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 [a]Division of Endocrinology, Department of Pediatrics, Boston Children's Hospital, [b]Departments
6 of Pediatrics (A.S., Y.M.C.) and Medicine (D.I.C.), Harvard Medical School, [c]Division 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. ORCID: 0000-0003-4342-5343

14

15 **Email addresses:**

16 Yee-Ming Chan: yee-ming.chan@childrens.harvard.edu; Daniel I. Chasman:
17 dchasman@bwh.harvard.edu

18

23   **Abstract**

25   **Background:** In women, both earlier and later age at menarche (AAM) are associated with
26   increased risk of coronary artery disease (CAD). This study sought to determine if the
27   relationship of AAM with CAD and CAD risk factors differs for different underlying sources of
28   variation in AAM – specifically, variation attributable to common genetic variants as represented
29   by a polygenic score (PGS) vs. variation in AAM adjusted for the PGS.

31   **Methods:** Primary analyses were conducted on data from 201,037 women in the UK Biobank
32   and validation studies on data from 23,268 women in the Women's Genome Health Study
33   (WGHS). For each individual, a PGS for AAM was calculated, then two variables were
34   estimated from linear regression models: the genetically predicted AAM (the estimated AAM for
35   each woman solely due to the effects of common genetic variants) and the PGS-adjusted AAM
36   (estimated AAM for each woman solely due to factors other than the PGS). Logistic regression
37   and linear splines were then used to study the relationships of these variables with CAD and
38   CAD risk factors.

40   **Results:** Genetically predicted AAM demonstrated a linear relationship with CAD and linear or
41   roughly linear relationships with CAD risk factors. In contrast, PGS-adjusted AAM
42   demonstrated a U-shaped relationship with CAD and with hemoglobin A1c, triglycerides, HDL-
43   C, and waist-hip ratio. Validation studies using WGHS data produced similar results.

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

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

1

51   Non-standard Abbreviations and Acronyms:

52   AAM: age at menarche

53   BMI: body-mass index

54   CAD: coronary artery disease

55   DBP: diastolic blood pressure

56   HbA1c: hemoglobin A1c

57   HDL-C: high-density lipoprotein cholesterol

58   LDL-C: low-density lipoprotein cholesterol

59   PGS: polygenic score

60   SBP: systolic blood pressure

61   WGHS: Women's Genome Health Study

INTRODUCTION

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

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

3

84   been associated with several CAD risk factors, there may be distinct mechanisms underlying the

85   association of later AAM with increased risk of CAD.

86

87   Variation in AAM can stem from several upstream sources, including genetics (both common

88   genetic variants and rare genetic variants), acquired factors such as chronic illness, chronic

89   stress, underweight, and undernutrition, and environmental factors such as family composition

90   (e.g., presence or absence of father) (17). It is possible that some of these sources of variation

91   may influence risk of CAD only through their influence on AAM, while others may directly

92   influence CAD risk and features of the metabolic syndrome (Figure 1). Thus, dissecting variation

93   in AAM based on underlying sources of variation could provide a clearer understanding of the

94   relationship of AAM with CAD.

95

96   Genetics is a major source of variation in AAM, with half to three-quarters of variation

97   attributable to genetics (18,19). A 2017 genome-wide association study (GWAS) on AAM

98   identified 389 independent single-nucleotide polymorphisms (SNPs) associated with AAM at

99   genome-wide significance (20). The results from this GWAS allow the calculation of a polygenic

100   score (PGS) for a given individual to reflect the cumulative contribution of common genetic

101   variants to AAM.

102

103   Previous studies have used PGSs to dissect the influence of genetics vs. environmental factors

104   (or other factors not captured by the PGS) that contribute to traits such as BMI and LDL-C

105   (21,22). These studies have found that associations of these traits with health outcomes differ

106   between genetically and environmentally influenced traits. For example, obesity driven by

107   environmental factors was associated with more harmful cardiovascular outcomes than

108   genetically predicted obesity (21), suggesting that dissecting effects based on underlying source

109   of variation can allow deeper insights into pathogenic mechanisms.

110

111   The aims of this study were two-fold: first, to determine if the association between later AAM

112   and increased risk of CAD depends on the underlying source of variation in AAM – specifically,

113   common genetic variation vs. other sources of variation; second, to study the relationships of

114   these different sources of variation in AAM with CAD risk factors.

115

116   **METHODS**

117   To study how different sources of variation in the timing of menarche are associated with risk of

118   CAD, two variables were calculated: genetically predicted AAM (the estimated AAM

119   determined solely by the effects of common genetic influences, as estimated by a PGS for AAM)

120   and PGS-adjusted AAM (the estimated AAM determined solely by factors other than common

121   genetic variants, the PGS).

122

123   *Study cohorts*

124   This study used data from two cohorts: the UK Biobank for primary analyses and the Women's

125   Genome Health Study (WGHS) for validation analyses. The UK Biobank is a population-based

126   cohort of over 500,000 men and women in the UK 40 years and older at the time of recruitment,

5

127   with extensive health-related phenotypic and laboratory data as well as individual-level genetic

128   data (23). This study analyzed data from 201,037 unrelated women in the UK Biobank of non-

129   Finnish European ancestry (as determined through principal component analysis) (24) who had

130   genetic and self-reported AAM data. During data collection in the UK Biobank, any value of

131   AAM <5 years or >25 years was rejected, and any AAM entered as <6 years or >20 years

132   required confirmation from participants (25). Women with missing AAM data were excluded

133   from analysis. The UK Biobank obtained the multiple ethical and regulatory approvals required

134   for recruitment and research procedures, and participants provided written consent (23). The

135   WGHS is a cohort of initially healthy American women aged 45 years and older at enrollment

136   with genetic and phenotypic data, followed over 26-28 years for cardiovascular and other

137   outcomes (26). This analysis studied data from 23,268 women in the WGHS who had genetic

138   and self-reported AAM data. Self-reported AAM ≤9 years or ≥ 17 years were entered as 9 years

139   or 17 years, respectively. The WGHS was approved by the institutional review board of Brigham

140   and Women's Hospital, and participants consented to ongoing analyses (26).

141

142   *Polygenic score calculation*

143   A PGS was calculated for each woman in the above cohorts in two steps. The first step used the

144   PRS-CS algorithm, which allows the inclusion of all available SNPs from the GWAS on AAM

145   (20), not just those that meet a given *p*-value threshold, and weights the SNP effect sizes based

146   on their significance and adjusts for linkage disequilibrium (27). The use of this algorithm has

147   the potential to explain more variability in AAM than algorithms that use only the SNPs that

148   meet a given significance threshold. The second step calculated the PGS using PRSice-2

149   (without clumping or thresholding) to sum the weighted effect sizes for all SNPs in each

150   individual with the ability to incorporate the probabilistic genotype dosages generated by

151   imputation (28).

152

153   *Subdividing variation in age at menarche*

154   Using each full cohort, regression of self-reported AAM was performed against the PGS for

155   AAM, with the first 10 genetic principal components, assessment center and technical variables

156   such as array number as covariates. The following two variables were then calculated for each

157   individual:

1.   Genetically predicted AAM: This represents the estimated AAM that each woman would
      have had if her AAM were determined solely by her common genetic variants, as
      estimated by the PGS. Statistically, it is the AAM predicted by the regression of AAM
      against the PGS (Supplemental Figure 1, (29)).

2.   PGS-adjusted AAM: This represents the estimated AAM that each woman would have
      had if there were no effect of her common genetic variants (as represented by the PGS),
      i.e., her AAM were determined solely by sources of variation other than common genetic
      variants. Statistically, this was the residual of the regression of AAM against the PGS for
      each individual, added to the AAM corresponding to the mean PGS (to simulate a
      scenario in which the contribution of the PGS is the same for all women) (Supplemental
      Figure 1, (29)).

169

170   *Outcomes and analytical methods*

7

171 The study's primary outcome variable was CAD risk. In the UK Biobank, prevalent CAD at

172 baseline was determined as previously described using a combination of self-report, ICD-9/10

173 codes, and procedure codes (24). The WGHS recruited middle-aged female healthcare

174 professionals with no history of CAD at baseline and identified validated incident CAD during

175 26-28 years of follow-up as described previously (26). Secondary outcome variables were CAD

176 risk factors at baseline for both cohorts: hemoglobin A1c (HbA1c), triglycerides, high-density

177 lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), systolic blood

178 pressure (SBP), diastolic blood pressure (DBP), and BMI; data on waist-hip ratio was available

179 at baseline in the UK Biobank and 6 years after recruitment in the WGHS.

180

181 The relationship of genetically predicted AAM and PGS-adjusted AAM was studied with logistic

182 regression for prevalent CAD, with Cox proportional hazards models for incident CAD, and with

183 linear splines for each continuous variable (HbA1c, triglycerides, HDL-C, LDL-C, SBP, DBP,

184 BMI, waist-hip ratio), with a knot at 12.94 years, which is the AAM corresponding to the mean

185 PGS. Because we observed nonlinear relationships, we also used linear splines to separately

186 analyze values of genetically predicted and PGS-adjusted AAM earlier and later than the mean.

187 For analyses with LDL-C, results were corrected for self-reported use of cholesterol-lowering

188 medications – these were specific LDL-lowering medications in the UK Biobank, (statins,

189 ezetimibe, and bile-acid sequestrants) and collective cholesterol-lowering medications in the

190 WGHS; additional analyses included only women not taking these medications.

191 Covariates of age and age^2 were used for all analyses. Analyses were conducted using R v.4.3.1.

192 A significance threshold of 0.05 was used.

193

194   To determine if results in the UK Biobank were biased by overfitting as a result of the UK

195   Biobank having contributed to the GWAS for AAM, analyses of CAD were repeated using a

196   PGS calculated using an earlier GWAS that did not include the UK Biobank (30). Furthermore,

197   to determine if outlier values of AAM were disproportionately affecting results, analyses were

198   repeated after excluding women with extreme values of AAM such that up to 0.1% of women

199   were excluded at each extreme.

200

201   **RESULTS**

202   To understand potential different effects of different sources of variation in AAM, we

203   constructed two hypothetical scenarios. In the first scenario, a woman's AAM is determined

204   solely by the effects of common genetic variants. We refer to the estimated AAM in this scenario

205   as "genetically predicted AAM." In the second scenario, a woman's AAM is determined solely

206   by the effects of factors other than common genetic variants (or more precisely, by factors other

207   than the PGS). We refer to the estimated AAM in this scenario as "PGS-adjusted AAM."

208

209   To examine the differential associations of these two sources of variation with CAD risk, we

210   analyzed data from 201,037 unrelated, non-Finnish European women in the UK Biobank. We

211   first calculated a PGS for AAM for each woman, then regressed self-reported AAM against the

212   PGS for AAM to calculate genetically predicted AAM and to derive PGS-adjusted AAM

213   (Supplemental Figure 1, (29)). For instance, for an individual with a self-reported AAM of 16

214   years and a PGS of 0.357, the regression provided a genetically predicted AAM of 13.42 years

215   and a residual of 2.58 years. In the regression, the AAM corresponding to the mean PGS was

216   12.95 years, and hence the PGS-adjusted AAM was 2.58 + 12.95 = 15.53 years.

217

218   For women in the UK Biobank, the regression of self-reported AAM against the PGS for AAM

219   demonstrated that the PGS accounted for 15.8% of the variation in AAM. Genetically predicted

220   AAM had a mean ± standard deviation of 12.95 ± 0.64 years, and PGS-adjusted AAM had a

221   mean ± standard deviation of 12.94 ± 1.47 years with a standard deviation of 1.47 years

222   (Supplemental Figure 2, (29)). We then studied the associations of genetically predicted AAM

223   and PGS-adjusted AAM with risk of CAD and with CAD risk factors.

224

225   *Risk of coronary artery disease*

226   Risk of CAD demonstrated a linear relationship with genetically predicted AAM but a non-

227   linear, U-shaped relationship with PGS-adjusted AAM (Figure 2). In the linear relationship of

228   genetically predicted AAM with risk of CAD; each 1-year increase in genetically predicted

229   AAM was associated with an odds ratio (OR) for CAD of 0.91. In other words, for every 1 year

230   that AAM was later due solely to the effects of common genetic variants (as estimated by the

231   PGS), the odds of CAD were lower by 9% (Table 1). To assess for non-linear relationships,

232   linear spline analyses were done which showed no difference in slopes when genetically

233   predicted AAM was earlier vs. later than the mean (Table 1). This lack of difference in slopes

234   indicates that the linear relationship of genetically predicted AAM and risk of CAD extends

235   across all values of genetically predicted AAM.

236

10

237   In contrast, the association between PGS-adjusted AAM and risk of CAD was U-shaped, with

238   both earlier and later values associated with increased CAD risk (Figure 2). For PGS-adjusted

239   AAM that was earlier than the mean, each 1-year increase (causing AAM to be less early) was

240   associated with an OR for CAD of 0.91 (Table 1). In other words, in women whose PGS-

241   adjusted AAM was earlier than the mean, for every 1 year that AAM was less early due to

242   factors other than the PGS, the odds of CAD were lower by 9%. In contrast, for PGS-adjusted

243   AAM later than the mean, each 1-year increase (causing AAM to be even later) was associated

244   with an OR for CAD of 1.11, i.e., the odds of CAD were higher by 11% (Table 1).

245

246   Because the 2017 AAM GWAS meta-analysis included data from the UK Biobank, which could

247   introduce bias in the above analyses, we conducted sensitivity analyses using the results from an

248   AAM GWAS meta-analysis published in 2014 that did not include the UK Biobank. These

249   analyses showed similar results for CAD (Supplemental Figure 3, Supplemental Table 1, (29)).

250   We also conducted sensitivity analyses excluding women with extreme values of AAM (at the

251   upper and lower 0.1%) and obtained similar results.

252

253   *Hemoglobin A1c*

254   We then examined the associations of genetically predicted AAM and PGS-adjusted AAM with

255   CAD risk factors. Just as the relationship between genetically predicted AAM and CAD was

256   linear, we observed linear (or roughly linear) relationships between genetically predicted AAM

257   and most CAD risk factors. In contrast, for PGS-adjusted AAM we observed mostly non-linear

258   relationships with CAD risk factors.

11

259

260 For hemoglobin A1c, an indicator of average blood glucose over the preceding 3 months,

261 genetically predicted AAM demonstrated a roughly negative linear relationship, with later

262 genetically predicted AAM associated with lower HbA1c (Figure 3). Linear spline analyses

263 indicated that while both slopes were negative, for values of genetically predicted AAM earlier

264 than the mean, the slope of the association with HbA1c was slightly steeper than for values of

265 genetically predicted AAM later than the mean (-0.034 and -0.022 %/year, respectively; *p* for

266 difference between slopes: 7×10^{-5}; Table 1). In contrast, PGS-adjusted AAM demonstrated a U-

267 shaped relationship with HbA1c: both earlier and later PGS-adjusted AAM were associated with

268 an increase in HbA1c (for values earlier than the mean: slope = -0.022 %/year, for values later

269 than the mean: slope = 0.007%/year; *p* for difference between slopes $< 2 \times 10^{-16}$; Figure 3, Table

270 1). This was similar to the U-shaped relationship of PGS-adjusted AAM with CAD risk.

271

272 Comparing genetically predicted AAM and PGS-adjusted AAM, for values earlier than the

273 mean, the slopes for the associations with HbA1c were comparable (Table 1). However, for

274 values of AAM later than the mean, different slopes were seen for genetically predicted vs. PGS-

275 adjusted AAM; as noted above, for every 1 year that AAM was delayed, HbA1c *decreased* by

276 0.01% for genetically predicted AAM but *increased* by 0.007% for PGS-adjusted AAM.

277

278 *Lipids*

279 Genetically predicted AAM demonstrated linear relationships with triglycerides and HDL-C

280 (Figure 3), with later genetically predicted AAM associated with lower triglycerides and higher

12

281   HDL-C (Figure 3) and no significant difference in slopes between earlier and later genetically

282   predicted AAM (*p* for difference in slopes for triglycerides = 0.83 mg/dL/year, for HDL-C =

283   0.38 mg/dL/year; Table 1). In contrast, PGS-adjusted AAM demonstrated non-linear

284   relationships with HDL-C and triglycerides – U-shaped for triglycerides and inverted-U-shaped

285   for HDL-C (Figure 3) - with both increasingly earlier and later PGS-adjusted AAM associated

286   with higher triglycerides and lower HDL-C (Figure 3, Table 1).

287

288   For LDL-C, associations showed a less clear pattern. Earlier genetically predicted AAM showed

289   no association with LDL-C, and later genetically predicted AAM showed a negative linear

290   relationship (Table 1). In contrast, for PGS-adjusted AAM, earlier values showed a negative

291   linear relationship, but later PGS-adjusted AAM showed no significant association (Figure 3,

292   Table 1). Similar results were seen when analyses excluded those taking LDL-lowering

293   medications (Table 1).

294

295   *Blood pressure*

296   Later genetically predicted AAM was associated with both lower SBP and lower DBP (Figure

297   3), with no significant difference between the slopes for values of genetically predicted AAM

298   earlier vs. later than the mean (Table 1). Later PGS-adjusted AAM was also associated with

299   lower SBP and DBP (Figure 3), but the slopes of the associations were steeper for values of

300   PGS-adjusted AAM that were earlier vs. later than the mean (Table 1).

301

302   *Adiposity*

13

303   We studied two estimates of adiposity: BMI and waist-hip ratio. For BMI, we found a roughly

304   negative linear relationship between genetically predicted AAM and BMI, with later genetically

305   predicted AAM associated with lower BMI (Figure 3, Table 1). The slope of the association with

306   BMI was slightly steeper for values of genetically predicted AAM earlier vs. later than the mean

307   (Table 1). This pattern was similar to the relationship seen between genetically predicted AAM

308   and HbA1c.

309

310   For PGS-adjusted AAM, a different relationship was seen with BMI (Figure 3). While later PGS-

311   adjusted AAM was consistently associated with lower BMI, the slope of the association was ten-

312   fold steeper for values earlier vs. later than the mean (-0.80 kg/m2/year vs. -0.079 kg/m2/year

313   respectively; p for difference between slopes 2×10^{-16}; Table 1). This resembled the patterns seen

314   with SBP and DBP.

315

316   For waist-hip ratio, genetically predicted AAM showed a negative linear association, with no

317   significant difference in slopes for the association when genetically predicted AAM was earlier

318   vs. later than the mean (Figure 3, Table 1). In contrast, the relationship of PGS-adjusted AAM

319   with waist-hip ratio was U-shaped, with values both earlier and later than the mean associated

320   with higher waist-hip ratio (Figure 3, Table 1), similar to the associations seen with CAD,

321   HbA1c, triglycerides, and HDL-C.

322

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

14

324   Validation studies using data from the WGHS produced results similar to the above results for

325   the UK Biobank. In the WGHS, genetically predicted AAM showed linear relationships with

326   HbA1c, triglycerides, HDL-C, SBP, DBP, BMI and waist-hip ratio and no clear relationship with

327   LDL-C (Figure 4). Just as in the UK Biobank, PGS-adjusted AAM showed U-shaped

328   relationships with triglycerides and HbA1c, an inverted-U-shaped relationship with HDL-C, and

329   a not completely linear relationship with BMI in the WGHS (Figure 4).

330

331   There were two differences between results from the WGHS and the UK Biobank: 1) for SBP

332   and DBP, PGS-adjusted AAM showed roughly U-shaped relationships in the WGHS (Figure 4)

333   rather than the not fully linear relationships seen in the UK Biobank (Figure 3), and 2) for waist-

334   hip ratio, PGS-adjusted AAM showed a negative linear relationship for values earlier than the

335   mean but no significant relationship for values later than the mean in the WGHS (Figure 4)

336   compared to the clear U-shaped relationship seen in the UK Biobank (Figure 2).

337

338   For risk of CAD itself, the relationship of genetically predicted AAM with CAD in the WGHS

339   was similar to that in the UK Biobank (Figure 5, Supplemental Table 1, (29)), with a negative

340   linear relationship in both cohorts. For the relationship of PGS-adjusted AAM with CAD,

341   analyses in the WGHS demonstrated a reverse-J shaped relationship (Figure 5, Supplemental

342   Table 1, (29)), slightly different from the U-shaped relationship seen in the UK Biobank.

343

344   **DISCUSSION**

15

345   In this study, we dissected variation in AAM into variation attributable to common genetic

346   variants, as estimated by a PGS, (as reflected by genetically predicted AAM), and variation

347   adjusted for the PGS (as reflected by PGS-adjusted AAM), and we found different relationships

348   with these two sources of variation in AAM with risk of CAD and with CAD risk factors,

349   particularly when causing AAM to occur later.

350

351   In general, later AAM showed favorable associations when attributable to common genetic

352   variation (as estimated by the PGS) and harmful or neutral associations when attributable to

353   other yet-to-be-identified sources of variation. If all sources of variation were affecting CAD and

354   CAD risk factors wholly through AAM itself, the associations would be similar regardless of the

355   source of variation studied. Our finding that these associations varied based on the underlying

356   source of variation driving later AAM therefore indicates that it is not later AAM itself that

357   causes increased risk of CAD. Rather, there appear to be PGS-independent factors that cause

358   both later AAM and increased risk of CAD and unfavorable cardiometabolic risk profiles. Such

359   factors could include environmental or acquired influences (such as chronic illnesses, chronic

360   stress, undernutrition) as well as genetic influences not captured by the PGS, and future studies

361   will identify these factors and determine how they contribute to CAD risk.

362

363   An alternative possibility is that later AAM itself does in fact cause increased risk of CAD,

364   regardless of underlying source of variation. For this to be true, given our finding that later

365   genetically predicted AAM is associated with lower CAD risk, there would have to be strong

366   direct (pleiotropic) effects of common genetic variants on risk of CAD, with later genetically

16

367   predicted AAM strongly protective against CAD, such that the net effect is the lower risk of

368   CAD that we observed with later genetically predicted AAM. However, previous MR studies

369   have suggested that such pleiotropic effects of common genetic variants for AAM on risk of

370   CAD are small (10,31–33). Hence, this alternative possibility is not supported by existing

371   evidence, and the more likely explanation is that PGS-independent factors that cause later AAM

372   have direct deleterious effects on CAD risk factors and CAD.

373

374   Earlier AAM, whether attributable to common genetic variants or other sources of variation, was

375   consistently associated with greater cardiometabolic risk. Thus, it seems that earlier AAM itself

376   is intrinsically associated with risk for CAD and worsening CAD risk factors; indeed, some

377   Mendelian randomization studies suggest that earlier AAM is causative of these negative

378   outcomes (11,31). However, other Mendelian randomization studies have suggested that,

379   because many loci that affect AAM also affect BMI, the associations with increased risk of CAD

380   are mainly through effects on BMI, and AAM itself may have only a small direct influence

381   (10,33).

382

383   The U-shaped relationship of PGS-adjusted AAM with CAD was mirrored by the relationships

384   with HbA1c, triglycerides, HDL-C, and waist-hip ratio. This raises the possibility that

385   dysglycemia, dyslipidemia, and central adiposity contribute to the relationship between later

386   AAM and increased risk of CAD; further studies will be required to formally evaluate these

387   CAD risk factors as potential mediators of this relationship. As noted above, previous studies

388   evaluating associations of later AAM with these CAD risk factors have had conflicting results.

17

389   While differences between study cohorts may have accounted for some of the differing results,

390   our findings raise the additional possibility that the relationships between later AAM and these

391   outcomes may have been obscured by opposing effects of genetically predicted and PGS-

392   adjusted variation in AAM. Of note, the associations with waist-hip ratio differed from those

393   with BMI, suggesting that central adiposity, reflected by waist-hip ratio, is more relevant than

394   BMI for CAD, as has been suggested by prior studies (34,35).

395

396   Given the well-known effects of increased adiposity on CAD and CAD risk factors such as

397   dysglycemia and dyslipidemia, it is possible that associations of PGS-adjusted variation in AAM

398   with CAD, dysglycemia, and dyslipidemia are largely mediated by associations with BMI and

399   central adiposity; however, we were unable to test this possibility with UK Biobank data. The

400   relationship between AAM and BMI is complex because childhood BMI – not available in the

401   UK Biobank – affects both AAM and adult BMI, and AAM is itself also associated with adult

402   BMI. Thus, adjusting for adult BMI would introduce collider bias. Future work could focus on

403   disentangling these associations by using cohorts with measures of childhood BMI and/or

404   genetic tools such as clustering analyses to generate partitioned polygenic scores (36,37).

405

406   Our analyses in the Women's Genome Health Study (WGHS) largely validated our results from

407   the UK Biobank. Genetically predicted AAM demonstrated linear relationships with CAD and

408   CAD risk factors, and PGS-adjusted AAM demonstrated mostly non-linear relationships with

409   these outcomes, supporting the conclusions described earlier. However, the two cohorts also

410   demonstrated some differences, most notably in the association of PGS-adjusted AAM with the

18

411   risk of CAD and waist-hip ratio. In the UK Biobank, PGS-adjusted *later* AAM was associated

412   with an increased risk of CAD, but there was no association in the WGHS. For waist-hip ratio,

413   PGS-adjusted AAM demonstrated a clear U-shaped association in the UK Biobank but a reverse

414   J-shaped relationship in the WGHS. There are several potential reasons for this difference. First,

415   it is possible that the smaller sample size and lower power in the WGHS affected the ability to

416   find significant associations with PGS-adjusted AAM and risk of CAD and waist-hip ratio.

417   Second, environmental influences affecting AAM (which would contribute to PGS-adjusted

418   AAM) could differ between the two cohorts. The WGHS recruited women in the United States

419   of America born in 1950 or earlier, while the UK Biobank recruited women in the UK who were

420   born between 1932 and 1969. The different impact of global events such as World War II on the

421   two countries could contribute to differences in the PGS-independent factors (which includes

422   environmental factors) and, in turn, to different associations with risk of CAD. Third, the WGHS

423   excluded women with a history of CAD at the time of enrollment whereas the UK Biobank did

424   not, and this may also have led to differences in PGS-adjusted factors between the cohorts.

425   Fourth, the participants in the WGHS were health professionals while the UK Biobank drew

426   from the general UK population, and this may have led to further differences in PGS-adjusted

427   factors, such as higher socioeconomic status, greater knowledge of CAD and its risk factors,

428   healthier diets and lifestyles, and use of preventative interventions, as well as potentially less

429   variation in these factors. While the analyses also differed in the CAD measure used from each

430   cohort – prevalence of CAD was analyzed in the UK Biobank compared to incidence of CAD in

431   the WGHS – it is unlikely to account for the difference in results, as an analysis of incident CAD

432   in the Million Women Study in the UK also showed a U-shaped relationship between AAM and

433   CAD (1). Despite these differences in the results between the two cohorts, results from both

19

434 cohorts consistently demonstrated differences between the associations of *earlier* vs. *later* PGS-

435 adjusted AAM with these outcomes.

436

437 In 2021, Liang et al. used a PGS to represent genetically predicted AAM and examined

438 associations with all-cause mortality, also using data from the UK Biobank. Interestingly, they

439 found a U-shaped association with mortality, with both earlier and later genetically predicted

440 AAM associated with higher risk; this finding contrasts with the linear relationship we found

441 between genetically predicted AAM and CAD. This difference suggests that later genetically

442 predicted AAM increases the risk of causes of mortality other than CAD, and future studies of

443 genetically predicted AAM are needed to identify these causes.

444

445 Prior studies that have also dissected influences on human traits into genetic vs. environmental

446 influences have found differences in the magnitude of effect of genetics vs. environmental

447 influences on health outcomes (21,22). Interestingly, our results show not only different

448 magnitudes of association of genetically predicted and PGS-adjusted AAM with risk of CAD,

449 but also opposite directionality. This further underscores the value of separating the effects of

450 genetics vs. other influences while studying human traits as they can have starkly different

451 effects.

452

453 Sensitivity analyses using a GWAS that did not include the primary cohort, the UK Biobank,

454 produced similar results and suggested that these results were not biased by overfitting.

455   Additionally, sensitivity analyses excluding women with extreme values of AAM also produced

456   similar results, suggesting that results were not heavily influenced by outliers.

457

458   One limitation of this study is that while the PGS represents genetic influences on AAM, it does

459   not represent all genetic factors that influence AAM. Our analyses demonstrated that the PGS

460   explains 15.8% of the variation in the observed AAM, but prior studies suggest that 49-73% of

461   variation in AAM is inherited (18,19); hence, a large amount of the variation in AAM due to

462   genetic factors remains unexplained. Another limitation of this study is that the PGS represents

463   just one method of capturing the effects of common genetic variants that affect AAM. There may

464   be several pathways causing later AAM represented within these common genetic variants, and

465   using a single PGS to represent all those effects may obscure relationships with each individual

466   pathway. Future studies may identify these different pathways by methods such as clustering

467   analyses (36,37), which would then allow an estimation of multiple polygenic scores, each

468   representing a different pathway, to study their associations with risk of CAD and CAD risk

469   factors.

470

471   Distinguishing between sources of variation in AAM has provided a novel lens through which to

472   study associations of AAM with CAD and has allowed us to uncover differences in the

473   associations of genetically predicted vs. PGS-adjusted AAM with CAD and CAD risk factors.

474   Because later puberty in women, but not men, has been associated with an increased risk of

475   CAD, studying these differences further may provide unique insights into mechanisms that affect

476   CAD risk specifically in women.

21

487

REFERENCES

1. Canoy D, Beral V, Balkwill A, Wright FL, Kroll ME, Reeves GK, et al. Age at menarche and risks of coronary heart and other vascular diseases in a large UK cohort. Circulation [Internet]. 2015 [cited 2021 Dec 11];131(3):237–44. Available from: https://pubmed.ncbi.nlm.nih.gov/25512444/

2. Zhu J, Chan YM. Adult Consequences of Self-Limited Delayed Puberty. Pediatrics [Internet]. 2017 Jun 1 [cited 2021 Aug 29];139(6). Available from: www.aappublications.org/news

3. Day FR, Elks CE, Murray A, Ong KK, Perry JRB. Puberty timing associated with diabetes, cardiovascular disease and also diverse health outcomes in men and women: the UK Biobank study. Sci Rep [Internet]. 2015 Jun 18 [cited 2021 Dec 11];5. Available from: https://pubmed.ncbi.nlm.nih.gov/26084728/

4. Lee JJ, Cook-Wiens G, Johnson BD, Braunstein GD, Berga SL, Stanczyk FZ, et al. Age at Menarche and Risk of Cardiovascular Disease Outcomes: Findings From the National Heart Lung and Blood Institute-Sponsored Women's Ischemia Syndrome Evaluation. J Am Heart Assoc [Internet]. 2019 Jun 18 [cited 2024 Jun 10];8(12). Available from: https://www.ahajournals.org/doi/abs/10.1161/JAHA.119.012406

5. Santos MP, Li Y, Bazzano LA, He J, Rexrode KM, Ley SH. Age at menarche, type 2 diabetes and cardiovascular disease complications in US women aged under 65 years: NHANES 1999-2018. BMJ Nutr Prev Health [Internet]. 2023 Dec 1 [cited 2024 Apr 10];6(2):293–300. Available from: https://pubmed.ncbi.nlm.nih.gov/38264363/

6. Janghorbani M, Mansourian M, Hosseini E. Systematic review and meta-analysis of age at menarche and risk of type 2 diabetes. Acta Diabetol [Internet]. 2014 [cited 2024 Apr 10];51(4):519–28. Available from: https://pubmed.ncbi.nlm.nih.gov/24671509/

7. Stöckl D, Meisinger C, Peters A, Thorand B, Huth C, Heier M, et al. Age at menarche and its association with the metabolic syndrome and its components: results from the KORA F4 study. PLoS One [Internet]. 2011 Oct 18 [cited 2021 Dec 11];6(10). Available from: https://pubmed.ncbi.nlm.nih.gov/22028807/

8. Liu D, Qin P, Liu Y, Sun X, Li H, Wu X, et al. Association of age at menarche with hypertension in rural Chinese women. J Hypertens. 2021 Mar 1;39(3):476–83.

9. Kheradmand M, Hamzehgardeshi Z, Shahhosseini Z, Mirjalili R, Moosazadeh M. The association between early menarche and higher-risk cardiometabolic profile: a dose-response analysis of the Tabari cohort at enrollment phase. Front Cardiovasc Med [Internet]. 2023 [cited 2024 Apr 10];10. Available from: https://pubmed.ncbi.nlm.nih.gov/37719973/

10. Cao M, Cui B. Negative Effects of Age at Menarche on Risk of Cardiometabolic Diseases in Adulthood: A Mendelian Randomization Study. J Clin Endocrinol Metab [Internet]. 2020 Jan 23 [cited 2022 Feb 23];105(2). Available from: https://pubmed.ncbi.nlm.nih.gov/31614369/

11. Gill D, Brewer CF, Del Greco M F, Sivakumaran P, Bowden J, Sheehan NA, et al. Age at menarche and adult body mass index: a Mendelian randomization study. International Journal of Obesity

525   2018 42:9 [Internet]. 2018 Feb 26 [cited 2023 Feb 12];42(9):1574–81. Available from:
526   https://www-nature-com.ezp-prod1.hul.harvard.edu/articles/s41366-018-0048-7

527 12. Prentice P, Viner RM. Pubertal timing and adult obesity and cardiometabolic risk in women and
528   men: a systematic review and meta-analysis. Int J Obes (Lond) [Internet]. 2013 Aug [cited 2021
529   Dec 11];37(8):1036–43. Available from: https://pubmed.ncbi.nlm.nih.gov/23164700/

530 13. Yang L, Li L, Millwood IY, Lewington S, Guo Y, Sherliker P, et al. Adiposity in relation to age at
531   menarche and other reproductive factors among 300 000 Chinese women: findings from China
532   Kadoorie Biobank study. Int J Epidemiol [Internet]. 2017 Jan 1 [cited 2022 Mar 17];46(2):502–12.
533   Available from: https://pubmed.ncbi.nlm.nih.gov/27524817/

534 14. Chen L, Zhang L, Chen Z, Wang X, Zheng C, Kang Y, et al. Age at menarche and risk of
535   hypertension in Chinese adult women: Results from a large representative nationwide
536   population. J Clin Hypertens (Greenwich) [Internet]. 2021 Aug 1 [cited 2023 Dec 7];23(8):1615–
537   21. Available from: https://pubmed.ncbi.nlm.nih.gov/34255920/

538 15. Liu D, Zhao Y, Liu Y, Sun X, Li H, Yin Z, et al. Adiposity and insulin resistance as mediators between
539   age at menarche and type 2 diabetes mellitus. Menopause [Internet]. 2020 May 1 [cited 2022
540   Apr 1];27(5):579–85. Available from: https://pubmed-ncbi-nlm-nih-gov.ezp-
541   prod1.hul.harvard.edu/32068689/

542 16. Elks CE, Ong KK, Scott RA, Van Der Schouw YT, Brand JS, Wark PA, et al. Age at menarche and
543   type 2 diabetes risk: the EPIC-InterAct study. Diabetes Care [Internet]. 2013 Nov [cited 2022 Apr
544   1];36(11):3526–34. Available from: https://pubmed-ncbi-nlm-nih-gov.ezp-
545   prod1.hul.harvard.edu/24159179/

546 17. Schlomer GL, Marceau K. Father absence, age at menarche, and genetic confounding: A
547   replication and extension using a polygenic score. Dev Psychopathol [Internet]. 2022 Feb 27
548   [cited 2024 May 13];34(1):355–66. Available from: https://pubmed.ncbi.nlm.nih.gov/33107423/

549 18. Morris DH, Jones ME, Schoemaker MJ, Ashworth A, Swerdlow AJ. Familial concordance for age at
550   menarche: Analyses from the breakthrough generations study. Paediatr Perinat Epidemiol. 2011
551   May;25(3):306–11.

552 19. Towne B, Czerwinski SA, Demerath EW, Blangero J, Roche AF, Siervogel RM. Heritability of age at
553   menarche in girls from the Fels Longitudinal Study. Am J Phys Anthropol. 2005 Sep;128(1):210–9.

554 20. Day FR, Thompson DJ, Helgason H, Chasman DI, Finucane H, Sulem P, et al. Genomic analyses
555   identify hundreds of variants associated with age at menarche and support a role for puberty
556   timing in cancer risk. Nat Genet [Internet]. 2017 Jun 1 [cited 2022 Jul 26];49(6):834. Available
557   from: /pmc/articles/PMC5841952/

558 21. Ojalehto E, Zhan Y, Jylhävä J, Reynolds CA, Dahl Aslan AK, Karlsson IK. Genetically and
559   environmentally predicted obesity in relation to cardiovascular disease: a nationwide cohort
560   study. EClinicalMedicine [Internet]. 2023 Apr 1 [cited 2024 Jun 9];58. Available from:
561   http://www.thelancet.com/article/S2589537023001207/fulltext

562   22.   Hawkes G, Yengo L, Vedantam S, Marouli E, Beaumont RN, Tyrrell J, et al. Identification and
563   analysis of individuals who deviate from their genetically-predicted phenotype. PLoS Genet
564   [Internet]. 2023 Sep 21 [cited 2024 Jun 9];19(9). Available from:
565   https://pubmed.ncbi.nlm.nih.gov/37733769/

566   23.   Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access
567   resource for identifying the causes of a wide range of complex diseases of middle and old age.
568   PLoS Med [Internet]. 2015 Mar 1 [cited 2022 Mar 10];12(3). Available from: https://pubmed-
569   ncbi-nlm-nih-gov.ezp-prod1.hul.harvard.edu/25826379/

570   24.   Zhu J, Pujol-Gualdo N, Wittemans LBL, Lindgren CM, Laisk T, Hirschhorn JN, et al. Evidence From
571   Men for Ovary-independent Effects of Genetic Risk Factors for Polycystic Ovary Syndrome. J Clin
572   Endocrinol Metab [Internet]. 2022 Mar 24 [cited 2022 Jul 27];107(4):e1577–87. Available from:
573   https://academic-oup-com.ezp-prod1.hul.harvard.edu/jcem/article/107/4/e1577/6432124

574   25.   UK Biobank age at menarche data collection:
575   https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2714.

576   26.   Ridker PM, Chasman DI, Zee RYL, Parker A, Rose L, Cook NR, et al. Rationale, design, and
577   methodology of the Women's Genome Health Study: a genome-wide association study of more
578   than 25,000 initially healthy american women. Clin Chem [Internet]. 2008 [cited 2023 Nov
579   2];54(2):249–55. Available from: https://pubmed.ncbi.nlm.nih.gov/18070814/

580   27.   Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and
581   continuous shrinkage priors. Nat Commun [Internet]. 2019 Dec 1 [cited 2022 Dec 7];10(1).
582   Available from: https://pubmed.ncbi.nlm.nih.gov/30992449/

583   28.   Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. Gigascience
584   [Internet]. 2019 Jul 1 [cited 2022 Dec 7];8(7). Available from:
585   https://pubmed.ncbi.nlm.nih.gov/31307061/

586   29.   Sonawalla A, Chasman D, Chan YM. Hold for placement of supplemental materials.

587   30.   Perry JRB, Day F, Elks CE, Sulem P, Thompson DJ, Ferreira T, et al. Parent-of-origin specific allelic
588   associations among 106 genomic loci for age at menarche. Nature [Internet]. 2014 Oct 10 [cited
589   2024 Jun 17];514(7520):97. Available from: /pmc/articles/PMC4185210/

590   31.   Zheng J, Chen K, Huang T, Shao C, Li P, Wang J, et al. Genetically Determined Lifestyle and
591   Cardiometabolic Risk Factors Mediate the Association of Genetically Predicted Age at Menarche
592   With Genetic Predisposition to Myocardial Infarction: A Two-Step, Two-Sample Mendelian
593   Randomization Study. Front Cardiovasc Med [Internet]. 2022 Apr 25 [cited 2023 Apr 20];9.
594   Available from: /pmc/articles/PMC9081496/

595   32.   Ardissino M, Slob EAW, Carter P, Rogne T, Girling J, Burgess S, et al. Sex-Specific Reproductive
596   Factors Augment Cardiovascular Disease Risk in Women: A Mendelian Randomization Study.
597   Journal of the American Heart Association: Cardiovascular and Cerebrovascular Disease
598   [Internet]. 2023 Mar 7 [cited 2024 Nov 17];12(5):e027933. Available from:
599   https://pmc.ncbi.nlm.nih.gov/articles/PMC10111460/

600   33.   Bell JA, Carslake D, Wade KH, Richmond RC, Langdon RJ, Vincent EE, et al. Influence of puberty
601   timing on adiposity and cardiometabolic traits: A Mendelian randomisation study. PLoS Med
602   [Internet]. 2018 Aug 1 [cited 2022 Feb 23];15(8). Available from:
603   https://pubmed.ncbi.nlm.nih.gov/30153260/

604   34.   Huxley R, Mendis S, Zheleznyakov E, Reddy S, Chan J. Body mass index, waist circumference and
605   waist:hip ratio as predictors of cardiovascular risk--a review of the literature. Eur J Clin Nutr
606   [Internet]. 2010 Jan [cited 2024 May 30];64(1):16–22. Available from: www.nature.com/ejcn

607   35.   Sahakyan KR, Somers VK, Rodriguez-Escudero JP, Hodge DO, Carter RE, Sochor O, et al. Normal-
608   Weight Central Obesity: Implications for Total and Cardiovascular Mortality. Ann Intern Med
609   [Internet]. 2015 Dec 1 [cited 2024 Jun 9];163(11):827–35. Available from:
610   https://pubmed.ncbi.nlm.nih.gov/26551006/

611   36.   Chasman DI, Giulianini F, Demler O V., Udler MS. Pleiotropy-Based Decomposition of Genetic Risk
612   Scores: Association and Interaction Analysis for Type 2 Diabetes and CAD. Am J Hum Genet
613   [Internet]. 2020 May 7 [cited 2024 Jun 9];106(5):646–58. Available from:
614   https://pubmed.ncbi.nlm.nih.gov/32302534/

615   37.   Udler MS, Kim J, von Grotthuss M, Bonàs-Guarch S, Cole JB, Chiou J, et al. Type 2 diabetes genetic
616   loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft
617   clustering analysis. PLoS Med [Internet]. 2018 Sep 1 [cited 2024 Jun 9];15(9). Available from:
618   https://pubmed.ncbi.nlm.nih.gov/30240442/

619

620

621   **TABLES**

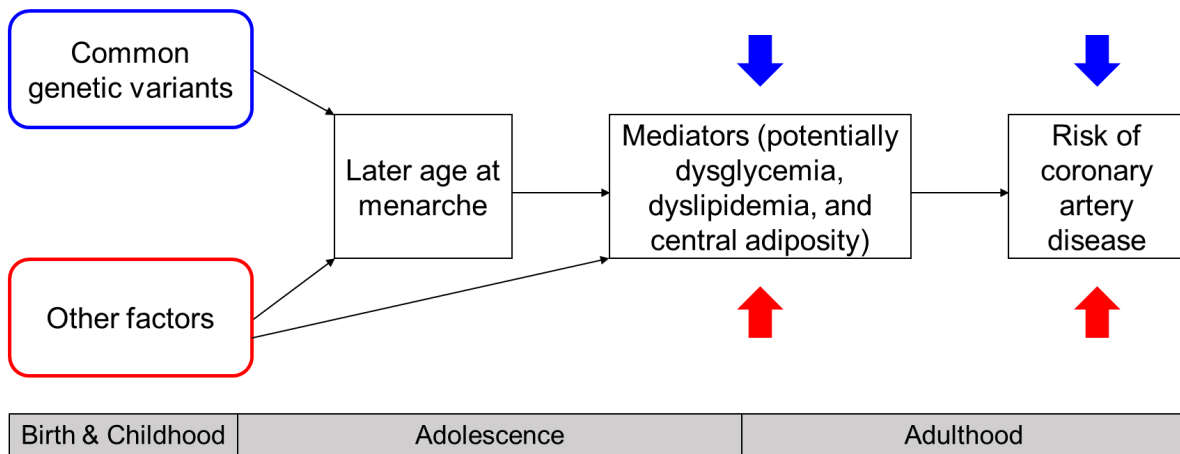| Coronary artery disease					
	Change per year increase in AAM				
	Values earlier than the mean		**Values later than the mean**		*p* for difference between slopes
	Odds ratio (95% CI)	*p*	Odds ratio (95% CI)	*p*	
Genetically predicted AAM	0.91 (0.83 to 0.99)	3×10^{-2}	0.92 (0.84 to 1)	0.063	0.87
PGS-adjusted AAM	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				
	Values earlier than the mean		**Values later than the mean**		*p* for difference between slopes
	Slope (95% CI)	*p*	Slope (95% CI)	*p*	
Hemoglobin A1c (%)					
Genetically predicted AAM	-0.034 (-0.041 to -0.027)	$<2\times10^{-16}$	-0.01 (-0.02 to -0.006)	3×10^{-3}	6×10^{-5}
PGS-adjusted AAM	-0.022 (-0.025 to -0.019)	$<2\times10^{-16}$	0.007 (0.004 to 0.01)	1.3×10^{-6}	$<2\times10^{-16}$
Triglycerides (mg/dL)					
Genetically predicted AAM	-2.82 (-3.84 to -1.79)	7×10^{-8}	-2.63 (-3.62 to -1.64)	2×10^{-7}	0.83
PGS-adjusted AAM	-4.25 (-4.7 to -3.8)	$<2\times10^{-16}$	1.2 (0.78 to 1.62)	2×10^{-8}	$<2\times10^{-16}$
High-density lipoprotein (HDL) cholesterol (mg/dL)					
Genetically predicted AAM	0.90 (0.69-1.11)	$<2\times10^{-16}$	0.74 (0.54-0.95)	8×10^{-13}	0.38

27

PGS-adjusted AAM	0.85 (0.76-0.95)	$<2\times10^{-16}$	-0.28 (-0.37 to -0.20)	2×10^{-10}	$<2\times10^{-16}$
Low-density lipoprotein (LDL) cholesterol (mg/dL)					
	Adjusted for LDL-lowering medications				
Genetically predicted AAM	0.27 (-0.15 to 0.69)	2×10^{-1}	-0.53 (-0.93 to -0.12)	1×10^{-2}	0.026
PGS-adjusted 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)				
Genetically predicted AAM	0.14 (-0.31 to 0.59)	0.054	-0.59 (-1.02 to -0.16)	8×10^{-3}	0.059
PGS-adjusted 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)					
Genetically predicted AAM	-0.73 (-0.99 to -0.47)	4×10^{-8}	-0.71 (-0.96 to -0.45)	5×10^{-8}	0.91
PGS-adjusted AAM	-0.69 (-0.81 to -0.58)	$<2\times10^{-16}$	-0.21 (-0.31 to -0.10)	2×10^{-4}	5×10^{-7}
Diastolic blood pressure (mmHg)					
Genetically predicted AAM	-0.49 (-0.63 to -0.34)	5×10^{-11}	-0.49 (-0.63 to -0.35)	9×10^{-12}	0.98
PGS-adjusted AAM	-0.52 (-0.59 to -0.46)	$<2\times10^{-16}$	-0.14 (-0.20 to -0.08)	5×10^{-6}	2×10^{-12}
Body-mass index (kg/m^2)					
Genetically predicted AAM	-0.98 (-1.05 to -0.91)	$<2\times10^{-16}$	-0.77 (-0.84 to -0.70)	$<2\times10^{-16}$	4×10^{-4}
PGS-adjusted AAM	-0.80 (-0.83 to -0.77)	$<2\times10^{-16}$	-0.079 (-0.11 to -0.051)	4×10^{-8}	$<2\times10^{-16}$
Waist-hip ratio					
Genetically predicted 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-adjusted AAM	-0.0046 (-0.0050 to -0.0042)	$<2\times10^{-16}$	0.0016 (0.0012 to 0.0020)	$<2\times10^{-16}$	$<2\times10^{-16}$

622

623   **Table 1:** Associations of genetically predicted and PGS-adjusted AAM with coronary artery

624   disease and its risk factors.

625   AAM: age at menarche; PGS: polygenic score

FIGURES

Figure 1: Relationships of different sources of variation in age at menarche (AAM) with

coronary artery disease (CAD) and CAD risk factors. Variation in AAM was dissected into

variation associated with common genetic variants (as estimated by a polygenic score for AAM)

and variation independent of the polygenic score, then associations of these sources of variation

with CAD and CAD risk factors were studied. When associated with common genetic variants,

later AAM was associated with a lower risk of CAD and favorable changes in CAD risk factors

(blue arrows). However, when occurring independently of the polygenic score, later AAM was

associated with higher risk of CAD and unfavorable changes in CAD risk factors (red arrows).

The "other factors" causing later AAM independent of the PGS have yet to be identified and

could potentially include chronic illness, underweight, and rare genetic variants.
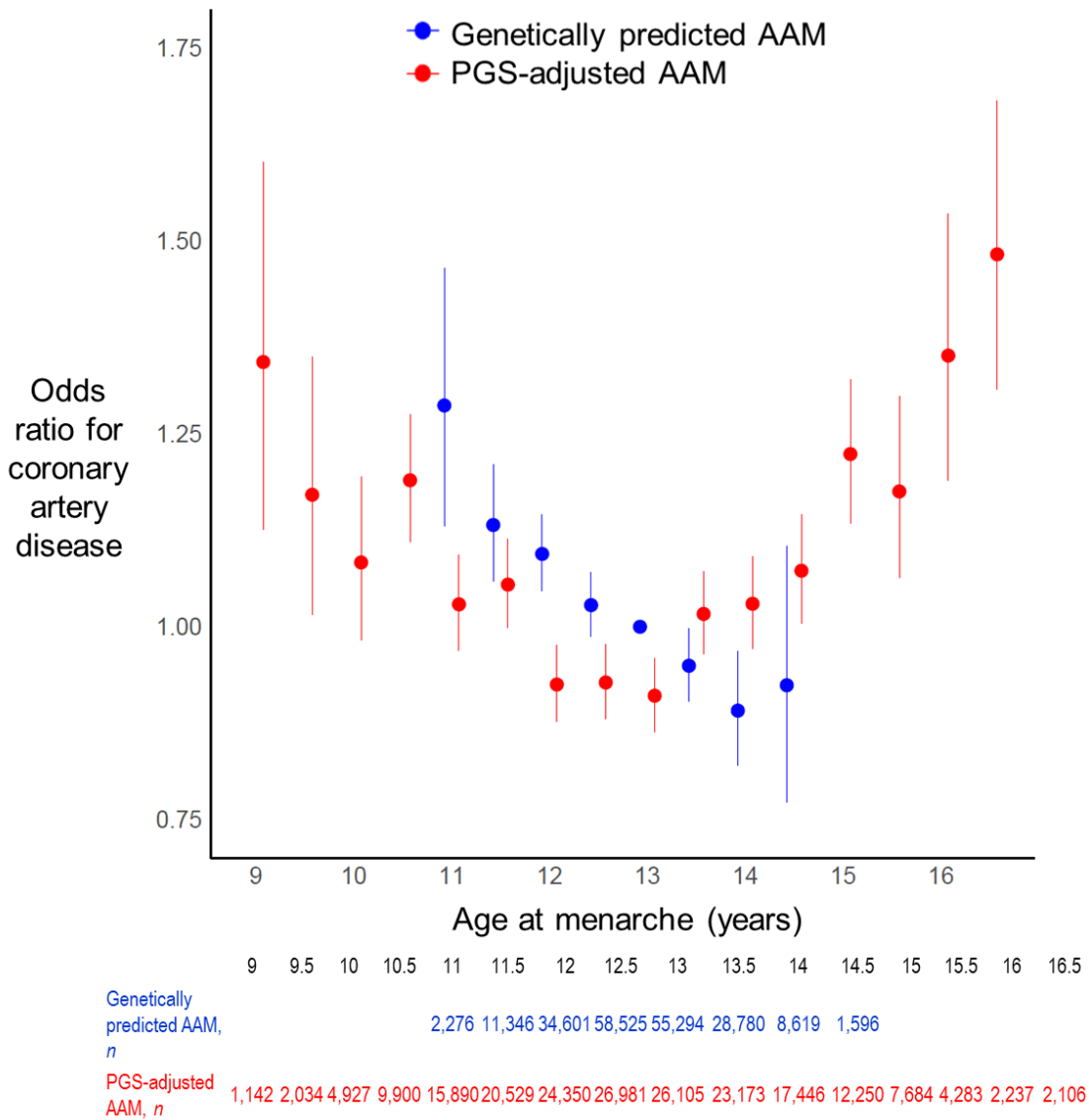
Figure 2: Associations of different sources of variation in age at menarche (AAM) with

odds of coronary artery disease (CAD) in the UK Biobank. For each individual in the cohort,

variation in AAM was dissected into variation attributable to common genetic variants as

estimated by a polygenic score (PGS) for AAM (represented by "genetically predicted AAM")

and variation attributable to sources other than common genetic variants (represented by "PGS-

adjusted AAM"), then associations with risk of CAD were studied. Genetically predicted AAM

showed a linear association with risk of CAD, whereas PGS-adjusted AAM showed a U-shaped

31

646   association. The different associations of genetically predicted AAM and PGS-adjusted AAM

647   with risk of CAD suggest that it's not later AAM itself that causes increased risk of CAD, but

648   that factors other than the PGS can cause both later AAM and increased risk of CAD.. To

649   achieve bin sizes of ≥1000 individuals, the first and last bins for each variable represent the

650   group of individuals with AAM ≤ or ≥ the value listed on the x-axis, respectively. Dots represent

651   estimates; bars represent standard errors.

652
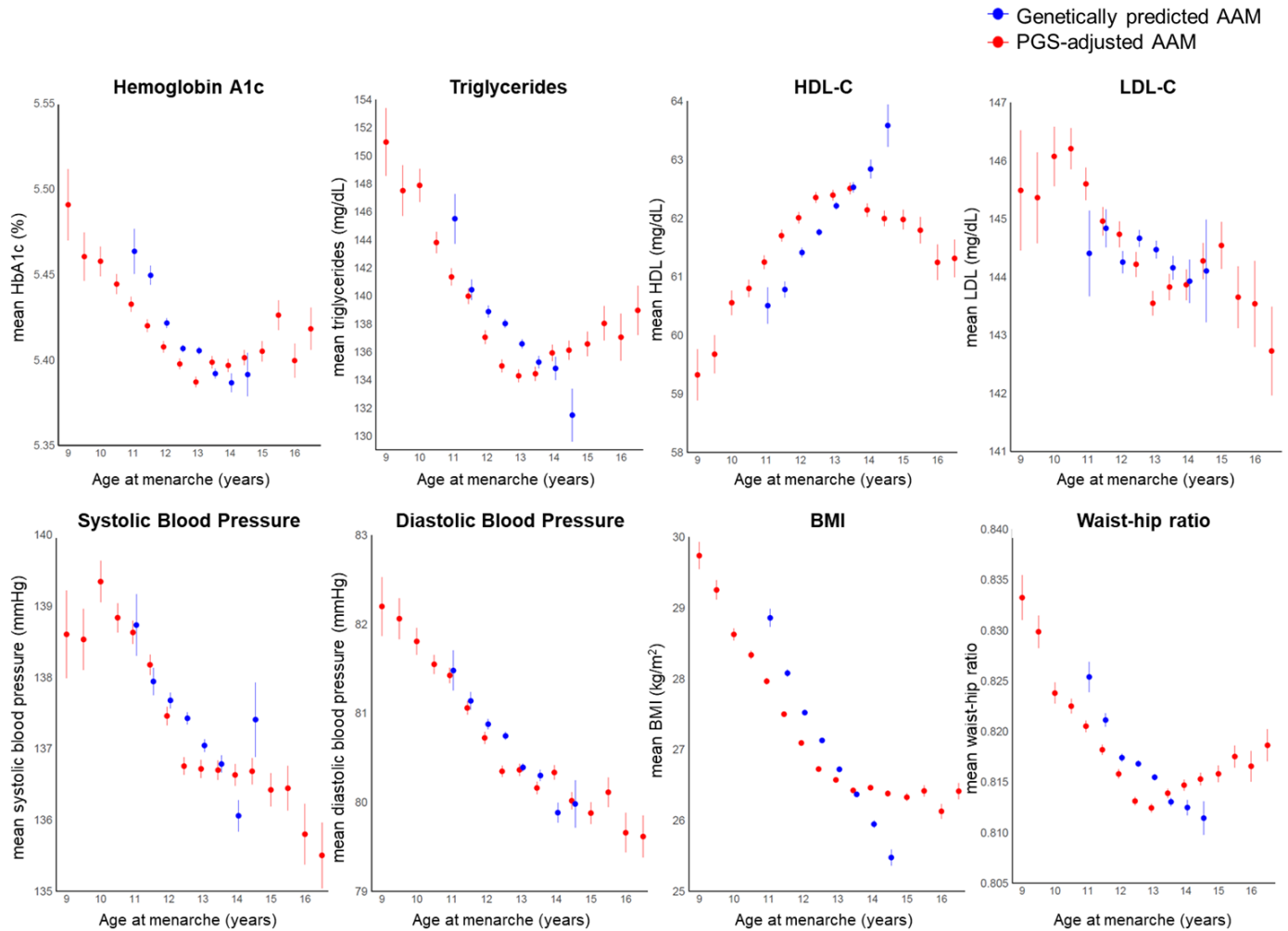
Figure 3: Associations of genetically predicted and PGS-adjusted AAM with risk factors for CAD in the UK Biobank. We observed linear (or roughly linear) relationships between genetically predicted AAM and most CAD risk factors. In contrast, for PGS-adjusted AAM we observed mostly non-linear relationships with CAD risk factors. The U-shaped relationship of PGS-adjusted AAM with CAD was mirrored by the relationships with HbA1c, triglycerides, HDL-C, and waist-hip ratio. This raises the possibility that dysglycemia, dyslipidemia and central adiposity may mediate the relationship between later AAM and increased risk of CAD. For LDL-C, n = 177,542 after excluding those taking LDL-C lowering medications. Dots represent estimates, bars represent standard errors.

663   AAM: age at menarche; BMI: body mass index; CAD: coronary artery disease; HDL-C: high-

664   density lipoprotein cholesterol; LDL-C: low-density lipoprotein cholesterol; PGS: polygenic

665   score
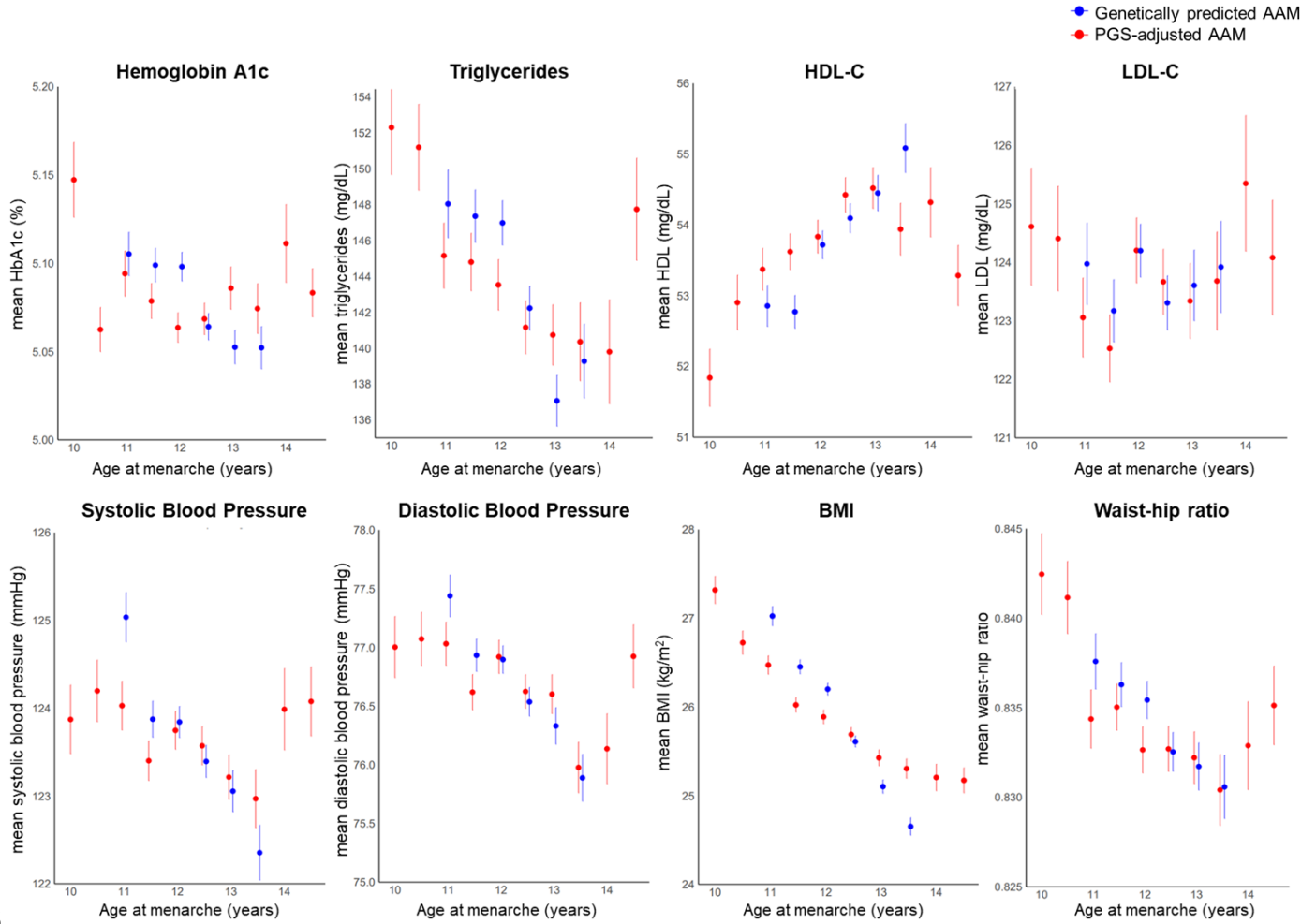
Figure 4: Associations of genetically predicted and PGS-adjusted AAM with risk factors for CAD in the WGHS. genetically predicted AAM showed linear or roughly linear associations with most CAD risk factors and PGS-adjusted AAM showed mostly non-linear associations with most CAD risk factors. Similar to results in the UK Biobank, hemoglobin A1c, triglycerides and HDL-C showed U-shaped associations with PGS-adjusted 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.

676   AAM: age at menarche; BMI: body mass index; CAD: coronary artery disease; HDL-C: high

677   density lipoprotein cholesterol; LDL-C: low density lipoprotein cholesterol; PGS: polygenic

678   score; WGHS: Women's Genome Health Study.
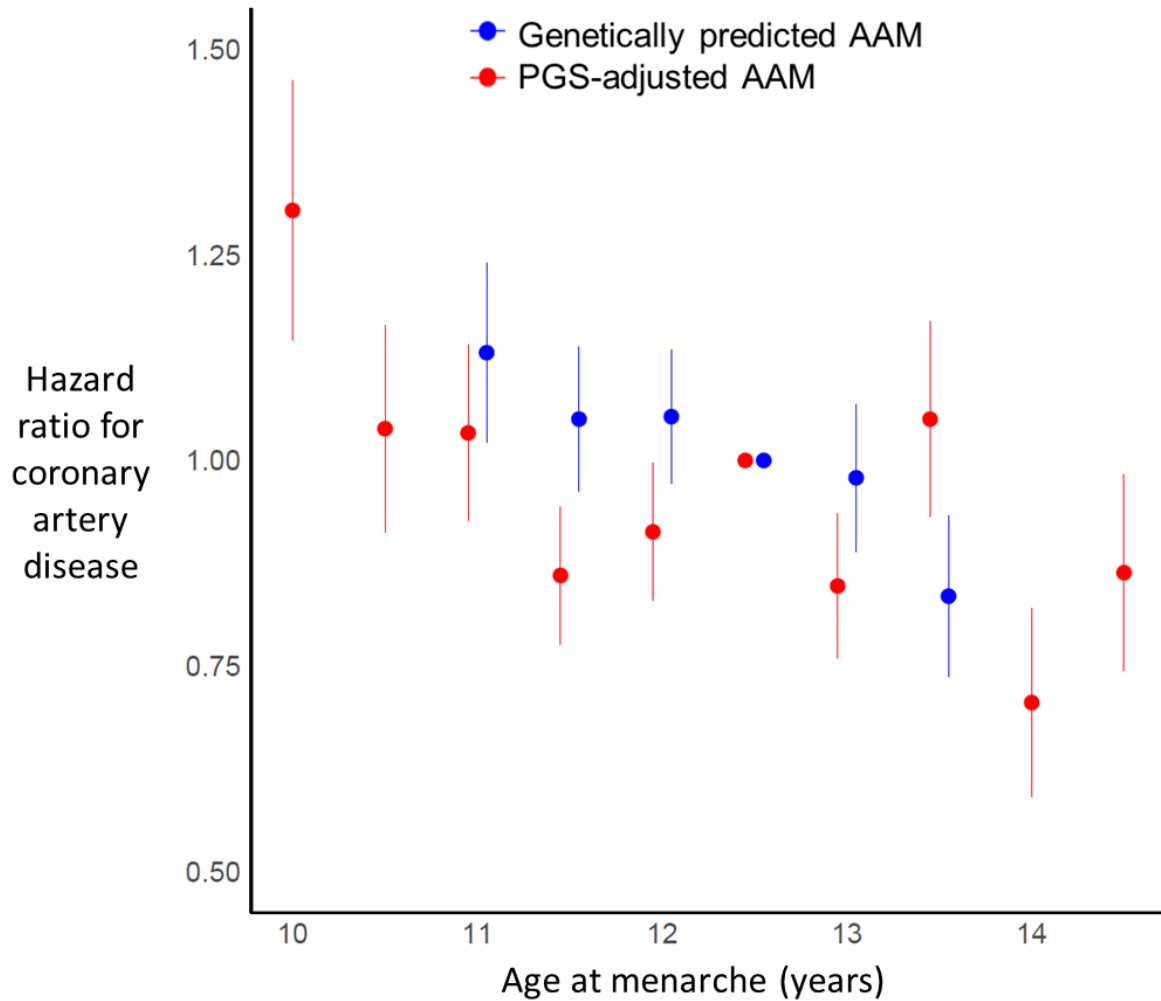
	10	10.5	11	11.5	12	12.5	13	13.5	14	14.5
Genetically predicted AAM, *n*			2,533	4,143	5,878	5,462	3,367	1,885		
PGS-adjusted AAM, *n*	1,275	1,557	2,445	3,527	4,070	3,826	1,691	2,779	938	1,160

679

Figure 5: Associations of genetically predicted and PGS-adjusted AAM with hazard ratio

for risk of CAD in the WGHS. Genetically predicted AAM demonstrated a negative linear

association with risk of CAD, similar to the results in the UKBB. PGS-adjusted AAM

demonstrated a reverse-J shaped relationship, slightly different from the U-shaped relationship

seen in the UK Biobank. Analyses in the WGHS largely validated the results from the UK

37

685   Biobank. The different associations of later PGS-adjusted AAM with risk of CAD may have

686   been due to differences in environmental factors between the two study cohorts.

687   AAM: age at menarche; ΔAAM: change in AAM; CAD: coronary artery disease; PGS:

688   polygenic score. Dots represent estimates, bars represent standard errors.