Social science genetics and fertility

Tropf, Felix

4. Mega-analysis of 31,396 individuals from 6 countries uncovers strong gene-environment interaction for human fertility[8]

Abstract: Family and twin studies suggest that up to 50 percent of individual differences in human fertility within a population might be heritable. However, it remains unclear whether the genes associated with fertility outcomes such as number of children ever born (NEB) or age at first birth (AFB) are the same across geographical and historical environments. By not taking into account this possibility, previous genetic studies implicitly assumed that the genetic effects are constant across time and space. We conducted a mega-analysis by applying whole genome methods on 31,396 unrelated men and women from six Western countries. Across all individuals and environments, common single-nucleotide polymorphisms (SNPs) explained only ~4 percent of the variance in NEB and AFB. For individuals belonging to the same population and demographic cohort (born before or after the 20th century fertility decline), SNP-based heritability was almost five times higher at 22 percent for NEB and 19 percent for AFB. We also found no evidence that genetic effects on fertility are shared across time and space. Our findings imply that the environment strongly modifies genetic effects on the tempo and quantum of fertility, that potentially ongoing natural selection is heterogeneous across environments, and that gene-environment interactions may partly account for missing heritability in fertility. Future research needs to combine genetic research with social science to better understand human fertility.

---

8 This chapter is based on a manuscript by Tropf, F. C., R. M. Verweij, P J. van der Most, G. Stulp, A. Bakshi, D. A. Briley, M. Robinson, A. Nyman, T. Esko, A. Metspalu, S. E. Medland, N. G. Martin, H. Snieder, S. H. Lee, M. C. Mills. A preprint is published on bioRxiv, doi:http://dx.doi.org/10.1101/049163
A modified version is currently under peer review in an international science journal.

4.1 Introduction

Twin and family studies from Western countries show that genetic factors may explain up to 50 percent of the differences in human fertility outcomes such as number of children ever born (NEB) or age at first birth (AFB) within a population (Byars et al., 2010; Kirk et al., 2001; Kohler et al., 1999, 2006; Mills & Tropf, 2016; Milot et al., 2011; Rodgers, Kohler, et al., 2001; Tropf, Barban, et al., 2015). It remains unknown, however, whether the same genes are important for fertility across different environments or whether gene-environment interaction modifies genetic effects on fertility. This is an important question for at least three reasons. First, the most successful and widely-used design to detect the approximate location of genetic variants associated with complex traits is a meta-analysis of genome-wide association studies (GWAS) from multiple populations (Visscher et al., 2012). This approach assumes genetic effects on a trait to be universal across environments. However, in the case of fertility, this assumption may not be valid given that environmental upheavals such as the introduction of the pill or educational expansion have substantially changed fertility behavior in the recent past (Balbo et al., 2013; Mills et al., 2011). A second and related point is that heritability studies relying on molecular genetic data result in lower estimates than family studies (Manolio et al., 2009) – and this is true also in fertility research (Kohler et al., 2006; Mills & Tropf, 2016; Polderman et al., 2015; Tropf, Barban, et al., 2015; Zaitlen et al., 2013). This discrepancy might be a consequence of the interaction between environment and genes. Family studies are conducted amongst members of the same populations, whereas GWAS exploit data from individuals across populations. If genes can explain variance in fertility within but not between populations, heritability estimates based on different populations will be smaller than within populations (Manolio et al., 2009; Visscher et al., 2008). Third, Fisher's fundamental theorem of natural selection predicts at environmental equilibrium, close to zero additive genetic effects on fitness-related traits such as fertility, because genes that reduce fitness are not expected to have been passed on to the next generation (Fisher, 1930). Nevertheless, additive genetic influences on fertility are well established and a potential explanation is that the genes that are important for fertility differ across environments (Hughes & Burleson, 2000).

Twin and family designs cannot be used to answer the question of whether different genes are important for fertility across populations or birth cohorts, since relatives usually live in the same country and twins always have the same age. However, with the advent of molecular genetic data and complementary analytical techniques and software, it has become possible to examine the genetic material of unrelated individuals from different historical populations. Testing whether the same genes influence a trait across diverse environments has therefore become possible (Purcell et al., 2007; Visscher et al., 2014, 2010; Yang, Lee, et al., 2011; Yang et al., 2010). In this study, we exploit these advances for the first time, by empirically assessing whether genetic effects on fertility differ across geographical and historical environments.

We pooled a series of large datasets consisting of 31,396 unrelated (~ second cousin, IBS<0.05, see Material and Methods) genotyped men (n = 10,489) and women (n = 20,907) from six countries and seven study populations for analysis (for the US: HRS, ARIC; for the Netherlands: LifeLines; For Sweden: STR/SALT; for Australia: QIMR; for Estonia: EGCUT; for the UK: TwinsUK) who are assumed to have completed their reproductive period ($age_{men} > 50; age_{women} > 45$). We first conducted a mega-analysis, which is based on individual information from different populations in contrast to a meta-analysis that uses summary statistics of analyses conducted within populations, and applied whole genome methods (Visscher et al., 2010; Yang et al., 2010) using GCTA software (Yang, Lee, et al., 2011) to estimate SNP-heritability (h^2_{SNP}). SNP-heritability is the proportion of total phenotypic variance that is explained by common genome-wide SNPs. Based on a previous study using data from women from the Netherlands and the UK, we expect h^2_{SNP} to be around 0.10 for number of children ever born and around 0.15 for AFB (Tropf, Stulp, et al., 2015).

Second, to investigate gene-environment interaction, we follow two strategies. The first one involves fitting multiple genetic relatedness matrices in our model, one global matrix for all individuals and more matrices indicating whether individuals lived in the same population and/or were part of the same birth cohort. The global matrix estimates the effects genes have across all environments, whereas the population/birth cohort specific matrices estimate context specific genetic effects (see Yang, Lee, et al., 2011 and Material and Methods for our specifications). The

second strategy involves fitting bivariate genetic models to investigate the moderating effect of the postponement transition on genetic effects on fertility (S. H. Lee et al., 2012 and Material and Methods for our specifications; see Visscher et al., 2014). This model allows us to estimate h^2_{SNP} separately for different birth cohorts as well as the genetic correlation across them. To maximize power in these models, we divided all populations into two demographic birth cohorts. A central turning point in the reproductive environment of the 20th century occurred when AFB experienced a massive postponement of up to 4-5 years in nearly all advanced societies, the so-called 'postponement transition' (Kohler, Billari, et al., 2002), or Second Demographic Transition (Balbo et al., 2013; Lesthaeghe, 1995; Mills et al., 2011; van de Kaa, 1987). The primary reasons proposed for fertility postponement have been women's increased educational attainment and their employment in the labour force, triggered by factors such as the availability of effective contraception (Balbo et al., 2013; Mills et al., 2011). Cultural transformations in terms of sexual freedom, family planning and the timing and role of children are also central (Lesthaeghe, 1995; van de Kaa, 1987). To investigate the moderating effect of fertility postponement we divide individuals into birth cohorts born either before or after this massive postponement in AFB in the past century (Balbo et al., 2013; Kohler, Billari, et al., 2002; Mills et al., 2011; Schmidt, Sobotka, Bentzen, & Andersen, 2012).

4.2 Data & method

4.2.1 Data

In this study we combined data from seven cohorts and six countries. For the US, we use data from ARIC, HRS, for Estonia from EGCUT, for Australia QIMR data from the Australia Twin and Family Register, for the Netherlands the LifeLines Cohort Study, for the United Kingdom TwinsUK and for Sweden the STR. All studies have received ethical approval.

ARIC

ARIC (Atherosclerosis Risk in Communities Study) is a community-based prospective cohort study of 15,792 adults, ages 45–64. Participants were identified by probability sampling from four U.S. communities (Forsyth County, North Carolina; Jackson, Mississippi; suburban Minneapolis, Minnesota; and Washington County, Maryland) and were enrolled between 1987 and 1989 (Investigators, 1989; Jackson et al., 1996; White & Folsom, 1996).

HRS

The Health and Retirement Study (HRS) is an ongoing cohort study of Americans, with interview data collected biennially on demographics, health behavior, health status, employment, income and wealth, and insurance status. The first cohort was interviewed in 1992 and subsequently every two years, with 5 additional cohorts added between 1994 and 2010. The full details of the study are described in (Juster & Suzman, 1995).

EGCUT

Estonian data come from of the Estonian Genome Center Biobank, University of Tartu (EGCUT, www.biobank.ee), a population-based database which comprises the health, genealogical and genome data of currently more than 51,530 individuals (Leitsalu et al., 2015). Each participant filled out a Computer Assisted Personal Interview including personal data (place of birth, place(s) of living, nationality etc.), genealogical data (family history, three generations), educational and occupational history and lifestyle data (physical activity, dietary habits, smoking, alcohol consumption, and quality of life).

QIMR

Data for Australia was received from the Queensland Institute for Medical Research (QIMR). The participants were drawn from cohorts of adult twin families that have taken part in a wide range of studies of health and well-being via questionnaire and telephone interview studies, and recruitment was extended to their relatives (parents, siblings, adult children and spouses).

LifeLines Cohort Study

The LifeLines Cohort Study (Klijs et al., 2015) is a multi-disciplinary, prospective population-based cohort study from the Netherlands, examining in a unique three-generation design the health and health-related behaviours of 167,729 persons living in the North of The Netherlands, which includes genotype information from more than 13,000 unrelated individuals. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioural, physical and psychological factors which contribute to the health and disease of the general population, with a special focus on multi-morbidity and complex genetics.

TwinsUK

For the UK, we use data from TwinsUK, the largest adult twin registry in the country with more than 12,000 respondents (Moayyeri et al., 2013). The TwinsUK Study recruited white monozygotic (MZ) and dizygotic (DZ) twin pairs from the TwinsUK adult twin registry, a group designed to study the heritability and genetics of age-related diseases (www.twinsuk.ac.uk). These twins were recruited from the general population through national media campaigns in the UK and shown to be comparable to age-matched population singletons in terms of clinical phenotype and lifestyle characteristics.

STR

The Swedish Twin Registry (STR) was first established in the late 1950s to study the importance of smoking and alcohol consumption on cancer and cardiovascular diseases whilst controlling for genetic propensity to disease. Between 1998 and 2002, the STR conducted telephone interview screening of all twins born in 1958 or earlier regardless of gender composition or vital status of the pair. This effort is known as Screening Across the Lifespan Twin study (SALT). A subsample of SALT (≈10,000) was genotyped as part of the TwinGene project (Benjamin et al., 2012; Lichtenstein et al., 2006) and we use the this information in the current study.

Fertility trends

Aggregate data to describe country specific fertility trends have been obtained from the Human Fertility Database (HFD, http://www.humanfertility.org/cgi-bin/main.php) and the Human Fertility Collection (HFC, http://www.fertilitydata.org/cgi-bin/index.php), where available. Both data collections are joint projects of the Max Planck Institute for Demographic Research (MPIDR) in Rostock, Germany and the Vienna Institute of Demography (VID) in Vienna, Austria. The projects provide access to detailed and high-quality data on period and cohort fertility. The HFD is entirely based on official vital statistics. The HFC incorporates a variety of valuable fertility data from diverse, not necessarily official, data sources. All data are freely available after registration. We focused on fertility information for cohorts that was aggregated for individuals older than 45.

For the UK, official data on birth order have only been collected within marriage, and might be biased. We therefore relied on estimates from the Office for National Statistics, Cohort fertility, Table 2 (Office National Statistics, 2013). For Estonia, data on completed reproduction by age 45 was only available until the year 1962. For subsequent cohorts, however, there was an estimate of AFB available based on official statistics at the age of 40. For Australia, no official data on a time series of cohort specific AFB was available, and the trends are based on the data used for analysis in this study (see Supporting Information 4-4 for distribution).

Genotypes

We received genotype data from all cohorts, which we imputed according to the 1000 genome panel – except for TwinsUK for which we already received the imputed data.

Genetic-relatedness-matrix (GRM)

To estimate the genetic relatedness-matrix (GRM), the HapMap3 imputation panel was used as a reference set as it was optimized to capture common genetic variation in the human genome (Consortium, 2010).We selected HapMap3 SNPs with an imputation score larger than 0.6. For quality control (QC), we excluded the SNPs with a larger missing rate than 5 percent after merging, lower minor allele frequency than 1 percent, and which failed the Hardy-Weinberg equilibrium test for a threshold of 10^{-6}. We merged the datasets subsequently applying QC again after merging each data set. 847,278 SNPs could be utilized to estimate the GRM between individuals. We used the software Plink (Purcell et al., 2007) for the quality control and merging of the datasets and GCTA (Yang, Lee, et al., 2011) to estimate the genetic relatedness matrix.

The GRM A_{ik} is estimated for each pair of individuals j and k:

$$A_{jk} = \frac{1}{N}\sum_{i=1}^{N}\frac{(x_{ij} - 2p_i)(x_{ik} - 2p_i)}{2p_i(1-p_i)}$$

where x_{ij} and x_{ik} is the number of copies of the reference allele for the i^{th} SNP of the j^{th} or k^{th} individual and p_i is the frequency of the reference allele and N the number of SNPs. If two individuals had a higher genetic relatedness than 0.05, one was excluded from the analyses to avoid bias due to environmental confounders amongst close relatives.

Phenotypes

Number of children ever born was available for all cohorts, but in ARIC and TwinsUK, however, only for women. NEB measures number of children a woman has given birth to or a man has fathered. It was either directly asked or we constructed it from questions on the date of birth of each child.

The measure is not perfectly harmonized across cohorts because some questionnaires explicitly exclude still-births (HRS, ARIC) while others remain undefined (TwinsUK asked in some waves: "How many children have you given birth to?"; EGCUT asked: "Do you have any biological children?", and subsequently: "Fill in their names, gender and date of birth). In STR, LifeLines, QIMR as well as most of the waves of the TwinsUK, information on both the date of birth and death of the child was asked. In LifeLines and TwinsUK, we compared the live birth measure with number of children ever born and, as expected given the diminishing mortality rate in both datasets, less than 0.2 percent of the children had not reached reproductive age, and the correlation of number of children ever born and number of children reaching reproductive age was >0.98. We therefore expect no large bias due to the fact that in some countries still-births are excluded.

Furthermore, the questionnaires were heterogeneous with respect to the maximum number of children that could be named. However, within each cohort, the maximum number of children in the respective questionnaires has never been named more often than in 0.5 percent of the cases and we do not anticipate that our results are influenced by this.

Information on AFB was available in all cohorts except for ARIC and the HRS. It was asked directly (in TwinsUK), or was constructed using the date of birth of the oldest child and the year of birth of the respondent.

Since fertility is strongly age dependent, we focus on women with a completed fertility history in reference to the phenotype. In general, the end of women's reproductive lifespan occurs around the age of 45, and the end of men's at the age of 50. Thus, we only included individuals beyond those ages in our analyses. Furthermore, in vitro fertilization (IVF) – which is often related to twinning and multiple births – can bias results if IVF compensates for genetically based infertility. However, in our TwinsUK sample, only 60 women reported using IVF, who we did not include in the final analyses. For all models, both phenotypes were standardized (Z-transformed) by cohort, year of birth and sex.

4.2.2 GREML Models

Common SNP heritability estimates (Model 1)

The genetic component underlying a trait is commonly quantified in terms of SNP-heritability (h^2_{SNP}) as the proportion of the additive genetic variance explained by common SNPs across the entire genome over the total phenotypic variance (σ^2_P) of the trait:

$$h^2_{SNP} = \frac{\sigma^2_g}{\sigma^2_P}.$$

The phenotypic variance is the sum of additive genetic and environmental variance, i.e. $\sigma^2_P = \sigma^2_g + \sigma^2_e$ where σ^2_g is the additive genetic variance explained by all SNPs across the genome and σ^2_e is the residual variance.

The methods we applied have been detailed elsewhere (S. H. Lee et al., 2012; Visscher et al., 2014, 2010; Yang, Lee, et al., 2011; Yang et al., 2010). Briefly, we applied a linear mixed model

$$\mathbf{y} = \boldsymbol{X\beta} + \mathbf{g} + \mathbf{e}$$

where **y** is an Nx1 vector of dependent variables, N is the sample size, **β** is a vector for fixed effects of the overall mean (intercept), **g** is the Nx1 vector with each of its elements being the total genetic effect of all SNPs for an individual, and **e** is an Nx1 vector of residuals. The variance covariance matrix of the observed phenotypes is:

$$\mathbf{V} = \mathbf{A}\sigma^2_g + \mathbf{I}\sigma^2_e$$

We have g~ $N(0, \mathbf{A}\sigma^2_g)$ and e~ $N(0, \mathbf{I}\sigma^2_e)$, **A** is the genetic relationship matrix (GRM) estimated from SNPs and **I** is an identity matrix. The variance components are estimated using the restricted maximum likelihood (REML) approach.

Genes x Population (Model 2)

The genes x demographic birth cohort interaction model is a joint model estimating global genetic effects for the fertility traits, effective between and within samples (σ_g^2) and the averaged additional genetic effects within cohorts (σ_{gxp}^2):

$$\mathbf{V} = \mathbf{A}\sigma_g^2 + \mathbf{A_{gxp}}\sigma_{gxp}^2 + \mathbf{I}\sigma_e^2$$

where $\mathbf{A}$ is the genetic relatedness matrix and $\mathbf{A_{gxp}}$ is a matrix only with values for pairs of individuals within populations 1-7:

$$\mathbf{A} = \begin{bmatrix} A_{p1p1}A_{p2p1}A_{p3p1}A_{p4p1}A_{p5p1}A_{p6p1}A_{p7p1} \\ A_{p1p2}A_{p2p2}A_{p3p2}A_{p4p2}A_{p5p2}A_{6p2p}A_{p7p2} \\ A_{c1p3}A_{p2p3}A_{p3p3}A_{p4p3}A_{p5p3}A_{p6p3}A_{p7p3} \\ A_{c1p4}A_{p2p4}A_{p3p4}A_{p4p4}A_{p5p4}A_{p6p4}A_{p7p4} \\ A_{c1p5}A_{p2p5}A_{p3p5}A_{p4p5}A_{p5p5}A_{p6p5}A_{p7p5} \\ A_{p1p6}A_{p2p6}A_{p3p6}A_{p4p6}A_{p5p6}A_{p6p6}A_{p7p6} \\ A_{p1p7}A_{p2p7}A_{p3p7}A_{p4p7}A_{p5p7}A_{p6p7}A_{p7p7} \end{bmatrix}$$

$$\mathbf{A_{gxp}} = \begin{bmatrix} A_{p1p1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & A_{p2p2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & A_{p3p3} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & A_{p4p4} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & A_{p5p5} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & A_{p6p6} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & A_{p7p7} \end{bmatrix}$$

Genes x Demographic birth cohort (Model 3)

The genes x demographic birth cohort interaction model is a joint model estimating the universal genetic effects for the traits, effective between and within samples (σ_g^2) and the averaged additional genetic effects within defined birth cohorts (σ_{gxb}^2):

$$\mathbf{V} = \mathbf{A}\sigma_g^2 + \mathbf{A_{gxd}}\sigma_{gxd}^2 + \mathbf{I}\sigma_e^2$$

whereas $\mathbf{A}$ is the genetic relatedness matrix and $\mathbf{A_{gxd}}$ is a matrix only with values for pairs of individuals within the same demographic birth cohorts b1-2:

$$\mathbf{A_{gxb}} = \begin{bmatrix} A_{d1d1} & 0 \\ 0 & A_{d2d2} \end{bmatrix}$$

Genes x Population x Demographic birth cohorts (Model 4)

Finally, we applied a model including both interaction terms from above and an additional interaction term $\mathbf{A_{gxcxb}}$ which is 0 for all pairs of individuals living in different time periods or in different cohorts

$$\mathbf{V} = \mathbf{A}\sigma_g^2 + \mathbf{A_{gxp}}\sigma_{gxp}^2 + \mathbf{A}\sigma_{gxd}^2 + \mathbf{A_{gxpxd}}\sigma_{gxpxd}^2 + \mathbf{I}\sigma_e^2$$

whereas $\mathbf{A}$ is the genetic relatedness matrix, $\mathbf{A_{gxp}}$ is a matrix only with values for pairs of individuals within populations 1-7 (Model 2), $\mathbf{A_{gxd}}$ is a matrix only with values for pairs of individuals within the same demographic periods b1-2 (Model 3) and $\mathbf{A_{gxpxd}}$ is a matrix only with values for pairs of individuals with both the same demographic periods and the same populations.

Bivariate Model

For bivariate analyses (S. H. Lee et al., 2012; Visscher et al., 2014), we split the data into individuals born before and after the turning point in fertility postponement in AFB (see also Supporting Information 4-2 & 4-7). Based on Model 2, we estimate a bivariate model with two GRMs:

$$\boldsymbol{V}\begin{bmatrix} \boldsymbol{f_b} \\ \boldsymbol{f_a} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A_b}\sigma_{g\_b}^2 + \mathbf{A_{gxp\_b}}\sigma_{gxp\_b}^2 + \mathbf{I}\sigma_{e\_b}^2 & \boldsymbol{A_{a\_b}}\sigma_{g\_a\_b}^2 + \mathbf{A_{gxp\_a\_b}}\sigma_{gxp\_a\_b}^2 \\ \boldsymbol{A_{a\_b}}\sigma_{g\_a\_b}^2 + \mathbf{A_{gxp\_a\_b}}\sigma_{gxp\_a\_b}^2 & \boldsymbol{A_b}\sigma_{g\_a}^2 + \mathbf{A_{gxp\_a}}\sigma_{gxp\_a}^2 + \mathbf{I}\sigma_{e\_a}^2 \end{bmatrix}$$

where as $\boldsymbol{f_b}$ **and** $\boldsymbol{f_a}$ are vectors of length N_b *and* $N_a$of fertility phenotypes (NEB or AFB) of individuals born **b**efore or **a**fter the postponement transition started, with N being the respective sample size of the subsets. Variance components refer to those from Model 2, whereas the lower index _b indicates that they are estimated in the subset of individuals born before and index _a born after the start of the postponement transition. The index _b_a denominates the covariance of variances components across subsets. The correlation of the genetic factors are estimated as:

$$r_{\sigma^2_{gxp\_a\_b}} = \sigma^2_{gxp\_a\_b} / \sqrt{\sigma^2_{gxp\_a} * \sigma^2_{gxp\_b}}$$

and

$$r_{\sigma^2_{g\_a\_b}} = \sigma^2_{g\_a\_b} / \sqrt{\sigma^2_{g\_a} * \sigma^2_{g\_b}}$$

4.3 Results

4.3.1 Descriptive findings

The descriptive statistics for NEB and AFB for all populations under study (LifeLines, TwinsUK, STR, Estonia, HRS, ARIC and QIMR) as well as the pooled data separate for men and women can be found in Supporting Information 4-1. The participants were born between 1903 and 1967. The mean number of children per woman is 2.0 in Estonia, Sweden and the UK and 3.3 in Australia. For men, the lowest reported number of children is in Sweden and Estonia with around 1.9 children per man, and is the highest in Australia at around 3.4. AFB was available for the Netherlands, UK, Sweden, Estonia and Australia. For both men and women, it was lowest in Estonia with an average of 24.6 for women and 27.7 for men and highest in Australia with 26.7 and 29.8 for men. Individuals who start reproducing at a later age have fewer children, with correlations between NEB and AFB ranging between -0.24 (Netherlands) and -0.38 (Australia; Supporting Information 4-2). This pattern is less consistent across countries; for example in Australia, the highest fertility levels are observed, despite there being the highest AFB. This reflects heterogeneity in fertility levels across countries, with Australia having traditionally higher fertility levels than other Western countries (for a trend comparison of the total fertility rate across countries see Supporting Information 4-3).

4.3.2 Demographic Trends

Figure 4-1 shows the trends in AFB during the 20th century for the countries in our study based on population data if available (see Data & Methods for details). We observe the well-established U-shaped pattern of AFB of a falling AFB in the first

half of the 20th century followed by a turning point and upturn in the trend of AFB towards older ages. This postponement transition in fertility timing was accompanied by a strong drop in completed fertility in most countries (Sobotka, 2004).

Sociocultural and technological changes, such as the introduction of effective contraception, educational expansion or changing norms in reference to sexuality and family planning, have largely driven these trends (Balbo et al., 2013; Mills et al., 2011). These environmental changes occurred in specific time periods in each country. In order to test for gene-environment interaction in our analyses, we split the data into birth cohorts born before and after the turning point of fertility postponement to reduce environmental heterogeneity amongst the individuals who are members of the same birth cohort. This turning point differs across countries (Figure 4-1), with Australia having the earliest onset of postponement (1939) and Estonia the latest (1962; see Supporting Information 4-5 for all turning points and details). Differences in the onset of the postponement transition are well established and can be due to political factors. This is true in the case of Estonia, for example, where early AFB had been strongly promoted through political incentives when the country was still of the Soviet Union prior to 1990 (Katus, Puur, & Põldma, 2004).

4.3.3 Genetic effects on fertility from the whole genome

Model 1: SNP heritability of AFB and NEB across environments

Not taking environmental differences into account, SNP based heritability (h^2_{SNP}) is significant and low for number of children ever born and age at first birth (Table 4-1). For NEB, h^2_{SNP} is 0.038 (SE = 0.0097, p-value = 2.0×10^{-5}) and for AFB it is 0.053 (SE = 0.019 p-value = 0.0020; these estimates are based on the full genetic relatedness matrix - see Material and Methods). These findings mean that around four percent of the variance in NEB and around five percent in AFB can be attributed to common, additive genetic effects in the pooled data. These estimates are much lower than those reported in other studies (Tropf, Stulp, et al., 2015).

Model 2: Genes x population interaction (g x p)

A potential reason why estimates are lower than expected is that the SNPs important for fertility have different effects across environments. Model 2 therefore adds an interaction term to Model 1 that captures the influence of genetic variance on fertility only within populations. The gene-population interaction models for NEB and AFB show that shared genetic effects across populations are much lower than genetic effects within populations. With respect to NEB, the shared genetic effects across populations are negligible (0.0070, SE = 0.011, p-value = 0.26), whereas within populations additional additive effects are estimated to be 0.15 (SE = 0.024, p-value = 6.0 x10^{-12}).

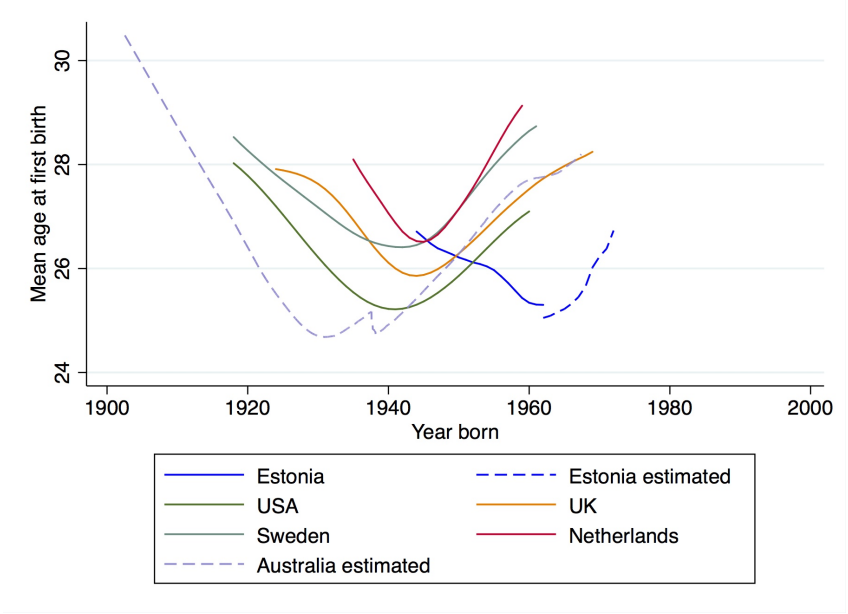

Figure 4-1 Trends in age at first birth in cohorts from the US, UK, Sweden, the Netherlands, Estonia and Australia (1903-1970).

Note: Trends in the mean age at first birth are moving averages based on aggregated data obtained from Human Fertility Database and the Human Fertility Collection (for details see Data and Methods). For Australia, no official data has been available and the trends have been estimated from the QIMR dataset. See Supporting Information 4-4 for the birth cohort specific average in the QIMR data.
Source: See Methods section for details

The same applies to AFB, where shared genetic effects are estimated to be only 0.024 (SE = 0.022, p-value = 0.14), whereas the within population effect is 0.10 (SE = 0.039, p-value=0.0037; Table 4-1; Model 2). These results show that there is little overlap in SNPs that influence fertility across populations, and that most of the SNPs influencing fertility are population specific.

Model 3: Genes x demographic birth cohort (g x d)

Similar to the case of Model 2, in which we modelled population specific effects, we also examined whether there were genetic influences on fertility specific to particular birth cohorts. We find that there is additional genetic variance explanation for individuals who live in the same demographic cohort. While h^2_{SNP} for all birth cohorts is estimated at zero for both NEB (SE = 0.013, p-value = 0.50) and AFB (SE = 0.03, p-value = 0.35), for individuals living in the same demographic cohort there is a significant additional genetic variance component of 0.097 (SE = 0.017, p-value = $3.3x10^{-16}$) for NEB and 0.084 (SE = 0.031, p-value = $4.6 x10^{-4}$) for AFB (Table 4-1; Model 3). Thus, as in the case of the different populations, we find that SNPs influencing fertility traits are specific to particular birth cohorts.

Model 4: Genes x population x demographic birth cohort (g x p x d)

Including a gene-environment interaction term that takes into account both the population and the demographic cohort simultaneously (Model 4), we observe that for NEB both the interaction with demographic cohort (0.064, SE = 0.020, p-value = $5.9x10^{-4}$) and the interaction with population and demographic cohorts (0.085, SE = 0.045, p-value = 0.0030) are significant. This suggests that living in the same demographic cohort increases h^2_{SNP} independent of whether individuals live in the same population, and that living in the same population and the same demographic period further increases h^2_{SNP}. For AFB, h^2_{SNP} is only significantly different from zero for individuals living in the same population and demographic cohort (0.18, SE = 0.077, p-value = 0.0032).

Overall SNP based heritability for each model

Subsequently, overall heritability estimates were calculated as the sum of the different components of each model. This was done in order to examine the increase in heritability estimates when including the different interaction terms (See Figure 4-2 corresponding to Supporting Information 4-6). The overall h^2_{SNP} for NEB increases almost fivefold, from 0.04 (SE = 0.01; Model 1) to 0.22 (SE = 0.026), when population and demographic cohort are taken into account. For AFB, the trend is very similar, with an h^2_{SNP} of 0.053 (SE = 0.019) in the baseline Model 1, and 0.19 (SE = 0.039) in the genes x population x demographic cohort interaction model—where population and demographic cohort are taken into account.

Sensitivity analysis: Genes x Sex

The analyses presented are based on pooled datasets of men and women. However, two data sources contain (almost) only women (TwinsUK and ARIC). To the extent that different genes influence fertility across sexes, this might drive the observed differences across populations. We therefore conducted a sensitivity analysis extending Model 3 to a genes x population x sex interaction model. We find that considering sex-differences does not significantly improve the model fit (p-value for AFB 0.5, for NEB 0.093). We are therefore confident that our findings do not result from sex-differences (Supporting Information 4-8).

We estimated a complementary bivariate model based on Model 2, by splitting data for demographic cohorts (see Material and Methods), which allows us to estimate genetic effects across (σ^2_g) and within (σ^2_{gxp}) populations separately for different demographic birth cohorts, and thereby investigate whether genetic effects are correlated across demographic birth cohorts. Table 4-2 shows that σ^2_{gxp} estimates for NEB within populations are significant for both demographic cohorts before ($\sigma^2_{gxp}/\sigma^2_p$ = 0.15, SE = 0.039, p-value = $9.6\text{x}10^{-6}$) and after ($\sigma^2_{gxp}/\sigma^2_p$ = 0.13, SE = 0.048, p-value = 0.0010) fertility postponement. It also shows a positive correlation of genetic effects with NEB across demographic cohorts and within populations (1.00, SE = 0.35, p-value = $1.3\text{x}10^{-5}$). In Model 4 of Table 4-1, this remained suggestive, since the genetic effects within populations but shared across demographic cohorts ($\sigma^2_{gxp}/\sigma^2_p$) were non-significant (0.073, SE = 0.036, p-value = 0.18).

Table 4-1 Heritability estimates of the full GREML model and gene environment interaction models for number of children ever born (NEB) and age at first birth (AFB)

	Number of children ever born							
Model	1		2		3		4	
	Estimate (SE)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value
σ_g^2 / σ_P^2	0.038 (0.0097)	2.0 x10^{-5}	0.0070 (0.011)	0.26	0.00 (0.013)	0.50	0.00 (0.015)	0.50
$\sigma_{gxp}^2 / \sigma_P^2$	--	--	0.15 (0.024)	6.0x10^{-10}	--	--	0.073 (0.036)	0.18
$\sigma_{gxd}^2 / \sigma_P^2$	--	--	--	--	0.097 (0.017)	3.3x10^{-8}	0.064 (0.020)	5.9x10^{-4}
$\sigma_{gxpxd}^2 / \sigma_P^2$	--	--	--	--	--	--	0.085 (0.045)	0.0030
N	31396							
	Age at first birth							
Model	1		2		3		4	
	Estimate (SE)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value
σ_g^2 / σ_P^2	0.053 (0.019)	0.0020	0.024 (0.022)	0.14	0.00 (0.030)	0.35	0.011 (0.028)	0.32
$\sigma_{gxp}^2 / \sigma_P^2$	--	--	0.10 (0.039)	0.0037	--	--	0.00 (0.062)	0.50
$\sigma_{gxd}^2 / \sigma_P^2$	--	--	--	--	0.084 (0.031)	4.6 x10^{-3}	0.00 (0.040)	0.50
$\sigma_{gxpxd}^2 / \sigma_P^2$	--	--	--	--	--	--	0.18 (0.077)	0.0032
N	16109							

Note: σ_g^2 / σ_P^2 = proportion of observed variance in the outcome associated with genetic variance across all environments, $\sigma_{gxp}^2 / \sigma_P^2$ = proportion of observed variance in the outcomes associated with additional genetic variance within populations, $\sigma_{gxd}^2 / \sigma_P^2$ = proportion of observed variance associated with additional genetic variance within demographic birth cohorts, $\sigma_{gxpxd}^2 / \sigma_P^2$ = proportion of observed variance associated with additional genetic variance within populations and demographic birth cohorts, p-values are based on likelihood-ratio test comparing the full model with the model with one constraining the particular effect to be zero, all analyses include the first 20 Principal Components, outcomes are standardized for sex, birth year and country.

Source: See Methods section for details

Bivariate analysis

The bivariate model for the AFB finds some evidence that in both demographic cohorts genetic effects are observed (before fertility postponement 0.099, SE = 0.073, p-value = 0.083; after fertility postponement 0.13, SE = 0.074, p-value = 0.070), although these effects were marginally significant. However, there is no evidence that genetic effects correlate across demographic birth cohorts (0.11, SE = 0.59, p-value = 0.27), which is well in line with the null-estimate $\sigma^2_{gxp} / \sigma^2_P$ (0.00, SE = 0.062, p-value = 0.50). Genetic effects shared across all populations (σ^2_g) are only significant for NEB and birth cohorts born before fertility postponement (σ^2_g / σ^2_P = 0.031, SE = 0.018, p-value 0.042). By contrast, they are non-significant for younger cohorts and for AFB.

4.4 Discussion

Using data from seven populations and six countries, we demonstrate that genetic effects on fertility outcomes –number of children ever born (NEB) and age at first birth (AFB) – differ across temporal and spatial environments. For NEB, genetic effects within populations are stronger than across populations, but correlate between individuals who were born before or after the turning point in fertility postponement during the 20th century. For AFB, genetic effects are only significant if individuals live in the same demographic cohort and the same population. Neither the full gene-environment interaction model (Model 4) nor the bivariate analyses provide evidence for shared genetic effects for each phenotype across populations and demographic birth cohorts. Our results show that different SNPs are associated with fertility traits in different populations and birth cohorts, and that very few genetic effects are consistently related to these traits across populations and cohorts. Our results uncover a strong interplay between genetic and environmental factors influencing human fertility.

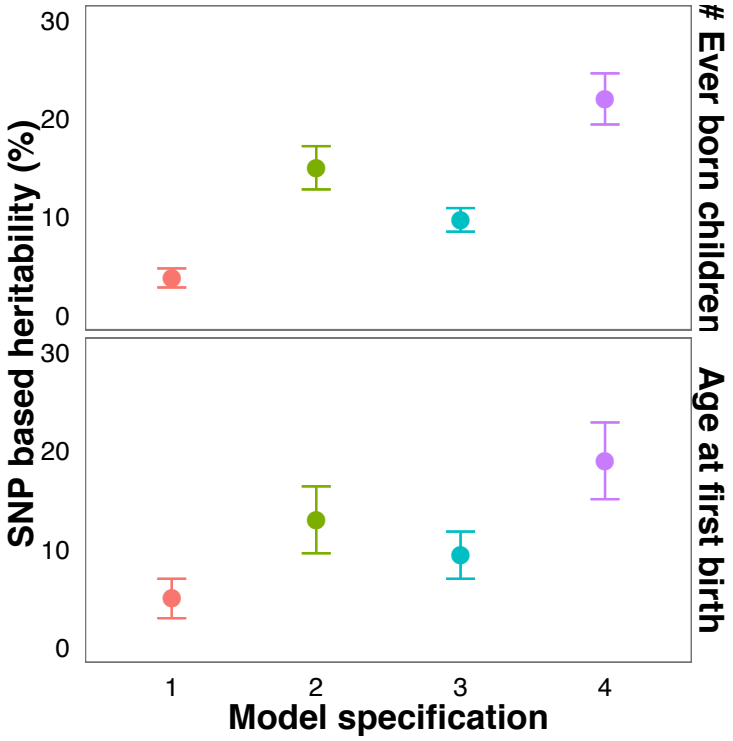

Figure 4-2 Bar Charts of the SNP-heritability estimates in number of children ever born (NEB) and age at first birth (AFB) for the different model specifications from Table 4-1

Note: SNP-heritability as the sum of genetic variance over the total variance in Model specification 1 = amongst all individuals, 2 = amongst individuals living within the same population, 3 = amongst individuals living within the same demographic birth cohort born either before or after fertility postponement, 4 = amongst individuals living in the same population and demographic birth cohort, dots = estimate, lines = estimate ± 1 SE, The corresponding table to Figure 4-2 an be found in SI 4-6.

Quantitative geneticists have been puzzled by low heritability estimates based on GWAS findings and whole-genome methods such as GREML, describing the phenomenon as the problem of 'missing heritability' (Manolio et al., 2009). Previous attempts to explain missing heritability partly by non-additive genetic effects remain largely empirically untested (Zuk & Hechter, 2012) or find only little support (Zhu et al., 2015). Our findings of strong gene-environment interaction imply that the detection of genetic variants associated with fertility traits is a major challenge when using meta-analyses of GWAS on individuals from different populations. Likewise, predictions out of the discovery sample might not be straightforward to test, because discovered SNPs might have different effects in different samples.

In addition, our findings imply that lower heritability estimates from GWAS studies, as compared to GREML approaches or family studies, might be due to the fact that genetic effects are (to some extent) not universal, but rather context specific. In the model considering gene-environment interaction across population and demographic cohort, we report heritability findings of 0.22 for NEB and 0.19 for AFB (see Figure 4-2 and Supporting Information 4-8), which are fourfold higher than across all contexts and approach heritability estimates from family models (Mills & Tropf, 2016; Polderman et al., 2015). It is therefore vital to understand the cultural and environmental factors that interact with human fertility as well as their origins across (family) environments in order—for example—to define missing heritability or validate the findings from twin studies. It should be noted that our findings are probably linked to the strong behavioural and social nature of fertility, which might be more sensitive to cultural and societal influence than morphological traits. A recent investigation by Yang et al. (2015) shows that missing heritability for the anthropometric traits height and body mass index is negligible when using whole genomic sequencing data in a new GREML model, and assuming that family models overestimate heritability.

Demonstrating that genetic effects on fertility outcomes differ across environments, our study contributes substantially to our current knowledge of the genetic architecture of human reproduction. Previous twin studies show for several countries and birth cohorts that fertility outcomes such as NEB and AFB are genetically influenced (Kohler et al., 2006; Mills & Tropf, 2016). However, it

remained unclear whether the same genes are associated with fertility across environments. Using molecular genetic data and GREML methods (Visscher et al., 2010; Yang, Lee, et al., 2011; Yang et al., 2010), we were able to study the genetic material of individuals across diverse environment,s and found that common SNPs explain substantially more variance within than between countries and birth cohorts for fertility traits.

Previous twin and family studies furthermore suggest that the heritability of fertility traits can change across time and space (Bras et al., 2013; Briley et al., 2015; Kohler et al., 1999; Tropf, Barban, et al., 2015). However, these differences could not be statistically validated. In the current study, we proposed a multi-matrix approach to test for gene-environment interaction but also applied bivariate GREML models across birth cohorts (S. H. Lee et al., 2012; Visscher et al., 2014). Bivariate GREML models allow estimation of SNP-heritability within two independent samples as well as the genetic correlation across them. We cannot confirm the suggestion that the level of heritability changed over time, but find that heritability levels are comparable before and after the strong fertility postponement during the past century.

Different levels of heritability have also been reported across countries (Mills & Tropf, 2016). Our multi-matrix GREML approach distinguishes between pairs of individuals who are living in the same or in different populations. The resulting within population estimate is therefore an average across all populations and we cannot compare different levels of heritability across populations. A more desirable study design would be a multivariate genetic modelling approach (similar to the one we presented in a bivariate design) to investigate differences across demographic birth cohorts. However, this approach was not possible in our study due to small sample sizes in each population, and consequent lack of statistical power (Visscher et al., 2014). However, it might become feasible in the future with better data availability.

Table 4-2 Bivariate analysis of Model 2 to estimate genetic correlations for gxp (genes x population) or global g (gene) component before and after fertility postponement

	Number of children ever born						
	Before postponement		After postponement		r(G)		
	Estimate (SE)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value	N
$\sigma^2_{gxp} / \sigma^2_P$	0.15 (0.039)	$9.6x10^{-6}$	0.13 (0.048)	0.0010	1.00 (0.35)	$1.3x10^{-5}$	17,969
σ^2_g / σ^2_P	0.031 (0.018)	0.042	0.0017 (0.026)	0.50	-1.00 (8.04)	0.50	13,427
	Age at first birth						
	Before postponement		After postponement		r(G)		
	Estimate (SE)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value	N
$\sigma^2_{gxp} / \sigma^2_P$	0.099 (0.073)	0.083	0.13 (0.074)	0.070	0.11 (0.59)	0.27	8,049
σ^2_g / σ^2_P	0.023 (0.04)	0.20	0.011 (0.048)	0.50	1.00 (2.89)	0.50	8,060

Note: σ^2_g / σ^2_P = proportion of observed variance in the outcome associated with genetic variance across all environments, $\sigma^2_{gxp} / \sigma^2_P$ = proportion of observed variance in the outcomes associated with additional genetic variance within populations, $\sigma^2_{gxd} / \sigma^2_P$ = proportion of observed variance associated with additional genetic variance within demographic birth cohorts, r(G) = genetic correlation, p-values are based on likelihood-ratio test comparing the full model with the model with one constraining the particular effect to be zero, all analyses include the first 20 Principal Components, outcomes are standardized for sex, birth year and country.

Our findings are of interest to scientists within the medical, biological and social sciences alike (Kohler et al., 2006; Mills & Tropf, 2016; Montgomery et al., 2014; Stearns et al., 2010). Research has successfully identified genetic variants associated with reproductive diseases and traits (Montgomery et al., 2014). However, it remains unknown how these affect realized fertility. We find no evidence that genetic effects underlying fertility in one country predict fertility outcomes in another one. Instead, genetic effects on fertility outcomes are strongly dependent on an individual's environment. Recently, social scientists have made large efforts to integrate molecular genetics into their research (Benjamin et al., 2012; J Bongaarts, 1983; Briley et al., 2015; Kohler et al., 2006; Mills & Tropf, 2016; Rietveld et al., 2014; Rietveld, Cesarini, et al., 2013; Rietveld, Medland, et al., 2013; Tropf, Stulp, et al., 2015). However, when it comes to reproductive health, environmental factors are also likely to be critical in understanding how genetic effects on fecundity and infertility may be modified.

For evolutionary biologists, our findings have at least two important implications. First, the number of children ever born has been used as a proxy for fitness, given the diminishing child mortality rate in contemporary societies (Byars et al., 2010; Stearns et al., 2010; Tropf, Stulp, et al., 2015). Additive genetic variance therefore indicates natural selection under environmental equilibrium within populations if all else being equal, genes that lead to a higher number of offspring will have a higher frequency in future generations. Due to natural selection, Fisher predicted additive genetic variance in fertility to be (close to) zero in the absence of gene environment interaction, since genes that reduce fitness do not tend to be passed on to the next generation, thereby becoming lower in frequency (Fisher, 1930). Nevertheless, we find significant additive genetic influences on fitness traits such as NEB and AFB – substantial yet lower than heritabilities observed for morphological traits such as height (Mousseau & Roff, 1987; Polderman et al., 2015; Tropf, Stulp, et al., 2015; Visscher et al., 2008). Finding significant genetic influences on these proxies of fitness suggests that, along with sociocultural changes surrounding fertility, genetic variants under selection have also changed (Briley et al., 2015; Courtiol, Rickard, & Lummaa, 2013; Hughes & Burleson, 2000; Kohler et al., 1999, for review see 2006; Kohler, Rodgers, et al., 2002; for review see Mills & Tropf, 2016; for comment see Milot & Pelletier, 2013; Tropf, Barban, et al., 2015). Gene-

environment interaction can explain why we find additive genetic variance in fitness related traits despite natural selection.

Second, previous research has uncovered ongoing natural selection in contemporary societies (Bolund, Hayward, Pettay, & Lummaa, 2015; Byars et al., 2010; Milot et al., 2011; Stulp et al., 2015; Tropf, Stulp, et al., 2015), and has even attempted to forecast changes in traits such as height and blood pressure across generations (Byars et al., 2010). In order for one to make valid evolutionary predictions about observable changes in traits across generations due to natural selection, fertility needs to be consistently heritable, the same genes need to be under selection across generations, and the direction of the selection needs be similar. Our results point to moderate genetic influences on fertility within populations, indicating potentially ongoing human evolution. However, this potential is limited in at least two ways. First, genetic effects on fertility strongly differ across countries and therefore may lead to heterogeneity across human populations rather than to universal changes in humans. Second, the finding that genetic effects underlying proxies of fitness vary so markedly across time periods suggests that substantial caution is needed when making long-term evolutionary predictions.

For social scientists, genetic influences had been originally thought of as biological constraints on human reproductive behavior (J Bongaarts, 1983). Yet some previous studies showed that genetic predispositions may underlie decision making processes governing fertility timing and motivation (W B Miller, 1992; Pasta & Miller, 2000; Rodgers, Kohler, et al., 2001; Tropf, Barban, et al., 2015). Some suspect that genetically based behavioural and psychological traits have become more important than physiological ones in the recent past (Briley et al., 2015; Kirk et al., 2001; Rodgers, Kohler, et al., 2001; Udry, 1996). This hypothesis remains to be tested, but our results confirm that genetic influences on fertility have evolved with social changes in the reproductive environment and therefore underscore the need to integrate social factors into genetic research on fertility.

Overall, our study uncovers great challenges for investigations into the genetic architecture of fertility, which can only be overcome by interdisciplinary work between both social scientists and geneticists using ever larger datasets, with combined information from genetics and social surveys (Stearns et al., 2010).