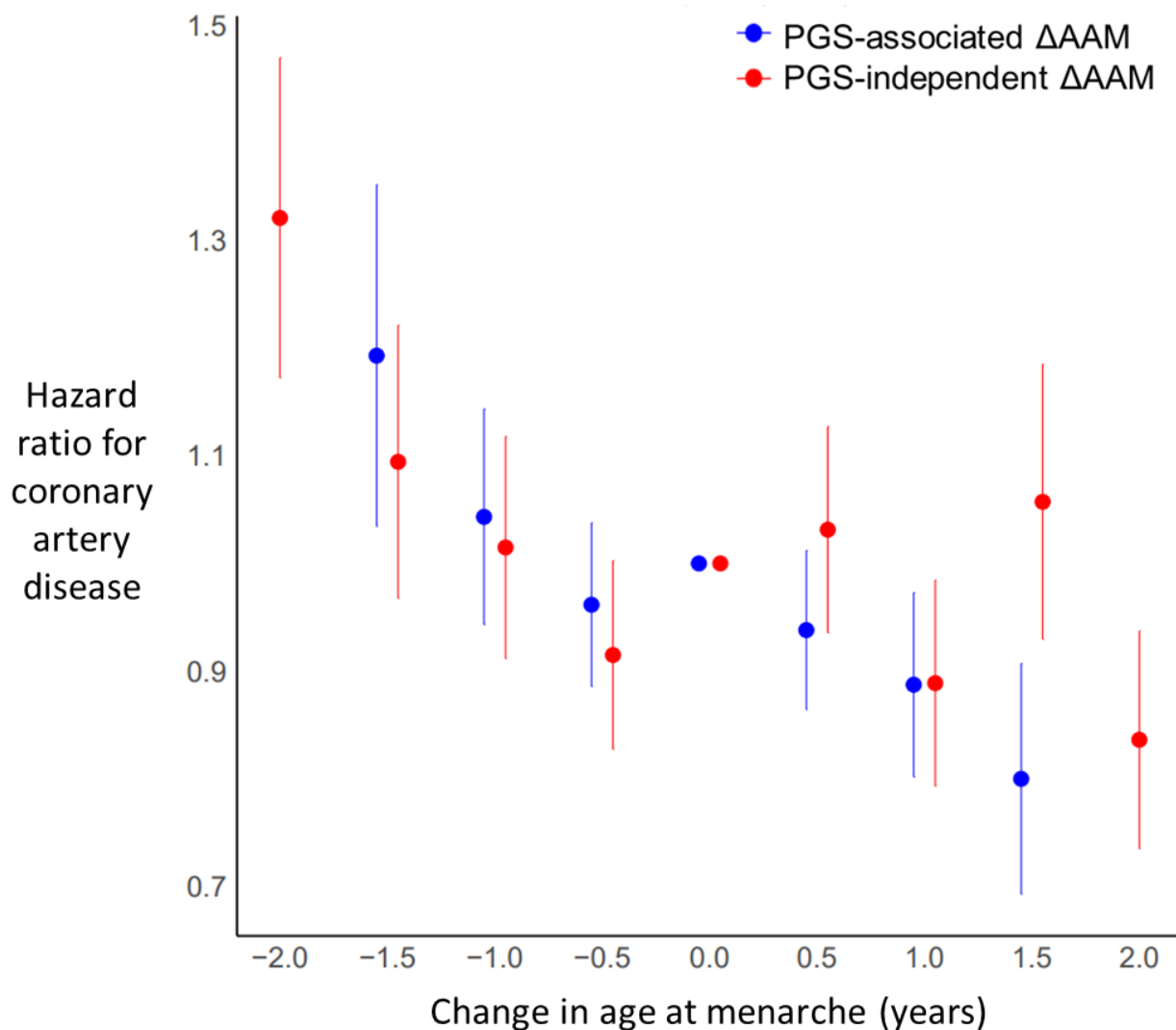


Figure 4: Associations of variation in AAM with risk factors for CAD in the WGHS. PGS-associated Δ AAM showed linear or roughly linear associations with most CAD risk factors and PGS-independent Δ AAM showed mostly non-linear associations with most CAD risk factors. Similar to results in the UK Biobank, hemoglobin A1c, triglycerides and HDL-C showed U-shaped associations with PGS-independent Δ AAM which support conclusions from the primary analyses that dysglycemia, dyslipidemia and central adiposity may mediate the relationship between later AAM and increased risk of CAD. For LDL-C, $n = 22,495$ after excluding those taking cholesterol-lowering medications. Dots represent estimates, bars represent standard errors.

704 AAM: age at menarche; Δ AAM: change in AAM; BMI: body mass index; CAD: coronary artery
705 disease; HDL-C: high density lipoprotein cholesterol; LDL-C: low density lipoprotein
706 cholesterol; PGS: polygenic score; WGHS: Women's Genome Health Study.



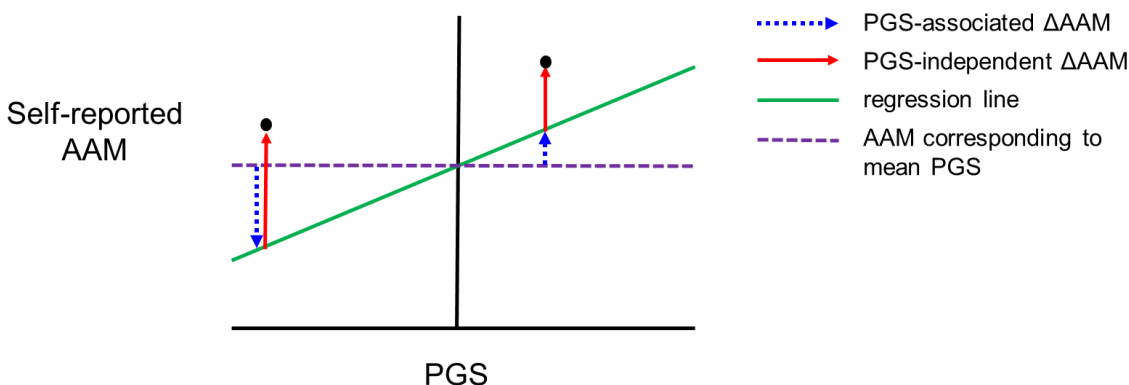
	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2
PGS-associated ΔAAM, <i>n</i>		954	2,454	4,784	5,945	5,016	2,784	1,331	
PGS-independent ΔAAM, <i>n</i>	1,620	1,752	2,736	3,755	4,055	3,614	2,475	1,470	1,791

707

708 **Figure 5: Associations of variation in AAM with hazard ratio for risk of CAD in the**
 709 **WGHS.** PGS-associated ΔAAM demonstrated a negative linear association with risk of CAD,
 710 similar to the results in the UKBB. PGS-independent ΔAAM demonstrated a reverse-J shaped
 711 relationship, slightly different from the U-shaped relationship seen in the UK Biobank. Analyses
 712 in the WGHS largely validated the results from the UK Biobank. The different associations of

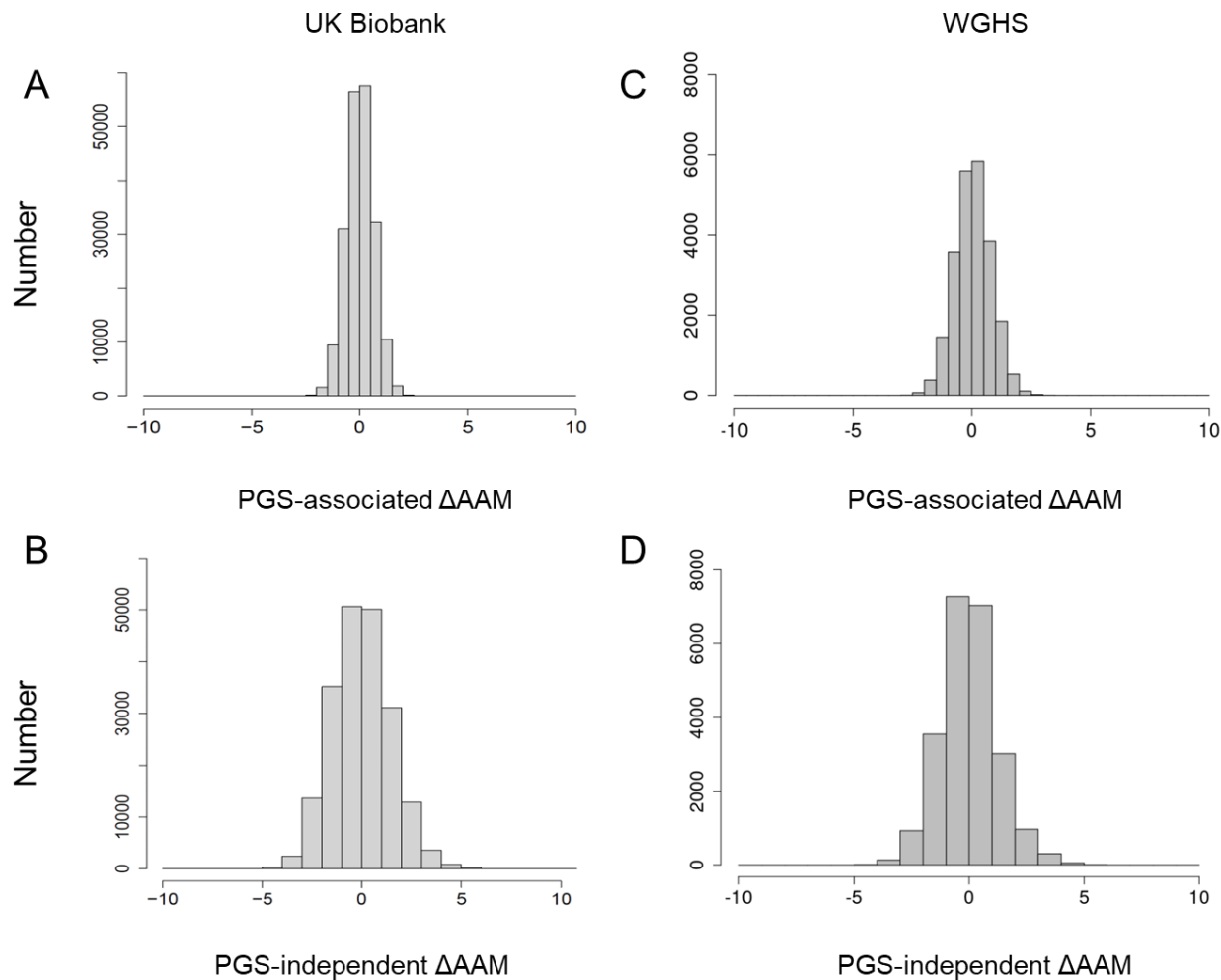
713 PGS-independent Δ AAM driving AAM later with risk of CAD may have been due to differences
714 in environmental factors between the two study cohorts.
715 AAM: age at menarche; Δ AAM: change in AAM; CAD: coronary artery disease; PGS:
716 polygenic score. Dots represent estimates, bars represent standard errors.

717 **SUPPLEMENTAL MATERIAL**



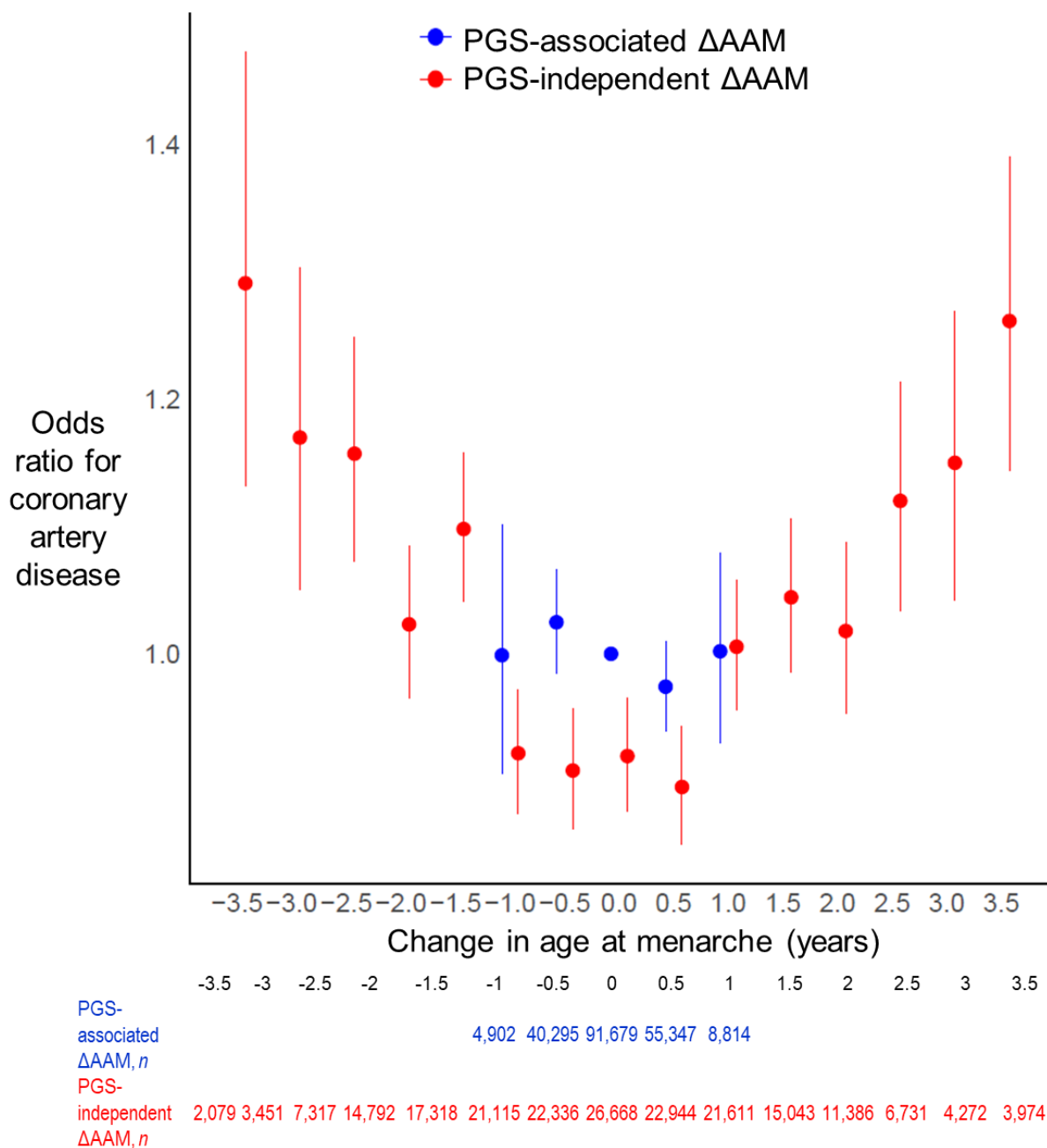
718

719 **Supplemental Figure 1: Derivation of PGS-associated and PGS-independent change in age**
720 **at menarche.** Self-reported age at menarche (AAM) was regressed against a polygenic score
721 (PGS) for AAM. The green line indicates the regression line for self-reported AAM against PGS,
722 the horizontal dashed line represents the observed AAM at the mean PGS for the cohort, the dots
723 represent two examples of women, the blue dotted arrow represents PGS-associated change in
724 AAM (Δ AAM), and the red dashed arrow represents PGS-independent Δ AAM. For the woman
725 represented by the dot on the left, AAM predicted by the PGS (regression) is below the mean for
726 the cohort, so PGS-associated Δ AAM is negative, and self-reported AAM is above the regression
727 line, so PGS-independent Δ AAM is positive. For the woman represented by the dot on the right,
728 AAM predicted by the PGS is above the mean for the cohort, so PGS-associated Δ AAM is
729 positive, and self-reported AAM is above the regression line, so PGS-independent Δ AAM is also
730 positive.



731
732 **Supplemental Figure 2: Distribution of PGS-associated Δ AAM in the UK Biobank (A),**
733 **PGS-independent Δ AAM in the UK Biobank (B), PGS-associated Δ AAM in the WGHS (C),**
734 **and PGS-independent Δ AAM in the WGHS (D).**

735 AAM: age at menarche; Δ AAM: change in AAM; PGS: polygenic score, WGHS: Women's
736 Genome Health Study.



737

738 **Supplemental Figure 3:** Association of variation in age at menarche with coronary artery
 739 disease in the UK Biobank using a PGS based on a 2014 genome-wide association study meta-
 740 analysis that did not include data from the UK Biobank.

741 AAM: age at menarche; ΔAAM: change in AAM; PGS: polygenic score. Dots represent
 742 estimates, bars represent standard errors.

	Change per year increase in AAM				<i>p</i> for difference between slopes
	Negative values of Δ AAM		Positive values of Δ AAM		
Coronary artery disease					
	Odds ratio (95% CI)	<i>p</i>	Odds ratio (95% CI)	<i>p</i>	
PGS-associated Δ AAM	0.89 (0.77 to 1.04)	0.14	0.98 (0.86 to 1.1)	0.79	0.43
PGS-independent Δ AAM	0.91 (0.88 to 0.94)	5×10^{-7}	1.09 (1.06 to 1.13)	4×10^{-7}	5×10^{-9}

743

744 **Supplemental Table 1:** Associations between variation in age at menarche and coronary artery
745 disease using a PGS based on a 2014 genome-wide association study meta-analysis that did not
746 include data from the UK Biobank.

747 AAM: age at menarche; Δ AAM: change in AAM; PGS: polygenic score