A model for co-occurrent assortative mating and vertical cultural transmission and its impact 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, France
(3) CESP Inserm, U1018, Université Paris-Saclay, UVSQ, Villejuif, France

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 modeling of this phenomenon, in the classical framework of the polygenic additive model. We then derive its consequences on typical genetic epidemiology study designs. We validate our calculations through simple forward-time simulations.

Introduction

Assortative mating (AM) describes the process of non-random mate choice that gives arise to 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 and Felsenstein, 1968), Wright (Wright, 1921) and Nagylaki (Nagylaki, 1978). Importantly, it has been shown that AM will result in long distance correlations between genetic variants that carry 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 and 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 et al., 2022a; Yamamoto et al., 2023; Yengo et al., 2018), validating some predictions of 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., 2022, 2021), heritability analysis (Border et al., 2022b), 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 and 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 environment. Its importance has long been recognized and is a motivation for the introduction of twin studies in estimating heritability. A similar notion 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 (Young et al., 2018; Zaitlen et al., 2013) or GWAS (Howe et al., 2022; 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 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 and their combined impact has so far not been described to our knowledge. In this work we explore exactly such a co-occurrence. We attain 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. The 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. Furthermore, we show that such inflation remains even when applying within-family based estimates. Simple forward-time simulations are presented to validate our calculations and demonstrate the impact of co-occurring AM and VCT.

Methods

The polygenic additive model with assortative mating and vertical transmission

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 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 $\mathrm{var}(A) = a^2$, $\mathrm{var}(E) = e^2$ and $\mathrm{var}(P) = \sigma^2$. 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 $h^2 = \frac{a^2}{\sigma^2}$.

Introducing homogamy

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 $r_{ho} = r_{P_1P_2}$ the correlation between the mates' phenotypes. This correlation implies that there is some 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 can be computed (see Supplementary Material and (Nagylaki, 1978)). The 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 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 entails a genetic variance $a^2 = 2\frac{1+r_{ga}}{1-r_{ga}}g_0^2$. Additionally, we give in the supplementary material equations for the evolution of these quantities across generations.

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 $r_{E_2E_3} = r_{E_1E_3}$.

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

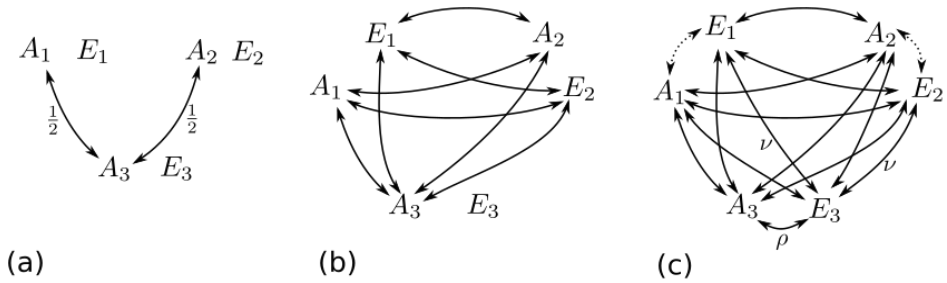

Figure 1. The correlations between genetic values and environments of parents (indexed by 1, 2) and offspring (indexed by 3), in models of increasing complexity. Panel (a) corresponds to the simple model of a panmictic population; (b) corresponds to a model of homogamy, in which the positive phenotypic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

<u>Introducing vertical cultural transmission</u>

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. The value of ρ will increase generation after generation, until reaching an equilibrium point. We give in the supplementary material evolution equations for all the quantities involved. Though there is no closed form for the equilibrium value of ρ, we show in supplementary material that it can be computed by solving numerically a functional equation 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.

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.

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

Consequences on marginal SNP association statistics

We consider a locus with an allelic 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 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 $\left(1+r_{ga}\right)$ to the correlation between the two gametes.

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

SNP-heritability

Considering pairs of unrelated individuals with a fixed value correlation between their genetic values, say $\mathrm{cor}(A_{i,1}, A_{i,2}) = u$, where the pairs are indexed by i, then one computes that $\mathrm{cor}(A_{i,1}, E_{i,2}) = \rho u$ and $\mathrm{cor}(E_{i,1}, E_{i,2}) = \rho^2 u$. From this, it comes that

$$\mathrm{cor}(P_{i,1}, P_{i,2}) = \frac{(a+\rho e)^2}{\sigma^2} u$$

and the SNP-heritability, which can be estimated as $\mathrm{cor}(P_{i,1}, P_{i,2})/\mathrm{cor}(A_{i,1}, A_{i,2})$ 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 while one would need to estimate separately the three 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.

Polygenic scores

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

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. It would moreover be necessary to model the fact that the variants included in the GPS are not in general the causal loci, but loci in (hopefully) tight linkage disequilibrium with those.

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. The genetic values A_{S_1} and A_{S_2} computed on these sets have variance

$$var(A_{S_i}) = 2g_0^2\, \alpha_i \left(1 + \alpha_i \frac{r_{ga}}{1 - r_{ga}}\right), \quad (i = 1,2)$$

and covariance

$$\mathrm{cov}(A_{S_1}, A_{S_2}) = 4g_0^2\, \alpha_1 \alpha_2 \frac{r_{ga}}{1 - r_{ga}},$$

leading to a correlation

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

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; in this case, one has simply $\mathrm{cor}\left(A_{S_1}, A_{S_2}\right) = r_{ga}$.

Consequences on sib-GWAS analysis

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 approach for capturing the 'direct' substitution effect of a SNP using parental genotypes can also be found in (Young et al., 2022).

Having shown the inflation of estimates of a GWAS $\hat{\beta}$ under our framework, we can also show the consequences on an approach that adjusts for mean parental genotype. Firstly, we can note that if the expected value of a GWAS $\hat{\beta}$ was $\gamma\beta$ (γ was derived previously and β was the effect size of the SNP in generation 0), it is simple to show that if we were to regress the un-transmitted genotypeagainst the phenotype, we would have an expected regression coefficient of $r_{ga}\gamma\beta$. In supplementary materials, we can use this result to show that the effect size of a SNP using such an adjustment of mean parental genotype will be $\left(1 - r_{ga}\right)\gamma\beta$. Therefore, a shrinkage in effect sizes, as reported by Howe et al, corresponding to $\left(1 - r_g\right)$ is indeed expected but the final estimate remains inflated by a factor of $\left(1 + \frac{\rho e}{a}\right)\left(1 + r_{ga}\right)$ compared to the 'true' substitution effect.

Results

We performed simple simulations to demonstrate the validity of our evolution and equilibrium calculations for a population of size M 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 $N = 1000$ independent causal SNPs and an architecture of $h^2 = 0.2$ (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 allelic effects were simulated as normally distributed.

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 on average 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) based using the parameter ν to ensure the correct covariance structure between E_1, E_2 and E_3 in each trio. We then 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 2).

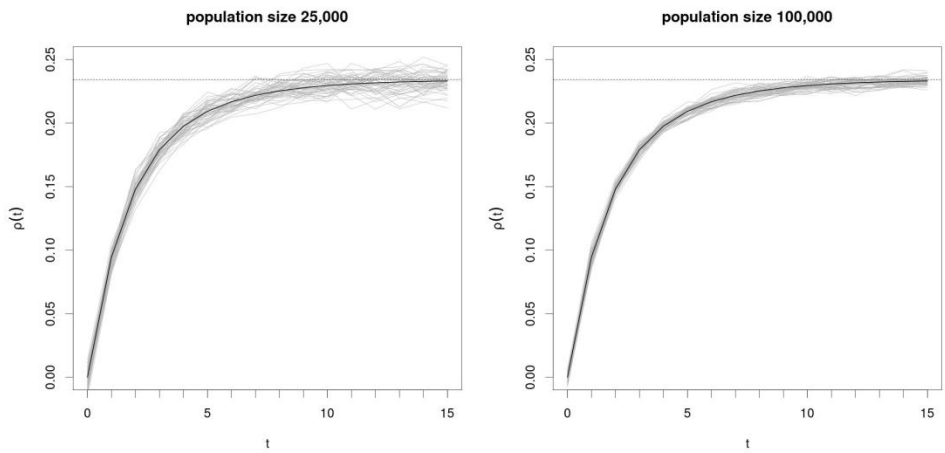

Figure 2: Simulation of the evolution of ρ , generations: t against observed ρ in generation t: $\rho(t)$. Dark line corresponds to the theoretical derivation of the evolution of ρ , the dotted line corresponds to the theoretical equilibrium point. Each light grey line shows the observe d evolution of ρ in a single simulation run (50 simulations plotted). AM was simulated with $r_{ho} = 0.6$ and VCT with $\nu = 0.6$.

Our predictions also held for other parameters (figure 3).

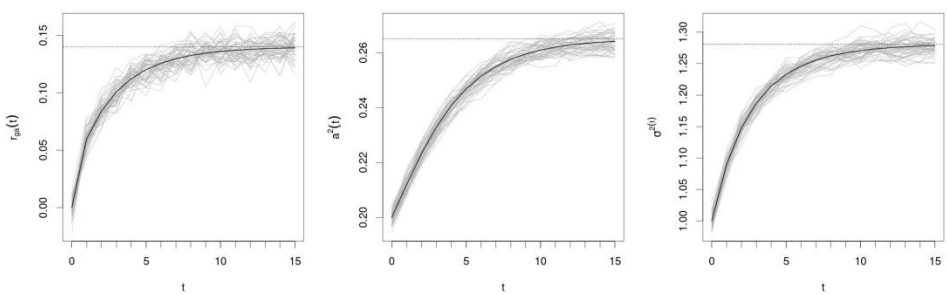

Figure 3: Evolution plots of a similar fashion to figure 2 for the parameters r_{ga}, a^2, and σ^2. Here, only a population of size $M = 25{,}000$ is shown.

Having established the validity of our calculations, we now demonstrate the equilibrium point of the important parameter ρ for a phenotype with $h^2 = 0.2$ 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 4).

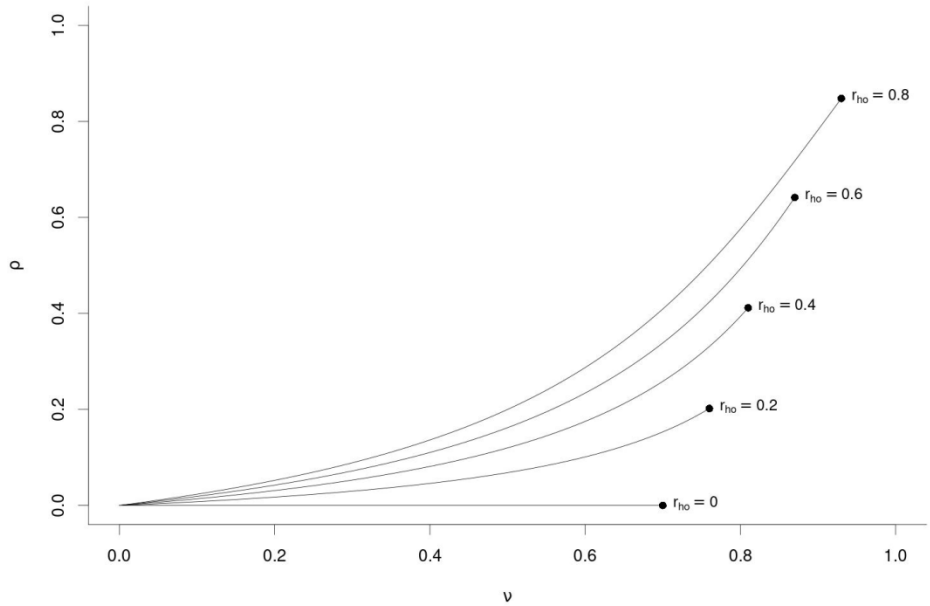

Figure 4: Equilibrium value of ρ as a function of r_{ho} and ν. Note that there is an upper bound to ν, depending on r_{ho}. For example, when $r_{ho} = 0$, and thus the environments of the mates are independent, ν cannot be higher that $\frac{\sqrt{2}}{2} \simeq 0.7$. Higher values of r_{ho} imply a positive correlation between the mates' environment, allowing ν to take higher values.

Moreover, we show the inflation of estimates of h^2_{SNP} and a given $\hat{\beta}$ in the presence of co-occurring AM and VCT as a function of r_{ho} and ν (figures 5 and 6).

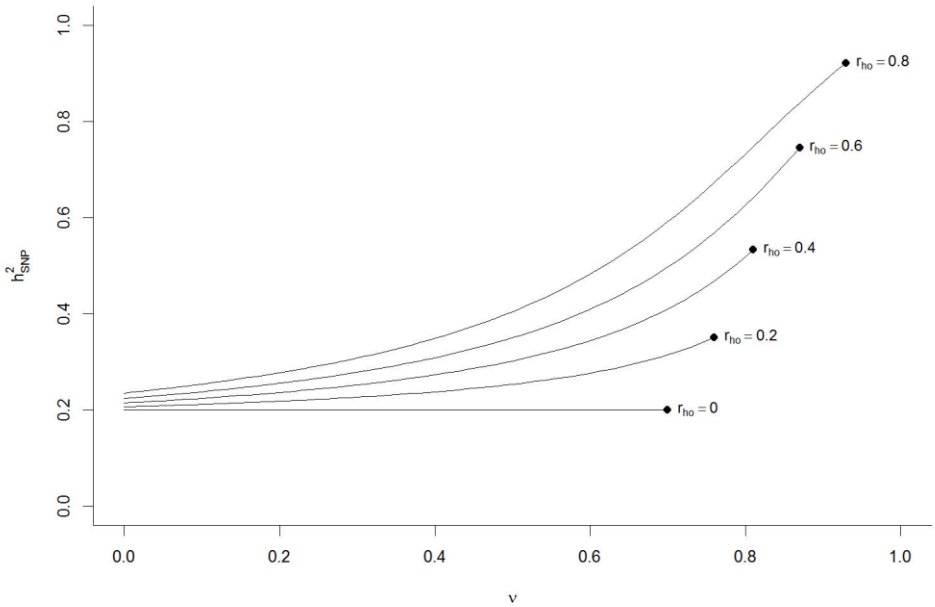

Figure 5: h^2_{SNP} at equilibrium as a function of r_{ho} and ν for a phenotype with $h^2 = 0.2$ in generation 0 (before AM and VCT).

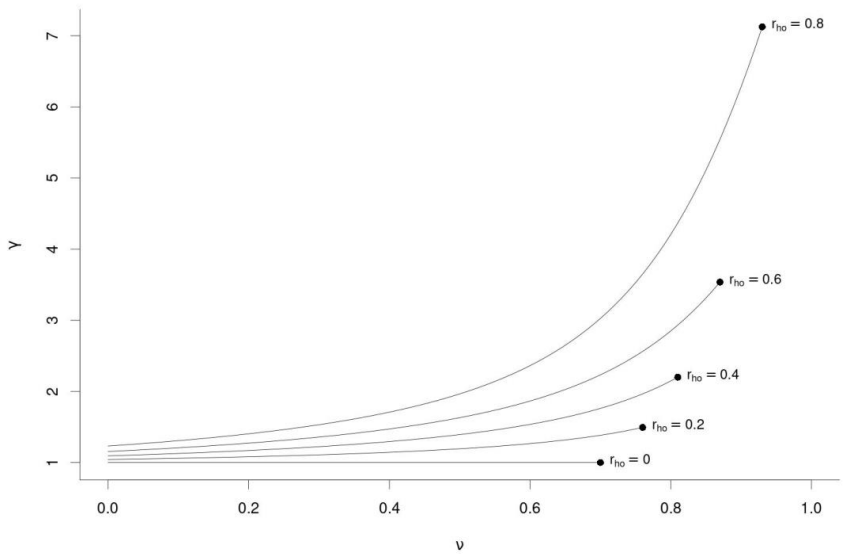

Figure 6: Inflation factor γ of SNP effect sizes at equilibrium as a function of r_{ho} and ν for a phenotype with $h^2 = 0.2$ in generation 0 (before AM and VCT).

Finally, we demonstrate 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 (figure 7); showing how estimates of SNP-heritability will be inflated and even capture a small part of the purely environmental variance.

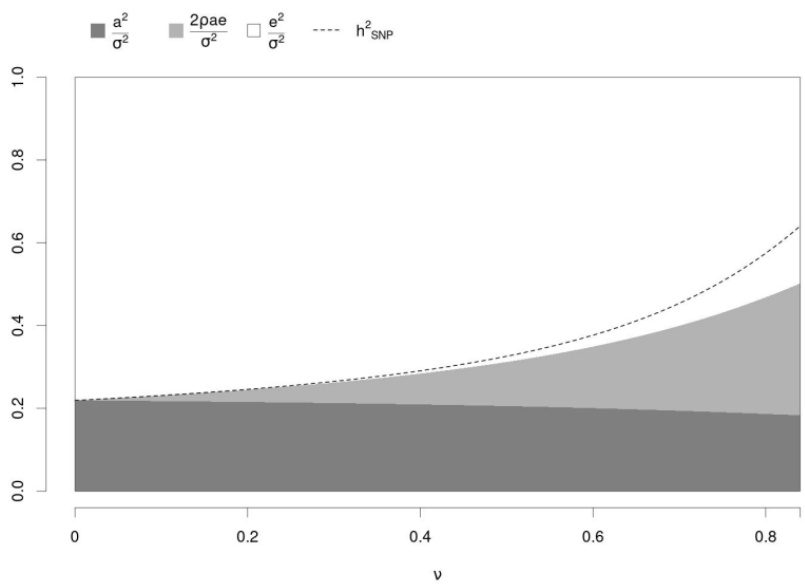

Figure 7: Decomposition of the variance as a function of ν for a population with heritability $h^2 = 0.2$ in absence of assortative mating; here $r_{ho} = 0.6$. The dark gray part corresponds to $\frac{a^2}{\sigma^2}$, the purely genetic component of the variance; the light gray part to $\frac{2\rho ae}{\sigma^2}$, 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 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.

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 additional 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. Through simulation, it was possible to confirm our calculations; after a small number of generations of both AM and VCT, we could determine the equilibrium point for the covariance structure of genetic and non-genetic components within a trio and hence the impact on estimations coming from population-based studies including inflations to estimates of SNP-heritability, GWAS effect sizes and hence polygenic scores. Furthermore, we showed that controlling for parental genotypes would not remove all sources of inflation. 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 do try to directly calculate gametic correlation but 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). Furthermore, is an PGS estimated PGS is capturing a part of the environments, using the PGS as a covariable in order to investigate environmental effects, as discussed in Uddin et al 2022 (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 proposed (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.

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.

Acknowledgements

AFH was funded by the POPGEN project as part of the Plan Médecine Génomique 2025 (FMG2025/POPGEN).

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.

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. N Engl J Med 381, 668–676. https://doi.org/10.1056/NEJMsr1809937

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., 2022a. Cross-trait assortative mating is widespread and inflates genetic correlation estimates. Science 378, 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., 2022b. Author Correction: Assortative mating biases marker-based heritability estimators. Nat Commun 13, 1903. https://doi.org/10.1038/s41467-022-29652-3

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., Morris, T., Li, S., Within-family Consortium, 23andMe Research Team, Auton, A., Windmeijer, F., Chen, W.-M., Bjørngaard, J.H., Hveem, K., Willer, C., Evans, D.M., Kaprio, J., Davey Smith, G., Åsvold, B.O., Hemani, G., Davies, N.M., 2020. Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. Nat Commun 11, 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, 203–209. https://doi.org/10.1038/s41586-018-0579-z

Crow, J.F., Felsenstein, J., 1968. The effect of assortative mating on the genetic composition of a population. Eugen Q 15, 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, 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. Hum Mol Genet 28, 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. Commun Biol 1, 163. https://doi.org/10.1038/s42003-018-0155-y

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, 399–433. https://doi.org/10.1017/S0080456800012163

Hartwig, F.P., Davies, N.M., Davey Smith, G., 2018. Bias in Mendelian randomization due to assortative mating. Genet Epidemiol 42, 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 325, 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. Nat Commun 13, 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 Genet 17, 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. Nat Commun 10, 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., Giannelis, A., Willoughby, E.A., Reynolds, C.A., Balbona, J.V., Andreassen, O.A., Ask, H., Baras, A., Bauer, C.R., Boomsma, D.I., Campbell, A., Campbell, H., Chen, Z., Christofidou, P., Corfield, E., Dahm, C.C., Dokuru, D.R., Evans, L.M., de Geus, E.J.C., Giddaluru, S., Gordon, S.D., Harden, K.P., Hill, W.D., Hughes, A., Kerr, S.M., Kim, Y., Kweon, H., Latvala, A., Lawlor, D.A., Li, L., Lin, K., Magnus, P., Magnusson, P.K.E., Mallard, T.T., Martikainen, P., Mills, M.C., Njølstad, P.R., Overton, J.D., Pedersen, N.L., Porteous, D.J., Reid, J., Silventoinen, K., Southey, M.C., Stoltenberg, C., Tucker-Drob, E.M., Wright, M.J., Social Science Genetic Association Consortium, Within Family Consortium, Hewitt, J.K., Keller, M.C., Stallings, M.C., Lee, J.J., Christensen, K., Kardia, S.L.R., Peyser, P.A., Smith, J.A., Wilson, J.F., Hopper, J.L., Hägg, S., Spector, T.D., Pingault, J.-B., Plomin, R., Havdahl, A., Bartels, M., Martin, N.G., Oskarsson, S., Justice, A.E., Millwood, I.Y., Hveem, K., Naess, Ø., Willer, C.J., Åsvold, B.O., Koellinger, P.D., Kaprio, J., Medland, S.E., Walters, R.G., Benjamin, D.J., Turley, P., Evans, D.M., Davey Smith, G., Hayward, C., Brumpton, B., Hemani, G., Davies, N.M., 2022. Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. Nat Genet 54, 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? Hum Mol Genet 28, 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. Eur Psychiatry 65, e53. https://doi.org/10.1192/j.eurpsy.2022.2304

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. Nat Genet 40, 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 359, 424–428. https://doi.org/10.1126/science.aan6877

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

Morris, T.T., Davies, N.M., Hemani, G., Smith, G.D., 2020. Population phenomena inflate genetic associations of complex social traits. Sci Adv 6, 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. J Epidemiol 27, 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, 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, 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, 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, 1125–1139. https://doi.org/10.1111/jcpp.13607

Stulp, G., Simons, M.J.P., Grasman, S., Pollet, T.V., 2017. Assortative mating for human height: A meta-analysis. Am J Hum Biol 29, 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. Nat Commun 13, 1108. https://doi.org/10.1038/s41467-022-28774-y

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. https://doi.org/10.1101/2023.02.26.530052

Wright, S., 1921. Systems of Mating. III. Assortative Mating Based on Somatic Resemblance. Genetics 6, 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. Nat Hum Behav 7, 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. Nat Hum Behav 2, 948–954. https://doi.org/10.1038/s41562-018-0476-3

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. Nat Genet 50, 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. Nat Genet 54, 897–905. https://doi.org/10.1038/s41588-022-01085-0

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 Genet 9, e1003520. https://doi.org/10.1371/journal.pgen.1003520