Impact of co-occurrent assortative mating and vertical cultural transmission on measures of genetic associations

Anthony F. Herzig (1,*), Camille Noûs (2), Aude Saint Pierre (1), Hervé Perdry (3,*)

(1) Univ Brest, Inserm, EFS, UMR 1078, GGB, F-29200 Brest, France
(2) Laboratoire Cogitamus, F-75005 Paris, France
(3) CESP Inserm, U1018, Université Paris-Saclay, UVSQ, F- 94807 Villejuif, France
* Corresponding Authors: anthony.herzig@inserm.fr, herve.perdry@universite-paris-saclay.fr

Abstract

Assortative mating for a given phenotype is the phenomenon by which mates select each other based on their phenotypic similarity. Other phenomena can create positive correlation between the parents' and the offspring's environment: vertical cultural transmission, or dynastic effects. When these phenomena occur together, they induce a gene-environment correlation at the population scale. It will impact genetic measures of associations such as SNP effect size and SNP-heritability.

In this paper, we provide a complete mathematical modelling of both assortative mating and vertical cultural transmission in the classical framework of the polygenic additive model. We establish for the first time the theoretical evolution and equilibrium values of all quantities of interest involved. We then derive its consequences on typical genetic epidemiology analyses; including both population and family-based study designs. We show the consequences on heritability estimation, Genome-Wide Association Studies (GWAS), and the variance explained by polygenic scores. We validate our calculations through simple forward-time simulations.

1 Introduction

Assortative mating (AM) is the process of non-random mate choice that results in a positive correlation of phenotypic values and was first described by RA Fisher in his seminal presentation of the polygenic model in 1919 (Fisher, 1919). The theoretical consequences of AM on the distribution of genetic variants in a population have been explored in detail, notably by Crow and Felsenstein (Crow & Felsenstein, 1968), Wright (Wright, 1921) and Nagylaki (Nagylaki, 1978). Importantly, it has been shown that AM results in long distance correlations between genetic variants that have an effect on the phenotype (under a polygenic model) and that this will result in an increase in both phenotypic and breeding (additive genetic)-value variances.

Correlated phenotypes between mates have been observed directly in human populations. A most simple example is the often-studied correlation between standing height of spouses (Pearson & Lee, 1903; Stulp et al., 2017). Correlated phenotypes have also been regularly estimated for a range of social and behavioural traits (Buss, 1985) including education level (Hugh-Jones et al., 2016; Mare, 1991) as well as for multifactorial disease (Hippisley-Cox et al., 2002; Yamamoto et al., 2023) and psychiatric disorders (de Jong et al., 2018; Jefsen et al., 2022; Nordsletten et al., 2016).

In the age of widely available genetic data and increasingly huge cross-sectional epidemiological studies including genetic data (All of Us Research Program Investigators et al., 2019; Bycroft et al., 2018; Nagai et al., 2017), it has become readily possible to search for the signs of AM on genetic variation in a population (Border, Athanasiadis, et al., 2022; Yamamoto et al., 2023; Yengo et al., 2018), validating some predictions made by early population geneticists almost 100 years later. Such findings have indicated that AM is likely prevalent and capable of distorting the inference of prevalent genetic epidemiology study design such as genome-wide association studies (GWAS) (Howe et al., 2021, 2022), heritability analysis (Border, O'Rourke, et al., 2022), polygenic score prediction (Torvik et al., 2022; Yengo et al., 2018) and Mendelian randomisation (Brumpton et al., 2020; Davies et al., 2019; Hartwig et al., 2018; Howe et al., 2019). Indeed, assortative mating has been recently described as one of the principal confounders of genetic association studies (Morris et al., 2020; Veller & Coop, 2023).

A further phenomenon that impacts studies in genetic epidemiological studies is that of vertical cultural transmission (VCT), in which environment is transmitted from parents to offspring. This can be simply modelled as a (positive) correlation between parents and offspring environments. Such VCT can intuitively produce correlations between phenotypes that can closely mimic the correlation induced by an additive genetic component (Cavalli-Sforza & Feldman, 1973). Indeed, the importance of considering the non-independence of non-genetic factors between closely related individuals has long been recognized and is a motivation for the introduction of twin studies in estimating heritability (Falconer, 1960).

A similar notion to VCT is that of a dynasty effect (DE) as described in (Brumpton et al., 2020) and in (Morris et al., 2020). VCT and DE respectively correspond to the influence of parental environments or parental phenotypes on offspring environment; both leading to a positive correlation between the non-genetic contributions to phenotypic variance between parents and offspring. Such effects were characterised by the term 'the Nature of Nurture' in (Kong et al., 2018) capturing the idea that parental genotypes (or even genotypes of more distant relatives) can affect the phenotype of the offspring though paths other than the actual sharing of alleles. Various methods have been put forward to control for such structure when estimating heritability (A. I. Young et al., 2018; Zaitlen et al., 2013) or GWAS (Howe et al.,

2022; A. I. Young et al., 2022) based on the idea of controlling for parental genotypes in order to try to capture the direct substitution effects of genetic factors.

Whilst both AM and VCT have therefore been widely addressed in terms of their impact of population based genetic epidemiology studies, there is no reason to suppose that AM and VCT may not co-occur. A general model that takes into account the combined consequences of AM and VCT was set out by Rice, Cloninger and Reich in three linked papers in 1978-79 (Cloninger et al., 1979a, 1979b; Rice et al., 1978). Rather than explicitly describing polygenic inheritance as in Crow and Felsenstein, 1968, path analysis is used by Rice, Cloninger and Reich with the simple assumption of the path coefficient between the genetic component of a parent and the offspring being ½ (Rice et al., 1978). Crucially, Rice, Cloninger and Reich describe the correlation between genetic and cultural components within an individual; w in their notation, ρ in ours. The possible presence of this correlation was first explored (to the best of our knowledge) in path analyses in (Rao et al., 1976). Gene-environment correlations involving models of complex gene-culture transmission and assortative mating are also presented in (Feldman & Cavalli-Sforza, 1977) and (Cavalli-Sforza & Feldman, 1978). Rice, Cloninger and Reich describe such gene-environment correlations as being either a primary correlation (a characteristic of the population) or one that can arise and be derived in the presence of AM and VCT (Cloninger et al., 1979a). This correlation was explored in real data by Rice, Cloninger and Reich for IQ scores and social-economic status data in the USA (Rice et al., 1980); estimating a value for w between 0.137 and 0.191. This correlation between genetics and cultural effects within individuals subsequently gained some traction in the field of behavioural genetics; for a few examples it is included in the model in (Carey & Rice, 1983) where it is explicitly described as being induced by AM and VCT; as well as in many subsequent works: for example in (Vogler & Fulker, 1983), (Martin et al., 1986), (Truett et al., 1994), and more recently in (Keller et al., 2009), (Vinkhuyzen et al., 2012), and (Balbona et al., 2022). It may be the case that such complicated models involving AM and VCT have found their niche in the analysis of highly complex social traits (e.g. intelligence and personality traits) where complex familial data (twin, siblings, adoption studies etc.) is often leveraged. The potential of an important impact on co-occurrent AM and VCT on twin studies in general was recently discussed in (Larrègue et al., 2023) for behavioural traits.

More broadly, in genetic epidemiological studies based on biobanks, any such complications are generally ignored and environmental/cultural factors are assumed to act independently from genetics. We therefore wished to revisit the question of co-occurring AM and VCT and explore the implication of the genetic-cultural correlation in modern study designs in genetic epidemiology. To this end, we derive here (and confirm through simulation) the evolution and equilibrium equations for the correlation structure between genetic and environmental values in a trio in the presence of both AM and VCT. Unlike previous work, we achieve this with an explicit model of polygenic inheritance including derivations of the eventual correlations that occur between all pairs of causal genetic variants in a population. Indeed, this allows us to calculate evolution and equilibrium values that to our knowledge have never been derived even in the simpler model involving only AM. Our key result is to show that the combination of AM and VCT leads to positive correlation between genetic and environmental values at the population level and that this has significant impact on subsequent study designs. Specifically, this leads to inflated SNP-heritability and SNP-effect estimates, and hence also on polygenic score (PGS)-based methods, due to the fact that estimated genetic effects will also capture non-genetic effects. We also investigate the impact on study designs based on analysing large-scale family data. Simple forward-time simulations are presented to validate our calculations and demonstrate the impact of co-occurring AM and VCT on the measures of association.

2 Methods

2.1 The polygenic additive model with assortative mating and vertical cultural transmission (AMVCT)

2.1.1 Model without homogamy

We consider Fisher's infinitesimal additive model, where a phenotype is determined by $P = A + E$ where the genetic value A is the sum of the effects of a large number N of genetic loci and E is the environmental effect. We consider two mates indexed by 1 and 2, and an offspring index by 3.

Denote $\text{var}(A) = a^2$, $\text{var}(E) = e^2$ and $\text{var}(P) = \sigma^2 = a^2 + e^2$. If the population is panmictic, and assuming E_1, E_2 and E_3 are independent, the correlation between the genetic values of the parent and the offspring is classically $r_{A_1A_3} = r_{A_2A_3} = \frac{1}{2}$ (Figure 1, panel a), leading to phenotypic correlations $r_{P_1P_3} = r_{P_2P_3} = \frac{1}{2}h^2$ where $h^2 = \frac{a^2}{\sigma^2}$, classically referred to as the (narrow-sense) heritability.

2.1.2 Introducing homogamy

We consider a panmictic population in which homogamy appears at a generation indexed by $t = 0$. Denote by g_0^2 the gametic variance in the absence of homogamy and the heritability in the absence of homogamy as $h_0^2 = \frac{2g_0^2}{2g_0^2+e^2}$. In the presence of homogamy, we denote $r_{ho} = r_{P_1P_2}$ the correlation between the mates' phenotypes. The presence of this correlation induces a correlation $r_{A_1A_2}$ between A_1 and A_2, so the genetic values of the gametes emitted by the parents have a positive correlation r_{ga}. This gametic correlation will, across generations, induce a small gametic disequilibrium between the causal loci, which will cause an increase in the gametic genetic variance g_t^2 (and of the individual genetic variance a_t^2), until an equilibrium is reached. At the same time, there is a correlation between A_1 and both A_2 and E_2, and also between E_1 and both A_2 and E_2. As the offspring genetic value A_3 is correlated to A_1 and A_2, which are respectively correlated to E_2 and E_1, this implies the presence of a correlation between A_3 and both E_1 and E_2 (Figure 1, panel b). The values of all these correlations at equilibrium will be computed (see below in 3.1.1, 3.1.2 and Supplementary Material sections 1-3).

2.1.3 Introducing vertical cultural transmission

We consider the same model as above, but with yet another parameter ν for the correlation between the parents' environment values E_1, E_2, and the offspring environment E_3 (Figure 1, panel c). As both E_1 and E_2 are correlated to A_3, this induces a correlation ρ between A_3 and E_3. Again, we will compute the se correlation in the Results and Supplementary Material sections 1-4. The value of ρ will increase generation after generation, until reaching an equilibrium point.

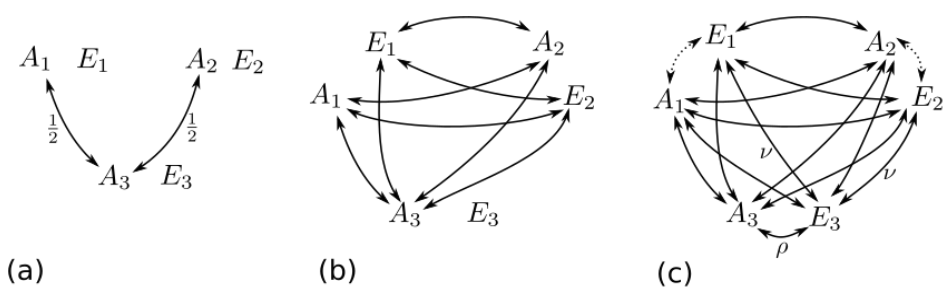

Figure 1. The correlations between genetic values and environments of parents (indexed by 1, 2) and offspring (indexed by 3), in models of increasing complexity. Panel (a) corresponds to the simple model of a panmictic population; (b) corresponds to a model of homogamy, in which the positive phenotypic correlations between the mates induces positive correlations between A_1

(respectively E_1) and both A_2 and E_2, inducing correlations between A_3 and all the four parental variables, but in which E_3 is independent of all other variables; and (c) corresponds to the introduction of an additional parameter ν for the correlation between the parental environment E_1(respectively E_2) and the offspring environment E_3, inducing positive correlations between all pairs of variables, except possibly $A_1$and E_1 (respectively A_2 and E_2), if this is the first generation of random mating.

2.2 Forward-time Simulations

We performed simple simulations in R to demonstrate the validity of our evolution and equilibrium equations (simulation scripts are provided, see Data Availability). We simulated populations of constant size $M = 25{,}000$ and $100{,}000$ over subsequent generations with co-occurring AM and VCT. This involved simulating the population at generation 0, with a quantitative phenotype with variance $\sigma^2 = 1$ determined by $N = 1{,}000$ unlinked causal SNPs and an architecture of $h_0^2 = 0.2$ (hence, $g_0 = \frac{1}{2}\sqrt{0.2}$ and $e = \sqrt{0.8}$). The gametic alleles were drawn with minor-allele frequencies of 0.5 and we applied equal allelic substitution effects for all causal SNPs.

AM was simulated by forming $M/2$ pairs in each generation, this was achieved by simulating a proxy variable for $M/2$ pairs from a bivariate normal distribution with a correlation of r_{ho}, and then using the ranks across all M observations to form pairs with the simulated phenotype data. Each pair would produce two offspring to keep the generation size stable. Each causal SNP was transmitted randomly and independently of the others, i.e., with no linkage. VCT was simulated by drawing environmental components for each offspring (E_3) using the parameter ν to ensure the correct covariance structure between E_1, E_2 and E_3 in each trio.

3 Results

3.1 Evolution and Equilibrium equations

3.1.1 Correlations under AM

We computed the values of all correlations between components, both genetic and environmental, in a trio at equilibrium (see Supplementary Material sections 1-3) using the same approach as (Nagylaki, 1978). The non-zero correlations between the genetic and environmental values of the family members are summarized in table 1, where a, e and σ are the standard deviations of A, E and P at equilibrium. The gametic correlation verifies $r_{ga} = \frac{r_{A_1A_2}}{2-r_{A_1A_2}}$, that is $r_{ga} = \frac{r_{ho}h^2}{2-r_{ho}h^2}$. The gametic disequilibrium between each pair of the N causal loci corresponds to a correlation of $\frac{1}{N}\frac{r_{ga}}{1-r_{ga}}$; the cumulated effects of these intra gametic correlation leads to a gametic variance $g^2 = \frac{g_0^2}{1-r_{ga}}$. In turn, this leads to a genetic variance $a^2 = 2\frac{1+r_{ga}}{1-r_{ga}}g_0^2$. Additionally, we give in the Supplementary Material (sections 4 and 5; in particular section 4.5) equations for the evolution and equilibrium values of all these quantities across generations. Expressions for the equilibrium values for r_{ga} and a^2 can be found that rely only on parameters of the initial population (r_{ho}, e, g_0^2); they are as follows:

$$r_{ga} = \frac{r_{ho}a^2}{(2-r_{ho})a^2+2e^2}\text{; where } a^2 = \frac{1}{2(1-r_{ho})}\left(\sqrt{(e^2+2g_0^2)^2 - 8r_{ho}e^2g_0^2} - (e^2-2g_0^2)\right)$$

Table 1. Correlations between the genetic and the environmental effects in presence of assortative mating, without a shared parent-offspring environment. Individuals 1 and 2 play symmetric roles, so we have $r_{A_2A_3} = r_{A_1A_3}$ and $r_{E_2E_3} = r_{E_1E_3}$.

	$\boldsymbol{A_2}$	$\boldsymbol{E_2}$	$\boldsymbol{A_3}$
$\boldsymbol{A_1}$	$r_{ho}h^2$	$r_{ho}\frac{ae}{\sigma^2}$	$\frac{1}{2}(1 + r_{ho}h^2)$
$\boldsymbol{E_1}$	$r_{ho}\frac{ae}{\sigma^2}$	$r_{ho}(1 - h^2)$	$\frac{1}{2}\left(1 + r_{ho}\frac{ae}{\sigma^2}\right)$

3.1.2 Correlations arising from vertical cultural transmission

We give in the Supplementary Material evolution equations for all the quantities of interest for a model including VCT. Though there is no closed form for the equilibrium value of ρ, we show in Supplementary Material (section 4.4) that it can be computed by solving numerically a system of equations involving ρ, ν, e (which does not change over time), and g_0^2. Considering that g_0^2 and e as fixed quantities, one gets that ρ is an increasing function of ν with $\rho = 0$ when $\nu = 0$.

The gametic variance still verifies $r_{ga} = \frac{r_{A_1A_2}}{2 - r_{A_1A_2}}$ (with $r_{A_1A_2}$ as in table 2 this time), and g^2 and a^2 depends on r_{ga} in the same way as before, thus in particular $a^2 = 2\frac{1+r_{ga}}{1-r_{ga}}g_0^2$. Table 2 summarizes the correlations between the genetic and environmental values of the family members, relying on the value of ρ.

Table 2. Correlations between the genetic and the environmental effects in presence of assortative mating, without shared parent-offspring environment. Individuals 1 and 2 play symmetric roles, thus for example $r_{A_2A_3} = r_{A_1A_3}$, etc.

	$\boldsymbol{A_2}$	$\boldsymbol{E_2}$	$\boldsymbol{A_3}$	$\boldsymbol{E_3}$
$\boldsymbol{A_1}$	$\frac{r_{ho}}{\sigma^2}(a + \rho e)^2$	$\frac{r_{ho}}{\sigma^2}(a + \rho e)(\rho a + e)$	$\frac{1}{2}\left(1 + \frac{r_{ho}}{\sigma^2}(a + \rho e)^2\right)$	ρ
$\boldsymbol{E_1}$	$\frac{r_{ho}}{\sigma^2}(a + \rho e)(\rho a + e)$	$\frac{r_{ho}}{\sigma^2}(\rho a + e)^2$	$\frac{1}{2}\left(1 + \frac{r_{ho}}{\sigma^2}(a + \rho e)(\rho a + e)\right)$	ν

In Figure 2, we plot the equilibrium value of the important parameter ρ for a phenotype with $h_0^2 = 0.2$ and for different values of the two parameters r_{ho} and ν that describe the respective strength of AM and VCT in the population.

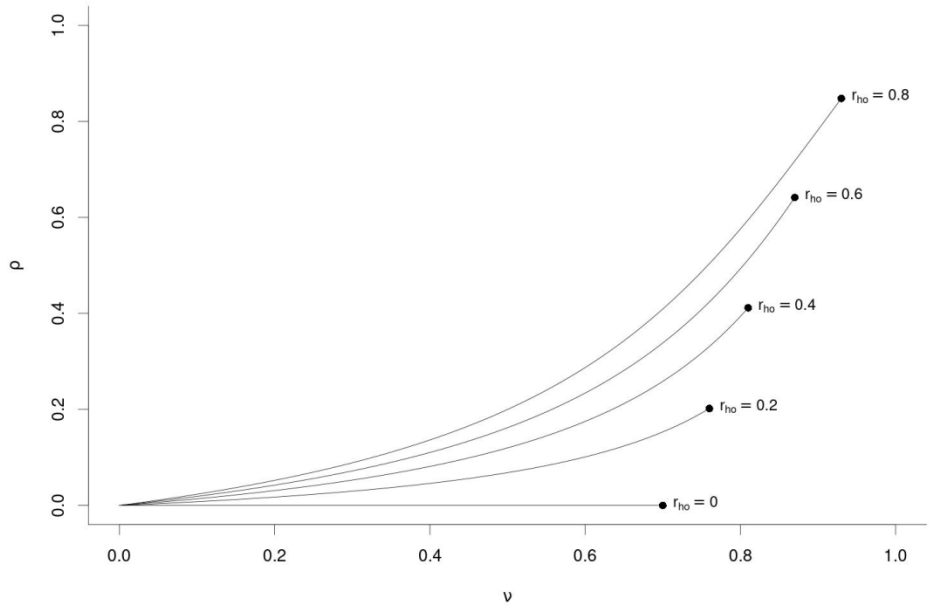

Figure 2: Equilibrium value of ρ as a function of r_{ho} and ν. Note that there is an upper bound to ν, depending on r_{ho} and h_0^2. For example, when $r_{ho} = 0$ and for any h_0^2, and thus the environments of the mates are independent, ν cannot be higher that $\frac{\sqrt{2}}{2} \simeq 0.7$. Higher values of r_{ho} imply a positive correlation between the mates' environment, allowing ν to take higher values.

Furthermore, we show the inflation of estimates of h_{SNP}^2 in the presence of co-occurring AM and VCT as a function of r_{ho} and ν (Figure 3).

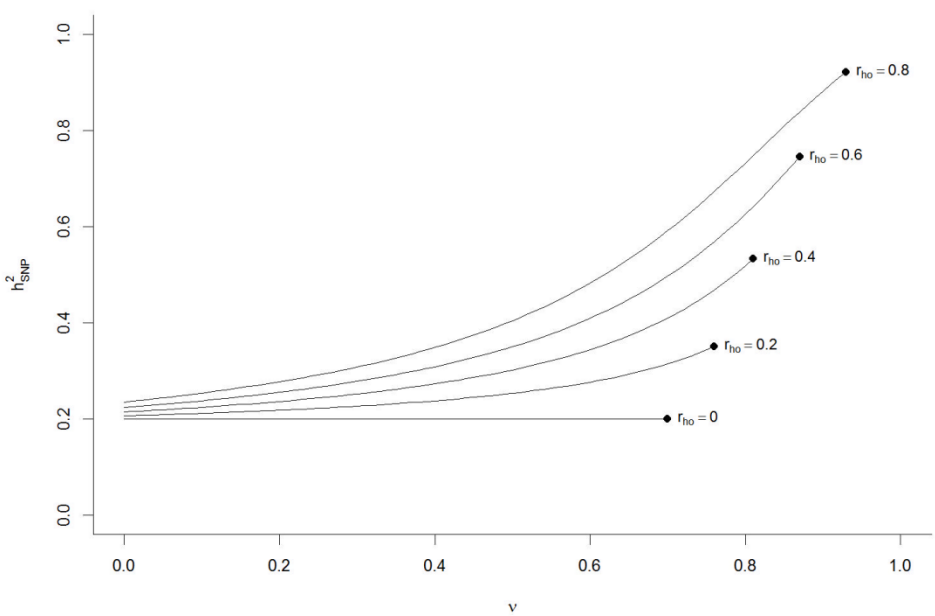

Figure 3: h_{SNP}^2 at equilibrium as a function of r_{ho} and ν for a phenotype with $h_0^2 = 0.2$ in generation 0 (before AM and VCT).

3.2 Consequences on global association statistics (SNP-heritability, GWAS, and Polygenic Scores)

In this section, we examine the impact of the AMVCT model of commonly used analyses in genetic epidemiology. The derivations of these results are given in full in section 6 in the Supplementary Material.

3.2.1 SNP-heritability

Considering pairs of unrelated individuals with a fixed value correlation between their genetic values, say $\text{cor}(A_{i,1}, A_{i,2}) = u$, where the pairs are indexed by i, then one computes $\text{cor}(P_{i,1}, P_{i,2}) = \frac{(a+\rho e)^2}{\sigma^2} u$.

Then the SNP-heritability, which can be estimated as $\text{cor}(P_{i,1}, P_{i,2})/\text{cor}(A_{i,1}, A_{i,2})$ in the spirit of Haseman-Elston regression, is equal to

$$h^2_{SNP} = \frac{(a+\rho e)^2}{\sigma^2} = \frac{1}{\sigma^2}(a^2 + 2\rho ae + \rho^2 e^2).$$

It is worth to note here that to unravel the genetic architecture of the trait, one would need to estimate separately the three phenotype variance components: a^2, $2\rho ae$ and e^2; the value of h^2_{SNP} not only includes the two first components, but even a part of the last one (full details in section 6.1 of the Supplementary Materials).

We depict the three components of total phenotypic variance that arise from the presence of the correlation ρ induced by the co-occurrence of AM and VCT in Figure 4; showing how estimates of SNP-heritability will be inflated and even capture a small part of the purely environmental variance.

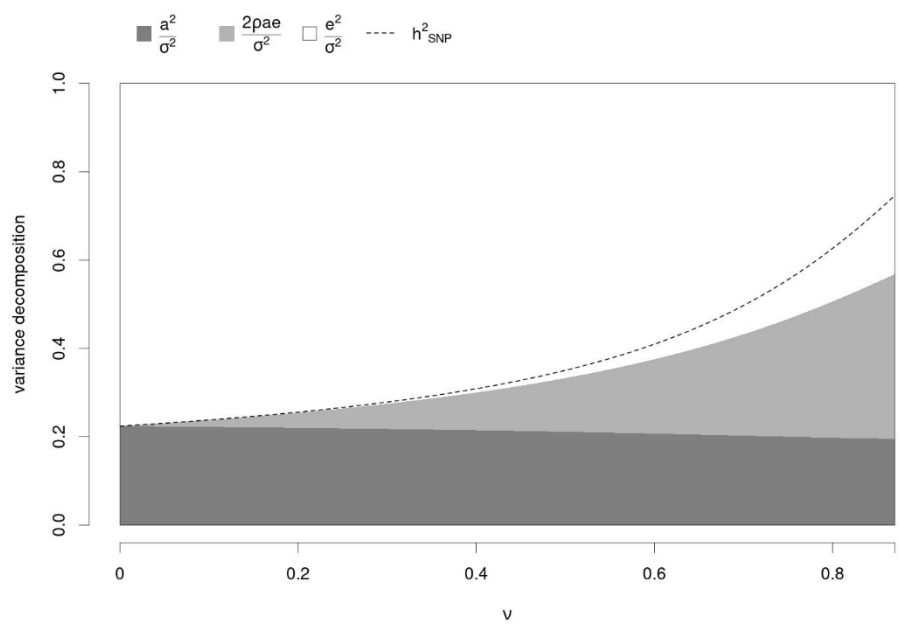

Figure 4: Decomposition of the variance as a function of ν for a population with heritability $h^2_0 = 0.2$ prior to assortative mating; here $r_{ho} = 0.6$. The dark gray part corresponds to $\frac{a^2}{\sigma^2}$, the purely genetic component of the variance; the light gray part to $\frac{2\rho ae}{\sigma^2}$, the component created by the positive correlation between genetic and environmental factors; the white part to $\frac{e^2}{\sigma^2}$, the purely environmental component. The three parts always sum to 1 as seen in the range on the y-axis. The dotted line is the value of h^2_{SNP}, which corresponds to the sum of the two first components, plus a small part of the third.

3.2.2 Consequences on marginal SNP-association statistics

We consider a locus with an allelic substitution effect β. The expected value of an estimate $\hat{\beta}$ obtained through a linear regression depends on the covariance between the phenotype P and the genotype at the locus. This covariance is impacted by the correlation between the causal loci on a given gamete, by the correlation r_{ga} between the two gametes, and by the correlation ρ between the genetic value A and the environmental effect E. We show in the Supplementary Materials (section 6.2) that $\mathrm{E}(\hat{\beta}) = \gamma \cdot \beta$ where the inflation factor γ is

$$\gamma = \left(1 + \frac{\rho e}{a}\right)\left(\frac{1 + r_{ga}}{1 - r_{ga}}\right).$$

It is worth noting that the multiplicative terms in γ all make sense separately: $\left(1 + \frac{\rho e}{a}\right)$ arises due to the gene-environment correlation ρ, $\left(\frac{1}{1-r_{ga}}\right)$ is due to the correlations between causal loci on a same gamete, and $(1 + r_{ga})$ to the correlation between the two gametes. We show the inflation factor gamma in the presence of co-occurring AM and VCT as a function of r_{ho} and υ (Figure 5).

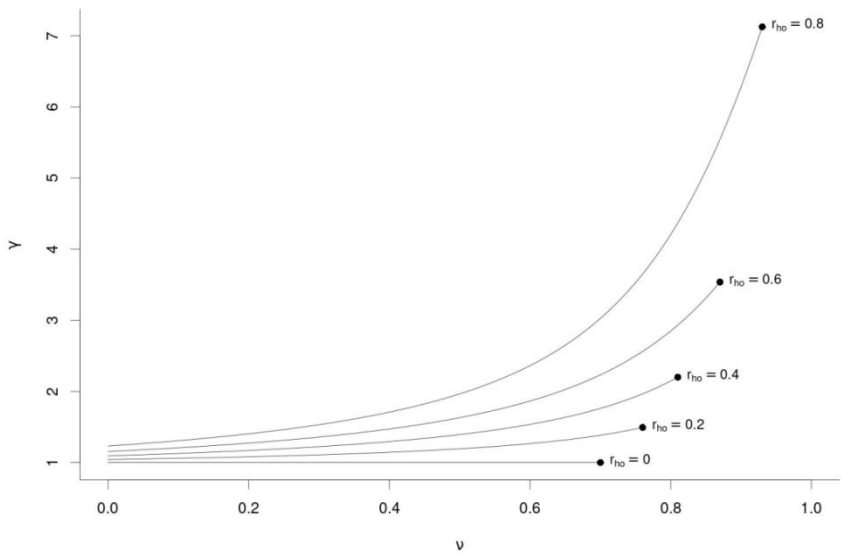

Figure 5: Inflation factor γ of SNP effect sizes at equilibrium as a function of r_{ho} and ν for a phenotype with $h_0^2 = 0.2$ in generation 0 (before AM and VCT).

3.3.3 Polygenic scores

The coefficients $\hat{\beta}$ are over-estimated, however the inflation factor γ is constant, thus the expected value of the PGS, $\hat{A}$, is (leaving aside the question of identifying the causal variants) $E(\hat{A}) = \gamma A$. The proportion of the phenotypic variance explained by the PGS will be (at most) $\mathrm{cor}(A, P)^2$, which is precisely $h^2_{SNP} = \frac{(a+\rho e)^2}{\sigma^2}$.

Of course, one must consider the fact that the values $\hat{\beta}$ are estimated on a finite sample, so this is an upper bound. As described in (Daetwyler et al., 2008), the PGS would explain less phenotypic variance than h^2_{SNP} by a factor depending on the size of the discovery sample on which the values $\hat{\beta}$ have been estimated as follows:

$$\mathrm{cor}(\hat{A}, P)^2 = \frac{h^2_{SNP}}{h^2_{SNP} + \frac{N}{M}(1 - r_{ho} h^2_{SNP})^2} h^2_{SNP}$$

where M is the size of the GWAS sample from which the $\hat{\beta}$ values are estimated. In the context on modern biobanks and large meta-analyses, the ratio $\frac{N}{M}$ could be argued to be close to 1 or even smaller (Yengo et al., 2022), and as M grows very large, this correlation will approach h^2_{SNP}. We give details of this calculation, equivalent to (Daetwyler et al., 2008) but under the AMVCT model, in section 6.3 of the Supplementary Materials. Here we do not model the fact that the variants included in the GPS are not in general the causal loci, but loci in tight linkage disequilibrium with those, and that some causal loci may be completely overlooked, which would further shrink the correlation.

3.3.4 Correlation between partial polygenic scores

One can also consider the correlation between PGS computed on different subsets of the genome. Let S_1 and S_2 be two disjoint sets of causal loci (e.g., the loci on even/odd chromosomes). Let $\alpha_1 = |S_1|/N$ and $\alpha_2 = |S_2|/N$ denote the proportion of causal variants included in these subsets. We show in section 6.4 in Supplementary Materials that the correlation of genetic values A_{S_1} and A_{S_2} computed on these sets is as follows:

$$\mathrm{cor}\left(A_{S_1}, A_{S_2}\right) = \frac{2\sqrt{\alpha_1 \alpha_2} r_{ga}}{\sqrt{\left(1 - (2\alpha_1 - 1) r_{ga}\right)\left(1 - (2\alpha_2 - 1) r_{ga}\right)}}$$

An important case is $\alpha_1 = \alpha_2 = \frac{1}{2}$, which should hold approximately when one of the sets is make of the odd numbered chromosomes, and the other one of the even numbered chromosomes (Yengo et al., 2018); in this case, one has simply $\mathrm{cor}\left(A_{S_1}, A_{S_2}\right) = r_{ga}$. In this particular case, if the partial genetic values are replaced by their estimates from polygenic scores, $\hat{A}_{S_1}$ and $\hat{A}_{S_2}$, then the correlation becomes:

$$\mathrm{cor}\left(\hat{A}_{S_1}, \hat{A}_{S_2}\right) = \frac{h^2_{SNP}}{h^2_{SNP} + \frac{N}{M}(1 + r_{ga})(1 - r_{ho} h^2_{SNP})^2} r_{ga}$$

3.3 Impact on large-scale family-based association studies

3.3.1 Mid-parent regression

In section 7 of the Supplementary Materials, we investigate further approaches for evaluating genetic architecture using family data. First, we considered estimating heritability via regression of offspring

phenotypes on mid-parent phenotypes in trios (Nagylaki, 1978) (section 7.2). This approach gives an unbiased estimate of heritability with a model including only AM; but we show that in the presence of both AM and VCT this gives an inflated estimate of heritability:

$$h^2 + \frac{(3 + r_{ho})\rho ae + 2\nu e^2}{(1 + r_{ho})\sigma^2}$$

Note that if $\nu = 0$ then $\rho = 0$ (see Figure 2) and there is no inflation in the estimate.

3.3.2 Family-based GWAS

In (Howe et al., 2022), the method of sib-GWAS for side-stepping the impact of assortative mating and other GWAS confounders was proposed. The concept is to adjust the regression of phenotype against genotype for the mean parental genotype; essentially looking to capture the true substitution effect of a SNP by examining the difference between genotypes and their expected value (mean parental genotype). As parental genotypes are often unavailable, this method uses a proxy of mean genotypes across sibships, playing on the abundance of sibships in the UK Biobank. A related and more general approach for capturing the 'direct' substitution effect of a SNP using parental genotypes can also be found in (A. I. Young et al., 2022). Having shown the inflation of estimates of a GWAS $\hat{\beta}$ under our framework, we show in section 7 of the Supplementary Materials that adjusting for mean parental genotype (section 7.3) or mean sibling genotype (section 7.4) will indeed produce unbiased estimators of β. Hence the shrinkage in effect sizes reported by (Howe et al., 2022) would correspond in our model to the inflation factor γ.

We also considered the technique of regressing sibling differences in phenotypes against sibling difference in (estimated) genetic values in section 7.5 of the Supplementary Material; a technique that can be applied (including in the presence of AM) to isolate the role of direct genetic effects from 'indirect' genetic effects coming from parental genotypes (Brumpton et al., 2020; A. I. Young et al., 2018). Under our model of AMVCT, we provide the calculation of the variance of sibling phenotype difference explained by difference in their genetic values; as has been previously investigated (Fletcher et al., 2024). Under our model of AMVCT, we arrive at the same result as (Fletcher et al., 2024): indeed, we observe that this technique isolates the direct genetic effects though it does not directly give an estimate of the 'true heritability' of the trait due to shared environmental effects. In our model, such shared environmental effects arise solely from VCT and so we parametrise them accordingly in our calculation. We also include the underestimation of the variance component for direct genetic effects in this framework when a previously estimated polygenic score is used rather than the true genetic values of the two siblings.

Finally, in section 7.6 of the Supplementary Material, we consider a further potential study design in family-based GWAS where parental alleles are used to query the 'Nature of Nurture' hypothesis (Kong et al., 2018). This proposes that the untransmitted alleles can have an (indirect) effect on the phenotype of the offspring. Here we do not explore the additional complexity that would arise from a model where the offspring phenotype truly has such an 'indirect genetic' component. We simply evaluated the expected results of comparing untransmitted alleles to the offspring phenotype under our AMVCT model for a population at equilibrium. In section 7.6 of the Supplementary Material, we first show that the correlation between the 'untransmitted polygenic score' in a trio with the offspring phenotype is as follows:

$$\mathrm{cor}(\hat{A}_u, P) = \sqrt{\frac{h^2_{SNP}}{h^2_{SNP} + \frac{N}{M}(1 - r_{ho}h^2_{SNP})^2}}\left(\frac{r_{A_1A_2}a}{\sigma} + \frac{\rho e}{\sigma}\right)$$

In this formulation, we can interpret the separate contributions from AM in the form of $\frac{r_{A_1A_2}a}{\sigma}$, of VCT in the form of $\frac{\rho e}{\sigma}$, and from the bias in the estimation of the genetic value with a polygenic score coming from the leading radical term; which was already detailed in a previous section above on 'Polygenic Scores'.

In section 7.6 of the Supplementary Material, we also show that if the untransmitted genotype is regressed against offspring phenotype for causal variant i with true allelic substitution effect β_i, then the expected regression coefficient of this 'untransmitted GWAS' will be $E(\hat{\beta}_{iu}) = (\gamma - 1)\beta_i$. Here, γ is the same shrinkage factor as detailed in section 3.2.2. This means that with a large number of trios, one could perform both a standard GWAS in the offspring and an 'untransmitted allele' GWAS and under our AMVCT model, the difference in the regression coefficients would have expected value $E(\hat{\beta}_i - \hat{\beta}_{iu}) = \gamma\beta_i - (\gamma - 1)\beta_i = \beta_i$; thus this provides another route to an unbiased estimate of the true allelic substitution effect under our model.

3.4 Forward-time Simulations

Using our simulated data (see section 2.2), we measured the observed correlation between genetic values and environmental values (i.e. A_3 and E_3) and compared it to the theoretical expectation for ρ until the 15th generation for populations of size $M = 25{,}000$ and $M = 100{,}000$ (Figure 6). Results show good concordance between the observed and the expected correlation across generations, for both population sizes.

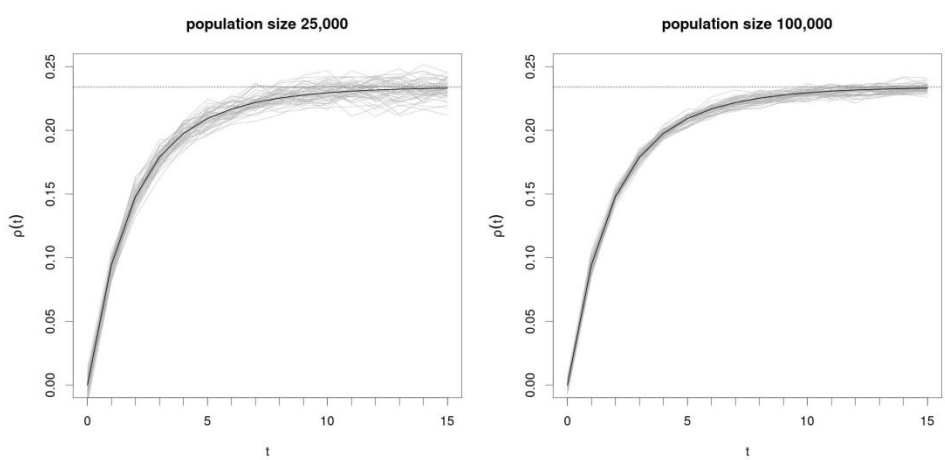

Figure 6: Simulation of the evolution of ρ, generations: t against observed ρ in generation t: $\rho(t)$. Dark line corresponds to the theoretical derivation of the evolution of ρ, the dotted line corresponds to the theoretical equilibrium point. Each light grey line shows the observed evolution of ρ in a single simulation run (50 simulation runs plotted). AM was simulated with $r_{ho} = 0.6$ and VCT with $\nu = 0.6$.

Our predictions also held for other parameters of interest (Figure 7).

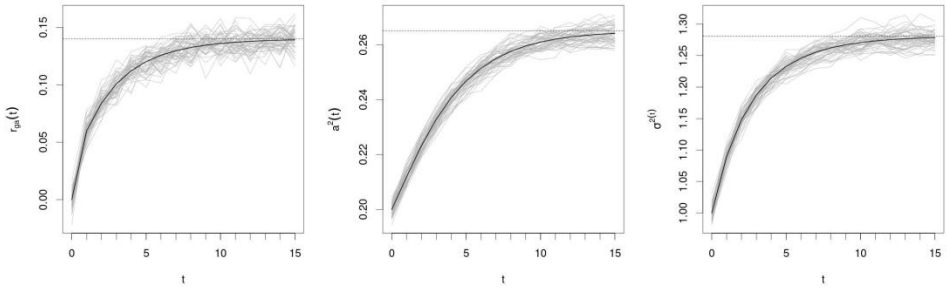

Figure 7: Evolution plots of a similar fashion to Figure 6 for the parameters r_{ga}, a^2, and σ^2. Here, only a population of size $M = 25{,}000$ is shown.

4 Discussion

Here we have explored the co-occurrence of two aspects of population architecture that have previously been shown to greatly impact genetic epidemiological studies in human populations: assortative mating and vertical cultural transmission under a classical polygenic additive model. We have set down the formal interplay of these two phenomena, crucially showing an emergent correlation ρ between genetic and non-genetic factors that leads to inflation of both heritability and genotype-phenotype association statistics.

After rising sharply for a small number of generations of both AM and VCT, all parameters eventually converge to the theoretical equilibrium points in roughly 15 generations. Through simulations, it was possible to confirm both our calculations of the equations of the evolution and the equilibrium values of ρ as well as other quantities of interest.

We were also able to evaluate the impact on estimations coming from population-based studies including inflations to estimates of SNP-heritability, GWAS effect sizes and hence polygenic scores. Whilst it may not be possible to attain a precise model for mate-choice or for the correlation of environmental factors, the huge breadth of recent genetic cross-sectional studies should allow for more nuanced approaches than assuming panmixia and completely random environmental effects.

The search for traces for assortative mating in genetic datasets has largely concentrated on observing long range correlation between variants detected through GWAS, notably using techniques such as observing genetic correlation between odd and even chromosomes (Yengo et al., 2018). An alternative method for observing genetic traces of AM in population-based studies could be to directly calculate gametic correlation though it is normally not possible to determine the required parent-of-origin information. However, recent advances in parent-of-origin determination in biobank-type datasets (Hofmeister et al., 2022) using ideas of surrogate parent groups (Kong et al., 2008) could allow such an approach to become practical.

The crucial observation in this work is that even in the absence of all confounders such as population/cultural stratification or ascertainment bias, genetic associations can still capture environmental factors. This would clearly add to the numerous difficulties in interpreting many GWAS associations and the predictive performance of PGS (Janssens, 2019; Pingault et al., 2022). In particular, if an estimated PGS is capturing a part of the environment effects, using the PGS as a covariable in order to investigate environmental effects, as discussed in (Uddin et al., 2022), would not be appropriate. Furthermore, whilst the impact of VCT and AM on Mendelian Randomization (MR) have previously been studied with methods for bias adjustment (Brumpton et al., 2020; Davies et al., 2019; Hartwig et al., 2018; Howe et al., 2019), their co-occurrence could be particularly disruptive for MR studies.

We confirmed that analysing family-data from trios or quads can provide unbiased estimates of allelic substitution effects in the presence of co-occurring AM and VCT; though we have not explored more complicated models for shared environmental effects in families beyond a simple model of positive covariance between environments between generations. In this work, we did not model 'indirect' genetic effects, where parental phenotypes (via parental genotypes) affect the phenotypes of the offspring; an in-depth exploration of this model in the presence of AM is detailed in (A. S. Young, 2023). We have however shown how the AMVCT model will lead to correlations between untransmitted alleles (in the form of single variants or as polygenic scores) with offspring phenotypes which could be interpreted as evidence of indirect genetic effects of parental genotypes; whereas these correlations arise in our model solely from the correlations between parental genetic values and between genetic and environmental components within an individual.

5 Conclusions

The list of strong model assumptions that underlie the classic additive polygenic model is extensive, of which random mating and random environmental factors are just two. We have set out a framework for the co-occurrence of AM and VCT, but it must be noted that non-random mating would be unlikely to depend on a single trait and be independent of population/cultural stratification. Such stratification could also reflect differences in the degree of VCT and so we have likely only scratched the surface of one of the processes that leads to genetic and non-genetic factors become correlated in human populations.

Author Contributions

HP and AFH derived the theoretical framework and performed the analyses. All authors participated in the review of the literature, the design of our study and the writing of the manuscript.

Data availability

Scripts for replicating the simulation studies presented in this work and for computing the evolution and equilibrium values of all parameters under the AMVCT model are available at https://github.com/HervePerdry/AMVCTpaper. No real human genetic data was used in this study.

Conflict of Interest

The authors have no conflicts to declare.

References

All of Us Research Program Investigators, Denny, J. C., Rutter, J. L., Goldstein, D. B., Philippakis, A., Smoller, J. W., Jenkins, G., & Dishman, E. (2019). The 'All of Us' Research Program. *The New England Journal of Medicine*, *381*(7), 668–676. https://doi.org/10.1056/NEJMsr1809937

Balbona, J. V., Kim, Y., & Keller, M. C. (2022). The estimation of environmental and genetic parental influences. *Development and Psychopathology*, 1–11. https://doi.org/10.1017/S0954579422000761

Border, R., Athanasiadis, G., Buil, A., Schork, A. J., Cai, N., Young, A. I., Werge, T., Flint, J., Kendler, K. S., Sankararaman, S., Dahl, A. W., & Zaitlen, N. A. (2022). Cross-trait assortative mating is widespread and inflates genetic correlation estimates. *Science (New York, N.Y.)*, *378*(6621), 754–761. https://doi.org/10.1126/science.abo2059

Border, R., O'Rourke, S., de Candia, T., Goddard, M. E., Visscher, P. M., Yengo, L., Jones, M., & Keller, M. C. (2022). Assortative mating biases marker-based heritability estimators. *Nature Communications*, *13*(1), 660. https://doi.org/10.1038/s41467-022-28294-9

Brumpton, B., Sanderson, E., Heilbron, K., Hartwig, F. P., Harrison, S., Vie, G. Å., Cho, Y., Howe, L. D., Hughes, A., Boomsma, D. I., Havdahl, A., Hopper, J., Neale, M., Nivard, M. G., Pedersen, N. L., Reynolds, C. A., Tucker-Drob, E. M., Grotzinger, A., Howe, L., … Davies, N. M. (2020). Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. *Nature Communications*, *11*(1), 3519. https://doi.org/10.1038/s41467-020-17117-4

Buss, D. M. (1985). Human mate selection. *American Scientist*, *73*, 47–51.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., & Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203–209. https://doi.org/10.1038/s41586-018-0579-z

Carey, G., & Rice, J. (1983). Genetics and personality temperament: Simplicity or complexity? *Behavior Genetics*, *13*(1), 43–63. https://doi.org/10.1007/BF01071743

Cavalli-Sforza, L. L., & Feldman, M. W. (1973). Cultural versus biological inheritance: Phenotypic transmission from parents to children. (A theory of the effect of parental phenotypes on children's phenotypes). *American Journal of Human Genetics*, *25*(6), 618–637.

Cavalli-Sforza, L. L., & Feldman, M. W. (1978). The Evolution of Continuous Variation. III. Joint Transmission of Genotype, Phenotype and Environment. *Genetics*, *90*(2), 391–425. https://doi.org/10.1093/genetics/90.2.391

Cloninger, C. R., Rice, J., & Reich, T. (1979a). Multifactorial inheritance with cultural transmission and assortative mating. II. a general model of combined polygenic and cultural inheritance. *American Journal of Human Genetics*, *31*(2), 176–198.

Cloninger, C. R., Rice, J., & Reich, T. (1979b). Multifactorial inheritance with cultural transmission and assortative mating. III. Family structure and the analysis of separation experiments. *American Journal of Human Genetics*, *31*(3), 366–388.

Crow, J. F., & Felsenstein, J. (1968). The effect of assortative mating on the genetic composition of a population. *Eugenics Quarterly*, *15*(2), 85–97. https://doi.org/10.1080/19485565.1968.9987760

Daetwyler, H. D., Villanueva, B., & Woolliams, J. A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PloS One*, *3*(10), e3395. https://doi.org/10.1371/journal.pone.0003395

Davies, N. M., Howe, L. J., Brumpton, B., Havdahl, A., Evans, D. M., & Davey Smith, G. (2019). Within family Mendelian randomization studies. *Human Molecular Genetics*, *28*(R2), R170–R179. https://doi.org/10.1093/hmg/ddz204

de Jong, S., Diniz, M. J. A., Saloma, A., Gadelha, A., Santoro, M. L., Ota, V. K., Noto, C., Major Depressive Disorder and Bipolar Disorder Working Groups of the Psychiatric Genomics Consortium, Curtis, C., Newhouse, S. J., Patel, H., Hall, L. S., O Reilly, P. F., Belangero, S. I., Bressan, R. A., & Breen, G. (2018). Applying polygenic risk scoring for psychiatric disorders to a large family with bipolar disorder and major depressive disorder. *Communications Biology*, *1*, 163. https://doi.org/10.1038/s42003-018-0155-y

Falconer, D. (1960). Introduction to quantitative genetics. In *Introduction to quantitative genetics.* Oliver & Boyd, Edinburgh & London.; CABI Databases.

Feldman, M. W., & Cavalli-Sforza, L. L. (1977). The evolution of continuous variation. II. Complex transmission and assortative mating. *Theoretical Population Biology*, *11*(2), 161–181. https://doi.org/10.1016/0040-5809(77)90024-7

Fisher, R. A. (1919). XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Earth and Environmental Science Transactions of The Royal Society of Edinburgh*, *52*(2), 399–433. https://doi.org/10.1017/S0080456800012163

Fletcher, J., Wu, Y., Li, T., & Lu, Q. (2024). Interpreting polygenic score effects in sibling analysis. *PLOS ONE*, *19*(2), e0282212. https://doi.org/10.1371/journal.pone.0282212

Hartwig, F. P., Davies, N. M., & Davey Smith, G. (2018). Bias in Mendelian randomization due to assortative mating. *Genetic Epidemiology*, *42*(7), 608–620. https://doi.org/10.1002/gepi.22138

Hippisley-Cox, J., Coupland, C., Pringle, M., Crown, N., & Hammersley, V. (2002). Married couples' risk of same disease: Cross sectional study. *BMJ (Clinical Research Ed.)*, *325*(7365), 636. https://doi.org/10.1136/bmj.325.7365.636

Hofmeister, R. J., Rubinacci, S., Ribeiro, D. M., Buil, A., Kutalik, Z., & Delaneau, O. (2022). Parent-of-Origin inference for biobanks. *Nature Communications*, *13*(1), 6668. https://doi.org/10.1038/s41467-022-34383-6

Howe, L. J., Battram, T., Morris, T. T., Hartwig, F. P., Hemani, G., Davies, N. M., & Smith, G. D. (2021). Assortative mating and within-spouse pair comparisons. *PLoS Genetics*, *17*(11), e1009883. https://doi.org/10.1371/journal.pgen.1009883

Howe, L. J., Lawson, D. J., Davies, N. M., St Pourcain, B., Lewis, S. J., Davey Smith, G., & Hemani, G. (2019). Genetic evidence for assortative mating on alcohol consumption in the UK Biobank. *Nature Communications*, *10*(1), 5039. https://doi.org/10.1038/s41467-019-12424-x

Howe, L. J., Nivard, M. G., Morris, T. T., Hansen, A. F., Rasheed, H., Cho, Y., Chittoor, G., Ahlskog, R., Lind, P. A., Palviainen, T., van der Zee, M. D., Cheesman, R., Mangino, M., Wang, Y., Li, S., Klaric, L., Ratliff, S. M., Bielak, L. F., Nygaard, M., … Davies, N. M. (2022). Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. *Nature Genetics*, *54*(5), 581–592. https://doi.org/10.1038/s41588-022-01062-7

Hugh-Jones, D., Verweij, K. J. H., St. Pourcain, B., & Abdellaoui, A. (2016). Assortative mating on educational attainment leads to genetic spousal resemblance for polygenic scores. *Intelligence*, *59*, 103–108. https://doi.org/10.1016/j.intell.2016.08.005

Janssens, A. C. J. W. (2019). Validity of polygenic risk scores: Are we measuring what we think we are? *Human Molecular Genetics*, *28*(R2), R143–R150. https://doi.org/10.1093/hmg/ddz205

Jefsen, O. H., Nudel, R., Wang, Y., Bybjerg-Grauholm, J., Hemager, N., Christiani, C. A. J., Burton, B. K., Spang, K. S., Ellersgaard, D., Gantriis, D. L., Plessen, K. J., Jepsen, J. R. M., Thorup, A. A. E., Werge, T., Nordentoft, M., Mors, O., & Greve, A. N. (2022). Genetic assortative mating for schizophrenia and bipolar disorder. *European Psychiatry: The Journal of the Association of European Psychiatrists*, *65*(1), e53. https://doi.org/10.1192/j.eurpsy.2022.2304

Keller, M. C., Medland, S. E., Duncan, L. E., Hatemi, P. K., Neale, M. C., Maes, H. H. M., & Eaves, L. J. (2009). Modeling Extended Twin Family Data I: Description of the Cascade Model. *Twin Research and Human Genetics : The Official Journal of the International Society for Twin Studies*, *12*(1), 8–18. https://doi.org/10.1375/twin.12.1.8

Kong, A., Masson, G., Frigge, M. L., Gylfason, A., Zusmanovich, P., Thorleifsson, G., Olason, P. I., Ingason, A., Steinberg, S., Rafnar, T., Sulem, P., Mouy, M., Jonsson, F., Thorsteinsdottir, U., Gudbjartsson, D. F., Stefansson, H., & Stefansson, K. (2008). Detection of sharing by descent, long-range phasing and haplotype imputation. *Nature Genetics*, *40*(9), 1068–1075. https://doi.org/10.1038/ng.216

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E., Benonisdottir, S., Oddsson, A., Halldorsson, B. V., Masson, G., Gudbjartsson, D. F., Helgason, A., Bjornsdottir, G., Thorsteinsdottir, U., & Stefansson, K. (2018). The nature of nurture: Effects of parental genotypes. *Science (New York, N.Y.)*, *359*(6374), 424–428. https://doi.org/10.1126/science.aan6877

Larrègue, J., Lebaron, F., Perdry, H., & Robette, N. (2023). Eurythmics or Xenakis? Cultural Tastes (Are Not Made of Genes): Comment on Jæger and Møllegaard,. *Sociological Science*, *10*, 454–466. https://doi.org/10.15195/v10.a15

Mare, R. D. (1991). Five decades of educational assortative mating. *American Sociological Review*, *56*, 15–32. https://doi.org/10.2307/2095670

Martin, N. G., Eaves, L. J., Heath, A. C., Jardine, R., Feingold, L. M., & Eysenck, H. J. (1986). Transmission of social attitudes. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(12), 4364–4368. https://doi.org/10.1073/pnas.83.12.4364

Morris, T. T., Davies, N. M., Hemani, G., & Smith, G. D. (2020). Population phenomena inflate genetic associations of complex social traits. *Science Advances*, *6*(16), eaay0328. https://doi.org/10.1126/sciadv.aay0328

Nagai, A., Hirata, M., Kamatani, Y., Muto, K., Matsuda, K., Kiyohara, Y., Ninomiya, T., Tamakoshi, A., Yamagata, Z., Mushiroda, T., Murakami, Y., Yuji, K., Furukawa, Y., Zembutsu, H., Tanaka, T., Ohnishi, Y., Nakamura, Y., & Kubo, M. (2017). Overview of the BioBank Japan Project: Study design and profile. *Journal of Epidemiology*, *27*(3 Suppl), S2–S8. https://doi.org/10.1016/j.je.2016.12.005

Nagylaki, T. (1978). The correlation between relatives with assortative mating. *Annals of Human Genetics*, *42*(1), 131–137. https://doi.org/10.1111/j.1469-1809.1978.tb00938.x

Nordsletten, A. E., Larsson, H., Crowley, J. J., Almqvist, C., Lichtenstein, P., & Mataix-Cols, D. (2016). Patterns of Nonrandom Mating Within and Across 11 Major Psychiatric Disorders. *JAMA Psychiatry*, *73*(4), 354–361. https://doi.org/10.1001/jamapsychiatry.2015.3192

Pearson, K., & Lee, A. (1903). On the Laws of Inheritance in Man: I. Inheritance of Physical Characters. *Biometrika*, *2*(4), 357–462. https://doi.org/10.2307/2331507

Pingault, J.-B., Allegrini, A. G., Odigie, T., Frach, L., Baldwin, J. R., Rijsdijk, F., & Dudbridge, F. (2022). Research Review: How to interpret associations between polygenic scores, environmental risks, and phenotypes. *Journal of Child Psychology and Psychiatry*, *63*(10), 1125–1139. https://doi.org/10.1111/jcpp.13607

Rao, D. C., Morton, N. E., & Yee, S. (1976). Resolution of cultural and biological inheritance by path analysis. *American Journal of Human Genetics*, *28*(3), 228–242.

Rice, J., Cloninger, C. R., & Reich, T. (1978). Multifactorial inheritance with cultural transmission and assortative mating. I. Description and basic properties of the unitary models. *American Journal of Human Genetics*, *30*(6), 618–643.

Rice, J., Cloninger, C. R., & Reich, T. (1980). Analysis of behavioral traits in the presence of cultural transmission and assortative mating: Applications to IQ and SES. *Behavior Genetics*, *10*(1), 73–92. https://doi.org/10.1007/BF01067320

Stulp, G., Simons, M. J. P., Grasman, S., & Pollet, T. V. (2017). Assortative mating for human height: A meta-analysis. *American Journal of Human Biology: The Official Journal of the Human Biology Council*, *29*(1), e22917. https://doi.org/10.1002/ajhb.22917

Torvik, F. A., Eilertsen, E. M., Hannigan, L. J., Cheesman, R., Howe, L. J., Magnus, P., Reichborn-Kjennerud, T., Andreassen, O. A., Njølstad, P. R., Havdahl, A., & Ystrom, E. (2022). Modeling assortative mating and genetic similarities between partners, siblings, and in-laws. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-28774-y

Truett, K. R., Eaves, L. J., Walters, E. E., Heath, A. C., Hewitt, J. K., Meyer, J. M., Silberg, J., Neale, M. C., Martin, N. G., & Kendler, K. S. (1994). A model system for analysis of family resemblance in extended kinships of twins. *Behavior Genetics*, *24*(1), 35–49. https://doi.org/10.1007/BF01067927

Uddin, M. J., Hjorthøj, C., Ahammed, T., Nordentoft, M., & Ekstrøm, C. T. (2022). The use of polygenic risk scores as a covariate in psychological studies. *Methods in Psychology*, *7*, 100099. https://doi.org/10.1016/j.metip.2022.100099

Veller, C., & Coop, G. (2023). *Interpreting population and family-based genome-wide association studies in the presence of confounding* (p. 2023.02.26.530052). bioRxiv. https://doi.org/10.1101/2023.02.26.530052

Vinkhuyzen, A. A. E., van der Sluis, S., Maes, H. H. M., & Posthuma, D. (2012). Reconsidering the Heritability of Intelligence in Adulthood: Taking Assortative Mating and Cultural Transmission into Account. *Behavior Genetics*, *42*(2), 187–198. https://doi.org/10.1007/s10519-011-9507-9

Vogler, G. P., & Fulker, D. W. (1983). Familial resemblance for educational attainment. *Behavior Genetics*, *13*(4), 341–354. https://doi.org/10.1007/BF01065772

Wright, S. (1921). Systems of Mating. III. Assortative Mating Based on Somatic Resemblance. *Genetics*, *6*(2), 144–161. https://doi.org/10.1093/genetics/6.2.144

Yamamoto, K., Sonehara, K., Namba, S., Konuma, T., Masuko, H., Miyawaki, S., BioBank Japan Project, Kamatani, Y., Hizawa, N., Ozono, K., Yengo, L., & Okada, Y. (2023). Genetic footprints of assortative mating in the Japanese population. *Nature Human Behaviour*, *7*(1), 65–73. https://doi.org/10.1038/s41562-022-01438-z

Yengo, L., Robinson, M. R., Keller, M. C., Kemper, K. E., Yang, Y., Trzaskowski, M., Gratten, J., Turley, P., Cesarini, D., Benjamin, D. J., Wray, N. R., Goddard, M. E., Yang, J., & Visscher, P. M. (2018).

Imprint of assortative mating on the human genome. *Nature Human Behaviour*, *2*(12), 948–954. https://doi.org/10.1038/s41562-018-0476-3

Yengo, L., Vedantam, S., Marouli, E., Sidorenko, J., Bartell, E., Sakaue, S., Graff, M., Eliasen, A. U., Jiang, Y., Raghavan, S., Miao, J., Arias, J. D., Graham, S. E., Mukamel, R. E., Spracklen, C. N., Yin, X., Chen, S.-H., Ferreira, T., Highland, H. H., … Hirschhorn, J. N. (2022). A saturated map of common genetic variants associated with human height. *Nature*, *610*(7933), 704–712. https://doi.org/10.1038/s41586-022-05275-y

Young, A. I., Frigge, M. L., Gudbjartsson, D. F., Thorleifsson, G., Bjornsdottir, G., Sulem, P., Masson, G., Thorsteinsdottir, U., Stefansson, K., & Kong, A. (2018). Relatedness disequilibrium regression estimates heritability without environmental bias. *Nature Genetics*, *50*(9), 1304–1310. https://doi.org/10.1038/s41588-018-0178-9

Young, A. I., Nehzati, S. M., Benonisdottir, S., Okbay, A., Jayashankar, H., Lee, C., Cesarini, D., Benjamin, D. J., Turley, P., & Kong, A. (2022). Mendelian imputation of parental genotypes improves estimates of direct genetic effects. *Nature Genetics*, *54*(6), 897–905. https://doi.org/10.1038/s41588-022-01085-0

Young, A. S. (2023). Estimation of indirect genetic effects and heritability under assortative mating. *bioRxiv: The Preprint Server for Biology*, 2023.07.10.548458. https://doi.org/10.1101/2023.07.10.548458

Zaitlen, N., Kraft, P., Patterson, N., Pasaniuc, B., Bhatia, G., Pollack, S., & Price, A. L. (2013). Using Extended Genealogy to Estimate Components of Heritability for 23 Quantitative and Dichotomous Traits. *PLoS Genetics*, *9*(5), e1003520. https://doi.org/10.1371/journal.pgen.1003520