Age at menarche and cardiovascular risk factors using Mendelian randomization in the Guangzhou Biobank Cohort Study

Au Yeung, Shiu Lun; Jiang, Chaoqiang; Cheng, Kar Keung; Xu, Lin; Zhang, Weisen; Lam, Tai Hing; Leung, Gabriel Matthew; Schooling, C. Mary

DOI:
10.1016/j.ypmed.2017.06.006

Age at menarche and cardiovascular risk factors using Mendelian randomization in the Guangzhou Biobank Cohort Study

Shiu Lun Au Yeung, Chaoqiang Jiang, Kar Keung Cheng, Lin Xu, Weisen Zhang, Tai Hing Lam, Gabriel Matthew Leung, C. Mary Schooling

PII: S0091-7435(17)30210-4
DOI: doi: 10.1016/j.ypmed.2017.06.006
Reference: YPMED 5050

To appear in: *Preventive Medicine*

Received date: 10 January 2017
Revised date: 25 April 2017
Accepted date: 5 June 2017

Please cite this article as: Shiu Lun Au Yeung, Chaoqiang Jiang, Kar Keung Cheng, Lin Xu, Weisen Zhang, Tai Hing Lam, Gabriel Matthew Leung, C. Mary Schooling , Age at menarche and cardiovascular risk factors using Mendelian randomization in the Guangzhou Biobank Cohort Study, *Preventive Medicine* (2017), doi: 10.1016/j.ypmed.2017.06.006

Type of manuscript: Research Paper

Age at menarche and cardiovascular risk factors using Mendelian randomization in the Guangzhou Biobank Cohort Study

Shiu Lun Au Yeung[a], Chaoqiang Jiang[b], Kar Keung Cheng[c], Lin Xu[a], Weisen Zhang[b], Tai Hing Lam[a,b], Gabriel Matthew Leung[a], C Mary Schooling[a,d]

Affiliations: [a]School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China; [b]Guangzhou Number 12 Hospital, Guangzhou, China;[c]Department of Public Health and Epidemiology, University of Birmingham, UK; [d]City University of New York, Graduate School of Public Health and Health Policy, New York, NY, USA

Address correspondence to: Tai Hing Lam, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, G/F, Patrick Manson Building, 7 Sassoon Road, Hong Kong SAR, China, Phone: (852) 3917 9287, E-mail: hrmrlth@hku.hk

Abstract: 224

Main text: 3,475

Keywords: Menarche; cardiovascular risk factors; Chinese; Mendelian randomization

Table: 3

Appendix: 2

Email addresses:

Shiu Lun Au Yeung: ryanaysl@connect.hku.hk

Chaoqiang Jiang: cqjiang@hkucc.hku.hk

Kar Keung Cheng: k.k.cheng@bham.ac.uk

Lin Xu: linxu@hku.hk

Weisen Zhang: zwsgzcn@163.com

Tai Hing Lam: hrmrlth@hku.hk

Gabriel Matthew Leung: gmleung@hku.hk

C Mary Schooling: cms1@hkucc.hku.hk

Abstract

Observational studies show earlier age at menarche associated with higher risk of cardiovascular disease although these studies could be confounded by childhood obesity or childhood socioeconomic position. We hypothesized that earlier age at menarche is associated with poorer cardiovascular risk factors using a Mendelian randomization design. We conducted a Mendelian randomization study in a large Southern Chinese cohort, the Guangzhou Biobank Cohort Study (n=12,279), to clarify the causal role of menarche in cardiovascular disease risk factors including blood pressure, lipids, fasting glucose, adiposity and type 2 diabetes. A genetic allele score was obtained from single nucleotide polymorphisms associated with age at menarche using stepwise regression and with cross validation. Estimates of the association of age at menarche with cardiovascular disease risk factors were obtained using two stage least square regression. Height was included as a positive control outcome. The F-statistic for the allele score (rs17268785, rs1859345, rs2090409, rs4452860 and rs4946651) was 19.9. Older age at menarche was associated with lower glucose (-0.39 mmol/L per year older menarche, 95% confidence interval (CI) -0.78 to -0.001) but not clearly with any other cardiovascular risk factors. Older age at menarche was also associated with taller height. Age at menarche did not appear to affect cardiovascular disease risk factors except for glucose in an inverse manner. However, these results need to be confirmed in larger Mendelian randomization studies.

List of abbreviations

2SLS – 2 stage least squares

ANOVA – Analysis of variance

BGI – Beijing Genomics Institute

BMI – Body mass index

GBCS – Guangzhou Biobank Cohort Study

GHHARE – The Guangzhou Health and Happiness Association for the Respectable Elders

HDL – High density lipoprotein

LDL – Low density lipoprotein

SNP – Single nucleotide polymorphism

WHR – Waist hip ratio

Introduction

Earlier age at menarche is associated with a poorer cardiovascular profile in different settings (Chang et al., 2011; Feng et al., 2008; Gallagher et al., 2011; Heys et al., 2007; Lakshman et al., 2009). Younger age at menarche may be associated with cardiovascular disease, and its risk factors, as a result of particular childhood conditions prone to a higher cardiovascular risk in later life, such as physical inactivity and higher body mass index (BMI) in childhood (Chavarro et al., 2004; Morris et al., 2010) Younger age of menarche is also associated with higher adult BMI, so associations with cardiovascular disease might be mediated by adult BMI (Lakshman et al., 2009; Stockl et al., 2011). Alternatively, longer life time exposure to sex hormones might be a factor in cardiovascular disease, which is another consequence of earlier menarche (Howard and Rossouw, 2013; Vigen et al., 2013; Xu et al., 2013).

To reduce confounding by childhood BMI, examining the relation of age at menarche with cardiovascular risk in settings where childhood obesity is relatively recent, such as China, provides a possible way forward (Ji and Cooperative Study on Childhood Obesity: Working Group on Obesity in, 2008). Our previous observational study in China found young age of menarche associated with the metabolic syndrome in older women (Heys et al., 2007), although residual confounding by childhood socioeconomic position or infant infections is possible (McDonald et al., 2016). Mendelian randomization provides an alternative approach when randomized controlled trials are not available or not possible. Since genetic endowment is randomly allocated at conception, this is analogous to the randomization process in randomized controlled trials, and hence is less susceptible to confounding than observational studies (Lawlor

et al., 2008). To our knowledge no Mendelian randomization study has examined the role of age at menarche in cardiovascular risk factors in older populations. A small Mendelian randomization study from the US only focused on age at menarche and peri-pubertal body mass index (Johnson et al., 2013). A genetic study showed shared genetic determinants of pubertal timing and health outcomes, including cardiovascular disease and diabetes, but did not formally estimate the effect of age of menarche on these outcomes (Day et al., 2015). To address the long term effect of age at menarche on cardiovascular health in an under-studied population, we conducted a Mendelian randomization study using a large older Southern Chinese cohort to examine the relation of age at menarche with cardiovascular risk factors, including blood pressure, lipids, glucose and adiposity traits, using genetic instruments identified from previous genome wide association studies (Dvornyk and Waqar-ul-Haq, 2012). Given later age at menarche is consistently associated with taller height (Onland-Moret et al., 2005), we considered height as a positive control to rule out the possibility of underpowered analyses (Lipsitch et al., 2010).

Methods

The Guangzhou Biobank Cohort Study (GBCS) is an ongoing collaboration of Guangzhou Number 12 Hospital, the Universities of Hong Kong and Birmingham, UK.(Jiang et al., 2006) Recruitment of participants was in 3 phases. All participants were permanent residents of Guangzhou and members of "The Guangzhou Health and Happiness Association for the Respectable Elders" (GHHARE), a community social and welfare association unofficially aligned with the municipal government. Membership is open to older people for a monthly fee of 4 Yuan (50 US cents). About 7% of permanent Guangzhou residents aged 50+ years are

members of GHHARE, of whom 11% (about 10,000 participants) enrolled for each of phases one, two and three. The inclusion criteria were that they were capable of consenting, ambulatory, and not receiving treatment modalities which, if omitted, may result in immediate life-threatening risk, such as chemotherapy or radiotherapy for cancer, or dialysis for renal failure. The methods of measurement have previously been reported (Jiang et al., 2006). In brief, a medical interview was conducted using a standardized structured questionnaire, which covered socioeconomic position (including occupation, income, education), lifestyle (including smoking, physical activity and alcohol use), and health related questions (including past medical history). Age at menarche was recorded in years (as per the Gregorian calendar and related interpretation of age), rounded during data collection to the nearest year (e.g., 13 years represents the onset of menarche from 12 years 6 months to 13 years 5 months) (Heys et al., 2007). Physical examinations included measurement of seated blood pressure as the average of the last 2 of 3 measurements, using the Omron 705 CP sphygmomanometer (Omron Corp, Kyoto, Japan), measurement of standing height, without shoes, to the nearest 0.1 cm, measurement of weight (in light clothing) to the nearest 0.1 kg, measurement of hip circumference at the greatest circumference round the buttocks below the iliac crest and measurement of waist circumference horizontally around the smallest circumference between the ribs and iliac crest, or at the navel for obese participants. Low density lipoprotein (LDL) cholesterol, high density lipoprotein (HDL) cholesterol, triglycerides, and glucose were determined with a Shimadzu CL-8000 clinical chemical analyzer (Shimadzu Corp, Kyoto, Japan) in the hospital laboratory.

The Guangzhou Medical Ethics Committee of the Chinese Medical Association approved GBCS, including the use of genetic data. All participants gave written, informed consent prior to participation.

DNA extraction and single nucleotide polymorphisms (SNP) analysis

DNA was extracted at Guangzhou Number 12 Hospital from buffy coat previously stored at -80°C using a standard magnetic bead extraction procedure (MagPure Blood DNA Mini Kit). All DNA concentrations were checked by Nanodrop (Thermoscientific, USA). For DNAs concentrations lower than 15 ng/µl, silica-based column method was also used to re-extract DNA manually (HiPure Blood DNA Mini Kit). 92% of the DNA samples passed quality control before genotyping. Genotyping was performed using the MassARRAY Sequenom platform (San Diego, CA, USA) at the Beijing Genomics Institute (BGI), Beijing. The average genotyping call rate of these SNPs was 98%.

Instruments

Replication of genetic predictors of age of menarche identified in Europeans in East Asians was not available at the time of conduct of study. As such, we a-priori selected the 12 SNPs previously reported to be strongly associated with age at menarche which were variant in East Asian populations, including SNPs from *SPOCK* (rs13357391, rs1859345, rs2348186, rs7701979),*CCDC85A* (rs17268785), the 9q31.2 region (rs2090409, rs4452860, rs7861820) and *LIN28B* (rs314276, rs369065, rs4946651, rs7759938).(Dvornyk and Waqar-ul-Haq, 2012) From

these 12 SNPs a parsimonious set predicting age of menarche in our setting were selected using stepwise regression, and were combined into a genetic allele score to reduce the likelihood of weak instrument bias (Lawlor et al., 2008).

Exposure

The exposure was age at menarche, recorded in years (to the nearest 0.5 years).

Outcome

The outcomes were cardiovascular disease risk factors. These included systolic and diastolic blood pressure (mmHg), LDL cholesterol (mmol/L), HDL cholesterol (mmol/L), log transformed triglycerides, fasting glucose (mmol/L), BMI, calculated as weight /height^2 (kg/m^2), waist hip ratio (WHR) and type 2 diabetes. Type 2 diabetes was defined as fasting plasma glucose ≥7.0 mmol/L or use of anti-diabetic medication.

Control outcome

Height was included as a control outcome.

Statistical analysis

We tested for Hardy-Weinberg equilibrium at the SNP locus on a contingency table of observed-versus-predicted frequencies with an exact test. SNPs which deviated from the equilibrium were

discarded. Correlations between these SNPs were evaluated from the r^2 measure of linkage disequilibrium obtained from SNP Annotation and Proxy Search (http://www.broad.mit.edu/mpg/snap/ldsearchpw.php) using the available HapMap (release #22, JPT+CHB) reference. For SNPs in linkage disequilibrium ($r^2 \geq 0.8$) the SNP with a larger p-value and/or a smaller effect size was discarded. We used stepwise linear regression to find a parsimonious set of SNPs which best predicted age at menarche. The significance level was set at 0.20 to ensure that the initial inclusion criterion was not too restrictive, as used in our previous paper.(Zhao et al., 2014) To reduce the likelihood of false positive in the selection of SNPs based on only one overall sample we also used 10-fold cross validation (k=10).(Schonlau, 2005) The F statistic for the regression of age at menarche on genetic score (combined from the selected SNPs identified in the stepwise regression) was obtained, where a value ≥10 indicates that weak instrument bias is unlikely (Lawlor et al., 2008). We used analysis of variance (ANOVA) to assess whether age at menarche was associated with potential confounders. We also used ANOVA to assess whether the genetically estimated age at menarche was associated with potential confounders. We conducted instrumental variable analysis using 2 stage least squares (2SLS), with the genetic score as the instrument, to assess the association of age at menarche with continuous cardiovascular risk factors and with the log odds of diabetes. For comparison, we also present the multivariable linear and logistic regression analysis for the association of age at menarche with cardiovascular risk factors and height, adjusting for potential confounders including education, and recruitment phase (model 1).We also presented an additional model which we additionally adjusted for any cause of exposure or outcome as a potential confounding, including age, smoking, alcohol use, physical activity, job type, and corresponding medications such as antihypertensive for blood pressure (model 2) (VanderWeele and Shpitser, 2011).

All statistical analyses were conducted using Stata 13.1 (StataCorp LP, College Station, Texas, USA).

Results

Among 22,054 women in GBCS, 12,827 women had at least one SNP. rs369065 and rs2348186 deviated from Hardy Weinberg equilibrium ($p=0.04$ and $p<0.001$ respectively) and were discarded. rs13357391, rs1859345 and rs7701979; and rs4946651 and rs7759938 were highly correlated ($r^2 \geq 0.8$) although the strength of associations with age at menarche differed. Based on the observed p values and/or effect sizes, rs1859345 and rs4946651 were retained. rs314276 was not included as only 8,068 women (63%) had this SNP. From the stepwise regression with cross validation, rs17268785 (*CCDC85A*), rs1859345 (*SPOCK*), rs2090409 (9q31.2 region), rs4452860 (9q31.2 region) and rs4946651 (*LIN28B*) were identified as predicting age at menarche as these SNPs had a p value of ≤0.20 in at least 5 of the datasets (Appendix 1) and were used in the genetic allele score (F statistic 19.9, n=12,290). Table 1 shows the distribution of age at menarche across different sociodemographic and lifestyle factors. Women with older age at menarche tended to be older, current smokers, and never drinkers, as well as being less educated and having a manual job. Table 2 shows that genetically estimated age at menarche was not associated with age, smoking, physical activity, education, or job type. Although genetically estimated age at menarche was associated with alcohol use, the difference was very small.

Table 3 shows that observationally older age at menarche was associated with lower LDL cholesterol, log transformed triglycerides, glucose, BMI and risk of type 2 diabetes and was associated with higher waist hip ratio and taller height in model 1 which only included common causes of exposure and outcome. However, when additionally adjusted for any cause of exposure or outcome (model 2), older age at menarche was only associated with lower LDL cholesterol, log transformed triglycerides, body mass index, lower risk of type 2 diabetes and remained associated with taller height. In the Mendelian randomization analysis older age at menarche was associated with lower glucose (-0.39 mmol/L per year, 95%CI -0.78 to -0.001) but was not clearly associated with any other cardiovascular risk factor, although the wide confidence interval preclude definitive refutation. For example, the estimates for blood pressure, BMI and diabetes were all in the direction of older age of menarche being associated with lower point estimates, but the confidence intervals included null. Older age at menarche was associated with the positive control, i.e. height. Repeating the analyses using an allele score without rs2090409 and rs4452860, which did not replicate well in the cross validation (Appendix 1), showed directionally similar results although estimates had wider confidence intervals (Appendix 2). Repeating the analysis excluding samples with re-extraction (10% of the original samples have their DNA re-extracted) did not change the conclusion for Mendelian randomization (data not shown).

Discussion

This is the first Mendelian randomization study on age at menarche with cardiovascular disease risk factors where we found age at menarche inversely associated with glucose, and positively associated with height as expected.

The inverse association of age at menarche with cardiovascular risk factors has been consistently seen in different settings (Chang et al., 2011; Feng et al., 2008; Gallagher et al., 2011; Heys et al., 2007; Lakshman et al., 2009). Our findings suggest that some of the observed associations could be confounded such as childhood obesity, early life infections (Kwok et al., 2011), or other drivers of growth. A previous Mendelian randomization study (n=8,156) showed that girls with higher childhood BMI had a higher absolute risk of early menarche (<12 years) (Mumby et al., 2011). Childhood BMI is also a strong predictor of adult obesity where a recent meta-analysis showed that obese children are 5 times more likely to be obese in adulthood compared to non-obese children (Simmonds et al., 2016). As such, childhood BMI may confound the association of age at menarche with cardiovascular disease via its relation with adulthood BMI which in turn causes cardiovascular disease (Kivimaki et al., 2008; Nordestgaard et al., 2012). However, the women in our study grew up in China in the 1940s and 1950s during periods of significant hardship, making this an unlikely explanation for the inverse association in our observational analysis.. Similarly, a recent study from China showed that the association of age at menarche with cardiovascular disease was more evident in younger cohorts than older cohorts (Yang et al., 2017). Such differences could be explained by cohort effects but alternatively could be a result of confounding since more recent generations of Chinese women grew up in a more obesogenic environment, making the confounding effect of childhood obesity more influential. Early life

infections are associated with later age at menarche (McDonald et al., 2016), but are not thought to protect against cardiovascular disease risk factors.

Earlier uptake of unhealthy lifestyle with earlier puberty may explain findings from previous studies although such pathway was not supported by our previous study (Heys et al., 2007). Alternatively, earlier age at menarche generates greater exposure to endogenous sex hormones, which may explain the association of younger age of menarche with type 2 diabetes (He et al., 2010). Consistent with this explanation, both estrogen and younger age at menarche are associated with higher risk of breast cancer. (Collaborative Group on Hormonal Factors in Breast, 2012; Rossouw et al., 2002). However, contrary to this explanation, randomized controlled trials also showed that estrogen reduces the risk of diabetes (Margolis et al., 2004), which is inconsistent with an association of older age at menarche with lower glucose operating through such a pathway. Such discrepancies suggest the relation of age at menarche with cardiovascular risk may not simply be a reflection of the effects of lifelong estrogen exposure. Alternatively, some drivers of growth, such as growth hormone, may increase the risk of diabetes (Cutfield et al., 2000). Therefore, slower growth could underlie the inverse association of menarche and fasting glucose. Growth hormone is associated with lower blood pressure and LDL cholesterol in clinical trials (Maison et al., 2004), which suggests it should decrease cardiovascular risk, but these trials were among people with growth hormone deficiency which may not generalize to the general population.

The genetic predictors we used for age at menarche included SNPs from *CCDC85A*, *SPOCK*, 9q31.2 region and *LIN28B*. SNPs near *CCDC85A* have been associated with diabetes (Imamura et al., 2016). *SPOCK* is related to proteoglycan and inhibits matrix metalloproteinase -2 activations (Liu et al., 2009), but its function is not clearly understood. The 9q31.2 region is associated with breast cancer (Orr et al., 2015). *LIN28B* encodes a developmentally regulated RNA binding protein (He et al., 2009), and is associated with several cancers (Viswanathan et al., 2009). Given the function of these gene regions remains to be fully clarified, we may have included SNPs which have direct effects on glucose other than via menarche, and hence violated the exclusion restriction assumption. The SNP with the strongest effect on fasting glucose was rs1859345 (*SPOCK*) instead of rs17268785 (*CCDC85A*). Over 100 genetic loci have been associated with age at menarche, so it is possible that our genetic instrument did not encompass all the different domains, such as those in hormone synthesis and bioactivity, and energy homeostasis and growth (Perry et al., 2014), that age at menarche represents and hence only found an association with fasting glucose. However, we also found the expected relation for height.

We used a Mendelian randomization study which is less susceptible to confounding, nevertheless limitations exist. First, Mendelian randomization studies have stringent assumptions which cannot be empirically verified. However, the findings from this study are mainly null and null findings are potentially more reliable in the context of Mendelian randomization studies (VanderWeele et al., 2014). Age at menarche is invariant and is less susceptible to the violation of assumption where the predictors of the exposures only act on the outcomes through the exposure (exclusion-restriction assumption) compared to time varying exposures , such as body

mass index, that may directly affect the outcome at some times (VanderWeele et al., 2014). Age at menarche being invariant also means that genetic associations with age of menarche are less likely to represent correlations with other factors associated with the exposure that may accumulate over the lifespan. Although there remains a possibility of misclassification as age at menarche was based on recall, a previous study comparing recalled and original age at menarche showed reasonable correlations (Must et al., 2002). We used height as a positive control outcome and a large sample size to reduce the likelihood of false negatives in the main analyses (Freeman et al., 2013). However, some estimates had wide confidence intervals and hence larger Mendelian randomization studies are needed to verify our findings since there might be small cardiovascular effects of age at menarche which could still be meaningful from a public health perspective. Second, the SNPs predicting age at menarche were from Western populations but did not replicate well in our study (Dvornyk and Waqar-ul-Haq, 2012), as well as in another Asian study (Shi et al., 2016). This discrepancy may have arisen because of the difference between the populations studied in the environmental factors driving age at menarche, such as obesity. Although the small number of genetic variants was compensated for by a large sample size, as reflected by an F statistic of 19.9, weak instrument bias due to winner's curse may bias our estimates towards the confounded observational estimate (Lawlor, 2016). However, similar associations were seen when we only selected instruments after cross validation, suggesting the association seen in our analyses are not primarily driven by weak instrument bias. . Third, cardiovascular disease events are not high in our cohort (e.g. we only had 410 participants with self-reported ischemic heart disease among women included in this study) and are still accumulating so we could not draw a direct link between age at menarche and cardiovascular disease prospectively in this study, which would require a very large sample size with enough

cases given the low variance in age at menarche explained by genetics (2.7% for 123 SNPs) (Perry et al., 2014). The use of summary statistics from genome wide association studies may circumvent some of these limitations (Lawlor, 2016). However, existing genome wide association studies do not give sex-specific summary estimates and thus may not be suited for answering research questions which are sex specific, such as the health effects of age at menarche. Fourth, we cannot rule out the possibility that the association of age at menarche with cardiovascular risk has a U shape relation based on previous studies (Canoy et al., 2015; Yang et al., 2017). Although conducting Mendelian randomization by stratum of age at menarche may delineate potential non-linear relation, this would inevitably lead to imprecise stratum-specific estimates which could obscure the true exposure-outcome relation, as well as potentially generating weak instrumental biases due to reduced sample size in each stratified analysis (Burgess et al., 2014). Nevertheless, the potential U shape relation could be examined in future studies with larger cohorts with genetic data such as the UK Biobank or the China Kadoorie Biobank (Allen et al., 2014; Chen et al., 2005). Lastly, we did not have SNPs for all GBCS participants as DNA was only available for those who attended the follow up. However, our estimates would only be biased if the genotypes determined follow up, which is unlikely.

Conclusion

Our study found little evidence that age at menarche affects traditional cardiovascular disease risk factors except an inverse relation with glucose. As such, the secular trend of decreasing age at menarche and increasing cardiovascular disease in settings undergoing rapid economic development, such as Asia and Africa, or previous observational studies showing an association

between age at menarche and cardiovascular disease could be confounded by factors such as childhood obesity. Adequately powered Mendelian randomization studies with cardiovascular outcomes and diabetes will be needed to ascertain the relation of age at menarche with these outcomes allowing for the possibility that childhood obesity drives age at menarche.

Conflict of Interest

In addition to Kar Keung Cheng's appointment at University of Birmingham, he is affiliated to Department of General Practice at Peking University Health Science Centre. The latter receives support from Pfizer China to support the training of family doctors (approximately US$100,000 a year for 2014-16). The authors have indicated they have no other potential conflicts of interest to disclose.

Financial support

The Guangzhou Biobank Cohort Study was funded by the University of Hong Kong Foundation for Development and Research (Hong Kong, China); the University of Hong Kong University Research Committee Strategic Research Theme of Public Health (Hong Kong, China); Bureau of Guangzhou Science and Technology (Grant 2012J5100041; 2013J4100031), and the University of Birmingham (Birmingham, United Kingdom). This sub-study was funded by the Health and Medical Research Fund (#12132281), Food and Health Bureau, Hong Kong SAR, People's Republic of China. The funders had no role in the study design, data collection and analysis, the decision to publish, or preparation of the manuscript.

Author contributions

Shiu Lun Au Yeung and C Mary Schooling had substantial contributions to conception and design of this study, and analysis and interpretation of data, drafted the manuscript, revised it critically for important intellectual content, and approved the final manuscript as submitted.

Lin Xu helped with the design of this study, interpretation of the data, revised the manuscript critically for important intellectual content, and approved the final manuscript as submitted.

Chaoqiang Jiang, Tai Hing Lam, and Kar Keung Cheng had substantial contributions to conception and design of the original study, revised the manuscript critically for important intellectual content, and approved the final manuscript as submitted.

Gabriel Matthew Leung helped with the conception and design of the original study, revised the manuscript critically for important intellectual content, and approved the final manuscript as submitted.

Zhang Weisen helped with the acquisition of the original data, revised the manuscript critically for important intellectual content, and approved the final manuscript as submitted.

Acknowledgment

The Guangzhou Biobank Cohort Study investigators include: Guangzhou No. 12 Hospital--Dr. Zhang WS, Prof. Jiang CQ (Co-Principal Investigator (PI)); University of Hong Kong-Dr. C. M. Schooling, , Prof. R Fielding, Prof. GM Leung, Prof. TH Lam (Co-PI); University of Birmingham--Dr. G. N. Thomas, Prof. P Adab, Prof. KK Cheng (Co-PI).

References

Allen, N.E., Sudlow, C., Peakman, T., Collins, R., Biobank, U.K., 2014. UK biobank data: come and get it. Science translational medicine 6:224ed4.
Burgess, S., Davies, N.M., Thompson, S.G., Consortium, E.P.-I., 2014. Instrumental variable analysis with a nonlinear exposure-outcome relationship. Epidemiology 25:877-85.
Canoy, D., Beral, V., Balkwill, A., Wright, F.L., Kroll, M.E., Reeves, G.K., Green, J., Cairns, B.J., Million Women Study, C., 2015. Age at menarche and risks of coronary heart and other vascular diseases in a large UK cohort. Circulation 131:237-44.
Chang, H.S., Odongua, N., Ohrr, H., Sull, J.W., Nam, C.M., 2011. Reproductive risk factors for cardiovascular disease mortality among postmenopausal women in Korea: the Kangwha Cohort Study, 1985-2005. Menopause (New York, N.Y.) 18:1205-12.
Chavarro, J., Villamor, E., Narvaez, J., Hoyos, A., 2004. Socio-demographic predictors of age at menarche in a group of Colombian university women. Annals of human biology 31:245-57.
Chen, Z., Lee, L., Chen, J., Collins, R., Wu, F., Guo, Y., Linksted, P., Peto, R., 2005. Cohort profile: the Kadoorie Study of Chronic Disease in China (KSCDC). Int J Epidemiol 34:1243-9.
Collaborative Group on Hormonal Factors in Breast, C., 2012. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. Lancet Oncol 13:1141-51.
Cutfield, W.S., Wilton, P., Bennmarker, H., Albertsson-Wikland, K., Chatelain, P., Ranke, M.B., Price, D.A., 2000. Incidence of diabetes mellitus and impaired glucose tolerance in children and adolescents receiving growth-hormone treatment. Lancet 355:610-3.
Day, F.R., Bulik-Sullivan, B., Hinds, D.A., Finucane, H.K., Murabito, J.M., Tung, J.Y., Ong, K.K., Perry, J.R., 2015. Shared genetic aetiology of puberty timing between sexes and with health-related outcomes. Nature communications 6:8842.
Dvornyk, V., Waqar-ul-Haq, 2012. Genetics of age at menarche: a systematic review. Hum Reprod Update 18:198-210.
Feng, Y., Hong, X., Wilker, E., Li, Z., Zhang, W., Jin, D., Liu, X., Zang, T., Xu, X., 2008. Effects of age at menarche, reproductive years, and menopause on metabolic risk factors for cardiovascular diseases. Atherosclerosis 196:590-7.
Freeman, G., Cowling, B.J., Schooling, C.M., 2013. Power and sample size calculations for Mendelian randomization studies using one genetic instrument. Int J Epidemiol 42:1157-63.
Gallagher, L.G., Davis, L.B., Ray, R.M., Psaty, B.M., Gao, D.L., Checkoway, H., Thomas, D.B., 2011. Reproductive history and mortality from cardiovascular disease among women textile workers in Shanghai, China. Int J Epidemiol 40:1510-8.
He, C., Kraft, P., Chen, C., Buring, J.E., Pare, G., Hankinson, S.E., Chanock, S.J., Ridker, P.M., Hunter, D.J., et al., 2009. Genome-wide association studies identify loci associated with age at menarche and age at natural menopause. Nature genetics 41:724-8.
He, C., Zhang, C., Hunter, D.J., Hankinson, S.E., Buck Louis, G.M., Hediger, M.L., Hu, F.B., 2010. Age at menarche and risk of type 2 diabetes: results from 2 large prospective cohort studies. Am J Epidemiol 171:334-44.
Heys, M., Schooling, C.M., Jiang, C., Cowling, B.J., Lao, X., Zhang, W., Cheng, K.K., Adab, P., Thomas, G.N., et al., 2007. Age of menarche and the metabolic syndrome in China. Epidemiology 18:740-6.
Howard, B.V., Rossouw, J.E., 2013. Estrogens and cardiovascular disease risk revisited: the Women's Health Initiative. Curr Opin Lipidol.
Imamura, M., Takahashi, A., Yamauchi, T., Hara, K., Yasuda, K., Grarup, N., Zhao, W., Wang, X., Huerta-Chagoya, A., et al., 2016. Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes. Nature communications 7:10531.
Ji, C.Y., Cooperative Study on Childhood Obesity: Working Group on Obesity in, C., 2008. The prevalence of childhood overweight/obesity and the epidemic changes in 1985-2000 for Chinese school-

age children and adolescents. Obesity reviews : an official journal of the International Association for the Study of Obesity 9 Suppl 1:78-81.
Jiang, C., Thomas, G.N., Lam, T.H., Schooling, C.M., Zhang, W., Lao, X., Adab, P., Liu, B., Leung, G.M., et al., 2006. Cohort profile: The Guangzhou Biobank Cohort Study, a Guangzhou-Hong Kong-Birmingham collaboration. Int J Epidemiol 35:844-52.
Johnson, W., Choh, A.C., Curran, J.E., Czerwinski, S.A., Bellis, C., Dyer, T.D., Blangero, J., Towne, B., Demerath, E.W., 2013. Genetic risk for earlier menarche also influences peripubertal body mass index. Am J Phys Anthropol 150:10-20.
Kivimaki, M., Lawlor, D.A., Smith, G.D., Elovainio, M., Jokela, M., Keltikangas-Jarvinen, L., Vahtera, J., Taittonen, L., Juonala, M., et al., 2008. Association of age at menarche with cardiovascular risk factors, vascular structure, and function in adulthood: the Cardiovascular Risk in Young Finns study. Am J Clin Nutr 87:1876-82.
Kwok, M.K., Leung, G.M., Lam, T.H., Schooling, C.M., 2011. Early life infections and onset of puberty: evidence from Hong Kong's children of 1997 birth cohort. Am J Epidemiol 173:1440-52.
Lakshman, R., Forouhi, N.G., Sharp, S.J., Luben, R., Bingham, S.A., Khaw, K.T., Wareham, N.J., Ong, K.K., 2009. Early age at menarche associated with cardiovascular disease and mortality. J Clin Endocrinol Metab 94:4953-60.
Lawlor, D.A., 2016. Commentary: Two-sample Mendelian randomization: opportunities and challenges. Int J Epidemiol 45:908-15.
Lawlor, D.A., Harbord, R.M., Sterne, J.A.C., Timpson, N., Davey-Smith, G., 2008. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Stat Med 27:1133-63.
Lipsitch, M., Tchetgen Tchetgen, E., Cohen, T., 2010. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology 21:383-8.
Liu, Y.Z., Guo, Y.F., Wang, L., Tan, L.J., Liu, X.G., Pei, Y.F., Yan, H., Xiong, D.H., Deng, F.Y., et al., 2009. Genome-wide association analyses identify SPOCK as a key novel gene underlying age at menarche. PLoS Genet 5:e1000420.
Maison, P., Griffin, S., Nicoue-Beglah, M., Haddad, N., Balkau, B., Chanson, P., Metaanalysis of Blinded, R.P.-C.T., 2004. Impact of growth hormone (GH) treatment on cardiovascular risk factors in GH-deficient adults: a Metaanalysis of Blinded, Randomized, Placebo-Controlled Trials. J Clin Endocrinol Metab 89:2192-9.
Margolis, K.L., Bonds, D.E., Rodabough, R.J., Tinker, L., Phillips, L.S., Allen, C., Bassford, T., Burke, G., Torrens, J., et al., 2004. Effect of oestrogen plus progestin on the incidence of diabetes in postmenopausal women: results from the Women's Health Initiative Hormone Trial. Diabetologia 47:1175-87.
McDonald, J.A., Eng, S.M., Dina, O.O., Schooling, C.M., Terry, M.B., 2016. Infection and pubertal timing: a systematic review. Journal of developmental origins of health and disease 7:636-51.
Morris, D.H., Jones, M.E., Schoemaker, M.J., Ashworth, A., Swerdlow, A.J., 2010. Determinants of age at menarche in the UK: analyses from the Breakthrough Generations Study. Br J Cancer 103:1760-4.
Mumby, H.S., Elks, C.E., Li, S., Sharp, S.J., Khaw, K.T., Luben, R.N., Wareham, N.J., Loos, R.J., Ong, K.K., 2011. Mendelian Randomisation Study of Childhood BMI and Early Menarche. J Obes 2011:180729.
Must, A., Phillips, S.M., Naumova, E.N., Blum, M., Harris, S., Dawson-Hughes, B., Rand, W.M., 2002. Recall of early menstrual history and menarcheal body size: after 30 years, how well do women remember? Am J Epidemiol 155:672-9.
Nordestgaard, B.G., Palmer, T.M., Benn, M., Zacho, J., Tybjaerg-Hansen, A., Davey Smith, G., Timpson, N.J., 2012. The effect of elevated body mass index on ischemic heart disease risk: causal estimates from a Mendelian randomisation approach. PLoS Med 9:e1001212.
Onland-Moret, N.C., Peeters, P.H., van Gils, C.H., Clavel-Chapelon, F., Key, T., Tjonneland, A., Trichopoulou, A., Kaaks, R., Manjer, J., et al., 2005. Age at menarche in relation to adult height: the EPIC study. Am J Epidemiol 162:623-32.

Orr, N., Dudbridge, F., Dryden, N., Maguire, S., Novo, D., Perrakis, E., Johnson, N., Ghoussaini, M., Hopper, J.L., et al., 2015. Fine-mapping identifies two additional breast cancer susceptibility loci at 9q31.2. Hum Mol Genet 24:2966-84.
Perry, J.R., Day, F., Elks, C.E., Sulem, P., Thompson, D.J., Ferreira, T., He, C., Chasman, D.I., Esko, T., et al., 2014. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. Nature 514:92-7.
Rossouw, J.E., Anderson, G.L., Prentice, R.L., LaCroix, A.Z., Kooperberg, C., Stefanick, M.L., Jackson, R.D., Beresford, S.A., Howard, B.V., et al., 2002. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. JAMA 288:321-33.
Schonlau, M., 2005. Boosted regression (boosting): An intoductory tutorial and a Stata plugin. The Stata Journal 5:330-54.
Shi, J., Zhang, B., Choi, J.Y., Gao, Y.T., Li, H., Lu, W., Long, J., Kang, D., Xiang, Y.B., et al., 2016. Age at menarche and age at natural menopause in East Asian women: a genome-wide association study. Age (Dordr).
Simmonds, M., Llewellyn, A., Owen, C.G., Woolacott, N., 2016. Predicting adult obesity from childhood obesity: a systematic review and meta-analysis. Obesity reviews : an official journal of the International Association for the Study of Obesity 17:95-107.
Stockl, D., Meisinger, C., Peters, A., Thorand, B., Huth, C., Heier, M., Rathmann, W., Kowall, B., Stockl, H., et al., 2011. Age at menarche and its association with the metabolic syndrome and its components: results from the KORA F4 study. PLoS One 6:e26076.
VanderWeele, T.J., Shpitser, I., 2011. A new criterion for confounder selection. Biometrics 67:1406-13.
VanderWeele, T.J., Tchetgen Tchetgen, E.J., Cornelis, M., Kraft, P., 2014. Methodological challenges in mendelian randomization. Epidemiology 25:427-35.
Vigen, R., O'Donnell, C.I., Baron, A.E., Grunwald, G.K., Maddox, T.M., Bradley, S.M., Barqawi, A., Woning, G., Wierman, M.E., et al., 2013. Association of testosterone therapy with mortality, myocardial infarction, and stroke in men with low testosterone levels. JAMA 310:1829-36.
Viswanathan, S.R., Powers, J.T., Einhorn, W., Hoshida, Y., Ng, T.L., Toffanin, S., O'Sullivan, M., Lu, J., Phillips, L.A., et al., 2009. Lin28 promotes transformation and is associated with advanced human malignancies. Nature genetics 41:843-8.
Xu, L., Freeman, G., Cowling, B.J., Schooling, C.M., 2013. Testosterone therapy and cardiovascular events among men: a systematic review and meta-analysis of placebo-controlled randomized trials. Bmc Med 11.
Yang, L., Li, L., Millwood, I.Y., Peters, S.A., Chen, Y., Guo, Y., Bian, Z., Chen, X., Chen, L., et al., 2017. Age at menarche and risk of major cardiovascular diseases: Evidence of birth cohort effects from a prospective study of 300,000 Chinese women. Int J Cardiol 227:497-502.
Zhao, J., Jiang, C., Lam, T.H., Liu, B., Cheng, K.K., Xu, L., Au Yeung, S.L., Zhang, W., Leung, G.M., et al., 2014. Genetically predicted testosterone and cardiovascular risk factors in men: a Mendelian randomization analysis in the Guangzhou Biobank Cohort Study. Int J Epidemiol 43:140-8.

Table Legends

Table 1: Age at menarche by lifestyle and socio-demographics among 12,705 Southern Chinese older women in the Guangzhou Biobank Cohort Study (2003-8)

Table 2: Genetically estimated age at menarche by lifestyle and socio-demographics among 12,410 Southern Chinese older women in the Guangzhou Biobank Cohort Study (2003-8)

Table 3: Association of age at menarche (years) and cardiovascular risk factors, type 2 diabetes, and height among 12,692 Southern Chinese older women in the Guangzhou Biobank Cohort Study (2003-8) using Mendelian randomization and multivariable regression analysis

Appendix 1: P values for each single nucleotide polymorphism (SNP) in each training+validation pair in the 10-fold cross validation

Appendix 2: Association of age at menarche (years) and cardiovascular risk factors, type 2 diabetes, and height among 12,410 Southern Chinese older women in the Guangzhou Biobank Cohort Study (2003-8) using Mendelian randomization excluding less consistent single nucleotide polymorphisms (rs2090409 and rs4452860) in the cross validation

Table 1: Age at menarche by lifestyle and socio-demographics among 12,705 Southern Chinese older women in the Guangzhou Biobank Cohort Study (2003-8)

Characteristic		Sample size	Age at menarche (SD)	[a]*P* value
Age group	50-54	3,337	14.3 (1.8)	<0.001
	55-59	3,625	14.8 (1.9)	
	60-64	2,405	15.3 (2.1)	
	65-69	2,068	15.4 (2.0)	
	70-74	1,015	15.8 (2.1)	
	75-79	213	15.9 (2.2)	
	80+	42	15.4 (2.2)	
Smoking status	Never smokers	12,344	14.9 (2.0)	0.005
	Former smokers	167	15.5 (2.0)	
	Current smokers	180	15.0 (2.1)	
Alcohol status	Never drinkers	9,866	15.0 (2.0)	<0.001
	Former drinkers	330	14.9 (2.3)	
	Current drinkers	2,432	14.8 (2.0)	
[b]Physical activity (IPAQ)	Inactive	1,041	14.9 (2.1)	0.51
	Minimally active	4,890	15.0 (2.0)	
	HEPA active	6,774	15.0 (2.1)	
Education	Less than primary	1,271	16.2 (2.2)	<0.001
	Primary	4,219	15.4 (2.0)	
	Junior middle	3,448	14.7 (1.9)	
	Senior middle	2,977	14.3 (1.8)	
	Junior college	568	14.1 (1.7)	
	College	220	14.5 (2.1)	
[c]Longest-held occupation	Manual	8,411	15.1 (2.1)	<0.001
	Non-manual	2,463	14.6 (1.9)	
	Others	1,752	14.7 (1.9)	

[a]*P* value obtained from analysis of variance (ANOVA)

[b]HEPA: Health-enhancing physical activity (i.e. vigorous activity at least 3 days a week achieving at least 1500 metabolic equivalent (MET) minutes per week or activity on 7 days of the week, achieving at least 3000 MET minutes per week (IPAQ: International Physical Activity Questionnaire).

[c]Manual occupations are agricultural worker, factory work or sales and service; non-manual are administrator/ manager, professional/technical, military/disciplined.

Table 2: Genetically estimated age at menarche by lifestyle and socio-demographics among 12,410 Southern Chinese older women in the Guangzhou Biobank Cohort Study (2003-8)

Characteristic		Sample size	Genetically estimated age at menarche	[a]*P* value
Age group	50-54	3,271	14.96 (0.08)	0.59
	55-59	3,517	14.96 (0.08)	
	60-64	2,348	14.96 (0.08)	
	65-69	2,015	14.96 (0.08)	
	70-74	1,006	14.96 (0.08)	
	75-79	208	14.96 (0.09)	
	80+	45	14.94 (0.08)	
Smoking status	Never smokers	12,051	14.96 (0.08)	0.10
	Former smokers	163	14.95 (0.09)	
	Current smokers	181	14.95 (0.08)	
Alcohol status	Never drinkers	9,626	14.96 (0.08)	0.02
	Former drinkers	325	14.96 (0.08)	
	Current drinkers	2,382	14.95 (0.08)	
[b]Physical activity (IPAQ)	Inactive	1,014	14.96 (0.09)	0.55
	Minimally active	4,762	14.96 (0.08)	
	HEPA active	6,634	14.96 (0.08)	
Education	Less than primary	1,243	14.96 (0.08)	0.77
	Primary	4,136	14.96 (0.08)	
	Junior middle	3,355	14.96 (0.08)	
	Senior middle	2,912	14.96 (0.08)	
	Junior college	548	14.96 (0.09)	
	College	214	14.96 (0.08)	
[c]Longest-held occupation	Manual	8,233	14.96 (0.08)	0.66
	Non-manual	2,393	14.96 (0.08)	
	Others	1,704	14.96 (0.08)	

[a]*P* value obtained from analysis of variance (ANOVA)

[b]HEPA: Health-enhancing physical activity (i.e. vigorous activity at least 3 days a week achieving at least 1500 metabolic equivalent (MET) minutes per week or activity on 7 days of the week, achieving at least 3000 MET minutes per week (IPAQ: International Physical Activity Questionnaire).

[c]Manual occupations are agricultural worker, factory work or sales and service; non-manual are administrator/ manager, professional/technical, military/disciplined.

Table 3: Association of age at menarche (years) and cardiovascular risk factors, type 2 diabetes, and height among 12,692 Southern Chinese older women in the Guangzhou Biobank Cohort Study (2003-8) using Mendelian randomization and multivariable regression analysis

	Mendelian randomization			[a]Multivariable regression (Model 1)			[a]Multivariable regression (Model 2)		
	n	β	95% CI	n	β	95% CI	n	β	95% CI
Systolic blood pressure (mmHg)	12,270	-2.46	-7.40 to 2.48	12,682	0.11	-0.08 to 0.30	12,517	0.07	-0.10 to 0.25
Diastolic blood pressure (mmHg)	12,270	-1.62	-4.13 to 0.88	12,683	-0.06	-0.15 to 0.04	12,518	0.06	-0.03 to 0.16
HDL cholesterol (mmol/L)	12,256	0.02	-0.07 to 0.10	12,667	-0.002	-0.006 to 0.001	12,496	-0.002	-0.005 to 0.002
LDL cholesterol (mmol/L)	12,232	0.06	-0.10 to 0.21	12,642	-0.015	-0.02 to -0.008	12,471	-0.016	-0.02 to -0.009
Log transformed triglycerides	12,260	-0.05	-0.17 to 0.06	12,670	-0.01	-0.01 to -0.005	12,499	-0.01	-0.02 to -0.006
Fasting glucose (mmol/L)	12,229	-0.39	-0.78 to -0.001	12,640	-0.02	-0.03 to -0.002	12,468	-0.003	-0.02 to 0.009
Body mass index (kg/m^2)	12,270	-0.24	-0.96 to 0.49	12,683	-0.12	-0.15 to -0.09	12,524	-0.12	-0.15 to -0.09
Waist hip ratio	12,259	-0.003	-0.02 to 0.01	12,670	0.0009	0.0003 to 0.001	12,511	0.00	-0.0003 to 0.0008
Presence of diabetes (odds ratio)	12,231	0.65	0.32 to 1.33	12,642	0.94	0.91 to 0.97	12,484	0.92	0.89 to 0.95
Height (cm)	12,279	1.36	0.04 to 2.69	12,692	0.13	0.08 to 0.17	12,533	0.19	0.14 to 0.24

[a]Model 1 adjusted for education level and recruitment phase; Model 2 additionally adjusted for age, smoking, alcohol use, physical activity, job type, and corresponding medications such as antihypertensive for blood pressure.
HDL: High density lipoprotein; LDL: Low density lipoprotein

Appendix 1: P values for each single nucleotide polymorphism (SNP) in each training+validation pair in the 10-fold cross validation

	Training+Validation pair in the 10-fold cross validation										
SNP	1	2	3	4	5	6	7	8	9	10	Number of datasets which $P \leq 0.2$ for the corresponding SNP
rs17268785	**0.11**	**0.18**	0.30	**0.14**	**0.16**	**0.15**	**0.07**	**0.06**	**0.06**	**0.05**	9
rs1859345	**0.09**	**0.07**	0.28	**0.04**	**0.16**	**0.04**	**0.18**	**0.10**	**0.09**	**0.16**	9
rs2090409	**0.16**	0.23	0.21	**0.13**	0.46	**0.19**	**0.11**	0.36	0.56	**0.07**	5
rs4452860	**0.10**	0.21	**0.15**	**0.07**	0.31	**0.14**	**0.07**	0.27	0.38	**0.05**	6
rs4946651	**<0.001**	**<0.001**	**0.001**	**<0.001**	**0.002**	**0.001**	**0.001**	**0.001**	**<0.001**	**<0.001**	10
rs7861820	0.64	0.35	0.41	0.55	0.56	0.32	0.56	**0.20**	0.46	0.40	1

P value ≤0.20 were bolded

Appendix 2: Association of age at menarche (years) and cardiovascular risk factors, type 2 diabetes, and height among 12,410 Southern Chinese older women in the Guangzhou Biobank Cohort Study (2003-8) using Mendelian randomization excluding less consistent single nucleotide polymorphisms (rs2090409 and rs4452860) in the cross validation

	Excluding rs2090409 and rs4452860 (F statistics: 16.9)		
	n	β	95% CI
Systolic blood pressure (mmHg)	12,401	-2.40	-7.68 to 2.88
Diastolic blood pressure (mmHg)	12,401	-1.17	-3.79 to 1.45
High density lipoprotein cholesterol (mmol/L)	12,386	0.01	-0.08 to 0.10
Low density lipoprotein cholesterol (mmol/L)	12,361	0.06	-0.11 to 0.22
Log transformed triglycerides	12,390	-0.05	-0.17 to 0.07
Fasting glucose (mmol/L)	12,359	-0.43	-0.85 to -0.003
Body mass index (kg/m^2)	12,401	-0.28	-1.06 to 0.50
Waist hip ratio	12,389	-0.005	-0.02 to 0.01
Presence of diabetes (odds ratio)	12,361	0.60	0.27 to 1.30
Height (cm)	12,410	1.39	-0.03 to 2.82

Highlights

- Mendelian randomization (MR) study is more resistant to confounding.
- This MR study showed menarche inversely related to glucose.
- This MR study showed no strong evidence for menarche on blood pressure, lipids.
- Larger MR studies would be necessary to verify our findings.