FAMILY

"Running in the FAMILY – Understanding and predicting the intergenerational transmission of mental illness"

Grant Agreement number: 101057529

Deliverable 3.1

Report on the software pipelines based on simulation studies

Workpackage:	WP 3
Task:	T 3.1
Due Date:	30th March 2024 (M18)
Actual Submission Date:	27th May 2024 (M20)
Project Dates:	Project Start Date: October 01, 2022 Project Duration: 60 months
Responsible partner:	UCL
Responsible author:	Andrea Allegrini, Jean-Baptiste Pingault
Email:	j.pingault@ucl.ac.uk
Contributors:	Jean-Baptiste Pingault

Project funded by the European Commission within HORIZON-HLTH-2021-STAYHLTH-01-02: Towards a molecular and neurobiological understanding of mental health and mental illness for the benefit of citizens and patients'		
Dissemination Level		
PU	Public — fully open (automatically posted online)	X
SEN	Sensitive — limited under the conditions of the Grant Agreement	

SUMMARY

This document provides a description on methodology and software for genetic nurture effects estimation using polygenic scores (PGS), including recommendations based on simulation work and emerging literature. Example code for PGS generation, simulations and models discussed is provided in appendix.

TABLE OF CONTENTS

1 INTRODUCTION

1.1 Purpose and Scope

This document provides an overview of methods for estimating genetic nurture effects using polygenic scores within families. Throughout the document, we highlight recent advances in this active area of research, discussing issues of bias and statistical power for indirect genetic effects in the two-generation (parent-offspring trios) design while signposting relevant resources. We provide recommendations based on the literature and report on empirical work we conducted in the Norwegian Mother, Father and Child Cohort Study (MoBA) using the trio design, as well as on power simulations based on realistic estimates of indirect genetic effects for childhood neuropsychiatric related traits. Finally, we provide code for generating PGS, quantify statistical power and estimate direct and indirect genetic effects in the trio design.

1.2 References to other FAMILY Documents

- FAMILY DoA

1.3 Definitions, Abbreviations and Acronyms

Table 1 List of Abbreviations and Acronyms

Abbreviation/ Acronym	DEFINITION
PGS	Polygenic scores
DGE	Direct Genetic Effects
IGE	Indirect Genetic effects
AM	Assortative Mating

2 SUMMARY OF METHODOLOGY

2.1 PGS methods to estimate genetic nurture

Several methods employing polygenic scores in multi-generational (e.g., parent-offspring trios) and other family-level data (e.g., sibling data) have been developed to separately estimate direct genetic effects (DGE) and indirect genetic effects (IGE). DGE capture the intergenerational process of genetic transmission, which leads to causal links between genotypes and phenotypes within a person (Gc* -> Yc*, Figure 1). In family settings, IGE (partly) capture an intergenerational process of so-called vertical (cultural) transmission, leading to genotypes causally influencing phenotypes between persons within families. This arises because children 'inherit' phenotypes associated with their parents' genotypes, leading to a (passive) gene-environment correlation (Gc* <- Gm* -> Em*; Figure 1). While IGE may happen between unrelated individuals, within a multigenerational setting they are often referred to as genetic nurture, as they are thought to capture the indirect link from parents' genomes to child phenotypes via the rearing environment where children grow up (GM* -> Em* -> Yc*; Figure 1). Figure 1 is a depiction of this process (adapted from Pingault et al., 2022).

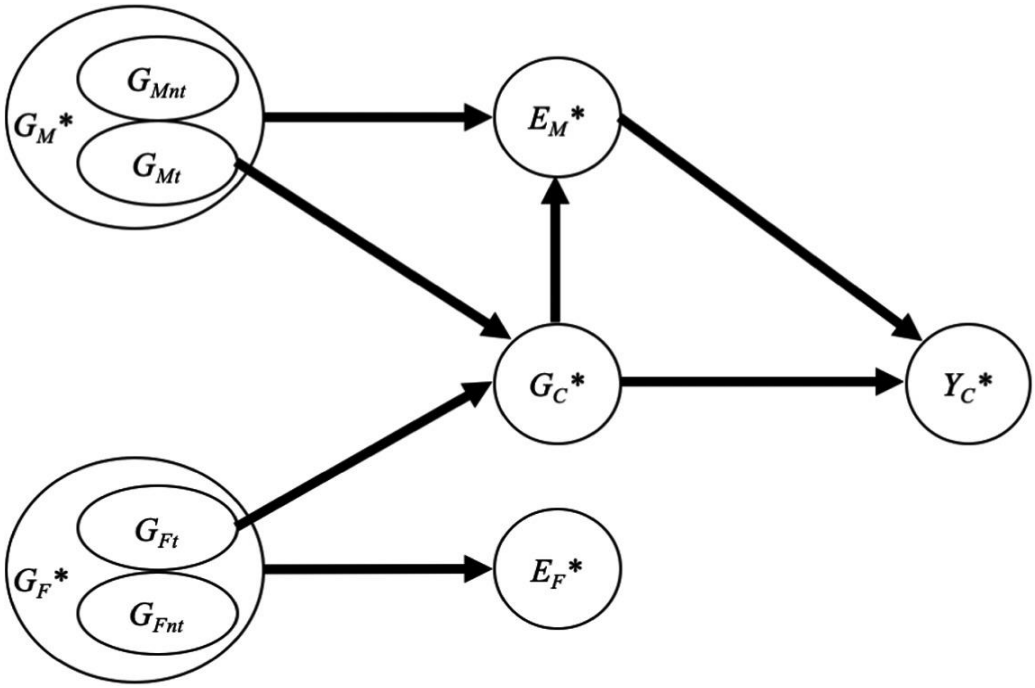

Figure 1. Genetic nurture. Genetic nurture (or familial genetic effects) occurs when parental genetics influence offspring outcomes via environmental pathways, for example, G_M*→E_M*→Y_C* for mothers. Note: child genetics (G_C*), mother genetics (G_m*), father genetics (G_f*). *E** parental nurturing envieonment. *Yc** child phenotype. Adapted from Pingault et al., 2022.

Some methods split individual polygenic scores based on transmitted and non-transmitted alleles, which in turn are related to child phenotypes to estimate these direct and indirect pathways of transmission. For example, PGS based on non-transmitted alleles can relate to child phenotypes only via indirect pathways (i.e., they cannot reflect a process of genetic transmission). Other models equivalently employ non-split polygenic scores within multivariable models, for example, jointly estimating parent-offspring PGS effects on child

phenotypes. Here, the only way in which parent PGS can affect child phenotypes, conditional on child genotypes (i.e. controlling for genetic transmission), is again via indirect pathways of transmission, such as the family environment.

Importantly, estimates of DGE and IGE may be affected by genetic and environmental confounding (including assortative mating and population stratification). While we note that all these models are more or less affected by these types of confounding, different models break down in different ways depending on the underlying data generating mechanism and the given bias at play (Balbona et al., 2022; Demange et al., 2022). Table 1 summarizes these designs with key references and assumptions/biases.

Design	**Method**	**Type of data**	**Key references / Example empirical study**	**Bias of indirect genetic effect**
T/NT	Estimation of PGS effects based on transmitted vs non-transmitted alleles on child phenotypes	Genotyped Parent-offspring trios + child phenotypes	Kong et al., 2020 Veller & Coop, 2024	May be particularly subject to selection bias (participation in cohorts of complete genotyped trios may be non-random)
Trio design	Multivariable model jointly estimating effects of parent – offspring PGS	Genotyped Parent offspring trios + child phenotypes	Tubbs et al., 2020 Tubbs et al., 2021	As in the T/NT design, may be subject to selection bias. Biased estimates in duo analyses when missing parental genotypes.
Duo design	Multivariable model jointly estimating effects of one parent's (usually maternal) PGS and the offspring PGS	Genotyped Parent + offspring duos + child phenotypes	Warrington et al., 2018 Tubbs et al., 2021	May be subject to selection bias. If paternal effects exist and are not accounted for this can lead to biased estimates of DGE and IGE.
Adoption - PGS	Comparison of PGS effects in adopted and unadopted individuals	Genotyped adopted and unadopted individuals	Cheesman et al., 2020	Assumptions: 1. Prenatal environment of biological mother does not play a role 2. Individuals are randomly adopted into families (no third variable confounding).

				3. Representative of the general population
Sibling differences	Deviation of a sibling PGS from the mean family PGS (mean across siblings).	Genotyped siblings	Selzam et al., 2019 Fletcher 2024	1. More biased by siblings' indirect genetic effects than other methods. 2. Direct genetic effects may be biased when using population based GWAS PGS.

Table 1. Summary of family-based designs employing PGS to infer IGE.

Previous work has shown how different designs using PGS to estimate IGE are differentially affected by bias, including sibling indirect effects, prenatal environment, population stratification, and assortative mating (Fletcher 2024, Demange et al., 2022). Because no single approach is immune to all biases, where possible, estimates from different designs should be compared to triangulate findings.

For example, the adoptees design is less affected by bias from assortative mating and population stratification (Demange et al., 2022, Balbona et al., 2021). However, this design makes several assumptions (Table 1) that may not be realistic in practice for many traits of interest. Furthermore, this type of data is typically more difficult to accrue in large sample sizes, which complicates issues of power, especially for childhood psychopathology-related traits, where PGS estimates are typically very small.

Genotyped siblings, on the other hand, are much more readily available across developmental cohorts, and therefore within-family methods based on siblings might be easier to implement. However, the sibling design may be particularly susceptible to bias of both direct and indirect genetic effects when employing population-based GWAS estimates to construct PGS (i.e., PGS not based on within-family GWAS estimates), as the mixture of direct and indirect effects is not straightforward to tease apart in this design (Fletcher 2024). The trio and the T/NT design estimate equivalent quantities in principle and should be affected by bias in similar ways. One main consideration (next to others related to power, see the relevant section below), is that the NT design needs the additional step of phasing to create pseudo-controls from non-transmitted alleles. The added benefit of these two-generation designs (trio and T/NT) over the others is the ability to estimate separately maternal and paternal effects.

While the direct genetic effects are unbiased in the trio design (Veller & Coop. 2024), indirect genetic effects don't estimate only genetic nurture but are an amalgam of different confounding and indirect processes at play, including environmental and genetic confounding (such as assortative mating and residual population stratification). Assortative mating, the non-random assortment of individuals on phenotypic traits, is problematic as it can increase genetic similarity within families, biasing GWAS estimates and therefore PGS analyses. In fact, AM can completely account for estimates of indirect genetic effects obtained within the trio design, and we discuss emergent methodology in this regard in turn.

2.1.1 Separating genetic nurture from assortative mating

Several studies in large population-based and registry data have shown pervasive assortative mating across human complex traits (Border et al., 2020; Border et al., 2021; Torvik et al., 2022; Sunde et al., 2024). Assortative mating induces correlations on variants within and across parental haplotypes within individuals, which in turn can inflate estimates of genetic effects (indexed by PGS in our case) within and across traits (Veller & Coop, 2024). Within-family designs, such as the trio design (two-generation model including parent-offspring trios), can obtain unbiased estimates of DGE. This design effectively adjusts PGS-trait associations for assortative mating (along with other biases affecting between-family variation such as population stratification) because, conditional on the parental genotypes, child genotypes vary only due to the random segregation of variants at birth. However, these between-family effects will be absorbed by the indirect genetic effect estimate in the trio design, which will therefore not necessarily only reflect genetic nurture.

A number of different approaches have been developed to attempt to quantify and separate assortative mating from genetic nurture effects. Here, we focus on the recently emerging literature leveraging PGS and multigenerational models that can be employed across FAMILY samples (Table 2).

One important caveat is that assortative mating is often thought to happen only via phenotypic assortment on a trait of interest but can also arise due to matching on 1) correlated phenotypes and 2) cultural and ancestry matching—for example, matching on familial characteristics of the spouse (Young 2023). A second aspect to consider is whether assortment happened over many generations and is thus said to be at equilibrium, or if it's relatively recent (at disequilibrium). For example, in the seminal paper by Kong et al. (2020), which popularized the estimation of genetic nurture effects with polygenic scores using the T/NT design, it was estimated that AM contributed little to the indirect genetic effect; however, only one generation of AM was assumed to be at play. Depending on whether assortative mating happens due to phenotypic, environmental, or genetic reasons, and whether it is at equilibrium or not, different designs break down in different ways, and this could bias estimates of IGE (Balbona et al., 2021). In Table 2, we summarize these two-generation study designs that can be employed across FAMILY cohorts.

Design	**method**	**Type of data**
Balbona et al., 2022	SEM decomposition	Trio genotypes, child and parent phenotypes
Nivard et al., 2024	Multilevel model	Trio and genotyped siblings in the parent generation, child phenotype
Young 2023	Estimating equations	Trio genotypes, child phenotype

Table 2. Two-generation study designs employing PGS to estimate DGE and IGE while teasing a part a number of quantities of interest including separating genetic nurture form Assortative mating

Within an SEM framework, the approach of Balbona et al., (SEM-PGS) allows for the separation of DGE and IGE from assortative mating (AM) and other quantities of interest such as vertical (cultural) transmission - the total contributions of a parental trait on the offspring trait mediated by the offspring rearing environment (Balbona et al., 2021). This is a very flexible method, but requires parental phenotypes in addition to parent and offspring genotypes and child phenotypes. In addition, the required predictive power of PGS to be employed in these models is much higher than what is typically found for many children emotional and behavioural traits at present ($r^2 \sim .01$). Furthermore, by default it relies only on phenotypic AM, and focuses solely on within-trait, not cross-trait IGE effects (e.g. parental polygenic score for depression affecting child depressive symptoms via parental phenotypic depression). The model can be extended in many ways to formally test and incorporate different types of assortative mating, but this comes at the expense of model complexity.

A different technique to separate genetic nurture from AM, as well as confounding due to different types of stratification, has been proposed by Nivard. This technique combines two different designs: the trio design with the sibling difference design (in the parent generation). In this design, any association of the parental PGS with the child phenotypes would indicate genetic nurture. This is because the DGE component would be captured by the child PGS, and between-family IGE induced by any type of stratification would be captured by the mean family PGS (mean sibling PGS) in the parent generation. In other words, this design separates IGE within families (genuine genetic nurture effects) from IGE between families, due to different types of stratification, including multigenerational AM and social stratification. These can induce correlations between parental genomes and child phenotypes as in the case of passive gene-environment correlation, but are not a result of a 'nurturing' process within the family.

As a caveat, the sibling-based part of the model may be especially subject to bias when employing population based GWAS PGS (see key references in Table 1). Furthermore, the special type of pedigree data this model requires may be unavailable in practice or difficult to obtain at large enough sample sizes for the small indirect genetic effects expected in childhood psychopathology, thus undermining statistical power.

Young (2023) proposes a method to separate a number of quantities of interest, including IGE from (any type of) AM, given: 1) An estimate of the parental correlation in a given PGS, 2) Coefficients from a two-generation model of DGE and IGE, and 3) An estimate of narrow-sense heritability from twin studies or Relatedness Disequilibrium Regression.

The method has several strengths, including that it doesn't assume a particular AM mechanism or a specific phenotype through which cultural transmission happens. However, it relies on estimates of heritability, and the accuracy of parameter estimates for quantities of interests such as IGE, will reflect the precision of those heritability estimates. In addition, when the fraction of heritability explained by the PGS is very small, which is often the case for many childhood psychopathology PGS estimates, the method can yield biased parameter estimates and inaccurate standard errors.

One final option could be to employ methods that adjust for assortative mating at the GWAS level (e.g., Bilghese et al., 2023), to then use these 'debiased' summary statistics to generate PGS and employ them in within-family analyses to estimate genetic nurture. However, no research on the validity of this approach has been conducted yet, and the feasibility of this suggestion remains to be tested.

Recommendation

A number of methods to separate assortative mating (AM) from genetic nurture are emerging, but efficient implementation may not be feasible across all phenotypes of interest in families. For example, owing to a lack statistical power or the precision of PGS estimates. It is important for studies to consider designs on a case-by-case basis, taking into account the phenotype of interest and the type of assortative mating at play. This will, of course, also depend on the type of data at hand. A three-generation model may be the best way to address this problem, but inefficient in practice due to data availability.

2.1.2 Missing parental genotypes

2.1.2.1 Bias

Because of problems with statistical power and selection bias, an important consideration across FAMILY cohorts is whether and how to impute missing parental genotypes. Not accounting for both parent PGS, for example in duo models (e.g., mother-offspring only models), can lead to bias in direct and indirect genetic effects (Tubbs et al., 2021). For example, child PGS effects are biased by 2/3 of the paternal indirect genetic effect when this is not properly accounted for.

This type of collider effect and consequent bias in parameter estimates will, of course, depend on the data generation mechanisms at play. For example, missing paternal genotype data is unlikely to lead to bias if the focus is on the effects of maternal genetic effects on child birth weight. In this case a duo model may just be the better option in terms of statistical power compared to the trio design (Tubbs et al., 2021).

Recommendation:
A recommendation for FAMILY cohorts is to perform analyses on all available data (family duos), but compare estimates with complete family trios and base conclusions from careful consideration of evidence from both designs.

In addition, one option is to attempt to adjust the duo model estimates based on the known bias. For example, assuming equal maternal and paternal effects, one could adjust downwardly the child PGS effects by 2/3 of the indirect effect and upwardly the maternal indirect effects by 1/3. These can, in turn, be compared to the estimates obtained from the trio model.

2.1.2.2 Imputation

Another option is to perform imputation of parental genomic data. A number of groups have developed methodology to this end. Table 3 summarises key references in this space.

Method	Software	Type of data	Key reference
Mendelian imputation	SNiPar (https://github.com/AlexTISYoung/snipar)	Nuclear family	Young et al., 2022

Imputation form sibling pairs	Impish (https://evansgroup.di.uq.edu.au/IMPISH/)	Sib/half sibling pairs	Hwang et al., 2020
SNP/PGS SEM imputation		Nuclear family	Tubbs et al., 2021

Table 3. Methods and software for imputation of parental genetic data.

While no systematic comparison is available at present, these methods are equivalent in principle, but in practice differ in the actual estimation procedure, computational demands, and type of data used (e.g. siblings vs nuclear family).

In the most basic way, imputation may be performed at the level of PGS imputing the missing parent PGS using the PGS for the child and the other parent. In this scenario for example the father PGS would be obtained from the residual of the child PGS regressed on the mother PGS.

Alternatively, handling missingness of PGS using Full Information Maximum Likelihood (FIML) in SEM could be considered. However, the extent to which this latter approach is valid and provides unbiased estimates while recovering power is not clear.

Recommendation

Depending on the data at hand, choose at least one imputation method and compare estimates retrieved to those from the complete trio data, if available. A power calculator under different imputation scenarios and types of data available (e.g., duos vs trios) is provided here: https://evansgroup.di.uq.edu.au/power-calculators.html

In the appendix, we provide code to perform power simulations and estimate DGE and IGE under the simple scenario where genotyped trio data is available.

3 STATISTICAL POWER AND EMPIRICAL APPLICATION

Little research has been conducted on the statical power between different approaches to detect indirect genetic effects (e.g. T/NT vs Trio). Power calculations are also complicated by considerations around biases such as assortative mating. Hence any comparison of statical power for these approaches will have to necessarily assume absence of assortative mating (and other confounding).

3.1 Review of simulation studies

In a simulation comparison (Tubbs et al., 2020) between two different applications of the T/NT approach and the trio design, all designs equivalently provided unbiased estimates of genetic nurture effects. However, the trio design was found to outperform the T/NT designs in terms of statistical power, specifically showing a 1/3 increase in power to detect genetic nurture over the split scores.

The SEM PGS (Balbona et al., 2020) is a more sophisticated version of the trio design, but it requires sample sizes well above 16,000 family trios (with genetic and phenotypic data in the parent generation) to reach a standard error below 0.05 and thus adequate power for the effect sizes expected for genetic nurture effects in childhood psychopathology. For reference see empirical estimates derived from the MoBA cohort (below).

In sum, the most feasible approach in terms of power across FAMILY cohorts will be the duo/trio design (a two-generation model employing PGS) over other methods.

3.1.1 An empirical application of the trio design along with power simulations in the MoBa cohort

In WP3 we conducted empirical simulations using measures of childhood psychopathology from the MoBa cohort across a range of realistic estimates of genetic nurture effects. We report on these simulations in detail in this online preprint: https://osf.io/preprints/psyarxiv/w4psd

Briefly, using the largest available set of complete genotyped trios with available phenotypic data on several measures of psychopathology in MoBA (N = 15,000), we performed power analyses for parent-offspring polygenic score (PGS) effects across 10,000 iterations at our fixed sample size. Parameters for the child and mother PGS were fixed to explain 0.2% and 0.1% of the variance respectively, while the father PGS was varied across three parameters (betas = .03, .01, .001). We show that at the current available sample size across these measures we had at least 80% power to detect effect sizes as small as beta = .03 for genetic nurture. Code to reproduce these analyses and summary-level data of these simulations is available here: https://github.com/AndreAllegrini/IRISK-p.

We then applied the trio design to test direct and indirect genetic effects of neuropsychiatric-related traits on different psychopathology measures, including the p-factor (a latent factor constructed from different childhood emotional and behavioral difficulty domains). We detected pervasive shrinkage in PGS estimates for neuropsychiatric-related traits when adjusting for parental PGS, suggesting indirect genetic effects at play. By and large we were adequately powered for the effect sizes detected for parental PGS.

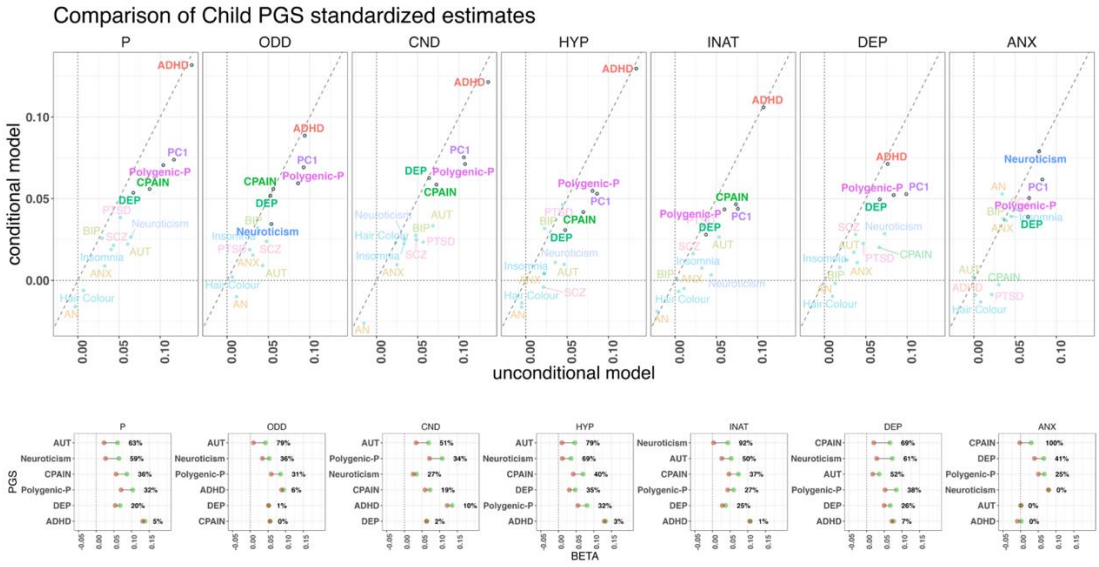

Figure 2. Polygenic scores contributions across emotional and behavioural difficulties domains. Top panel: Comparison of standardized regression coefficients for offspring PGS effects from conditional to unconditional models, showing relative importance of PGS contributions across emotional and behavioural difficulties. **Faded**: does not survive correction for multiple testing/not selected over the null model. Bottom panel: Shrinkage of standardized effects for the offspring PGS from unconditional (i.e. green) to conditional (i.e. red) models across emotional and behavioural difficulties domains, restricted to models favoured over the null-model (PC1 PGS not shown for clarity). Estimates are plotted in descending order of shrinkage.
Note: ADHD = attention deficit hyperactivity disorder, AUT = autism, BIP = bipolar disorder, SCZ = schizophrenia, AN = anorexia nervosa, ANX = anxiety, PTSD = post-traumatic stress disorder, DEP = broad depression, CPAIN = chronic pain, PC1 = first unrotated principal component of all neuropsychiatric (and related) PGS, PC1 psych = first unrotated principal component of all neuropsychiatric PGS. **Facets:** ODD = Oppositional Defiant Disorder; CND = Conduct Disorder; HYP = Hyperactivity; INA = Inattention; DEP = Depression; ANX = Anxiety. Adapted from Allegrini et al., 2023.

4 LITERATURE REFERENCES

Allegrini, A., Hannigan, L. J., Frach, L., Barkhuizen, W., Baldwin, J., Andreassen, O. A., ... & Pingault, J. B. (2023). Intergenerational transmission of polygenic predisposition for neuropsychiatric traits on emotional and behavioural difficulties in childhood.

Balbona, J. V., Kim, Y., & Keller, M. C. (2021). Estimation of parental effects using polygenic scores. Behavior Genetics, 51(3), 264-278.

Balbona, J. V., Kim, Y., & Keller, M. C. (2022). The estimation of environmental and genetic parental influences. Development and Psychopathology, 34(5), 1876-1886.

Bilghese, M., Manansala, R., Jaishankar, D., Jala, J., Benjamin, D. J., Kimball, M., ... & Turley, P. (2023). A General Approach to Adjusting Genetic Studies for Assortative Mating. BioRxiv.

Border, R., O'Rourke, S., de Candia, T., Goddard, M. E., Visscher, P. M., Yengo, L., ... & Keller, M. C. (2022). Assortative mating biases marker-based heritability estimators. *Nature communications*, *13*(1), 660.

Border, R., Athanasiadis, G., Buil, A., Schork, A. J., Cai, N., Young, A. I., ... & Zaitlen, N. A. (2022). Cross-trait assortative mating is widespread and inflates genetic correlation estimates. *Science*, *378*(6621), 754-761.

Cheesman, R., Hunjan, A., Coleman, J. R., Ahmadzadeh, Y., Plomin, R., McAdams, T. A., ... & Breen, G. (2020). Comparison of adopted and nonadopted individuals reveals gene–environment interplay for education in the UK Biobank. Psychological Science, 31(5), 582-591.

Demange, P. A., Hottenga, J. J., Abdellaoui, A., Eilertsen, E. M., Malanchini, M., Domingue, B. W., ... & Cheesman, R. (2022). Estimating effects of parents' cognitive and non-cognitive skills on offspring education using polygenic scores. Nature Communications, 13(1), 4801.

Fletcher, J., Wu, Y., Li, T., & Lu, Q. (2024). Interpreting polygenic score effects in sibling analysis. PLoS One, 19(2), e0282212.

Hwang, L. D., Tubbs, J. D., Luong, J., Lundberg, M., Moen, G. H., Wang, G., ... & Evans, D. M. (2020). Estimating indirect parental genetic effects on offspring phenotypes using virtual parental genotypes derived from sibling and half sibling pairs. PLoS Genetics, 16(10), e1009154.

Kim, Y., Balbona, J. V., & Keller, M. C. (2020). Bias and precision of parameter estimates from models using polygenic scores to estimate environmental and genetic parental influences.

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E., ... & Stefansson, K. (2018). The nature of nurture: Effects of parental genotypes. *Science*, *359*(6374), 424-428.

Nivard, M. G., Belsky, D. W., Harden, K. P., Baier, T., Andreassen, O. A., Ystrøm, E., ... & Lyngstad, T. H. (2024). More than nature and nurture, indirect genetic effects on children's academic achievement are consequences of dynastic social processes. Nature Human Behaviour, 1-8.

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.

Selzam, S., Ritchie, S. J., Pingault, J. B., Reynolds, C. A., O'Reilly, P. F., & Plomin, R. (2019). Comparing within-and between-family polygenic score prediction. *The American Journal of Human Genetics*, *105*(2), 351-363.

Sunde, H. F., Eftedal, N. H., Cheesman, R., Corfield, E. C., Kleppesto, T. H., Seierstad, A. C., ... & Torvik, F. A. (2024). Genetic similarity between relatives provides evidence on the presence and history of assortative mating. *Nature Communications*, *15*(1), 2641.

Torvik, F. A., Eilertsen, E. M., Hannigan, L. J., Cheesman, R., Howe, L. J., Magnus, P., ... & Ystrom, E. (2022). Modeling assortative mating and genetic similarities between partners, siblings, and in-laws. *Nature Communications*, *13*(1), 1108.

Tubbs, J. D., Hwang, L. D., Luong, J., Evans, D. M., & Sham, P. C. (2021). Modeling parent-specific genetic nurture in families with missing parental genotypes: application to birthweight and BMI. Behavior Genetics, 51(3), 289-300.

Tubbs, J. D., Porsch, R. M., Cherny, S. S., & Sham, P. C. (2020). The genes we inherit and those we don't: maternal genetic nurture and child BMI trajectories. Behavior Genetics, 50, 310-319.

Veller, C., & Coop, G. M. (2024). Interpreting population-and family-based genome-wide association studies in the presence of confounding. PLoS Biology, 22(4), e3002511.

Warrington, N. M., Freathy, R. M., Neale, M. C., & Evans, D. M. (2018). Using structural equation modelling to jointly estimate maternal and fetal effects on birthweight in the UK Biobank. International Journal of Epidemiology, 47(4), 1229–1240.

Young, A. I., Nehzati, S. M., Benonisdottir, S., Okbay, A., Jayashankar, H., Lee, C., ... & Kong, A. (2022). Mendelian imputation of parental genotypes improves estimates of direct genetic effects. Nature Genetics, 54(6), 897-905.

Young, A. S. (2023). Estimation of indirect genetic effects and heritability under assortative mating. bioRxiv.

5 APPENDICES

1. Pipeline to generate polygenic scores in LDpred2:

2. Code to estimate the trio model within SEM and perform power simulations based on this design across a number of sample sizes. Based on Tubbs et al., 2020.

3. (Power) simulation code based on largest sample size and set of psychopathology measures in the MoBa cohort.

ACKNOWLEDGEMENT
Funded by the European Union, the Swiss State Secretariat for Education, Research and Innovation (SERI) and the UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee'. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union, or the European Health and Digital Executive Agency (HADEA), the SERI or the UK Research and Innovation (UKRI). Neither the European Union nor the granting authorities can be held responsible for them.