Published text:

Familial concordance for age at menarche: analyses from the Breakthrough Generations study

Danielle H. Morris\textsuperscript{a}, Michael E. Jones\textsuperscript{a}, Minouk J. Schoemaker\textsuperscript{a}, Alan Ashworth\textsuperscript{b}, Anthony J. Swerdlow\textsuperscript{a}

\textsuperscript{a}Section of Epidemiology, Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK.

\textsuperscript{b}Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, London SW3 6JB, UK.

Abbreviated title: Familial concordance for age at menarche.

Address all correspondence to:

Miss Danielle Morris, Section of Epidemiology, Sir Richard Doll Building, Institute of Cancer Research, 15 Cotswold Rd, Sutton, Surrey SM2 5NG, UK.

E-mail: danielle.morris@icr.ac.uk

Tel: 0208 722 4192

Fax: 0208 722 4019
SUMMARY
Age at menarche is correlated within families, but estimates of the heritability of menarcheal age have a wide range (0.45-0.95). We examined the familial resemblance for age at menarche and the extent to which this is due to genetic and shared environmental factors. Between 2003 and 2010, data were retrospectively collected by questionnaire from participants within the UK-based Breakthrough Generations study. These analyses included 25,970 female participants aged 16-98 with at least one female relative who was also a study participant. A woman’s menarche was significantly delayed for each yearly increase in the menarcheal age of her monozygotic twin (average increase = 7.2 months, \( P < 0.001 \)), dizygotic twin (average increase = 3.0 months, \( P = 0.03 \)), older sister (average increase = 3.3 months, \( P < 0.001 \)), mother (average increase = 3.4 months, \( P < 0.001 \)), maternal grandmother (average increase = 1.5 months, \( P = 0.04 \)), maternal aunt (average increase = 1.4 months, \( P < 0.001 \)) and paternal aunt (average increase = 3.0 months, \( P < 0.001 \)). There was not a significant association between the menarcheal ages of half-sister pairs and of paternal grandmother-granddaughter pairs, based on small numbers. Heritability was estimated as 0.57 [95% confidence interval = 0.53 to 0.61]. Shared environmental factors did not have an effect in the model. In conclusion, approximately half of the variation in age at menarche was attributable to additive genetic effects with the remainder attributable to non-shared environmental effects.

Key terms: Adolescents; Female reproduction; Heritability; Menarche; Puberty.

Word count: Summary 237; Main text 2631.
Abbreviations: AIC, Akaike’s Information Criterion; BGS, Breakthrough Generations study; DZ, Dizygotic; MZ, Monozygotic; RMSEA, Root mean square error of approximation; SES, Socio-economic status.
INTRODUCTION

Age at menarche can have important implications for a woman’s health since it is associated with the risk of chronic diseases including breast cancer\(^1\) and type 2 diabetes.\(^2\) Therefore, better understanding of the determinants of age at menarche could improve knowledge of the aetiology of a number of diseases.

It is generally accepted that age at menarche is correlated within families and that this is partly due to genetic factors.\(^3\) It remains uncertain however to what extent genetic factors affect age at menarche since existing estimates of the narrow-sense heritability (\(h^2\); defined as the proportion of total phenotypic variance attributable to additive genetic effects) of age at menarche have a wide range between approximately 0.45 and 0.95.\(^3\) These estimates are based mostly on small studies that included at most a few thousand pairs.\(^3\)\(^-\)\(^7\)

A number of factors that affect the timing of menarche have been identified including childhood body size,\(^8\) exercise,\(^9\) socio-economic status (SES),\(^10\) and birth year.\(^11\) It is possible that the familial concordance for age at menarche is due in part to a tendency for family members to have similar environmental and behavioural experiences. Although most studies have found no evidence that shared environment and behaviour significantly influence familial resemblance for age at menarche,\(^3\)\(^-\)\(^6\)\(^,\)\(^7\) the majority of these studies comprised first degree relatives and so similarities between relatives that were due to shared environment or behaviour may have been incorrectly attributed to the genetic effect estimates.\(^3\)

We therefore investigated familial associations for menarcheal age in a large study, encompassing a wide range of familial relationships with many different combinations of genetic and environmental similarities.

METHODS

Description of the study

These analyses were based on participants of the Breakthrough Generations Study (BGS), a cohort study that began in 2003 primarily to investigate breast cancer aetiology. All women aged 16 or older who live in the United Kingdom are eligible and more than 111,000 have
joined. The initial recruits were registered supporters of the charity Breakthrough Breast Cancer, or were women who referred themselves through the study’s recruitment website and telephone line. Most participants nominated, and were nominated by, other women to join the BGS; as a result, a large number of BGS participants have other family members in the study. The study has been approved by the South Thames Multicentre Research Ethics Committee. The participants provided informed consent.

Subjects
Participants were potentially eligible for the current analyses if they had a first-degree relative (mother, daughter or sister) or second-degree relative (half-sister, grandmother, granddaughter, aunt or niece) who was also a BGS participant. Relatives were primarily identified through the nomination system and by asking participants to provide details of relatives who they believed had also joined the BGS.

We excluded those women who had a history of breast cancer or ductal carcinoma in situ (who were considerably over-represented due to the nature of the study), those who had never menstruated, and those whose age at menarche was missing in the questionnaire or occurred after age 20 years. The latter exclusion was used because these women often had their menses initiated by exogenous hormone treatment due to primary amenorrhoea. If only one family member remained eligible after these exclusions were applied, then this participant was not included in the analyses.

Variables
Most of the variables used in the analyses were self-reported by the participant on a questionnaire completed at study entry. The exception was the participant’s ACORN score which was used as a residential area measure of SES. An ACORN score ranging from 1 (highest) to 5 (lowest) and derived from Census and other information is assigned to each UK postcode, except those in the Isle of Man and the Channel Islands.
Age at menarche was reported in completed years. The questionnaire asked only about childhood exercise outside of school hours since exercise in UK schools is timetabled and so children generally participate in similar levels of exercise while at school. Weight and height at age seven years were reported in comparison with other girls of a similar age.

**Statistical analyses**

The familial resemblance for age at menarche was measured using linear regression in SAS 9.1, with the age at menarche of the younger relative as the outcome and the age at menarche of the older relative as the explanatory variable. The results of this show the estimated effect of a year’s delay of an older relative’s menarche on a woman’s own age at menarche. For twin pairs, it was not possible from our records to determine which twin was older, and so the twin who was younger at study entry was taken as the younger relative. Additionally, estimates were adjusted for both the subject’s and the relative’s weight at age seven (ordinal categories: ‘much thinner’, ‘a little thinner’, ‘about the same’, ‘a little heavier’, ‘much heavier’), height at age seven (ordinal categories: ‘much shorter’, a little shorter’, ‘about the same’, ‘a little taller’, ‘much taller’), average number of hours exercise per week during childhood (continuous), SES (ordinal categories: 1 to 5), and year of birth (continuous).

The effects of genetic and environmental factors on age at menarche were estimated using structural equation modelling. The model allowed for age at menarche to be determined by a combination of additive genetic factors (A; dominant genetic factors were assumed to be zero), childhood environmental factors that are common to sisters (C), environmental factors common to all members of the same family regardless of their relationship to one another (S), and unique childhood or adult environmental factors including random variation (E). It was assumed that A was correlated 1.0 for monozygotic (MZ) twins, 0.5 for other first degree relatives and 0.25 for second degree relatives. Using these assumptions the following equations were derived:

\[
\text{Correlation (MZ twins) = A + S + C + E}
\]
Correlation ( dizygotic/DZ twins, full sisters) = 0.5A + S + C + E
Correlation (mother-daughter) = 0.5A + S + E
Correlation (half-sisters) = 0.25A + S + C + E
Correlation (grandmother-granddaughter, aunt-niece) = 0.25A + S + E

The covariance in age at menarche for each type of relative was estimated in SAS 9.1, both with and without adjustment for the factors listed above. The covariances were then entered into the structural equation models in Mx.

The total variance (V) was estimated as $A^2 + S^2 + C^2 + E^2$ and $h^2$ was estimated as $A^2/V$.

The best fitting, most parsimonious model was selected as the model with the lowest Akaike’s information criterion (AIC), which penalises for the number of estimated parameters. The root mean square error of approximation (RMSEA) was also used as a measure of goodness of fit with an RMSEA of 0.05 or less indicating a very good fit.

**RESULTS**

There were 34,397 participants who had at least one first or second degree relative who was a BGS participant, and hence were potentially eligible for the analyses. Of these, 2085 women were excluded because they had a history of breast cancer or ductal carcinoma in situ, 3178 because their age at menarche was not reported, 18 because they had never menstruated, 5 because their age at menarche was older than 20 years, and subsequently 3141 because they no longer had an eligible relative in the study. Data were therefore analysed for 25,970 individuals.

The average age at study entry among these individuals was 46.4 years (range = 16 to 98 years) and nearly all described themselves as white (99.5%). Overall, the mean age at menarche was 12.7 years (standard deviation = 1.4 years, range = 7 to 19 years).

There were 8703 mother-daughter, 5843 non-twin sister, 95 MZ twin, 88 DZ twin, 157 half-sister, 232 grandmother-granddaughter and 2427 aunt-niece pairs (Table 1).

For each yearly increase in an MZ twin’s age at menarche, her co-twin’s menarche was delayed by an average of 7.2 months ($P < 0.001$). The resemblance was weaker for other
first-degree relatives but remained highly significant: a yearly increase in a DZ twin’s, older sister’s or mother’s age at menarche was associated with a 3.0 month ($P = 0.03$), 3.3 month ($P < 0.001$) and 3.4 month ($P < 0.001$) delay in menarche, respectively.

The associations for second-degree relatives were weaker than those of first-degree relatives. For each one year delay of the maternal grandmother’s menarche, the granddaughter’s menarche was 1.5 months later on average ($P = 0.04$). A yearly increase in a paternal or maternal aunt’s age at menarche was associated with an average delay of 3.0 months ($P < 0.001$) and 1.4 months ($P < 0.001$), respectively. Based on small numbers, the ages at menarche of half-sisters ($P = 0.78$) and of paternal grandmothers and granddaughters ($P = 0.24$) were not significantly associated with each other. Furthermore, the menarcheal ages of neither maternal nor paternal half sisters were significantly associated with each other when analysed separately (data not shown).

After adjustment for known risk factors for age at menarche, most of the associations were attenuated slightly, except for the associations between the menarcheal ages of DZ twins and of maternal aunts and nieces, which increased slightly. The associations that were significant prior to adjustment remained significant after adjustment, with the exception of the association between the menarcheal ages of maternal grandmother’s and granddaughter’s which was no longer significant.

In unadjusted analyses, the best fitting model for the variation in age at menarche included terms for additive genetic variance ($A$) and unique environmental ($E$) factors (AIC = 62.1, RMSEA = 0.04, Table 2). In this model, $h^2$ was 0.57 [95% confidence interval = 0.53 to 0.61] and the proportion of variance attributable to unique environmental factors was 0.43 [95% confidence interval = 0.40 to 0.46]. Total ($S$) and childhood ($C$) shared environmental or behavioural factors did not have a significant effect and were not included in the best fitting model. The AE model was also the best fitting model in adjusted analyses with a similar heritability estimate ($h^2 = 0.54$, 95% confidence interval = 0.50 to 0.57).
DISCUSSION

Accurate estimation of heritability is an important first step in genetic epidemiological analysis of a given trait because such studies are costly and time-consuming. Furthermore, the proportion of total heritability that is explained by the known loci can be estimated and this might indicate the extent to which further loci are involved.\(^{17}\)

Our data suggest that 57\% of the variation in age at menarche in our population can be attributed to additive genetic effects. Our finding that the resemblance for age at menarche between relatives increased with the degree of genetic similarity also accords with a genetic effect.

Previous estimates of the heritability of age at menarche from small studies range between approximately 45\%-95\%, but the larger of these studies tend to estimate heritability to be between 50\%-70\%,\(^{3, 5, 18}\) as do those that included second and higher degree relatives.\(^{3}\) Our heritability estimate, based on larger numbers than any of the previous studies, is therefore consistent with the previous literature, although at the lower end of the range. This might be because age at menarche was reported in whole years in our study but in years and months in some other studies;\(^{3-5, 18}\) these studies might therefore have been able to better explain more of the total variance, although only if women are able to recall their age at menarche with such precision.

Heritability estimates should be interpreted with caution since they explain the phenotypic variation in a particular study population and so there are limitations on the extent to which these estimates can be generalised to other populations. Participants in our study were volunteer recruits, and so could potentially have been less heterogeneous in terms of environmental and behavioural factors than in a population-based random sample of women. All women living in the UK over the age of 16 are however eligible to join the BGS, and in practice they have done so from the full range of geographical regions, socio-economic groups and ages, and so a reasonable cross-section of the UK population are represented in
In confidence this study. Hence our result could potentially be applied to a wider population than those from most other studies.

In recent years, several loci that are associated with age at menarche have been identified, but these explain at most ~2% of the variation in menarcheal age (assuming that they act additively). This means that the vast majority of the heritability of age at menarche is not due to the identified loci. Previous genome-wide association studies have included tens of thousands of women and so it seems likely that the remaining loci are either rare or explain a very small amount of variation. Nevertheless, research to find other loci remains important.

It is possible that a parent-of-origin effect acts on age at menarche, because there was greater concordance between the ages at menarche of paternal aunt-niece pairs than maternal aunt-niece pairs. However, we are only able to speculate about the presence of such an effect because the confidence intervals for the paternal and maternal aunt effects overlap and we had few data for paternal grandmothers and for half-sisters and so were unable to determine whether this difference was present for other familial relationships.

It was expected that shared environment would influence the familial similarity for age at menarche because some environmental risk factors for age at menarche, such as body size, are likely to cluster in families. However, the effects of shared environment were not significant in the structural equation models, which accords with findings from some, but not all, previous studies. The interpretation of this finding is difficult because it is a necessary simplification in the model that environmental factors are assumed to be either common to all relatives, or shared only by sisters, or solely individual, rather than the infinite possibilities in between. These assumptions may not be valid since, for example, it is plausible that DZ twins share more childhood experiences than non-twin sisters. There was however no evidence that DZ twins were more similar for age at menarche than non-twin sisters.

A limitation of this study is that it relies on recalled age at menarche and risk factors, and although it has been shown that women are able to accurately recall their age at menarche, this could affect comparisons between generations if recall is associated with age.
We used subjective measures of body size because it is difficult to obtain accurate childhood body size measurements retrospectively from such a large cohort; nevertheless, these measurements had a strong inverse association with age at menarche; i.e. heavier girls reached menarche much earlier than their lighter peers.

Each two year delay of menarche is associated with an estimated 10% reduction in relative risk of breast cancer. Our finding that approximately half of the variation in age at menarche is due to additive genetic factors suggests that a material proportion of the breast cancer risk is attributable to genetic factors determining menarcheal age. However, these would appear to be different genes, yet to be discovered, than the currently known breast cancer loci, which are not associated with age at menarche.

In conclusion, using data from a large number of related pairs with different combinations of shared genetic, environmental and childhood factors, we found that approximately half of the variation in age at menarche was attributable to genetic effects with the remainder attributable to unique (i.e. non-shared) environmental effects. There was no evidence that shared environmental factors had a significant effect on the familial resemblance for age at menarche.

ACKNOWLEDGEMENTS

This work was supported by Breakthrough Breast Cancer, the Sir John Fisher foundation, and the Institute of Cancer Research, who acknowledge funding to the National Institute for Health Research Biomedical Research Centre.

We thank the women who participated in the study, and our colleagues in the Breakthrough Generations Study Team, in particular Dr. Nicholas Orr for his helpful comments on the manuscript.
Table 1. Average change in age at menarche for each year delay of menarche in an older relative.

<table>
<thead>
<tr>
<th>Type of relative</th>
<th>Number of pairs</th>
<th>Unadjusted Change, 95% confidence interval</th>
<th>Adjusted change, 95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Change, months</td>
<td>95% confidence interval</td>
</tr>
<tr>
<td>Monzygotic twin&lt;sup&gt;b&lt;/sup&gt;</td>
<td>95</td>
<td>7.16 [5.28, 9.04] ***</td>
<td>6.83 [4.76, 8.90] ***</td>
</tr>
<tr>
<td>Dizygotic twin&lt;sup&gt;b&lt;/sup&gt;</td>
<td>88</td>
<td>3.03 [0.25, 5.81] *</td>
<td>3.19 [0.61, 5.76] *</td>
</tr>
<tr>
<td>Full sister</td>
<td>5843</td>
<td>3.33 [3.05, 3.61] ***</td>
<td>3.10 [2.81, 3.39] ***</td>
</tr>
<tr>
<td>Half sister</td>
<td>157</td>
<td>-0.24 [-1.96, 1.48]</td>
<td>0.01 [-1.59, 1.92]</td>
</tr>
<tr>
<td>Paternal grandmother</td>
<td>23</td>
<td>-2.52 [-6.81, 1.77]</td>
<td>-1.81 [-6.03, 2.42]</td>
</tr>
<tr>
<td>Maternal grandmother</td>
<td>209</td>
<td>1.51 [0.04, 2.98] *</td>
<td>1.34 [-0.22, 2.90]</td>
</tr>
<tr>
<td>Maternal aunt</td>
<td>2044</td>
<td>1.39 [0.88, 1.90] ***</td>
<td>1.42 [0.89, 1.96] ***</td>
</tr>
</tbody>
</table>

* P < 0.05; *** P < 0.001.

<sup>a</sup> Adjusted for both the proband’s and the relative’s weight at age 7, height at age 7, childhood exercise, socio-economic status, and year of birth.

<sup>b</sup> The twin who was older at study entry was arbitrarily taken as the predictor relative.
Table 2. Structural equation models for variation in age at menarche.

A. Unadjusted models.

<table>
<thead>
<tr>
<th>Model</th>
<th>Additive genetic (A)</th>
<th>Shared environment (S)</th>
<th>Childhood environment (C)</th>
<th>Unique environment (E)</th>
<th>AIC&lt;sup&gt;b&lt;/sup&gt;</th>
<th>RMSEA&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASCE</td>
<td>0.58</td>
<td>0.00</td>
<td>0.00</td>
<td>0.42</td>
<td>65.1</td>
<td>0.04</td>
</tr>
<tr>
<td>ASE</td>
<td>0.58</td>
<td>0.00</td>
<td></td>
<td>0.42</td>
<td>63.1</td>
<td>0.04</td>
</tr>
<tr>
<td>ACE</td>
<td>0.58</td>
<td></td>
<td>0.00</td>
<td>0.42</td>
<td>63.1</td>
<td>0.04</td>
</tr>
<tr>
<td>SCE</td>
<td></td>
<td>0.26</td>
<td>0.02</td>
<td>0.72</td>
<td>146.4</td>
<td>0.09</td>
</tr>
<tr>
<td>AE&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.57</td>
<td></td>
<td></td>
<td></td>
<td>0.43</td>
<td>62.1</td>
</tr>
<tr>
<td>SE</td>
<td></td>
<td>0.26</td>
<td></td>
<td>0.74</td>
<td>146.8</td>
<td>0.09</td>
</tr>
<tr>
<td>CE</td>
<td></td>
<td></td>
<td>0.28</td>
<td>0.72</td>
<td>842.4</td>
<td>0.10</td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>1319.2</td>
<td>0.12</td>
</tr>
</tbody>
</table>

B. Adjusted models.<sup>d</sup>

<table>
<thead>
<tr>
<th>Model</th>
<th>Additive genetic (A)</th>
<th>Shared environment (S)</th>
<th>Childhood environment (C)</th>
<th>Unique environment (E)</th>
<th>AIC&lt;sup&gt;b&lt;/sup&gt;</th>
<th>RMSEA&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASCE</td>
<td>0.54</td>
<td>0.00</td>
<td>0.00</td>
<td>0.46</td>
<td>47.3</td>
<td>0.03</td>
</tr>
<tr>
<td>ASE</td>
<td>0.54</td>
<td>0.00</td>
<td></td>
<td>0.46</td>
<td>45.3</td>
<td>0.03</td>
</tr>
<tr>
<td>ACE</td>
<td>0.54</td>
<td></td>
<td>0.00</td>
<td>0.46</td>
<td>45.3</td>
<td>0.03</td>
</tr>
<tr>
<td>SCE</td>
<td></td>
<td>0.24</td>
<td>0.03</td>
<td>0.74</td>
<td>108.4</td>
<td>0.07</td>
</tr>
<tr>
<td>AE&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.54</td>
<td></td>
<td></td>
<td></td>
<td>0.46</td>
<td>43.3</td>
</tr>
<tr>
<td>SE</td>
<td></td>
<td>0.25</td>
<td></td>
<td>0.75</td>
<td>109.3</td>
<td>0.07</td>
</tr>
<tr>
<td>CE</td>
<td></td>
<td></td>
<td>0.26</td>
<td>0.74</td>
<td>707.1</td>
<td>0.09</td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>1138.1</td>
<td>0.11</td>
</tr>
</tbody>
</table>
a Variables that were included in the model. A = additive genetic factors, S = shared environment, C = childhood environment, E = unique environment.

b AIC = Akaike’s information criterion; RMSEA = Root mean square error of approximation.

c Best fitting model.

d The models were adjusted for both the proband’s and the relative’s weight at age seven years, height at age seven years, childhood exercise, socio-economic status, and year of birth.
Reference List


