Multivariate genetic analyses of developmental complex traits

Allegrini, Andrea

Awarding institution:
King's College London

Multivariate genetic analyses of developmental complex traits

Andrea G Allegrini

PhD Thesis, July 2020

Social, Genetic and Developmental Psychiatry centre

King's College London

Submitted for the degree of Doctor of Philosophy in Statistical Genetics

Abstract

Consistent evidence from the quantitative genetic literature points towards a substantial genetic component underlying variation and covariation across complex traits. Novel genomic methods coupled with large population-based samples, afford the possibility to leverage this information to tackle important developmental questions. This thesis focuses on multivariate genetic and genomic approaches as applied to polygenic prediction and inference of trait associations across development, with a focus on cognitive- and psychopathology-related traits. The thesis follows two main themes:

Polygenic prediction

Methods that leverage the covariance structure of genetically correlated traits to increase power for variant discovery can in turn be used to boost predictive power of polygenic scores. In a first study I compare several multi-trait genomic approaches in the context of polygenic prediction of cognitive-related traits (general intelligence and educational achievement) in childhood and adolescence (Chapter 2). As genetic predictors become more powerful, they can be employed to further our understanding of the gene-environment interplay underlying variation in common complex traits. In a second study (Chapter 3), I focus on the longitudinal prediction of educational achievement by constructing penalized multivariable prediction models integrating multiple polygenic scores and environmental predictors, gaining insights into their multivariate interplay (gene-environment correlation and interaction).

Developmental co-occurrence of psychopathology

Akin to the concept of general intelligence, the co-occurrence of traits related to mental health during development suggests a general dimension of psychopathology underlying the emergence and co-morbidity of problem behaviours in childhood (the p-factor). In a third study (Chapter 4) I systematically investigate the manifestation of the p-factor across childhood and adolescence by means of multivariate genetic methods, showing that this co-occurrence is partly explained by a common genetic aetiology. There are at least two plausible processes that can account for the co-occurrence of psychopathology traits in childhood. First, as investigated in Chapter 4 the correlation between psychopathologies could be the product of individual differences *between* people on stable traits attributable to a heritable p-factor. Second, the developmental co-occurence of psychopathology could emerge from a causal process *within* people where the temporal state on one variable causally influences the state of another variable, inducing correlation between them. To this end, in a fourth study I investigate longitudinal reciprocal effects between problem behaviours (Chapter 5),

separating between vs within person effects in two longitudinal population-based cohorts. Extending this model to family-level data, I further investigate reciprocal directional influences between siblings over time, separating them from similarities between siblings that arise through shared (genetic or environmental) influences that exist in a family.
The thesis concludes (Chapter 6) with a discussion of future prospects for multivariate genomic research, opportunities for integrating emerging methods, challenges and limitations.

Contents

List of Figures

List of Tables

Acknowledgements

I would first like to thank my supervisor Robert Plomin, for the patient mentoring and for inspiring me to embark in a career in behavioural genetics. Thanks for supporting my ideas, and for being such a brilliant supervisor. I'd also like to thank Jean-Baptiste Pingault, my second supervisor, for the methodological help, and for always encouraging critical thinking. I'm grateful to the TEDS twins and for the amazing work of the core TEDS team, Rachel Ogden, Louise Webster and Andy MacMillan, without which the research in this thesis would have not been possible.

Throughout my PhD I have met very brilliant and fun people that made all the difference. It has been such a privilege working at the SGPD centre. I'd especially like to thank several people that contributed to my research with their insights: Gerome Breen, Joni Coleman, Thalia Eley, Tom McAdams, Paul O'Reilly, Saskia Selzam, and Sophie von Stumm; as well as my colleagues and friends with which I shared my PhD journey: Chris, Rosa and Tom; Ville and Julian for our stats chats; and Kaili and Margherita.

As part of my PhD I had the absolute luck of being part of a research consortium, CAPICE, that put me in touch with very talented people throughout Europe. All the projects in this thesis have received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 721567.

I would also like to thank Meike Bartels and Michel Nivard for their supervision during my secondment, and for hosting me at the Biological Psychology Department at Vrije Universiteit Amsterdam, where I felt very much at home.

My academic journey started long before my PhD, and I'm indebted to many people that encouraged a career in science with their friendship and mentoring: my academic friends Felix, Stefania and Yayouk, and my supervisors Karin Verweij, Jacqueline Vink, Brittany Evans, Anja Huizink, and Hans Koot.

I dedicate this thesis to my mother, Adina – my greatest teacher, and to the memory of my father, Ettore. Also, to Marci, for making everything brighter.

Statement of authorship

Genotypic and/or phenotypic data were previously collected by the Twin Early Development Study and the Netherlands Twin Registry research teams. I was responsible for preparing all data for the projects included in this thesis and for executing analyses. To the best of my knowledge the contents presented here are original and resulted from my own work, except where differently acknowledged in the text.

Andrea G Allegrini

Publications resulting from chapters of this thesis

Published:

Allegrini, A. G.*, Cheesman, R.*, Rimfeld, K., Selzam, S., Pingault, J. B., Eley, T. C., & Plomin, R. (2019). The p factor: genetic analyses support a general dimension of psychopathology in childhood and adolescence. *Journal of Child Psychology and Psychiatry*, *61*(1), 30-39. (*= joint first authorship)

Allegrini, A. G., Selzam, S., Rimfeld, K., von Stumm, S., Pingault, J. B., & Plomin, R. (2019). Genomic prediction of cognitive traits in childhood and adolescence. *Molecular psychiatry*, *24*(6), 819.

Allegrini, A. G., Karhunen, V., Coleman, J. R., Selzam, S., Rimfeld, K., von Stumm, S., Pingault, J. B & Plomin, R. (2020). Multivariable GE interplay in the prediction of educational achievement. *PLoS genetics*, *16*(11), e1009153.

In preparation:

Allegrini, A.G., van Beijsterveldt, T., Boomsma, D., Rimfeld, K., Pingault, J.-B., Plomin, R., Bartels, M. & Nivard, M. G. (In preparation). Directional relationships between childhood psychopathology dimensions across development.

Other publications

Other papers that have been published over the course of the PhD:

Akingbuwa, W. A., Hammerschlag, A. R., Jami, E. S., **Allegrini, A. G.,** Karhunen, V., Sallis, H., ... Middledorp, C. M. (2020). Genetic Associations Between Childhood Psychopathology and Adult Depression and Associated Traits in 42 998 Individuals: A Meta-Analysis. JAMA psychiatry

Malanchini, M., Rimfeld, K., **Allegrini, A. G.,** Ritchie, S. J., & Plomin, R. (2020). Cognitive ability and education: how behavioural genetic research has advanced our knowledge and understanding of their association. Neuroscience & Biobehavioural Reviews, 111, 229-245.

Allegrini, A. G., Verweij, K. J., Abdellaoui, A., Treur, J. L., Hottenga, J. J., Willemsen, G., ... & Vink, J. M. (2018). Genetic vulnerability for smoking and cannabis use: associations with e-cigarette and water pipe use. *Nicotine and Tobacco Research*, *21*(6), 723-730.

Allegrini, A. G., Evans, B. E., de Rooij, S., Greaves-Lord, K., & Huizink, A. C. (2017). Gene× Environment contributions to autonomic stress reactivity in youth. *Development and psychopathology*, *31*(1), 293-307.

Chapter 1 - General introduction

After decades of quantitative genetic research, the question of whether it is genetic or environmental influences that makes us who we are, the nature–nurture debate, has shifted to the question of how these forces jointly shape human existence (Knopik, Neiderhiser, DeFries, & Plomin, 2016; Plomin, 2019). Developmental traits are extremely complex phenotypes, determined by a multitude of genetic and environmental influences, as well as their interplay. In addition, any given trait does not exist in a vacuum, but is embedded in a multi-dimensional (multivariate) space, partly defined by its relationship with other traits. This multivariate space arises from both genetic and environmental underpinnings, common causes underlying covariation between traits, as well as complex interdependencies between phenotypic dimensions leading to developmental differences. The present thesis focuses on the application of multivariate genetic and genomic approaches to study the developmental (co)occurrence of cognitive- and psychopathology-related traits. This first chapter serves as a general introduction to the concepts and methods discussed in depth in the chapters that follow. I first lay out some basic concepts in quantitative genetic research, such as heritability and estimation of genetic and environmental effects using twin data, particularly as they relate to multivariate analyses of behavioural (psychopathology related) traits in childhood. I then expand to multivariate genomic approaches in the context of prediction with a focus on educationally relevant traits and gene-environment interplay. I finally conclude with an overview of the topics discussed in the other thesis chapters.

Quantitative genetics & heritability

Since Sir Ronald Fisher reconciled the views of Mendelians (from Gregor Mendel) and biometricians (from Sir Francis Galton) starting the field of quantitative genetics, it was clear that genetic and environmental influences were two sides of the same quantitative coin. In a seminal paper Fisher showed that continuous traits follow the modes of inheritance of discrete traits such as those investigated by Mendel in pea plants. The key to this understanding was something that in statistics is known as the central limit theorem. Continuous trait variation could be reconciled with Mendelian inheritance if multiple genetic variants, each contributing small effects, were involved in quantitative traits variation yielding a bell-shaped frequency distribution. Quantitative traits were therefore complex traits, determined by a multitude of genetic factors, as well as environmental effects (Neale, Ferreira, Medland, & Posthuma, 2007). Fisher partitioned the variance for a single genetic locus into additive and dominance variance, examined how this related to resemblance between relatives, and extended the model to multiple loci and the environment. Quantitative genetic theory thus separates

phenotypic variation into genetic and environmental components of variance. The total variation underlying an observed trait (the phenotype) attributable to genetic variation is called heritability. We can distinguish different types of genetic variation (see below), but in its broadest sense heritability is the proportion of variation in a trait that is accounted for by genetic variation in a population (i.e. broad-sense heritability; H^2).

Developmental heritability

Importantly, estimates of heritability are specific to populations and times, such that they will change as circumstances change. For example, the heritability of intelligence increases across development (Plomin & Deary, 2015), such that heritability is higher in early adulthood than in childhood, as the developmental context changes. This is also true for many other developmental phenotypes investigated in the following chapters (Bergen, Gardner, & Kendler, 2007). A compelling explanation for this finding is that as individuals grow up they start selecting environments that are consistent with their genetic propensities (Plomin & Deary, 2015), a type of gene-environment correlation discussed in chapter 3 and in the 'gene-environment interplay' section below. This, in turn, might partly be the reason why genomic prediction of cognitive related traits increases from childhood to adolescence as shown in chapter 2. From the definition given in the paragraph above it should be clear that heritability refers to genetic contributions to individual differences, but not to the extent to which a phenotype for one particular person is genetically influenced (Knopik, Neiderhiser, DeFries, & Plomin, 2016; Moore & Shenk, 2017). With this in mind, perhaps the most important consideration about heritability is as follows: heritability does not imply genetic determinism, since by changing the context you can change the estimated heritability.

Genetic and Environmental variance components

Broad-sense heritability includes different types of genetic variation, additive, dominance and epistatic effects. We can estimate heritability by employing measured genetics methods (genomic methods such as those employed in **chapter 2 and 3** of this thesis) or inferring it from resemblance between relatives (**chapters 4 and 5**). The workhorse of the quantitative genetic literature, the twin design offers a powerful approach to infer genetic and environmental influences on trait variation. Since monozygotic (MZ) and dizygotic (DZ) twin pairs share 100% and 50% (on average) of their DNA respectively and the same environment (equal environment assumption), we can compare their resemblance to infer the relative shared genetics (additive or dominance variation), shared environmental, and unique environmental influences underlying variation in a particular trait. For example, a rough

estimate of additive genetic variation can be obtained by doubling the difference between MZ and DZ correlations. This is one of the formulae derived by Falconer (Falconer, Mackay, & Frankham, 1996; see **chapter 5** for other Falconer's derivations), which can be used to quantify the relative contributions of genes and environments.

Two types of genetic contributions can be distinguished: additive and dominance genetic effects (**chapter 5**). Additive effects on a trait occur when alternate forms of genetic variants (alleles) within and across loci on the genome add up, while dominance effects refer to interactions of two alleles at the same locus. This additive component of variance is also called narrow-sense heritability in that excludes dominance and epistatic effects, this is by far the largest proportion of genetic variability and is the focus of all genomic methods employed in this thesis (**chapter 2 and 3**). A more formal way to model MZ and DZ twin covariance is by using structural equation modelling approaches, model fitting techniques to partition genetic and environmental variance components and this is described more in detail in **chapter 4**.

Multivariate genetic analyses

We can extend model-fitting techniques to a multivariate framework by jointly analysing multiple traits and estimate genetic and environmental contributions that are shared among them. Furthermore, we can formally test hypotheses regarding the underlying structure explaining the relationship between traits of interest. This is the focus of much of **chapters 4 and 5**. By employing multivariate genetic analyses of psychopathology-related traits throughout childhood, it is possible to gain a better understanding of the developmental co-(co)occurrence of psychopathology. Multivariate genetic analyses of twin data can be used to investigate the extent to which traits are genetically correlated, and the relative contributions of genes and environment in their co-occurrence. In this regard, the abundant twin-design literature on the multivariate co-occurrence of disorders suggest that **the same genetic influences broadly underlie multiple disorders** (Knopik et al., 2016; Plomin, DeFries, Knopik, & Neiderhiser, 2016)**.** This is also known as **pleiotropy** (see below), the idea that the same genetic variants are associated with multiple traits concurrently.

In **chapter 4,** I analyse the genetic and environmental factors shared by multiple problem behaviours in childhood, with a systematic look at the manifestation of a common factor of psychopathology called the p-factor. Psychopathology-related traits in childhood tend to co-occur, and there is evidence that their intercorrelation is due to a common predisposition

accounting for individual differences between children (Martel et al., 2017). We can employ multivariate twin modelling techniques to explore the manifestation of this common factor and whether the relative genetic and environmental influences underlying this common component are stable across time.

From a different angle, **leveraging multivariate twin data** can also be a powerful approach to **control for** those **genetic and environmental influences** that make individuals more similar within families. We can interrogate multivariate data to dig deeper into the reasons why psychopathology traits are correlated. For example, are there direct relationships between psychopathology traits that are not explained by shared genetic and environmental aetiology? In this regard, are there reciprocal influences between siblings not accounted for by genetic or environmental similarities between them? **Chapter 5** investigates these questions by means of a multivariate longitudinal within-family design.

Modern-day quantitative genetics

The DNA code of individuals is 99.9% identical but differs at roughly 0.1% of the genome between people. Variation in this 0.1% of the DNA code partly accounts for similarities and differences in the manifestation of traits and characteristics across the population. We can look at these DNA differences across the genome to estimate the total genetic contributions to trait variation, or for associations of specific variants to particular traits. Importantly in contrast to the twin design, where we test for relative contributions of additive and dominance components of variance, typically genomic approaches are concerned with additive variation (although not exclusively). This additive genetic variation is the focus of all the genomic methods employed in the current thesis (**chapters 2, 3 and 4**).

GWAS

For example, one of the most commonly employed observational approaches of modern-day quantitative genetics are genome-wide association studies (GWAS; **chapter 2**), which model the average genetic component part of Fisher's equation for a single locus (Visscher & Goddard, 2019). In GWAS we separately test millions of genetic variants (single nucleotide polymorphisms; SNPs), with coding of 0, 1 or 2 (depending on the number of alleles an individual carries at a particular locus), for association with a phenotype. GWAS in a sense offers the possibility to test Fisher quantitative genetic model with the use of genomic data (Plomin, Haworth, & Davis, 2009; Visscher & Goddard, 2019). In fact, the most important take-home message from the GWAS literature is that virtually all human phenotypes,

particularly behavioural traits, are extremely polygenic, with thousands of SNPs of small effect underlying their variation (Visscher et al., 2017). This has been codified as the 4th law of behavioural genetics (Chabris, Lee, Cesarini, Benjamin, & Laibson, 2015). Two prominent examples are the most recent GWAS of educational attainment (Lee et al., 2018) and height (Yengo et al., 2018), which both discovered more than 1000 genetic variants implicated independently in the variation of these traits.

SNP heritability

By using these observed genetic effects we can come up with an estimate of the additive contributions of measured genetics to trait variation. This type of heritability is called SNP heritability (SNP h2) and falls short of the narrow sense heritability previously described (Plomin & von Stumm, 2018). Narrow sense heritability is an upper bound to SNP h2 because this is limited to observed genetic variants, and does not capture variation due to rare variants (single nucleotide variants with low minor allele frequency), non-SNP variation, such as indels (insertions and deletions less than 50 base pairs in length), and structural variation, (such as copy number variations). The gap between SNP heritability and narrow sense heritability derived from twin studies is known as the missing heritability problem (Manolio et al., 2009) and has practical and theorical implications for much of the work currently done in genomic research (Young, 2019).

Discovering single variants associated with trait variation can help illuminate the underlying biology of traits, and improve our understanding of disease (Stranger, Stahl, & Raj, 2011). However, taken singularly these variants have little value in themselves to predict whether a person will develop a common disorder or manifest a particular trait, especially for behavioural traits, because their effect sizes are typically very small (with some notable exceptions, such as the causal role of the FTO variant rs1421085 on Body Mass Index via increased expression of IRX3 and IRX5 genes; see Claussnitzer et al., 2015).

Polygenic scores

We can aggregate these tiny effects sizes from GWAS summary statistics in scores reflecting the genetic based predisposition carried by individuals, called polygenic scores. We can in turn employ polygenic scores to predict phenotypic traits and infer genetic roots shared among them. This was demonstrated in a landmark study in the field of psychiatric genetics (The International Schizophrenia et al., 2009) whereby a polygenic score for schizophrenia was shown to be associated with bipolar disorder, and thus, also indicating a shared genetic

component. This approach can also be extended to the multivariate framework by considering several polygenic scores in the same model (Krapohl et al., 2018). In **chapter 4** I discuss how an aggregate of adult polygenic risk scores for major psychiatric disorders associates with childhood problem behaviours indicating a multivariate link between childhood psychopathology dimensions and adult genetic liability to mental health disorders.

Uses of polygenic scores

Polygenic scores are becoming potent predictors of trait variation within populations (with some caveats; see Duncan et al., 2019; and chapter 6), with far-reaching implications for research. For example, **polygenic scores can be integrated in multivariable models including environmental factors to improve prediction** (chapter 3). We can use polygenic scores to infer relationships between traits where one of the traits of interests is not directly measured in the sample of interest, for example assessing relationships between adult psychiatric disorders and child problem behaviours (**chapter 4**). Finally, identifying groups of individuals with increased (or decreased) polygenic based predisposition can help understanding how specific traits or disorders (concurrently) vary in the population (e.g. Abdellaoui et al., 2019). More generally, polygenic scores can also help us think quantitatively in terms of psychiatric disorders, shifting the paradigm from discrete disorders to quantitative dimensions while considering the full spectrum of polygenic predisposition (Plomin et al., 2009). In **chapter 2** I discuss several ways of constructing polygenic scores testing their performance with respect to their predictive ability of cognitive-related traits in childhood and adolescence.

Polygenic score heritability

The fraction of the phenotypic variance that can be predicted by polygenic scores is also called polygenic score heritability (PGS h^2), the ceiling of which is SNP heritability. That is, predictive power of polygenic scores is limited by the additive variation captured by observed genotypes. Typically, polygenic h^2 is less than half the SNP h^2 of a trait (Plomin & von Stumm, 2018). This is in part attributable to noise attached to the estimates that we aggregate. Predictive power of polygenic scores can be improved by increasing sample size of GWAS from which SNP effects are estimated or reducing heterogeneity (de Vlaming et al., 2017; Dudbridge, 2013; Mostafavi et al., 2020). We can thus reduce the gap between SNP h^2 and PRS h^2 by improving the accuracy of GWAS estimates. One powerful method to achieve this improvement in accuracy is provided by a multivariate framework (see below and chapter 2).

Multi-trait methods

As already mentioned, shared genetic influences (or genetic correlations) across traits support the conclusion that many genetic effects are general across traits and disorders (Plomin et al., 2016). In fact, recent findings in the fields of psychiatric and medical genetics point to widespread pleiotropy – the extent to which the same genetic variants affect two or more traits – across many complex traits (psychiatric, metabolic or anthropometric; Pickrell et al., 2016). Emerging multivariate (multi-trait) genome-wide approaches (Grotzinger et al., 2019; Turley et al., 2018) leverage the covariance structure between (genetically) correlated traits to increase power to discover trait-related variants, as well as variants concurrently affecting multiple traits (pleiotropic variants). Many genetic variants involved in different complex traits are highly interconnected (Boyle, Li, & Pritchard, 2017), and this property can be leveraged to enhance the predictive power of polygenic scores. For example, as a byproduct of the enhanced power at the level of variant discovery in multi-trait GWAS, we can obtain more accurate SNP estimates, which in turn yield more powerful polygenic scores when aggregated. **Chapter 2** discusses several multi-trait genomic methods and leverages this approach to boost predictive power of cognitive-related polygenic scores.

Potent polygenic score predictors are emerging, such as the score derived from the recent GWAS of educational attainment (Lee et al., 2018). As these genetic predictors become stronger, they are also becoming instrumental in gaining important insights at the level of environmental exposures, and to better understand the gene-environment interplay underlying variation in common complex traits (Barcellos, Carvalho, & Turley, 2018; Belsky et al., 2018).

Gene-environment interplay

Quantitative genetic theory distinguishes two types of gene-environment interplay: interaction and correlation (Knopik et al., 2016). Gene-environment (GE) interactions refer to genetic influences that depend on levels of the environment, that is, whether the environment moderates genetic influence or vice versa. Conversely, when the environment exerts differential effects depending on a person's genetic makeup we talk about genetic moderation. Gene-environment correlation is a related, but different, concept and refers to the covariance between genes and environments. Three types of GE correlation are typically distinguished: passive, active and evocative (Plomin, 2014; Plomin, DeFries, & Loehlin, 1977). We have passive GE when a person's genotype is associated with their rearing environment, active GE when a person actively seeks environments according to their genetic makeup, and evocative

GE when the environment responds to a person's genetic predisposition (Chapter 3 describes these types of GE correlation with respect to current genomic research).

Polygenic scores by environment interplay

While historically GE interplay has been addressed by quantitative genetic designs (Knopik et al., 2016), more recently genomic methods afford the possibility to investigate GE interplay with measured genetics. In particular, polygenic scores offered a step forward in GxE research compared to so called candidate gene studies, where putative biologically plausible variants were used as a proxy for genetic liability for a particular trait. This approach garnered momentum before GWAS were commonly used, partly due to-wrong expectations about effect sizes that could be found. This led to a failure to replicate many such candidate genes studies (Border et al., 2019), mainly due to issues of power and poor study design. Polygenic scores on the other hand offer a powerful alternative to candidate genes in that they are agnostic to underlying assumptions about biological plausibility by relying on the aggregate effects on many genetic variants across the genome, and are inherently more powerful instruments indexing a person genetic-based predisposition (Colodro-Conde et al., 2018; Peyrot et al., 2014; Peyrot et al., 2018).

Importantly the correlation between genes and environments can confound GxE interactions (Purcell, 2002). Relatedly, and consistent with long standing evidence from the quantitative genetics literature, recent work using genomic data suggests that polygenic score prediction includes passive GE correlation. As such polygenic scores are not pure measures of genetic predisposition partly capturing environmental effects. Similarly, it is well known from twin studies that measures of the environment are heritable (Plomin et al., 2016). The extent and implications of GE correlation with regard to multivariable prediction models is explored in **chapter 3** of the thesis. Gene-environment interactions can be extended as well to a multivariate framework by considering several measured environmental factors and polygenic scores in multivariable models. This can be potentially very useful because on one hand the effect of multiple interactions is modelled concurrently, controlling for their reciprocal effects. Furthermore, we would expect total prediction to be improved over main effects of genes and environments separately. While the literature on polygenic scores by environment is accumulating this remains an active area of investigation. **Chapter 3** addresses the issues discussed with respect to longitudinal prediction of educational achievement.

Summary and overview of chapters

I have outlined a brief general introduction to quantitative genetic and genomic approaches touching upon multivariate methods and themes discussed at length in the thesis' chapters. A multivariate framework can help illuminate important questions, from prediction of complex developmental traits, to interrogating the underlying structure of traits and their covariance. Under an overarching multivariate theme, this thesis focuses on genetic and genomic approaches as applied to childhood psychological development following two main threads:

The first part of the thesis is concerned with polygenic prediction modelling of cognitive related phenotypes. In **chapter 2** I employ state-of-the-art **multi-trait genomic methods to improve prediction** of intelligence and educational achievement throughout childhood and adolescence. Here, intercorrelations between traits at the genomic level are leveraged to boost predictive power of polygenic scores. I present evidence that exploiting multivariate relationships between traits consistently increases predictive power of polygenic scores by improving accuracy of estimates at the genome-wide level. I also show that different multi-trait methods are comparable in terms of this boost in prediction. While polygenic scores methods differ in their predictive capacity based on how they handle information across the genome.

In **chapter 3,** I approach polygenic prediction from a different angle, by combining polygenic scores and environmental measures in **multivariable longitudinal prediction models** of educational achievement in adolescence, while investigating their multivariate interplay, including gene-environment interaction and correlation. As evidence is gathering suggesting that polygenic scores are not pure measures of a person's genetic predisposition, as they partly capture environmental effects, it is important to assess the implications that this have for prediction when jointly considering environmental predictors, which partly capture individual genetic predisposition. Furthermore, a question this chapter addresses is the role of gene-environment interactions in the prediction setting, when considered within a multidimensional space where several G-E interactions are modelled jointly. By the same token the role of gene-environment correlation in the context of multivariable prediction is investigated from two different angles. First, to what extent do polygenic scores and environmental prediction models of EA overlap, sharing the same information? Second, to what extent are environmental and polygenic score effects reciprocally mediated? This chapter investigates these questions within a **multivariable G-E interplay** framework.

The second part of the thesis is concerned with the **co-occurrence of problem behaviours in childhood,** investigating the reason for the positive intercorrelations observed at the level of psychopathology measures throughout development. In **chapter 4,** I systematically investigate the manifestation of a general psychopathology factor (p-factor) throughout development, as a **common cause** for the intercorrelation between psychopathology traits. Here I quantify the **genetic and environmental contributions** to this **general psychopathology** dimension by employing quantitative genetic approaches across several measures of child psychopathology from age 7 to 16. I further discuss findings on the longitudinal stability of the p-factor and associations with a polygenic index of adult psychopathology.

In **chapter 5,** I investigate the co-occurrence of child psychopathology from a different angle by asking whether the positive intercorrelation between psychopathology dimensions can be partly attributed to a temporal **network of directed influences** of one trait on the other. I do this by investigating child problem behaviours over time, separating between-person trait-like stable effects from within-person state-like temporal effects of one psychopathology dimension on the other. In this study I further extend this network model to include an important source of variation of complex behavioural traits in childhood: **sibling effects**. Here, I develop an extension of the network model to family level data by estimating reciprocal directional influences between siblings over time separating them from similarities between siblings that arise through shared (genetic or environmental) influences that exist in a family. Within this approach I parse out genetic and environmental components of variance at the level of time-invariant overarching stable traits, as well as age-specific effects, controlling for the fact that family members are related to each other while extending the network model from an individual to a family level.

References

Abdellaoui, A., Hugh-Jones, D., Yengo, L., Kemper, K. E., Nivard, M. G., Veul, L., . . . Wray, N. R. (2019). Genetic correlates of social stratification in Great Britain. *Nature human behaviour*, 1-21.

Barcellos, S. H., Carvalho, L. S., & Turley, P. (2018). Education can reduce health differences related to genetic risk of obesity. *Proceedings of the National Academy of Sciences, 115*(42), E9765-E9772. doi:10.1073/pnas.1802909115

Belsky, D. W., Domingue, B. W., Wedow, R., Arseneault, L., Boardman, J. D., Caspi, A., . . . Harris, K. M. (2018). Genetic analysis of social-class mobility in five longitudinal studies. *Proceedings of the National Academy of Sciences, 115*(31), E7275-E7284. doi:10.1073/pnas.1801238115

Bergen, S. E., Gardner, C. O., & Kendler, K. S. (2007). Age-related changes in heritability of behavioural phenotypes over adolescence and young adulthood: a meta-analysis. Twin Research and Human Genetics, 10(3), 423-433.

Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F., & Keller, M. C. (2019). No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples. *American Journal of Psychiatry, 176*(5), 376-387.

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An expanded view of complex traits: from polygenic to omnigenic. *Cell, 169*(7), 1177-1186.

Chabris, C. F., Lee, J. J., Cesarini, D., Benjamin, D. J., & Laibson, D. I. (2015). The fourth law of behaviour genetics. *Current directions in psychological science, 24*(4), 304-312.

Claussnitzer, M., Dankel, S. N., Kim, K. H., Quon, G., Meuleman, W., Haugen, C., ... & Abdennur, N. A. (2015). FTO obesity variant circuitry and adipocyte browning in humans. *New England Journal of Medicine*, *373*(10), 895-907.

Colodro-Conde, L., Couvy-Duchesne, B., Zhu, G., Coventry, W. L., Byrne, E. M., Gordon, S., . . . Ripke, S. (2018). A direct test of the diathesis–stress model for depression. *Molecular psychiatry, 23*(7), 1590-1596.

de Vlaming, R., Okbay, A., Rietveld, C. A., Johannesson, M., Magnusson, P. K., Uitterlinden, A. G., . . . Koellinger, P. D. (2017). Meta-GWAS Accuracy and Power (MetaGAP) calculator shows that hiding heritability is partially due to imperfect genetic correlations across studies. *PLoS genetics, 13*(1).

Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLoS genetics, 9*(3), e1003348.

Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., . . . Domingue, B. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nature communications, 10*(1), 1-9.

Falconer, D. S., Mackay, T. F., & Frankham, R. (1996). Introduction to quantitative genetics (4th edn). *Trends in Genetics, 12*(7), 280.

Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S. J., Mallard, T. T., Hill, W. D., ... Tucker- Drob, E. M. (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature human behaviour, 3*(5), 513.

Knopik, V. S., Neiderhiser, J. M., DeFries, J. C., & Plomin, R. (2016). *Behavioural genetics*: Macmillan Higher Education.

Krapohl, E., Patel, H., Newhouse, S., Curtis, C. J., von Stumm, S., Dale, P. S., . . . Plomin, R. (2018). Multi-polygenic score approach to trait prediction. *Molecular psychiatry, 23*(5), 1368-1374.

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., . . . Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet*. doi:10.1038/s41588-018-0147-3

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., . . . Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature, 461*(7265), 747-753.

Martel, M. M., Pan, P. M., Hoffmann, M. S., Gadelha, A., do Rosário, M. C., Mari, J. J., . . . Salum, G. A. (2017). A general psychopathology factor (P factor) in children: structural model analysis and external validation through familial risk and child global executive function. *Journal of abnormal psychology, 126*(1), 137.

Moore, D. S., & Shenk, D. (2017). The heritability fallacy. Wiley Interdisciplinary Reviews: Cognitive Science, 8(1-2), e1400.

Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J. K., & Przeworski, M. (2020). Variable prediction accuracy of polygenic scores within an ancestry group. *Elife, 9*, e48376.

Neale, B., Ferreira, M., Medland, S., & Posthuma, D. (2007). *Statistical genetics: gene mapping through linkage and association*: Taylor & Francis.

Peyrot, W. J., Milaneschi, Y., Abdellaoui, A., Sullivan, P. F., Hottenga, J. J., Boomsma, D. I., & Penninx, B. W. (2014). Effect of polygenic risk scores on depression in childhood trauma. *The British Journal of Psychiatry, 205*(2), 113-119.

Peyrot, W. J., Van der Auwera, S., Milaneschi, Y., Dolan, C. V., Madden, P. A., Sullivan, P. F., . . . Pennix, B.W.J.H. (2018). Does childhood trauma moderate polygenic risk for depression? A meta-analysis of 5765 subjects from the psychiatric genomics consortium. *Biological psychiatry, 84*(2), 138-147.

Pickrell, J. K., Berisa, T., Liu, J. Z., Ségurel, L., Tung, J. Y., & Hinds, D. A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. *Nature genetics, 48*(7), 709.

Plomin, R. (2014). Genotype-environment correlation in the era of DNA. *Behaviour Genetics, 44*(6), 629-638.

Plomin, R. (2019). *Blueprint: How DNA makes us who we are*: Mit Press.

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 replicated findings from behavioural genetics. *Perspectives on psychological science, 11*(1), 3-23.

Plomin, R., & Deary, I. J. (2015). Genetics and intelligence differences: five special findings. Molecular psychiatry, 20(1), 98-108.

Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype-environment interaction and correlation in the analysis of human behaviour. *Psychological bulletin, 84*(2), 309.

Plomin, R., Haworth, C. M., & Davis, O. S. (2009). Common disorders are quantitative traits. *Nature reviews genetics, 10*(12), 872-878.

Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature reviews genetics, 19*(3), 148.

Purcell, S. (2002). Variance components models for gene–environment interaction in twin analysis. *Twin Research and Human Genetics, 5*(6), 554-571.

Stranger, B. E., Stahl, E. A., & Raj, T. (2011). Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics, 187*(2), 367-383.

The International Schizophrenia, C., Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., . . . Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature, 460*, 748. doi:10.1038/nature08185

Turkheimer, E. (2000). Three laws of behaviour genetics and what they mean. *Current directions in psychological science, 9*(5), 160-164.

Turley, P., Walters, R. K., Maghzian, O., Okbay, A., Lee, J. J., Fontana, M. A., . . . Benjamin, D. J. (2018). Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nature genetics, 50*(2), 229-237.

Visscher, P. M., & Goddard, M. E. (2019). From RA Fisher's 1918 paper to GWAS a century later. *Genetics, 211*(4), 1125-1130.

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 years of GWAS discovery: biology, function, and translation. *The American Journal of Human Genetics, 101*(1), 5-22.

Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., . . . Visscher, P. M. (2018). Meta-analysis of genome-wide association studies for height and body mass index in~ 700000 individuals of European ancestry. *Human molecular genetics, 27*(20), 3641-3649.

Young, A. I. (2019). Solving the missing heritability problem. *PLoS genetics, 15*(6), e1008222.

Chapter 2 – Genomic prediction of cognitive traits in childhood and adolescence

This chapter is presented as a published paper. It is an exact copy of the following publication:

Allegrini, A. G., Selzam, S., Rimfeld, K., von Stumm, S., Pingault, J. B., & Plomin, R. (2019). Genomic prediction of cognitive traits in childhood and adolescence. *Molecular psychiatry*, *24*(6), 819.

Supplementary materials are included in Appendix 1.

Published in final edited form as:
Mol Psychiatry. 2019 June ; 24(6): 819–827. doi:10.1038/s41380-019-0394-4.

Genomic prediction of cognitive traits in childhood and adolescence

A.G. Allegrini[1,*], **S. Selzam**[1], **K. Rimfeld**[1], **S. von Stumm**[2], **J.B. Pingault**[3], **R. Plomin**[1]

[1]King's College London, Social, Genetic and Developmental Psychiatry Centre, London, United Kingdom

[2]London School of Economics and Political Science, Department of Psychological and Behavioural Science, London, United Kingdom

[3]University College London, Clinical Educational and Health Psychology, Division of Psychology and Language Sciences, London, United Kingdom

Abstract

Recent advances in genomics are producing powerful DNA predictors of complex traits, especially cognitive abilities. Here, we leveraged summary statistics from the most recent genome-wide association studies of intelligence and educational attainment, with highly genetically correlated traits, to build prediction models of general cognitive ability and educational achievement. To this end, we compared the performances of multi-trait genomic and polygenic scoring methods. In a representative UK sample of 7,026 children at ages 12 and 16, we show that we can now predict up to 11 percent of the variance in intelligence and 16 percent in educational achievement. We also show that predictive power increases from age 12 to age 16 and that genomic predictions do not differ for girls and boys. We found that multi-trait genomic methods were effective in boosting predictive power. Prediction accuracy varied across polygenic score approaches, however results were similar for different multi-trait and polygenic score methods. We discuss general caveats of multi-trait methods and polygenic score prediction, and conclude that polygenic scores for educational attainment and intelligence are currently the most powerful predictors in the behavioural sciences.

Introduction

Ever increasing sample sizes and methodological advances in polygenic methods have made it possible to powerfully predict complex traits such as cognitive abilities without knowing anything about the causal chain between genes and behaviour. Progress in predicting cognitive traits from inherited DNA variants has been rapid in the past five years and

[*]corresponding author: Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, Denmark Hill, London, United Kingdom, SE5 8AF, andrea.allegrini@kcl.ac.uk.

Author contributions

AGA and RP conceived and designed the study. AGA analysed and interpreted the data. SS performed quality control of genotype data. AGA and RP wrote the manuscript. SS, KR, SvS and JBP contributed to and critically reviewed the manuscript.

Competing interests

The authors declare no conflict of interest.

Supplementary information is available at MP's website

especially in the past year[1]. Three methodological advances have mainly been responsible for this progress: increasingly large genome-wide association (GWA) studies, genome-wide polygenic scores (GPS) and multivariate analytic tools. The key has been the recognition that the largest associations are extremely small, accounting for less than 0.05% of the variance[2]. To achieve sufficient power to detect such small effect sizes, samples in the hundreds of thousands are needed before GWA studies can begin to detect these tiny effects. Because the largest associations are so small, useful predictions of individual differences can only be made by aggregating the effects of thousands of DNA variants in GPS[3]. The third advance is the development of genomic methods that leverage genetic correlations between traits to boost power for variant discovery[4] and polygenic risk prediction[5].

Together, these three advances have greatly increased the ability to predict intelligence, educational attainment (years of schooling), and educational achievement (tested performance). For example, for intelligence, until 2017, no replicable associations were found in seven GWA studies[6–12], which we refer to collectively as 'IQ1'. These studies had sample sizes from 18,000 to 54,000, which seemed large at the time but were not sufficiently powered to detect effect sizes of 0.05%. GPS derived from these IQ1 GWA studies at most accounted for 1% of the variance in independent samples. Increasing GWA sample sizes to 78,000 (IQ2[13]) and then to 280,000 (IQ3[14]) paid off in increasing predictive power of GPS from 1% to 3% to 4%. Here we present results for IQ3.

Educational attainment has led the way in terms of increasing GWA sample size, from 125,000 in 2013 (EA1[15]) to 294,000 in 2016 (EA2[16]) to 1.1 million in 2018 (EA3[17]). The growing sample sizes increased the predictive power of GPS from 2% to 3% to 12% of the variance in educational attainment[1]. Similarly, in previous work we showed that EA GPS predicted an increasingly substantial amount of variance in tested educational achievement as sample size from replications of the EA GWAS increased over the years. EA1 predicted 3% of the variance in educational achievement at age 16 [18] and EA2 predicted 9% of the variance for overall educational achievement at age 16 [19].

Because 'years of education' is obtained as a demographic marker in most GWA studies, it was possible to accumulate samples sizes with the necessary power to detect very small effect sizes. It is more difficult to obtain very large sample sizes for intelligence, which needs to be assessed with a psychometric test administered to each individual, whereas years of education can be captured with a single self-reported item. Because of the large sample size available for EA GWA studies and the substantial genetic correlation between EA and intelligence, EA GPS predicted as much or more variance in intelligence than did GPS derived from GWAS of the target trait of intelligence itself. EA1 predicted 1% of the variance in intelligence [18, 20] and EA2 predicted 4% of the variance [16]. Here we present results for EA3.

Finding that EA GPS predict educational achievement and intelligence better than do GWA of the target traits themselves suggests the usefulness of multivariate approaches. In a previous study, a multivariate GPS approach involving regularized regression was applied to show that with EA2 and 80 other GPS 11% of the variance in educational achievement at age 16 and 5% of the variance in intelligence at age 12 could be predicted[21]. Although

adding 1–2% to the predictive power of GPS might not seem like much, it should be noted that five years ago the total variance that could be predicted in either trait was statistically indistinguishable from zero.

The aim of the present study is to estimate how much variance in intelligence and educational achievement can be predicted by applying several state-of-the-art multi-trait genomic approaches and leveraging highly powered GWA summary statistics. First we compare three polygenic score methods (PRSice[22], LDpred[23], and Lassosum[24]) and test how much variance the new IQ3 and EA3 GPS maximally predict. We then jointly analyse IQ3 and EA3 with three highly (genetically) correlated traits (Income[25], Age when completed full time education[26], Time spent using computer[26]) to boost predictive power and compare performance of three multi-trait methods (Genomic SEM[27], MTAG[4] and SMTpred[5]) using predictive power as our criterion.

We conducted these analyses in a sample of 7,026 unrelated individuals from the Twins Early Development study, which is representative of the UK population[28]. We analysed intelligence and educational achievement at the end of compulsory schooling in the UK at age 16; we also investigated developmental trends in genomic prediction from age 12 to 16. Based on previous research[19], we expected genomic predictions to increase from 12 to 16.

Materials and Methods

Sample

The sample was drawn from the Twins Early Development Study (TEDS[29]), an ongoing population-based longitudinal study. It consists of twins born in England and Wales between 1994 and 1996, who have been assessed on a variety of psychological domains. More than 10,000 twin pairs representative of the general UK population [28] remain actively involved in the study to date. Ethical approval for TEDS has been provided by the King's College London Ethics Committee (reference: PNM/09/10–104). Parental consent was obtained before data collection. Genotypes for 10,346 individuals (including 3,320 DZ twin pairs) were processed with stringent quality control procedures followed by SNP imputation using the Haplotype Reference Consortium (release 1.1) reference panels. Current analyses were limited to the genotyped and imputed sample of 7,026 unrelated individuals. Following imputation, we excluded variants with minor allele frequency < 0.5%, Hardy-Weinberg equilibrium p-values of $< 1\times10^{-5}$. To ease computational demands, we selected variants with an info score of 1, resulting in 515,000 SNPs used for analysis (see Supplementary Methods S1 for a full description of quality control and imputation procedures).

Outcome variables

The outcome variables were intelligence and educational achievement at ages 12 and 16. Intelligence was assessed as a composite of verbal and nonverbal web-based tests. Educational achievement was indexed by a mean of scores on the compulsory subjects of English, mathematics and science obtained from the UK National Pupil Database. A more detailed description of outcome variables is provided in the Supplementary Methods S2. Supplementary Table S1 includes descriptive statistics for the outcomes variables and

Supplementary Figure S1 shows phenotypic correlations. Phenotypes and polygenic scores were corrected for age, sex and 10 genetic principal components. The obtained standardised residuals were used in all subsequent analyses.

Discovery GWA summary statistics

We based our prediction models on beta weights derived from large, publicly available, GWA summary statistics. Of central importance for our analyses were the most recent GWA studies of educational attainment (EA3[17]) and intelligence (IQ3[14]). Because the original IQ GWA meta-analysis included TEDS as one of its samples, to avoid bias due to sample overlap with our target sample we used summary statistics from new GWA analyses that excluded TEDS. The EA3 summary statistics employed here do not include 23andMe data (~300k individuals) due to their data availability policy.

Polygenic score approaches

We used IQ3 and EA3 summary statistics to construct genome-wide polygenic scores (GPS) comparing three distinct approaches: PRSice2[22], a clumping/pruning + P-value thresholding (P+T) approach, with an in-built high-resolution option that returns the best-fit GPS for the trait of interest; LDpred[23], a Bayesian approach that uses a prior on the expected polygenicity of a trait (assumed fraction of non-zero effect markers) and adjusts for linkage disequilibrium based on a reference panel to compute SNPs weights; and Lassosum[24], a machine-learning approach which uses penalized regression on GWA summary statistics to produce more accurate beta weights.

A detailed description of the construction of these polygenic scores is included in Supplementary Methods S3.

Multi-trait approaches

In order to boost power of IQ3 (N = 266,453) and EA3 (N = 766,345) GWA results and thus precision of beta weights to construct more predictive IQ3 and EA3 polygenic scores, we jointly analysed these summary statistics with three cognitive and educationally relevant traits: “Income”[25] (N = 96,900), “Age when completed full time education[26] (N = 226,899) and “Time spent using computer”[29] (N = 261,987). The choice of these traits is consistent with a multi-trait framework, as these traits show the highest genetic correlations with IQ and educational attainment among publicly available GWA summary statistics, with pairwise-genetic correlations ranging from ~.5 to ~.9 (see Supplementary Figure S2). Summary statistics from these GWA studies are reported in Supplementary Table S2.

We used three recently developed multi-trait methods, one of which is specifically designed to boost polygenic score prediction: SMTpred[5], and two of which are strictly speaking multivariate GWA approaches, designed to boost power for discovery, but which have been shown to increase predictive power of polygenic scores created from multi-trait reweighted summary statistics: MTAG[4] and Genomic SEM[27]. Details about these methods are provided in Supplementary Methods S4. Briefly, SMTpred[5] is a multi-trait extension of the random effects model approach, which can be used to create multivariate best linear unbiased predictors based on summary statistics (wMT-SBLUP). MTAG is a generalization of

inverse-variance weighted meta-analysis, which jointly analyses univariate GWA summary statistics. It boosts power for discovery for each trait conditional on the effect size estimates of other traits and outputs trait-specific summary statistics. Genomic SEM is a two-stage structural equation modelling approach that can be applied in the context of multivariate GWA. In the form employed here (common factor GWAS), it directly tests effect of SNPs on a latent genetic factor defined by several indicators (i.e. traits) and outputs summary statistics for the common factor. We also compared these new multivariate approaches to a simple multiple regression on intelligence and on educational achievement using five GPS, each derived from the univariate GWA summary statistics used in multi-trait analyses.

Analyses

Univariate analyses

We first calculated polygenic scores for the IQ3 and EA3 GWA summary statistics using PRSice, LDpred and Lassosum. This was done to compare current state-of-the-art polygenic scores approaches and in order to obtain a benchmark against which to compare improvements in prediction accuracy due to multivariate GWA analyses. For each phenotype (i.e. intelligence and educational achievement at ages 12 and 16), we randomly split the sample into training and test sets (~50% training, ~50% test). Supplementary Table S1 shows descriptive statistics for each set. In the training sets, parameter optimization of GPS was performed, in which each GPS instrument (or p-value threshold in the case of PRSice, fraction of markers with nonzero effect in the case of LDpred, and tuning parameters in the case of Lassosum) was tested on each of the four phenotypes and the best instrument was selected with respect to prediction accuracy (as indexed by R^2). Performance of the optimized GPS instrument retained from the validation was then assessed in the test sample in order to evaluate how well the chosen predictors would perform in independent samples. We then proceeded to perform the multi-trait analyses.

Multi-trait analyses

We performed a multi-trait reweighting in SMTpred after transforming the ordinary least square betas from GWA studies of ‘IQ’, ‘EA’, ‘Income’, ‘Age completed full time education’ and ‘Time spent using computer’ in approximate Best Linear Unbiased Predictors (BLUP) using GCTA-Cojo [30]. We then used LDSC to calculate SNP h2 and genetic correlations between traits and proceeded to the multivariate weighting of traits as described in (Meier et al., 2018) to obtain multi-trait summary statistics BLUP (wMT-SBLUP; see also Supplementary Methods S4).

MTAG was run on the five GWA summary statistics (IQ, EA, Income, Age completed full time Education, Income) using standard settings. Because MTAG combines differently powered summary statistics (as indexed by the GWAS mean χ^2; see Supplementary Methods S4), as well as differing degrees of genetic overlap between traits, it can lead to an increased rate of false positives Type I error[4]. However, this is not an issue in the present study, which focuses on prediction accuracy rather than variant discovery. It has been shown[4] that MTAG estimates consistently have a lower genome-wide mean-squared error compared to single-trait GWA estimates, and, therefore, polygenic scores created from

MTAG perform better than those created at the univariate level. However, in order to control for type I error inflation, we used the recommended[4] false discovery rate (FDR) calculations (see Supplementary Methods S4).

The same five summary statistics were analysed using Genomic SEM. First a common factor model with the five summary statistics as indicators was fitted using a weighted least-square (WLS) estimator (default setting in Genomic SEM). Then a common factor GWA analysis with a WLS estimator was run, testing effects of single SNPs on the common factor. The WLS estimator was expected to yield lower standard errors and possibly increased prediction accuracy of GPS[30].

We then created polygenic scores from the MTAG EA, MTAG IQ and common factor GWA summary statistics across the three polygenic scores approaches, after splitting the sample into a training set to tune parameters and a testing set to assess model performance. In the case of SMTpred, polygenic scores for IQ3 and EA3 converted and reweighted indices (wMT-SBLUP) were calculated using PLINK[31]. These multi-trait predictors were then directly tested for model performance in the test set, as with the other GPS approaches. Based on previous power analyses for polygenic score prediction in the TEDS sample[19, 32] we did not expect any power issues for the current analysis plan.

For prediction estimates derived from both univariate and multi-trait models, we calculated bootstrapped confidence intervals with 1000 replications. Furthermore, we performed a comparison of R^2 estimates between models, by calculating bootstrapped confidence intervals for the R^2 pairwise mean differences. As such, for each model, bootstrap samples were generated by sampling with replacement from the data 1000 times. Each row of data for resampling consisted of all polygenic scores and phenotypes examined herein. This procedure yielded an R^2 distribution for each method tested. The R^2 difference between methods was then calculated for each bootstrap iteration. This generated a distribution of R2 differences, from which we calculated 95% confidence intervals.

Results

Polygenic score prediction of IQ and EA across GPS methods

Figure 1 shows variance in intelligence and educational achievement predicted by IQ3 GPS and EA3 GPS calculated following three polygenic score methods (PRSice, LDpred and Lassosum). Supplementary Table S3 presents associations in the training and test sets across all models.

For intelligence, IQ3 GPS predicted a maximum of 5.3% ($\beta = 0.221$, se = 0.023, $p < .0001$) of the variance at age 12 and 6.7% ($\beta = 0.266$, se = 0.032, $p < .0001$) at age 16. For educational achievement, EA3 GPS predicted a maximum of 6.6% ($\beta = 0.259$, se = 0.020, $p < .0001$) of the variance at age 12 and 14.8% ($\beta = 0.389$, se = 0.019, $p < .0001$) at age 16. EA3 GPS was also a powerful predictor of intelligence, predicting 7.2% ($\beta = 0.265$, se = 0.024, $p < .0001$) of the variance in intelligence at age 12 and 9.9% ($\beta = 0.321$, se = 0.031, $p < .0001$) at age 16.

Generally, Lassosum was the most powerful approach, predicting up to 1% more of the variance compared to LDpred and up to 2% more compared to PRSice. Supplementary Figure S4 shows a comparison of prediction estimates for each pair of approaches tested. Bootstrapped confidence levels calculated for pairwise comparisons indicated significant differences in prediction accuracy of IQ3 GPS within-trait between LDpred and PRSice at age 12 (MeanDiff = −0.014, 95% CIs [−0.024; −0.005]), and cross-trait at age 12 and 16. However, no significant differences were found for PRSice- vs LDpred-based EA3 GPS. Similarly, significant differences were also found for IQ3 GPS between Lassosum and PRSice at age 12 within trait (MeanDiff = −0.010, 95% CIs −0.021; −0.001]), and at age 12 and 16 cross-trait. Lassosum-based EA3 GPS performed better within trait at age 16 (MeanDiff = −0.020, 95% CIs −0.031; −0.009]).

No differences were found in prediction accuracy between LDpred- vs Lassosum-based IQ3 GPS or EA3 GPS, within or cross-trait. Supplementary Table S3a reports mean differences and CIs for these comparisons.

Multi-trait polygenic score prediction

Results of multivariate GWA analyses are reported in Supplementary Methods S4 and Supplementary Tables S5 and S6. Here we report results of polygenic score associations for our best predictive polygenic models after multi-trait approaches were applied to GWA summary statistics (Figure 3 and Figure S5). Figure S6 shows a comparison of variance predicted in intelligence and educational achievement at ages 12 and 16 in the test samples across polygenic score methods after multi-trait analyses. Supplementary Table S4 reports details of these results.

Figure 2 presents variance predicted in intelligence and educational achievement at age 16 by polygenic scores derived from multi-trait methods. For intelligence, variance predicted by IQ3 GPS increased from 6.7% (Figure 1) to a maximum of 10.0% ($\beta = 0.327$, se = 0.032, $p < .0001$) at age 16. For educational achievement, variance predicted by EA3 GPS increased from 14.8% to a maximum of 15.9% ($\beta = 0.403$, se = 0.018, $p < .0001$) at age 16. Again, EA3 GPS was generally the best performing predictor across phenotypes, predicting a maximum of 10.6% ($\beta = 0.332$, se = 0.031, $p < .0001$) in intelligence. Similar improvements in prediction were observed at age 12 (see Supplementary Table S4 and supplementary figure S4).

Supplementary Figure S6 shows a test of the differences in predictive performance of Lassosum-based scores between multi-traits methods tested at age 12 and 16. There were no significant differences between multi-trait methods for both IQ3 and EA3 GPS across all phenotypes. The only exceptions were the SMTpred IQ3 score, which tended to perform better than MTAG at age 16 cross-trait (MeanDiff = −0.011, 95% CIs [−0.022; −0.001]), and the MTAG EA3 score which tended to perform better than Genomic SEM at age 16 within trait (MeanDiff = −0.0077, 95% CIs [−0.0143; −0.002]). Supplementary Table S4 a reports mean differences and CIs for these comparisons.

Polygenic scores quantile differences—Figure 3 shows the results for the best predictive models at age 16 by GPS deciles. For both intelligence (panel a) and educational

achievement (panel b), the relationship with GPS deciles is linear and the lowest and highest deciles differ substantially. For intelligence, the mean difference (~1 SD) is comparable to 15 IQ points. For educational achievement, the mean difference corresponds to an average 'C' grade for the lowest decile and an average 'A' grade for the highest decile. However, the range of distributions in the lowest and highest deciles overlap considerably, as would be expected from GPS correlations of ~0.32 with intelligence and ~0.40 with educational achievement.

Sex differences—We tested associations for the best prediction model (i.e. MTAG EA3 GPS calculated in Lassosum) separately for males and females in the test set. For intelligence at age 16, the GPS predicted 10.7% of the variance (95% CIs [6.33;16.74]) in males (N= 369, $\beta = 0.334$, se = 0.049) and 10.5% (95% CIs [6.49;15.41]) in females (N = 558, $\beta = 0.329$, se = 0.040). For educational achievement in males (N = 1,105) the GPS predicted 14.2% (95% CIs [10.96;17.86]) of the variance ($\beta = 0.375$, se = 0.027); in females (N = 1,300) estimates were 17.2% (95% CIs [13.51;21.43]; $\beta = 0.420$, se = 0.025). To test the significance of these sex differences, we performed a Fisher's r to z transformation of corresponding correlation coefficients. Sex differences were not significant for intelligence (observed z = −0.066, p = 0.472) nor educational achievement (Observed z = −1.419, p= 0.077).

Multiple regression model—We compared the results from our multi-trait GPS analyses to a simple multiple regression using the five GPS from summary statistics of our multi-trait analyses (IQ, EA, income, age when completed full time education, time spent using computer) to predict intelligence and educational achievement. The multiple regression model predicted similar amounts of variance as the best single multi-trait GPS predictors. For intelligence, the adjusted R^2 was 8.6% at age 12 and 9.9% at age 16. For educational achievement, the adjusted R^2 was 9.6% at age 12 and 16.7% at age 16. Results are shown in Supplementary Table S7.

Discussion

Using summary statistics from the latest GWA studies of intelligence (IQ3[14]) and educational attainment (EA3[17]), we report the strongest polygenic prediction estimates for cognitive-related traits to date. Comparing standard polygenic score approaches, we showed that IQ3 GPS predicts a maximum of 6.73% of the variance in intelligence at age 16, while EA3 GPS predicts 14.78% of the variance in educational achievement at age 16.

In an attempt to boost predictive power, we compared results using state-of-the-art genomics methods that leverage the multivariate nature of traits in order to increase power of GWA summary statistics. We then tested boosted summary statistics across a number of polygenic score approaches, showing that we can predict 10.6% of the variance in intelligence and 15.9% of the variance in educational achievement, both at age 16. These results compare favourably with polygenic prediction estimates from the recent EA3 GWA analysis, whereby a polygenic score constructed from multi-trait summary statistics of educational attainment and three cognitive-related phenotypes predicted up to 13% of the variance in educational attainment and up to 10% in cognitive performance[17], this is especially notable given the

larger discovery sample size employed in that study (N ~ 1.1 million including 23andMe). We note that differences between these studies may be attributable to systematic differences at the level of trait measurement (e.g. accuracy of measurement) and sample characteristics (e.g. differences in ancestry; differences in heritability). Nevertheless, this is a good indication that a multi-trait approach to polygenic prediction replicates well across independent samples yielding robust prediction estimates.

We found that trait prediction increased from age 12 to age 16. Polygenic scores become more predictive with age, probably because as the sample approaches adulthood it is closer in age to the samples in which beta weights were estimated in the original GWA studies for IQ3 and EA3. Another possible reason for this finding is that given that heritability of intelligence increases with age[33], the variance that can be predicted by cognitive-related polygenic scores also increases. Lastly, we did not find significant differences in the predictive power of IQ3 and EA3 for males and females.

These results indicate the usefulness of taking into account the multivariate nature of complex traits in polygenic prediction, and add to the possibility of practical use of polygenic scores at the level of individuals[34]. It is important to note that we randomly split our sample (~50%) to validate our models and assessed performance of prediction models in the test sample in order to avoid overfitting. Because TEDS is a representative sample of the UK population, these prediction estimates are expected to be a close representation of how these models would perform in similar samples. Overall, multi-trait methods were successful in increasing variance predicted; compared to our 'baseline' predictions, estimates increased from 1% to 3%. Multi-trait methods were especially useful in increasing predictive power of the IQ3 GPS, which was constructed using less powerful summary statistics than the EA3 GPS. However, differences in prediction accuracy across the tested combinations of genomic methods seemed to reflect differences in polygenic score approaches rather than in multi-traits approaches. An indication of this intuition was also provided by a formal comparison of R^2 estimates, which showed no consistent differences across multi-trait methods. Yet, reassuringly, there were no dramatic differences in prediction accuracy across polygenic score approaches either, especially when considering approaches that do not perform clumping (thereby losing information across the genome).

One limitation that could affect the interpretability of our findings is that by jointly analysing traits with differing levels of power and genetic overlap, the multi-trait methods considered here might confound the genetic architecture of boosted traits with that of other traits. In this regard, genetic correlations between traits before and after multi-trait analyses and with a control trait, as those reported in Supplementary Methods S4, may indicate the degree to which the genetic architecture of one trait has 'shifted' towards that of others in the multi-trait analysis. This is an important post-hoc test to be considered by future studies employing multi-trait approaches in the context of polygenic prediction.

An ongoing debate concerns the causal mechanisms by which polygenic scores predict phenotypes such as educational achievement and intelligence. Passive gene-environment correlation may be a mechanism underlying the association between polygenic scores and educational attainment. Given parent-child shared genetics (~50%), if EA trait-increasing

variants are correlated with rearing environments which in turn are contributing to attainment, GWAS estimates obtained for EA would be partly picking up genetic effects mediated via the environment. That is, GWAS effect estimates may be due to indirect genetic effects via rearing environments that could reflect both inherited and non-inherited parental DNA. Therefore, the association between an individual's EA polygenic score and cognitive traits could partly reflect an environmentally transmitted parental genetic effect[17, 35, 36]. Analyses relying on family-based designs have put forward evidence in this regard[17, 37, 38]. These studies confirmed what have long been acknowledged by twin and adoption studies on the nature of nurture[39,40]. Separating the different mechanisms of gene-environment interplay by which polygenic scores influence complex traits is an important area of research. However, prediction of individual differences in behavioural phenotypes from polygenic scores can be achieved without an underlying explanatory model.

Finally, a general limitation of all genomic analyses is that they only assess additive effects of common SNPs used on currently SNP arrays. SNP heritability is the ceiling for polygenic score prediction, which is about 20%[14] of the total variance for intelligence and 30% [41] for educational achievement. Viewed in this light, our best polygenic scores predict about half of the SNP heritability. With bigger and better GWA studies and other methodological advances like multivariate approaches, the missing SNP heritability gap will be narrowed. Polygenic scores will only reach their full potential when we are able to close the gap between SNP heritability (about 25%) and family study estimates of heritability (about 50%).

Nonetheless, these polygenic scores predictions are already among the strongest predictors in the behavioural sciences. Because inherited DNA variants do not change during development, polygenic scores are unique predictors in two ways. First, unlike other characteristics of the individual, DNA variants can predict individual differences in adult behaviour from birth. Second, unlike other correlations, associations between DNA variants and behaviour are causal from DNA to behaviour in the sense that there can be no backward causation from behaviour to DNA. These unique features will put genomic prediction of cognitive traits in the front line of the DNA revolution.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We gratefully acknowledge the ongoing contribution of the participants in the Twins Early Development Study (TEDS) and their families. TEDS is supported by a program grant to RP from the UK Medical Research Council (MR/M021475/1 and previously G0901245), with additional support from the US National Institutes of Health (AG046938). The research leading to these results has also received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ grant agreement n° 602768 and ERC grant agreement n° 295366. RP is supported by a Medical Research Council Professorship award (G19/2). This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 721567.

References

1. Plomin R, von Stumm S. The new genetics of intelligence. Nature reviews Genetics 2018; 19(3): 148–159.
2. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. American journal of human genetics 2017; 101(1): 5–22. [PubMed: 28686856]
3. Pasaniuc B, Price AL. Dissecting the genetics of complex traits using summary association statistics. Nature reviews Genetics 2017; 18(2): 117–127.
4. Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. Nature genetics 2018; 50(2): 229–237. [PubMed: 29292387]
5. Maier RM, Zhu Z, Lee SH, Trzaskowski M, Ruderfer DM, Stahl EA et al. Improving genetic prediction by leveraging genetic correlations among human diseases and traits. Nature Communications 2018; 9(1): 989.
6. Benyamin B, Pourcain B, Davis OS, Davies G, Hansell NK, Brion MJ et al. Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. Molecular psychiatry 2014; 19(2): 253–258. [PubMed: 23358156]
7. Butcher LM, Davis OS, Craig IW, Plomin R. Genome-wide quantitative trait locus association scan of general cognitive ability using pooled DNA and 500K single nucleotide polymorphism microarrays. Genes, brain, and behavior 2008; 7(4): 435–446.
8. Davies G, Armstrong N, Bis JC, Bressler J, Chouraki V, Giddaluru S et al. Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53949). Molecular psychiatry 2015; 20(2) : 183–192. [PubMed: 25644384]
9. Davies G, Marioni RE, Liewald DC, Hill WD, Hagenaars SP, Harris SE et al. Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). Molecular psychiatry 2016; 21(6): 758–767. [PubMed: 27046643]
10. Davies G, Tenesa A, Payton A, Yang J, Harris SE, Liewald D et al. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. Molecular psychiatry 2011; 16(10): 996–1005. [PubMed: 21826061]
11. Plomin R, Hill L, Craig IW, McGuffin P, Purcell S, Sham P et al. A genome-wide scan of 1842 DNA markers for allelic associations with general cognitive ability: a five-stage design using DNA pooling and extreme selected groups. Behavior genetics 2001; 31(6): 497–509. [PubMed: 11838529]
12. Trampush JW, Yang ML, Yu J, Knowles E, Davies G, Liewald DC et al. GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: a report from the COGENT consortium. Molecular psychiatry 2017; 22(3): 336–345. [PubMed: 28093568]
13. Sniekers S, Stringer S, Watanabe K, Jansen PR, Coleman JRI, Krapohl E et al. Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. Nature genetics 2017; 49 (7): 1107–1112. [PubMed: 28530673]
14. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, de Leeuw CA et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nature genetics 2018; 50(7): 912–919. [PubMed: 29942086]
15. Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science (New York, NY) 2013; 340(6139): 1467–1471.
16. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature 2016; 533(7604): 539–542. [PubMed: 27225129]
17. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nature genetics 2018.

18. Krapohl E, Plomin R. Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. Molecular psychiatry 2016; 21(3): 437–443. [PubMed: 25754083]
19. Selzam S, Krapohl E, von Stumm S, O'Reilly PF, Rimfeld K, Kovas Y et al. Predicting educational achievement from DNA. Molecular psychiatry 2018; 23(1): 161. [PubMed: 28948970]
20. Rietveld CA, Esko T, Davies G, Pers TH, Turley P, Benyamin B et al. Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. Proceedings of the National Academy of Sciences of the United States of America 2014; 111(38): 13790–13794. [PubMed: 25201988]
21. Krapohl E, Patel H, Newhouse S, Curtis CJ, von Stumm S, Dale PS et al. Multi-polygenic score approach to trait prediction. Molecular psychiatry 2018; 23(5): 1368–1374. [PubMed: 28785111]
22. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. Bioinformatics 2015; 31(9): 1466–1468. [PubMed: 25550326]
23. Vilhjalmsson BJ, Yang J, Finucane HK, Gusev A, Lindstrom S, Ripke S et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. American journal of human genetics 2015; 97(4): 576–592. [PubMed: 26430803]
24. Mak TSH, Porsch RM, Choi SW, Zhou X, Sham PC. Polygenic scores via penalized regression on summary statistics. Genetic epidemiology 2017; 41(6): 469–480. [PubMed: 28480976]
25. Hill WD, Hagenaars SP, Marioni RE, Harris SE, Liewald DCM, Davies G et al. Molecular Genetic Contributions to Social Deprivation and Household Income in UK Biobank. Current Biology 2016; 26(22): 3083–3089. [PubMed: 27818178]
26. Seed C. Hail: An Open-Source Framework for Scalable Genetic Data. 2017.
27. Grotzinger AD, Rhemtulla M, de Vlaming R, Ritchie SJ, Mallard TT, Hill WD et al. Genomic SEM Provides Insights into the Multivariate Genetic Architecture of Complex Traits. bioRxiv 2018.
28. Haworth CM, Davis OS, Plomin R. Twins Early Development Study (TEDS): a genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. Twin research and human genetics : the official journal of the International Society for Twin Studies 2013; 16(1): 117–125. [PubMed: 23110994]
29. Oliver BR, Plomin R. Twins' Early Development Study (TEDS): a multivariate, longitudinal genetic investigation of language, cognition and behavior problems from childhood through adolescence. Twin research and human genetics : the official journal of the International Society for Twin Studies 2007; 10(1): 96–105. [PubMed: 17539369]
30. Yang J, Ferreira T, Morris AP, Medland SE, Madden PA, Heath AC et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nature genetics 2012; 44(4): 369–375, s361–363. [PubMed: 22426310]
31. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics 2007; 81(3): 559–575. [PubMed: 17701901]
32. Krapohl E, Euesden J, Zabaneh D, Pingault JB, Rimfeld K, von Stumm S et al. Phenome-wide analysis of genome-wide polygenic scores. Molecular psychiatry 2015; 21: 1188. [PubMed: 26303664]
33. Haworth CMA, Wright MJ, Luciano M, Martin NG, de Geus EJC, van Beijsterveldt CEM et al. The heritability of general cognitive ability increases linearly from childhood to young adulthood. Molecular psychiatry 2010; 15(11): 1112–1120. [PubMed: 19488046]
34. Plomin R. Blueprint: How DNA Makes Us Who We Are. Allen Lane/Penguing Press: London, 2018.
35. Fletcher JM, Lehrer SF. Genetic lotteries within families. Journal of health economics 2011; 30(4): 647–659. [PubMed: 21664708]
36. Pingault J-B, O'Reilly PF, Schoeler T, Ploubidis GB, Rijsdijk F, Dudbridge F. Using genetic data to strengthen causal inference in observational research. Nature Reviews Genetics 2018; 19(9): 566–580.

37. Belsky DW, Domingue BW, Wedow R, Arseneault L, Boardman JD, Caspi A et al. Genetic analysis of social-class mobility in five longitudinal studies. Proceedings of the National Academy of Sciences 2018; 115(31): E7275–E7284.
38. Kong A, Thorleifsson G, Frigge ML, Vilhjalmsson BJ, Young AI, Thorgeirsson TE et al. The nature of nurture: Effects of parental genotypes. Science (New York, NY) 2018; 359(6374): 424–428.
39. Plomin R, Bergeman CS. The nature of nurture: Genetic influence on "environmental" measures. Behavioral and Brain Sciences 2011; 14(3): 373–386.
40. Plomin R. Geneticsand experience: The interplay between nature and nurture. Sage Publications: Thousand Oaks, CA, 1994.
41. Krapohl E, Plomin R. Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. Molecular psychiatry 2015; 21: 437. [PubMed: 25754083]

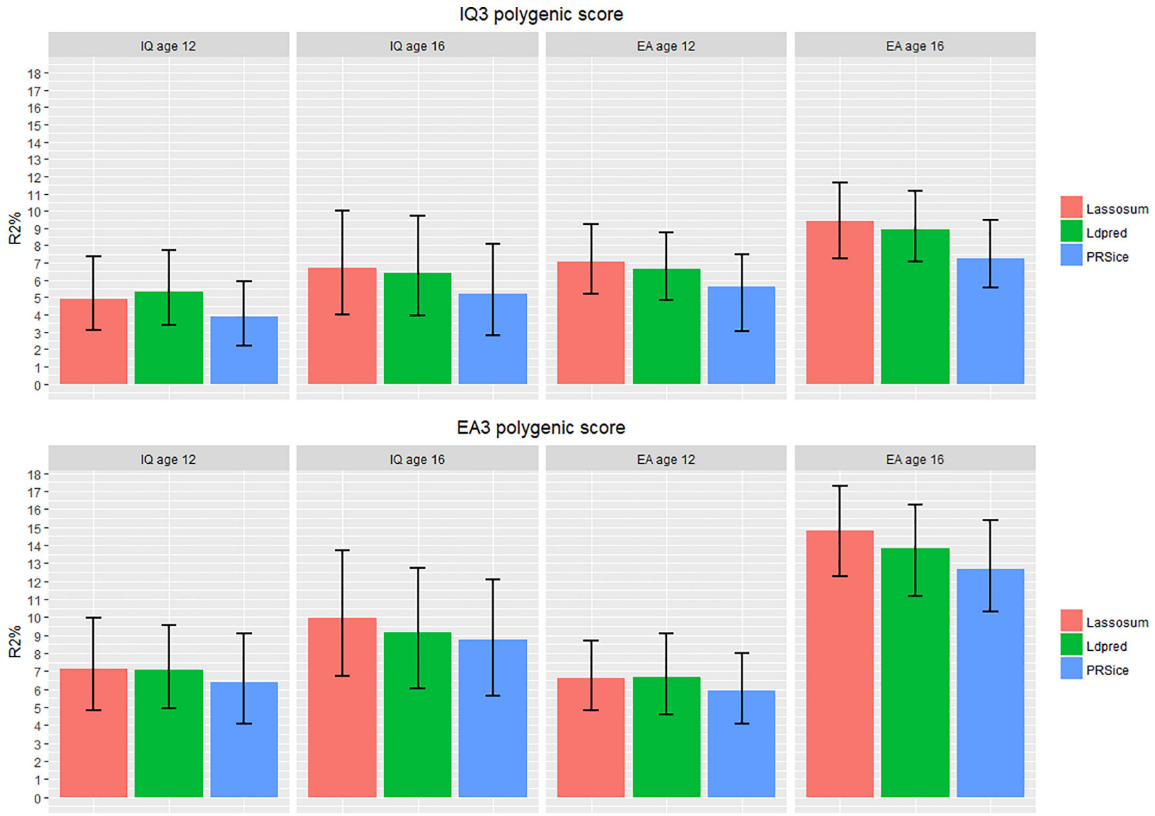


Figure 1.
Polygenic score prediction of intelligence (IQ) and educational achievement (EA) at age 12 and 16. Figure shows polygenic prediction accuracy across polygenic score methods. Error bars are bootstrapped 95% confidence intervals based on 1,000 replications.

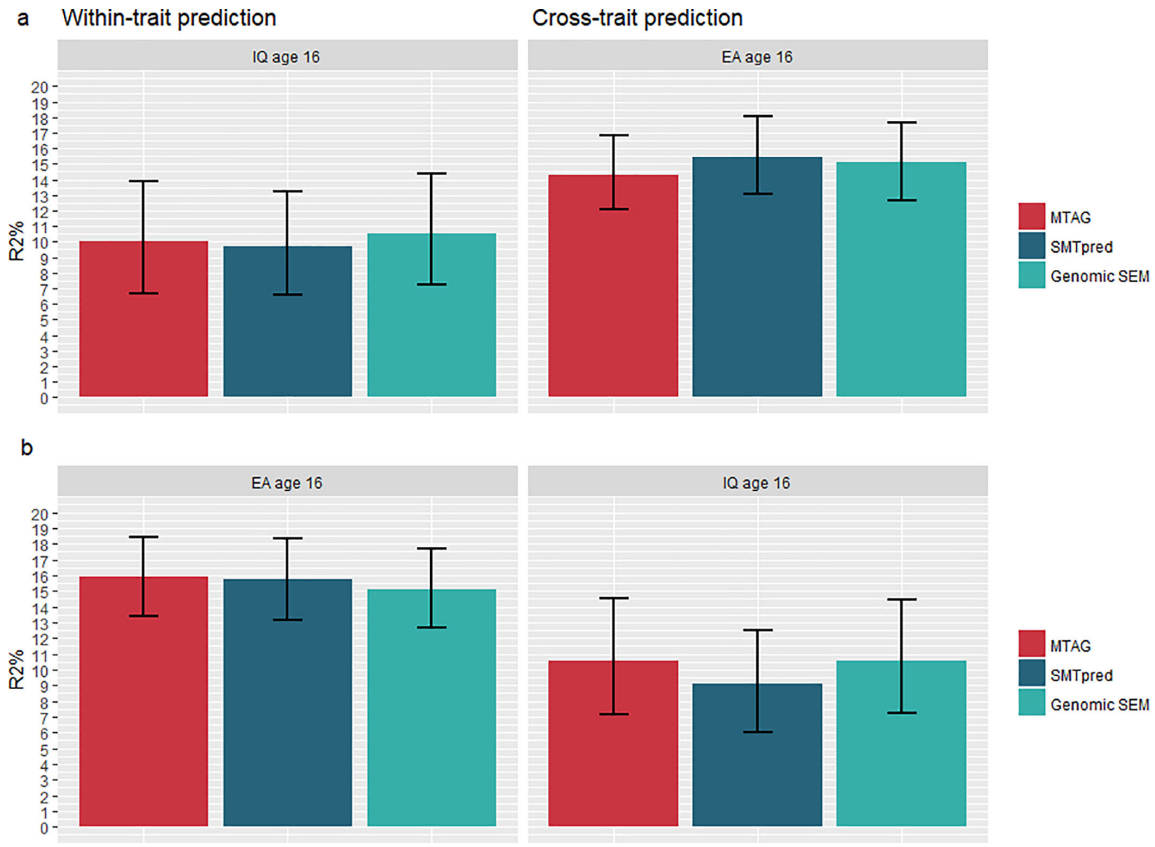


Figure 2.

Within-trait and cross-trait polygenic score prediction of intelligence and educational achievement at age 16 across multi-trait methods.

Note. MTAG = MTAG IQ3 (panel **a**)/ MTAG EA3 (panel **b**) polygenic scores constructed in Lassosum; SMTpred = IQ3 (panel **a**)/EA3 (panel **b**) wMT-SBLUP predictors; Genomic SEM = Common Factor polygenic score constructed in Lassosum (panel **a** and **b**). Error bars are bootstrapped 95% confidence intervals based on 1,000 replications.

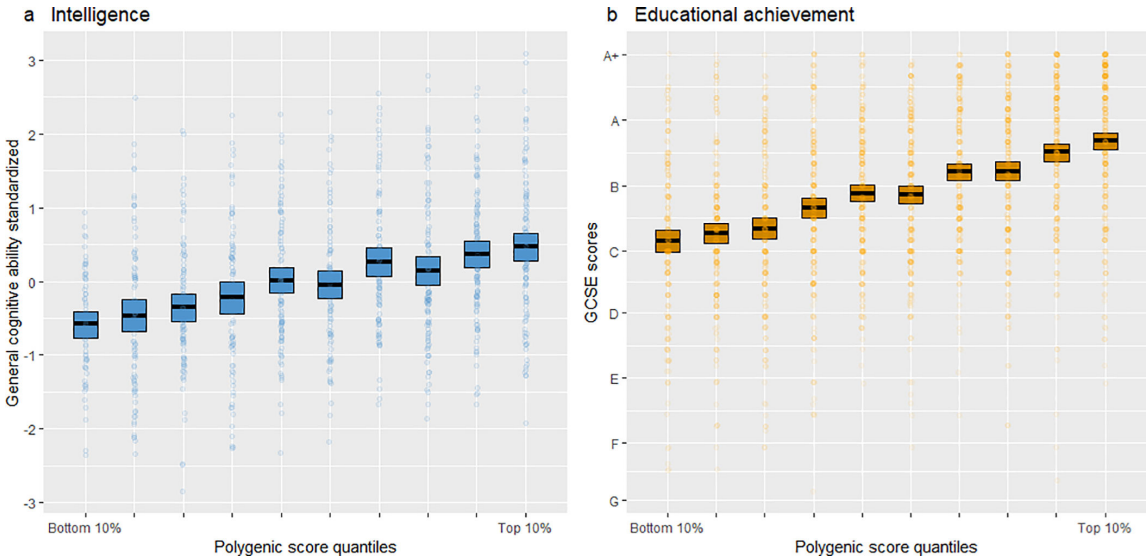


Figure 3.

Mean intelligence scores (panel **a**) and mean educational achievement (panel **b**; GCSE grades) at age 16 by GPS deciles for the best polygenic predictors in the test set. Bars represent bootstrapped 95% confidence intervals. Coloured dots represent individual data points.

Chapter 3 – Multivariable G-E interplay in the prediction of educational achievement

This chapter is an adapted version of a manuscript published in PLOS Genetics:

Allegrini, A. G., Karhunen, V., Coleman, J. R., Selzam, S., Rimfeld, K., von Stumm, S., Pingault, J. B & Plomin, R. (2020). Multivariable GE interplay in the prediction of educational achievement. *PLoS genetics*, *16*(11), e1009153.

Supplementary materials are included in Appendix 2.

Allegrini, A.G.[1*], Karhunen, V. [2], Coleman, J. R. I. [1,3], Selzam S.[1], Rimfeld K.[1], von Stumm, S.[4], Pingault, J.-B. [1,5], Plomin, R[1].

[1]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK

[2] Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, UK

[3]NIHR Maudsley Biomedical Research Centre, King's College London, UK

[4] Department of Education, University of York, UK

[5]Division of Psychology and Language Sciences, University College London, UK

Abstract

Polygenic scores are increasingly powerful predictors of educational achievement. It is unclear, however, how sets of polygenic scores, which partly capture environmental effects, perform jointly with sets of environmental measures, which are themselves heritable, in prediction models of educational achievement.

Here, for the first time, we systematically investigate gene-environment correlation (rGE) and interaction (GxE) in the joint analysis of multiple genome-wide polygenic scores (GPS) and multiple environmental measures as they predict tested educational achievement (EA). We predict EA in a representative sample of 7,026 16-year-olds, with 20 GPS for psychiatric, cognitive and anthropometric traits, and 13 environments (including life events, home environment, and SES) measured earlier in life. Environmental and GPS predictors were modelled, separately and jointly, in penalized regression models with out-of-sample comparisons of prediction accuracy, considering the implications that their interplay had on model performance.

Jointly modelling multiple GPS and environmental factors significantly improved prediction of EA, with cognitive-related GPS adding unique independent information beyond SES, home environment and life events. We found evidence for rGE underlying variation in EA (rGE = .38; 95% CIs = .30, .45). We estimated that 40% (95% CIs = 31%, 49%) of the polygenic scores effects on EA were mediated by environmental effects, and in turn that 18% (95% CIs =12%, 25%) of environmental effects were accounted for by the polygenic model, indicating genetic confounding. Lastly, we did not find evidence that GxE effects significantly contributed to multivariable prediction. Our multivariable polygenic and environmental prediction model suggests widespread rGE and unsystematic GxE contributions to EA in adolescence.

Introduction

Education is compulsory in nearly all countries because it provides children with the skills, such as literacy and numeracy, that are essential for successfully participating in society. How well children perform at school, indicated by their educational achievement (EA; not to be confused with educational attainment, which is a measure of years spent in education), predicts many important life outcomes, especially further education and occupational status (1). Quantitative genetic research based on twin studies showed that EA is 60% heritable throughout the school years (2, 3). These studies also suggested that about 20% of the variance of EA and other learning-related traits can be ascribed to shared environmental factors, for example growing up in the same family and going to the same school. However, the picture became more complicated with the discovery that ostensible measures of the environment associated with educational achievement showed genetic influence – most notably, parents' educational attainment, socio-economic status (SES) and aspects of the home environment (4).

Quantitative genetic theory distinguishes two types of interplay between genetic and environmental effects, genotype-environment correlation (rGE) and genotype-environment interaction (GxE) (5). rGE occurs when an individual's genotype covaries with environmental exposures. There are three types of rGE: passive, active and evocative. Passive rGE results from the inheritance of both genetic propensities and environments linked to parental genotypes. That is, individuals inherit from parents a genetic predisposition to a particular trait, but parental genotypes are also associated with rearing environments that, in turn, increase the likelihood of developing a particular trait. For example, individuals with stronger genetic predispositions to educational attainment tend to grow up in higher socioeconomic status families (6). Evocative rGE happens when individuals' genetic propensities evoke a response from the surrounding environment; for example children's predisposition to higher food intake might elicit restrictive food behaviours from their parents (7). Or, in the context of

education, children's learning difficulties might yield parents to provide more support for learning. Active rGE results from individuals actively selecting environments that are linked to their genetic propensity; for example, individuals with a higher genetic predisposition to educational attainment tend to migrate to economically prosperous regions that offer greater educational opportunities (8).

GxE, on the other hand, refers to genetic moderation of environmental effects. That is, when the effects of environmental exposures on phenotypes depend on individuals' genotypes. Equivalently, environmentally moderated genetic effects occur when genetic effects on a phenotype depend on environmental exposures. Importantly, however, rGE may confound GxE effects (9). For example, if a genetic predisposition for a particular trait is found in a particular environment, it is difficult to know whether this represents rGE between the trait and the environment or true GxE. This picture becomes even more complicated when we consider that environments are themselves heritable (4).

Research on GxE was rejuvenated when it became possible to include measured genetic and environmental factors in statistical models. Hundreds of studies were published purporting to show interactions between candidate genes and environmental measures as they predict behavioural traits. For example, a seminal GxE study in the field (10) showed that carriers of two copies of the short serotonine allele on the 5HTT gene exposed to adversity had an increased the risk for depression compared to their genetic counterpart. However, GxE effects such as these have a poor replication history (11, 12). The main problem with this approach is that it ignores the high polygenicity of complex traits, with a reductionist focus on single 'candidate' variants. This combined with typically small sample sizes, underpowered to detect the very small effects that can be expected for GxE, lead to a replication failure (13).
In complex traits, very few individual variants capture more than a tiny fraction of trait variance (14). Genome-wide polygenic scores (GPS) are the missing piece for investigating

the interplay between genes and environment because they can theoretically capture genetic influences up to the limit of SNP-based heritability, which is usually 25-50% of the total heritability for behavioural traits. GPS are indices of an individual's genetic propensity for a trait and are typically derived as the sum of the total number of trait-associated alleles across the genome, weighted by their respective association effect size estimated through genome-wide association analysis (15). A GPS derived from a genome-wide association study of educational attainment (years of schooling) (16) predicts up to 15% of the variance of EA (17). As more powerful GPS become available, they have begun to be used widely in research on GxE (18-23) and rGE (7, 24-27).

Recently it has been possible to dissect the role of parental genetics on child achievement by splitting the parental genome into transmitted alleles (indexing passive rGE) and non-transmitted alleles (indexing environmentally transmitted parental genetic effects). The latter demonstrated that parental genotypes are associated with the environment they provide for the child (28, 29). In fact, a growing body of evidence is showing the importance of considering gene-environment correlation when assessing polygenic effects on trait variation (30, 31), especially for educationally relevant traits. Paralleling previous findings from the quantitative genetics literature, a key point is that environmental measures are themselves heritable and GPS effects can be mediated by the environment, while environmental effects can be accounted for by genetics (genetic confounding). In this sense, polygenic scores for cognitive traits are not pure measures of genetic predisposition: their predictive power also captures environmental effects. For the same reason, environmental measures are not pure measures of the environment.

Rather than examining rGE and GxE for single polygenic scores and environmental measures, here we look at sets of GPS (32) and environmental measures. A multivariable approach is especially warranted for EA because twin analyses show that the high heritability (60%) of EA reflects many genetically influenced traits, including personality and behaviour problems

in addition to cognitive traits (33, 34). Correspondingly, EA GPS is associated with a wide range of traits, including psychiatric, anthropometric and behavioural traits (35). Similarly, environmental predictors of EA are also intercorrelated (e.g. SES and home environment). However, it is not yet clear how sets of polygenic scores, partly capturing environmental effects, perform jointly with sets of environmental measures, which are themselves heritable, and the effect that their interplay (rGE and GxE) might have on prediction.

Here for the first time we systematically investigate the interplay of GPS and environmental measures in the multivariable prediction of tested educational achievement. We jointly analyse multiple GPS and multiple environmental measures, considering the effect of their interplay in out-of-sample prediction. Specifically, we test the joint prediction of 20 well-powered GPS for psychiatric, cognitive and anthropometric traits and 13 proximal and distal measured environments including life events, home environment and SES (see methods for descriptions of all measures). First, we model polygenic scores (henceforth G model) and environmental measures (henceforth E model), separately and jointly (full model), to predict educational achievement in penalized regression models (36) with out-of-sample tests of prediction accuracy. Penalized methods are especially warranted when dealing with multiple correlated predictors as they can overcome problems of multicollinearity and overfitting. To investigate the relative contributions of the employed predictors to the full model, we carry out post-selection estimation (37) of partial regression coefficients, testing independent effects of single GPS and environmental measures. Secondly, we separate direct from mediated effects of the multivariable G and E models on EA and assess rGE defined in terms of the GPS and environmental measures employed. Finally, we assess GxE using a hierarchical group-lasso technique (38) to systematically discover two-way interactions between all GPS and environmental measures, and test their improvement in prediction of EA.

Results

Joint modelling of GPS and environmental effects

In a first step we tested three models for association with EA: all genetic factors (polygenic scores; G model), all environmental factors (measured environments; E model), and a joint model of all factors (full model; G+E). The G+E model achieved the best hold-out sample prediction compared to the G or E models considered separately. The full model predicted 36% of the variance (95% CI = 30.5, 41.6) in EA (Figure 1 panel B, Table S2), 6% more than the E model (30.1%; 95% CI = 24.3, 35.6; Figure S1) and up to 18% more compared to the G model alone (18.3%; 95% CI = 12.7, 23.6; Figure S2). Nested comparisons of the G+E model vs the G and E models separately indicated that the difference in hold-out set prediction accuracy between models (Figure 1 panel D, Table S2b) was significant for both the G+E model vs E the model (median R^2 diff = 5.9%; 95% CI = 2.8, 9.1) and the G+E model vs the G model (median R^2 diff = 17.7%; 95% CI = 13.2, 22.3). This suggested the presence of genetic effects on EA not mediated via environmental effects, and vice versa of environmental effects not accounted for by the genetic effects. Next, we untangled the specific independent contributions of GPS and measured environments to variation in EA.

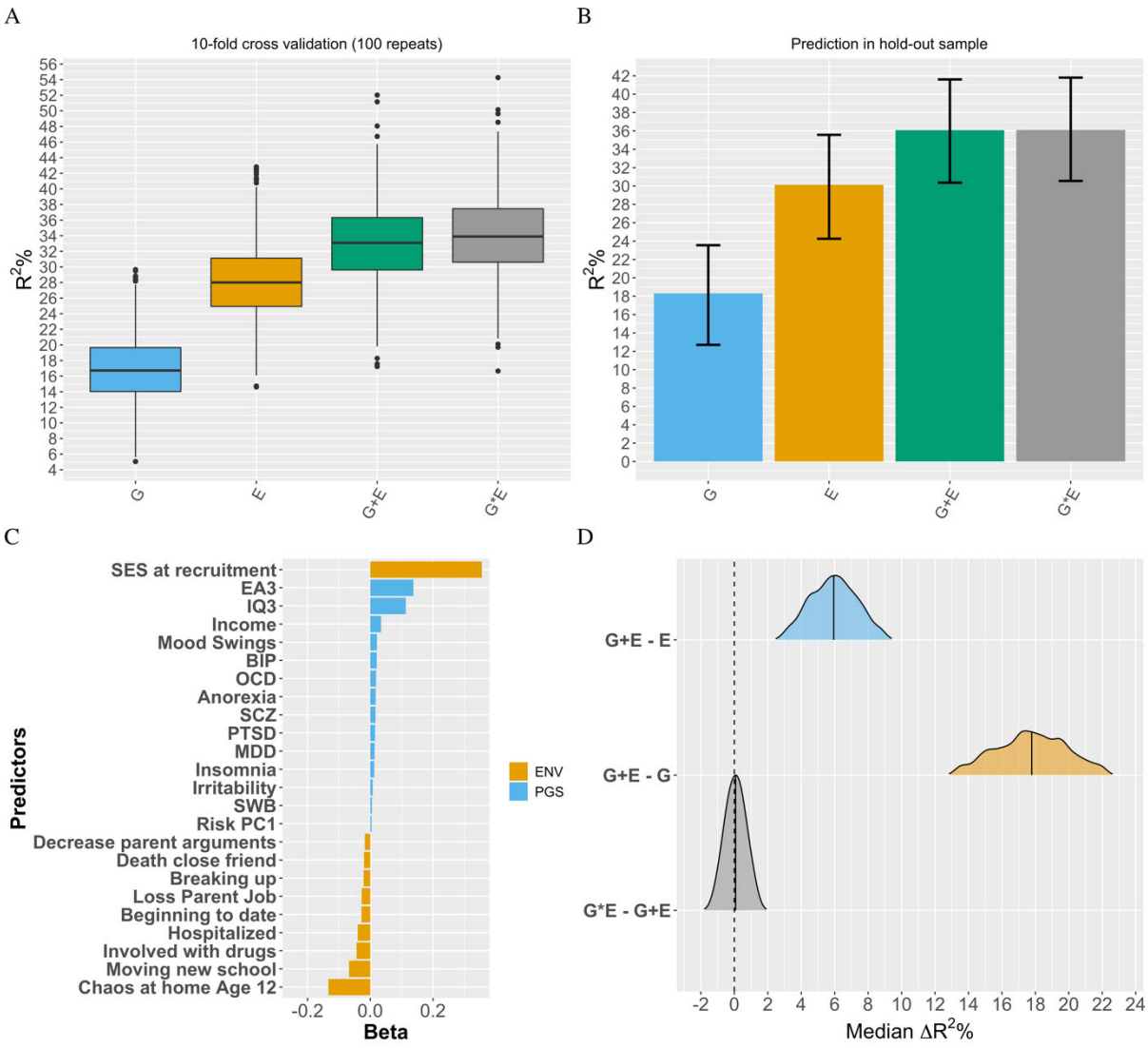


Figure 1. Out of sample prediction of educational achievement. **Panel A** = repeated 10-fold cross validation in training set, for the environmental (E), multi-polygenic score (G), joint (G+E), and interaction (G*E) prediction models. **Panel B** = Hold-out set prediction of EA for best models obtained via repeated cross validation in training set. Error bars are 95% bootstrapped confidence intervals. **Panel C** = G+E model used in hold-out set prediction. Figure shows variables selected via repeated cross-validation in the training set, and relative importance. **Panel D** = Comparison of prediction accuracy for models tested as bootstrapped R^2 difference between nested models in the hold-out set. Distributions represent independent (non-mediated) genetic effects (G+E - E), environmental effects (G+E - G), and G*E effects (G*E – G+E). **Note.** PGS = polygenic scores, ENV = Environmental measures. ASD = Autism Spectrum Disorder, BIP = Bipolar Disorder, BMI = Body Mass Index, EA3 = educational attainment, IQ3 = intelligence, OCD = Obsessive Compulsive Disorder, PTSD = Post-Traumatic Stress Disorder, SCZ = Schizophrenia.

Best-model and coefficient estimation

The best G+E model selected via 10-fold repeated (100 repeats; Figure 1 panel A) cross-validation in the training set included 24 predictors, 14 of which were GPS (blue) while 10 were environments (orange) (Figure 1, panel C). Of these top EA-increasing variables were SES in early life, followed by the GPS for educational attainment (EA3 GPS) and the GPS for intelligence (IQ3 GPS), while the top trait decreasing variable was chaos at home at age 12. In terms of coefficient estimation, partial regression coefficients in post-selection inference analyses (Figure 2 and Table S3) showed that EA3 GPS (β = 0.13; 95% CI = 0.09, 0.17; p = 8.45E-7) and IQ3 GPS (β = 0.12; 95% CI =0.08, 0.15; p = 1.33E-7) remained significant in the model after adjusting for the other predictors. SES was by far the most powerful predictor in the conditional model (β = 0.37; 95% CI =0.35, 0.41; p = 2.30E-60). Other environmental exposures that remained significant were 'chaos at home' at age 12 (β = -0.14; 95% CI =-0.17, -0.12; p = 3.93E-15) and two life events experienced in the past year (all trait decreasing), including 'moving to a new school' (β = -0.07; 95% CI -0.10, -0.04; p = 2E-5) and 'involved with drugs' (β = -0.06; 95% CI -0.09, -0.03; p = 2E-3). SES, EA3 GPS, IQ3 GPS and 'chaos at home' were significant in all three models (i.e. naive, hold-out and conditional).

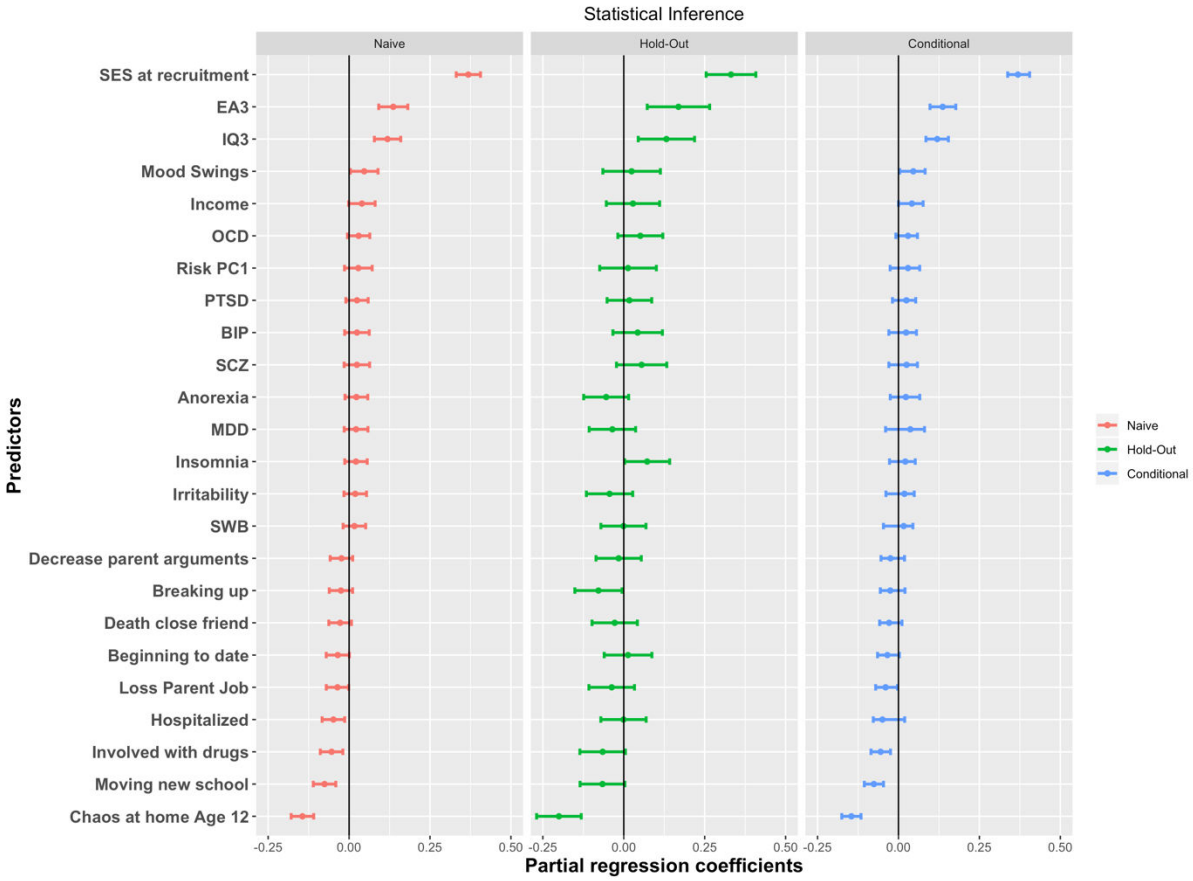


Figure 2. Relative contributions of model selected variables for the G+E model in the prediction of educational achievement. Figure shows partial regression coefficients, and 95% CIs around estimates. Naive = partial regression coefficients from multiple regression of selected variables in Training set; Hold-out = partial regression coefficients of selected variables in the hold-out set; Conditional = partial regression coefficients of training set for selected variables estimated with a conditional probability from a truncated distribution (see method section). **Note.** ASD = Autism Spectrum Disorder, ADHD = Attention-Deficit Hyperactivity Disorder, BIP = Bipolar Disorder, EA3 = educational attainment, IQ3 = intelligence, MDD = Major Depressive Disorder, SWB = Subjective Well-Being, OCD = Obsessive Compulsive Disorder, PTSD = Post-Traumatic Stress Disorder, Risk PC1 = first principal component of risky behaviours, SCZ = Schizophrenia.

rGE and mediated environmental vs GPS effects

Table S2a shows prediction model estimates for all models considered, and Table S2b reports nested comparisons of hold-out set prediction accuracy (R^2) for the full model vs. E and the full model vs. G. We tested the correlation between the EA predicted values from the G model (G_{ea}) and the E model (E_{ea}) in the hold-out-set. This was r = 0.38 (95% CIs = 0.30,

0.45), indicating the extent of overlapping information between the G and E models in hold-out set prediction or, in other words, of rGE (as defined by the variables employed) underlying variation in EA. Then we proceeded to test the extent to which G and E effects on EA were reciprocally mediated (see methods). Table S4 shows results of mediation analyses**.** We found evidence for environmentally mediated genetic effects (indirect path: $\beta = 0.17$; bootstrapped 95% CI 0.13, 0.21) and genetically mediated environmental effects (indirect path: $\beta = 0.10$; bootstrapped 95% CI 0.07, 0.13). The effects of G_{ea} on EA ($\beta = 0.43$; bootstrapped 95% CIs = 0.36, 0.50) were reduced by 40% after introduction of the E_{ea} mediator in the model ($\beta = 0.45$; bootstrapped 95% CIs = 0.38, 0.51); these effects can be interpreted as the direct G model contributions to EA not accounted for by the E model. In other words, 40% of G effects on EA were explained by environmental mediation. Similarly, the direct E_{ea} effects on EA ($\beta = 0.55$; bootstrapped 95% CIs = 0.50, 0.60) were subject to a reduction of 18% ($\beta = 0.45$; bootstrapped 95% CIs = 0.39, 0.51) after introduction of G_{ea} as a mediator in the model, indicating partial genetic mediation of environmental effects (i.e. genetic confounding).

GxE effects and multivariable prediction

We finally tested all possible two-way interactions jointly modelled by means of a hierarchical group lasso procedure using glinternet. Out of the possible 528 two-way interactions between all study variables (i.e. interactions between and within sets of GPS and environmental measures), 32 two-way interactions were detected by the hierarchical group-lasso technique (glinternet, Table S5), 15 of which were GxE interactions. Figure S4 depicts an interaction network from the trained glinternet model (10-fold cross validation). Hold-out set prediction accuracy was only slightly improved ($R^2 = 36.4\%$; 95% CI = 29, 41) over the joint G and E model ($R^2 = 36\%$). We then introduced the 15 GxE interactions found in the full elastic net model (Figure S3) to test whether they improved the prediction of EA over the full

model that had only considered additive effects of GPS and environmental measures. There was no improvement in hold-out set prediction accuracy (R^2 = 36.1%; 95% CI = 30.5, 41.8), and the difference in prediction with the G+E model was not significant (median R^2 diff = 0.1%; 95% CI = -1.2, 1.3). Table S2 shows fit statistics for the glinternet and elastic net models. Table S5 reports GxE interactions detected by the hierarchical lasso model.

Discussion

We tested the joint prediction accuracy of sets of multiple environmental measures and polygenic scores in prediction models of educational achievement and considered the effect of their interplay on model performance. Three main findings emerged from our analyses. First, the joint modelling of multiple GPS and related environmental exposures improved the prediction of EA, consistent with theory (39). Second, paralleling previous quantitative genetic findings, we found consistent evidence of rGE effects underlying variation in EA (rGE = 0.38; 95% CIs = 0.30, 0.48), with a substantial proportion of polygenic score effects mediated by the environmental effects (40%), and evidence for genetic confounding (18%). Lastly, we did not find evidence that GxE effects jointly contributed to the prediction of EA.

Our multivariable GPS model alone predicted 18.3% of the variance in EA. Integration of multiple polygenic scores in the same model can be expected to increase as sample size in genome-wide association studies (GWAS) increases (40). Here we constructed GPS in lassosum (41) based on previous observations that lassosum tends to perform better than more conventional approaches (17, 41) for educationally relevant traits. However, other methods for GPS construction can be expected to yield similar results when considering multivariable GPS penalized approaches, with performance of the relative approaches likely to converge as accuracy of GWAS estimates increases.

Of interest were the relative contributions of the single GPS to the best model selected via repeated cross-validation in the training set. In post-selection inference analyses, IQ3 and EA3 were the only GPS independently associated with variation in EA after adjusting for measured environments and polygenic scores. This indicated that both these GPS contributed unique predictive information beyond other related, proxy environmental predictors (e.g. SES, parental educational attainment), and polygenic scores (e.g. household income). Similarly, we found that several environments were independently predictive of EA. The best predictor was early life SES, a composite of parental educational attainment, employment status and

maternal age at first birth. Life events and chaos at home were also significant contributors to the model, with negative independent effects on EA. Polygenic scores, however, improved the prediction of EA on top of the environment with a 20% increase in accuracy (from 30% to 36%). It is noteworthy that EA3 and IQ3 GPS were both significant in post-selection inference models after adjusting for SES, home environment and proximal environmental effects, all of which also tag genetic variance partly overlapping with that captured by the GPS. This suggested that cognitive-relevant GPS independently captured variation beyond environmental variables and variance due to rGE in our model. While this was important to understand the model composition, it should be highlighted that these estimates are dependent on variables included in analyses, and can be expected to change as other variables are considered in the model (see below).

A central finding of the current study emerged when we separated direct and indirect effects of the GPS and environmental models by statistically testing for rGE. We found significant G mediation of the prediction of EA by the E model. This is in line with several quantitative genetics findings (42-44). However, since it would be unreasonable to assume a causal effect of E on G (i.e. E does not change DNA sequence), in the sense employed here G acts as a 'confounder' – in causal modelling parlance, 'third variable confounding' – of E effects on EA (E ← G → EA). That is, because our G model is associated with both the E model and EA, it partly induces an association between the E model and EA in addition to the independent effects of E on EA. This rGE effect explained 18% of the E effects on EA. By extension, this type of effect could arise because E is predicted by parental genotypes, and as parents share their genotypes with their offspring, this creates a link between E and (child) G. Different types of genetic confounding have been described in detail elsewhere (45). It should also be noted that here G does not represent directly genotypes, but a combination of estimated effect sizes from GWAS summary statistics. Since E could affect estimation of G (betas), in this general sense E could affect G.

We also found evidence of environmental mediation of the G model effects on EA. The E model explained 40% of the GPS model effects on EA. This result is also in line with previous research in quantitative genetics (27-29, 46). A growing body of evidence points to the rGE conclusion that genetic effects on cognitive trait variation are partly environmentally mediated (25), which is likely to be due to passive rGE. Passive rGE emerges because parents create a family environment that corresponds to their genotypes and, by extension, also correlates with the genotypes of their offspring. As previously described, alternative mechanisms include evocative and active rGE effects, which, as noted elsewhere (26), represent not mutually exclusive possibilities. Another related, but different, type of effect that could explain this finding is genetic nurture, whereby the parent genome exert an effect on the child phenotype via the environment, over and above shared parent-offspring genetics. However, in order to disentangle these effects, different study designs are needed, for example, looking within families at the effects of maternal and paternal non-transmitted genotypes on child outcomes. Disentangling the different underlying mechanisms to the predicted variance in this regard is an issue for future studies, but out of the scope of the present investigation. Here for the first time we show that reciprocal indirect effects between multivariable E and G prediction models explain a substantial proportion of variation of their direct effects on EA. These results provide converging evidence with recent research looking at rGE underlying parenting and children educational attainment (27, 47). Both genetic confounding and environmental mediation are important factors to take into account in the prediction of EA.

Lastly, we applied a hierarchical group-lasso model (glinternet) to automatically detect two-way interactions. This model helped us to identify GxE effects that show strong hierarchy, which would have otherwise been difficult to detect due to the great multiple-testing burden relative to the sample size of the present study. Furthermore, since glinternet performs shrinkage and grouping before testing for interaction effects, this enabled discovery of

interactions that would have been confounded by strong main effects of correlated predictors. In other words, because the coefficients of main effects have been regularized (that is shrunk, see Methods), their fit is reduced, which facilitates the discovery of interaction effects (38). However, neither the glinternet model including all discovered pairwise interactions, nor the elastic net model including two-way GxE effects, significantly improved hold-out set prediction over the G+E model. One possible explanation for this finding is that GxE effects are typically very small, and that the trade-off between true effect and variance introduced in the model, signal to noise ratio, was too small. It might be that even if two-way GxE effects were relevant the noise incurred in fitting their coefficients may outweigh the improvement in accuracy that they bring to the model. In this regard we note that in repeated cross-validation in the training set the model performance of both the elastic net based GxE model, and the glinternet model was substantially increased compared to the G+E model. Application of this method in larger datasets, or using different phenotypes with different genetic architectures, might be fruitful for hypothesis-free GxE discovery as well as for prediction.

This study must be considered in light of a few limitations. First, our results are subject to the constraint that we performed an apriori selection on variables to be employed in our analyses. For example, we modelled exposures that are typically defined as environmental; however, many other variables can be argued to capture environmental influences. In this regard estimates for non-mediated genetic effects for the model presently tested are likely upper-bounds, in the sense that if we were to include more E variables predictive of the outcome EA, the polygenic score contributions independent of E would either stay the same or decrease. Likewise, we included a broad range of polygenic scores that are currently available as the most predictive for cognitive, psychiatric and anthropometric traits. However, polygenic scores predictive power is in part a function of GWAS sample size (40), therefore as more powerful GWAS become available these prediction estimates are expected to

increase. This in turn suggests that the contributions of the G model are likely to be on the lower bound compared to future polygenic score work in this area.

Finally, we focused here on EA but predictive models of other complex traits are likely to yield different results, because EA shows comparatively great shared environmental influences (30). This suggests that rGE is likely to be stronger for EA than for other behavioural traits, such as personality traits and social-emotional competencies. Regarding our analytical approach, we focused on GxE interactions that obeyed strong hierarchy as identified by the group lasso technique. Future studies could relax this assumption and include interactions where one of the main effect sizes is not significant, as well as higher order interactions. Finally, although it is a strength of our study that we used measured environmental exposures, we note that methods for inferring GxE without measured environmental data are emerging that have reported GxE for some complex traits (48). The extent to which these effects are systematic, stable, and generalizable to EA remains to be determined.

As large multidimensional biobank datasets become increasingly available, the integration of multi-omics data with multiple environmental measures will become more common in prediction modelling. Here, we provide an indication of the effects of integrating multiple GPS and environmental measures in prediction models of EA and the effect that their interplay has on prediction accuracy in a population cohort of adolescents. In conclusion, we found consistent evidence for rGE in prediction models of EA that systematically tested the interplay between polygenic scores and measured environments within a hypothesis-free multivariable prediction framework. When integrating multiple GPS and environmental measures, their interplay must be taken into account. Separate effects of environmental and polygenic scores cannot just be assumed to add up because pervasive rGE affects prediction.

Material and Methods

Sample

We test our models using data from 16 year olds from the UK Twin Early Development Study (TEDS; 49), a large longitudinal study involving 16,810 pairs of twins born in England and Wales between 1994-1996, with DNA data available for 10,346 individuals (including 3,320 dizygotic twin pairs and 7,026 unrelated individuals). Ethical approval for TEDS has been provided by the King's College London Ethics Committee (reference: PNM/09/10–104). Parental consent was obtained before data collection. Genotypes for the 10,346 individuals were processed with stringent quality control procedures followed by SNP imputation using the Haplotype Reference Consortium (release 1.1) reference panels. Current analyses were limited to the genotyped and imputed sample of 7,026 unrelated individuals. Following imputation, we excluded variants with minor allele frequency <0.5%, Hardy-Weinberg equilibrium p-values of $<1 \times 10{-}5$. To ease computational demands, we selected variants with an info score of 1, resulting in 515,000 SNPs used for analysis (see the supplementary information for a full description of quality control and imputation procedures).

Measures

Dependent measure: Educational Achievement

Educational achievement was measured as the self-reported mean grade of three core subjects (English, math and science) scored by the individuals at age 16 in their standardized UK General Certificate of Secondary Education (GCSE) exams.

EA was operationalized as the mean grade of the three compulsory subjects, with results coded from 4 (G, or lowest grade) to 11 (A+, or highest grade). These self-report measures are highly replicable and show high genetic and phenotypic correlations with teacher scores(50). The variable distribution was slightly negatively skewed (similar to the national

average) and subject to a rank based inverse normal transformation to approximate a normal distribution.

Environmental measures

Socio economic status: SES at recruitment (mean age = 18 months) was calculated as a composite of mother and father qualification levels ranging from 1 = 'no qualifications' to 8 = 'postgraduate qualification', mother and father employment status (51), and mother's age at birth of first child.

Chaos at home: as a measure of home environment a shortened version of the Confusion, Hubbub and Order Scale (52) was used to measure children's perception of chaos in the family environment at age 12. Children rated the extent to which they agree (range: 'not true', 'quite true' or 'very true') to six items: 'I have a regular bedtime routine' (reversed coded), 'You can't hear yourself think in our home', 'It's a real zoo in our home', 'We are usually able to stay on top of things' (reversed coded), 'There is usually a television turned on somewhere in our home' and 'The atmosphere in our house is calm' (reverse coded). The Chaos score was computed as the mean of the rated items.

Life events: Self-reported life events experienced in the past year were measured (at age 16) using a shortened version of the Coddington life events (53). Individuals had to report on 20 items that might have happened in the past year, by responding yes (coded as 1) if the event had happened or no (coded as 0) if it didn't happen. Items included stochastic, proximal events such as "death of a close friend or relative", "being hospitalized", as well as family-wide events e.g. "loss of a parent job", "decrease in parental income". When considering prediction of educational achievement, educationally relevant items were removed from the models (i.e. "failing exam" and "outstanding achievement"). Items being endorsed by fewer

than 100 people were discarded from analyses. A total of 11 life events were retained in analyses. All items were considered separately in prediction models (i.e. they were not aggregated in a scale). Table S1 reports descriptive statistics for variables employed in this study, separately by training and hold-out sets.

Genome-wide polygenic scores (GPS)

GPS for 20 cognitive, anthropometric and psychopathological traits were constructed using Lassosum (41). Lassosum is a penalized regression approach applied to GWAS summary statistics. In lassosum we try to minimize the following loss function:

$$y^T y + (1 - s)\, \beta^T X_r^T X_r \beta - 2\beta^T r + s\beta^T \beta + 2\lambda ||\beta||^1 \qquad (1)$$

Where y is a vector of the phenotype, X is the matrix of genotypes, such that $X_r^T X_r$ is a matrix of correlations between SNPs, the LD matrix. r denotes the correlation between SNPs and the phenotype, $r = X^T y$. The subscript $_r$ in $X_r^T X_r$ indicates that SNPs employed to obtain the LD matrix (based on a reference panel, see below) will generally not correspond to SNPs used to infer the correlation with the phenotype.

In the equation λ controls the L1 penalty (L1 norm, (54)). The notation $||\beta||^1$ describes the L1 norm of a coefficient vector β, defined as

$$||\beta||^1 = \sum|\beta|. \qquad (2)$$

While *s* is another tuning parameter controlling the L2 penalty ($|\beta|^2$, the sum of the squared betas). Here *s* has the additional constraint of being between 0 and 1. When $\lambda = 0$ and $s = 1$ the problem becomes unconstrained.

Tuning parameters, λ and s, are chosen in the validation step (this is akin to optimization that can be performed in p-value thresholding methods). We used our training set to perform parameter tuning optimizing (with respect to R^2) polygenic scores against EA. LD was accounted for via a reference panel, here the same as the training set sample, and estimation of LD blocks was performed using LD regions defined in (55).

We created cognitive and educationally relevant polygenic scores for educational attainment (16), intelligence (56), and income (57). We also created polygenic scores for mental health-related traits: autism spectrum disorder (58), major depressive disorder (MDD; 59), bipolar disorder (BIP; 60), schizophrenia (SCZ; 61), attention deficit hyperactivity disorder (ADHD; 62), obsessive compulsive disorder (OCD; 63), anorexia nervosa (AN; 64), post-traumatic stress disorder (PTSD; 65), broad depression (66), mood swings (67), subjective well-being (68), neuroticism (69), irritability (67), insomnia (70), and risk taking (71). Finally, we created polygenic scores for height and BMI (72). Table S6 reports information on these summary statistics, while Table S7 reports parameter tunings for the lassosum GPS.

Analyses

All variables were first regressed on age, sex, 10 genetic principal components and genotyping chip. The obtained standardized residuals were used in all subsequent analyses.

Penalised regression

We fit elastic net regularization (73) models for EA. Elastic Net minimizes the residual sum of squares (RSS) subject to the L1 penalty, consisting of the sum of the absolute coefficients, which introduce sparsity allowing for parameters selection, and the L2 penalty, consisting of the sum of the squared coefficients, which allows for parameters shrinkage (73).

Elastic net tries to minimise the following loss function:

$||y - X\beta||^2 + \lambda(\alpha*|\beta|^1 + (1-\alpha)*|\beta|^2)$ (3)

where

$||y - X\beta||^2$ is the residual sum of squares

$|\beta|^2$ is the sum of the squared betas (the L2 penalty)

$|\beta|^1$ is the sum of the absolute betas (the L1 penalty)

Here X is an nxp ('n' observations and 'p' predictors) matrix of polygenic scores, environmental predictors or a combination of both (see below). α determines the mixing of penalties, where the first parameter introduces sparsity while the second shrinks correlated predictors towards each other. λ is a tuning parameter that control the effect of the penalty terms over the regression coefficients. When α = 1 the solution is equivalent to a LASSO regression, while when α = 0 the solutions is equivalent to a Ridge regression. For every α multiple λ exists, and the optimal combination of tuning parameters is determined by cross-validation, here a 10-fold cross-validation repeated 100 times. For every model tested we split the sample into an independent training set (80%) and a hold-out set (20%). In the training set we perform 10-fold repeated cross-validation to select the model that minimises the Root Mean Square Error (RMSE) – that is the tuning parameter for which the cross-validation error is the smallest. The model performance is then assessed by the variance explained (R^2) in the hold-out test set. The hold-out set R^2 was calculated as 1- $\frac{SSE}{SST}$ (SSE = sum of squared errors, SST = sum of square total).

Bootstrapping: For every model tested we sampled with replacement from the data (1000 times) to calculate bootstrapped confidence intervals for the hold-out set prediction accuracy (R^2). Rows of data for resampling included the phenotype under study and the predictors according to the model tested: either polygenic scores, environmental predictors or a combination of both. For each bootstrap sample drawn we calculated the hold-out set R^2, and

we took the difference in R^2 between nested models. This procedure yielded a distribution of R^2 for each model tested and a distribution of R^2 differences (ΔR^2) for each pairwise comparison. We then calculated 95% confidence levels as the 2.5th and 97.5th percentiles of these distributions. For nested comparisons, if the interval didn't contain 0 we concluded that the pairwise model ΔR^2 was significantly different from 0 with a α level of .05.

Post selection inference: For every model tested we conducted statistical inference of models coefficients after selection of most informative predictors performed by Elastic Net, that is effect sizes, p-values and confidence levels around the prediction estimates.
Post-selection inference (37) refers to the practice of attempting to accurately estimate prediction coefficients after a model selection has been performed. If we fit the optimal model's selected predictors in a multiple regression model in the training set (that is where the selection has been performed) our confidence in the estimates will tend to be over-optimistic. On the other hand, estimation of these parameters in a hold-out set would not be subject to this problem. The hold-out set, however, will typically be smaller than the training set, leading to wider confidence intervals. In addition, the results will be dependent on the random split (80-20) performed. A third way is to calculate P-values conditional to the selection that has been made in the training set. Briefly, after selection is performed, accurate estimation of a given partial regression coefficient can be approximated by a truncated normal distribution:

$$\hat{\beta} \sim TN^{a,b}(\beta, \boldsymbol{\tau}^2) \qquad (4)$$

With mean β, variance $\boldsymbol{\tau}^2$ and boundaries of the truncated normal distribution (TN) 'a' and 'b' given by the data and the selection procedure, in this case the predictors, the active set (the variables with non 0 coefficients selected by our model) and λ (37). We refer elsewhere to a thorough discussion of the topic (74), with a focus on lasso like approaches. Here we compare

results from the three procedures: the 'naive' estimation of partial regression coefficients in the training set, estimation of coefficients in the hold-out set, and the conditional estimation of p-values performed using the R package 'SelecitveInference' (75).

rGE

We quantified rGE in two ways. First, the hold-out set predicted EA values from the GPS (henceforth G_{ea}) and environmental (henceforth E_{ea}) models can be tested for correlation. In this sense the covariance between these variables would be an indication of overlapping information between E and G underlying EA, i.e.

$$r_{G,E} = cor(G_{ea}, E_{ea}). \quad (5)$$

Second, another way to quantify rGE is by modelling E and G effects in a mediation model (Figure 3), considering the indirect effects of G on EA via E, and vice versa the indirect effects of E on EA via G. We used the predicted EA values from the GPS and environmental models (i.e. G_{ea} and E_{ea}) to test mediation models in the hold-out set. We fit a structural equation model (SEM) in 'lavaan' (76) to test whether and to what extent E and G effects on EA were reciprocally mediated. Panel A (Figure 3) is a schematic representation of a mediation model, where βC is the effect of a predictor X on an outcome Y, βa the effect of X on the mediator (M), and βb the effect of M on Y after adjusting for X. βC' corresponds to the effects of the predictor on the outcome when controlling for the mediator (i.e. when the full equation is estimated). If the effects are reduced (partial mediation) or are not different from 0 (full mediation) then there is evidence for mediation. We quantify the proportion of the mediated effects as (βC- βC') / βC and test for significance of the indirect path using bootstrapping (with 1000 repetitions).

Figure 3 represent direct and indirect effects of the G model effects on EA mediated by E (panel B), and of the E model effects on EA mediated by G (panel C). While panel B represents a causal model where we estimate the environmentally mediated G effects on EA, panel C is a statistical abstraction since it would be unreasonable to assume a causal relationship of E on G. Here we model G as mediator to estimate the third variable confounding effects underlying the relationship between E and EA, as mediating and confounding effects have been shown to be equivalent in a linear context (77).

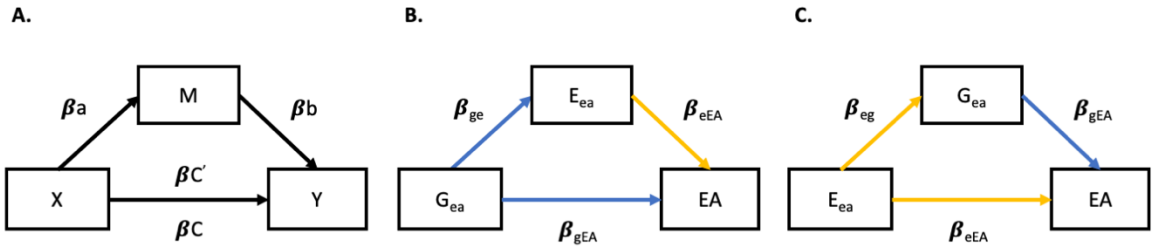


Figure 3. *Panel A* = schematic representation of mediation analysis; βC = effect of a predictor X on an outcome Y; βa = effect of X on a mediator (M); βb = effect of M on Y after adjusting for X; $\beta C'$ = effect of X on Y after adjusting for M. *Panel B* = Directed acyclic graph (DAG) showing E_{ea} mediated effects of G_{ea} on EA in the hold out-set; β_{ge} = causal path between G_{ea} and E_{ea} equivalent to $r_{G,E}$; β_{eEA} = direct independent E_{ea} effects on EA; β_{gEA} = total G_{ea} effects on EA. *Panel C* = DAG showing G_{ea} mediated effects on EA (genetic confounding, see methods and discussion); β_{eg} = causal path between E_{ea} and G_{ea} equivalent to $r_{G,E}$; β_{gEA} = direct independent G_{ea} effects on EA; β_{eEA} = total E_{ea} effects on EA. **Note.** Blue paths represent G model effects, yellow paths represent E model effects.

GxE

After fitting the joint GPS and environmental models, we apply a hierarchical lasso procedure to automatically search the feature space for interactions, and retrain our models introducing GxE interactions. With 33 predictors there is a total of 33(33-1)/2 = 528 possible 2-ways interactions. Testing all models separately would imply a multiple testing burden (e.g.

bonferroni correction .05/820 = 9E-5), in addition to the expected low signal to noise ratio for GxE effects. Here we employ a hierarchical group lasso approach to automatically search for two-way interactions, implemented in the R package 'glinternet' (38) (group-lasso interaction network). Glinternet leverages group lasso, an extension of LASSO, to perform variable selection on groups of variables, dropping or retaining them in the model at the same time, to select interactions. As noted above, the L1 regularization produces sparsity. Glinternet uses a group lasso for the variables and variable interactions, which introduces a strong hierarchy: an interaction between two variables can only be picked by the model if both variables are also selected as main effects. That is, interactions between two predictors are not considered unless both predictors have non-zero coefficients in the model. Once two-way interactions obeying strong hierarchy were identified, we selected GxE interactions (i.e. GPS that interact with environmental variables) and reintroduced them in our best elastic net models to test whether the hold-out set prediction accuracy improved beyond the full (E+G) prediction model.

Acknowledgements

We gratefully acknowledge the ongoing contribution of the participants in the Twins Early Development Study (TEDS) and their families. TEDS is supported by a programme grant to RP from the UK Medical Research Council (MR/M021475/1 and previously G0901245), with additional support from the US National Institutes of Health (AG046938). The research leading to these results has also received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007- 2013)/grant agreement n° 602768 and ERC grant agreement n° 295366. RP is supported by a Medical Research Council Professorship award (G19/2). This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 721567. JRIC is supported in part by the UK National Institute for Health Research (NIHR) as part of the Maudsley Biomedical Research Centre (BRC). This study represents independent research partly funded by the NIHR BRC at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. SvS is supported by a Jacobs Fellowship (2017-2019).

High performance computing facilities were funded with capital equipment grants from the GSTT Charity (TR130505) and Maudsley Charity (980).

References

1. Asbury K, Plomin R. G is for genes: what genetics can teach us about how we teach our children. Wiley, Oxford; 2013.
2. Rimfeld K, Malanchini M, Krapohl E, Hannigan LJ, Dale PS, Plomin R. The stability of educational achievement across school years is largely explained by genetic factors. NPJ science of learning. 2018;3(1):16.
3. Polderman TJC, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nature genetics. 2015;47:702.
4. Plomin R, Bergeman CS. The nature of nurture: Genetic influence on "environmental" measures. Behavioural and Brain Sciences. 1991;14(3):373-86.
5. Plomin R, DeFries JC, Loehlin JC. Genotype-environment interaction and correlation in the analysis of human behaviour. Psychological bulletin. 1977;84(2):309.
6. Belsky DW, Domingue BW, Wedow R, Arseneault L, Boardman JD, Caspi A, et al. Genetic analysis of social-class mobility in five longitudinal studies. Proceedings of the National Academy of Sciences. 2018;115(31):E7275-E84.
7. Selzam S, McAdams TA, Coleman JR, Carnell S, O'Reilly PF, Plomin R, et al. Evidence for gene-environment correlation in child feeding: Links between common genetic variation for BMI in children and parental feeding practices. PLoS genetics. 2018;14(11):e1007757.
8. Abdellaoui A, Hugh-Jones D, Yengo L, Kemper KE, Nivard MG, Veul L, et al. Genetic correlates of social stratification in Great Britain. Nature human behaviour. 2019:1-21.
9. Lau JY, Eley TC. Disentangling gene-environment correlations and interactions on adolescent depressive symptoms. Journal of Child Psychology and Psychiatry. 2008;49(2):142-50.
10. Caspi A, Sugden K, Moffitt TE, Taylor A, Craig IW, Harrington H, et al. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. Science. 2003;301(5631):386-9.
11. Border R, Johnson EC, Evans LM, Smolen A, Berley N, Sullivan PF, et al. No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples. American Journal of Psychiatry. 2019;176(5):376-87.

12. Dick DM, Agrawal A, Keller MC, Adkins A, Aliev F, Monroe S, et al. Candidate gene–environment interaction research: Reflections and recommendations. Perspectives on Psychological Science. 2015;10(1):37-59.
13. Duncan LE, Keller MC. A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. American Journal of Psychiatry. 2011;168(10):1041-9.
14. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 years of GWAS discovery: biology, function, and translation. The American Journal of Human Genetics. 2017;101(1):5-22.
15. Wray NR, Lee SH, Mehta D, Vinkhuyzen AA, Dudbridge F, Middeldorp CM. Research review: polygenic methods and their application to psychiatric traits. Journal of Child Psychology and Psychiatry. 2014;55(10):1068-87.
16. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nature genetics. 2018.
17. Allegrini A, Selzam S, Rimfeld K, von Stumm S, Pingault J, Plomin R. Genomic prediction of cognitive traits in childhood and adolescence. Molecular psychiatry. 2019;24(6):819.
18. Mullins N, Power R, Fisher H, Hanscombe K, Euesden J, Iniesta R, et al. Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. Psychological medicine. 2016;46(4):759-70.
19. Barr PB, Kuo SIC, Aliev F, Latvala A, Viken R, Rose RJ, et al. Polygenic Risk for Alcohol Misuse is Moderated by Romantic Partnerships. Addiction (Abingdon, England). 2019.
20. Barcellos SH, Carvalho LS, Turley P. Education can reduce health differences related to genetic risk of obesity. Proceedings of the National Academy of Sciences. 2018;115(42):E9765-E72.
21. Pasman JA, Verweij KJ, Vink JM. Systematic Review of Polygenic Gene–Environment Interaction in Tobacco, Alcohol, and Cannabis Use. Behaviour genetics. 2019:1-17.
22. Coleman JR, Krapohl E, Eley TC, Breen G. Individual and shared effects of social environment and polygenic risk scores on adolescent body mass index. Scientific reports. 2018;8(1):6344.

23. Peyrot WJ, Van der Auwera S, Milaneschi Y, Dolan CV, Madden PA, Sullivan PF, et al. Does childhood trauma moderate polygenic risk for depression? A meta-analysis of 5765 subjects from the psychiatric genomics consortium. Biological psychiatry. 2018;84(2):138-47.
24. Dobewall H, Savelieva K, Seppälä I, Knafo-Noam A, Hakulinen C, Elovainio M, et al. Gene–environment correlations in parental emotional warmth and intolerance: genome-wide analysis over two generations of the Young Finns Study. Journal of Child Psychology and Psychiatry. 2019;60(3):277-85.
25. Belsky DW, Domingue BW, Wedow R, Arseneault L, Boardman JD, Caspi A, et al. Genetic analysis of social-class mobility in five longitudinal studies. Proceedings of the National Academy of Sciences. 2018;115(31):E7275-E84.
26. Krapohl E, Hannigan L, Pingault J-B, Patel H, Kadeva N, Curtis C, et al. Widespread covariation of early environmental exposures and trait-associated polygenic variation. Proceedings of the National Academy of Sciences. 2017;114(44):11727-32.
27. Wertz J, Belsky J, Moffitt TE, Belsky DW, Harrington H, Avinun R, et al. Genetics of nurture: A test of the hypothesis that parents' genetics predict their observed caregiving. Developmental psychology. 2019.
28. Bates TC, Maher BS, Medland SE, McAloney K, Wright MJ, Hansell NK, et al. The nature of nurture: Using a virtual-parent design to test parenting effects on children's educational attainment in genotyped families. Twin Research and Human Genetics. 2018;21(2):73-83.
29. Kong A, Thorleifsson G, Frigge ML, Vilhjalmsson BJ, Young AI, Thorgeirsson TE, et al. The nature of nurture: Effects of parental genotypes. Science (New York, NY). 2018;359(6374):424-8.
30. Selzam S, Ritchie SJ, Pingault J-B, Reynolds CA, O'Reilly PF, Plomin R. Comparing within-and between-family polygenic score prediction. The American Journal of Human Genetics. 2019;105(2):351-63.
31. Cheesman R, Hunjan A, Coleman JR, Ahmadzadeh Y, Plomin R, McAdams TA, et al. Comparison of adopted and non-adopted individuals reveals gene-environment interplay for education in the UK Biobank. bioRxiv. 2019:707695.
32. Krapohl E, Patel H, Newhouse S, Curtis CJ, von Stumm S, Dale PS, et al. Multi-polygenic score approach to trait prediction. Molecular psychiatry. 2018;23(5):1368-74.
33. Krapohl E, Rimfeld K, Shakeshaft NG, Trzaskowski M, McMillan A, Pingault J-B, et al. The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. Proceedings of the National Academy of Sciences. 2014;111(42):15273-8.

34. Rimfeld K, Kovas Y, Dale PS, Plomin R. True grit and genetics: Predicting academic achievement from personality. Journal of personality and social psychology. 2016;111(5):780.
35. Krapohl E, Euesden J, Zabaneh D, Pingault JB, Rimfeld K, von Stumm S, et al. Phenome-wide analysis of genome-wide polygenic scores. Molecular psychiatry. 2015;21:1188.
36. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. Journal of statistical software. 2010;33(1):1.
37. Taylor J, Tibshirani RJ. Statistical learning and selective inference. Proceedings of the National Academy of Sciences. 2015;112(25):7629-34.
38. Lim M, Hastie T. Learning interactions via hierarchical group-lasso regularization. Journal of Computational and Graphical Statistics. 2015;24(3):627-54.
39. Dudbridge F, Pashayan N, Yang J. Predictive accuracy of combined genetic and environmental risk scores. Genetic epidemiology. 2018;42(1):4-19.
40. Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS genetics. 2013;9(3):e1003348.
41. Mak TSH, Porsch RM, Choi SW, Zhou X, Sham PC. Polygenic scores via penalized regression on summary statistics. Genetic epidemiology. 2017;41(6):469-80.
42. Plomin R, DeFries JC, Knopik VS, Neiderhiser JM. Top 10 replicated findings from behavioural genetics. Perspectives on psychological science. 2016;11(1):3-23.
43. Plomin R. Genotype-environment correlation in the era of DNA. Behaviour genetics. 2014;44(6):629-38.
44. Krapohl E, Plomin R. Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. Molecular psychiatry. 2015;21:437.
45. Pingault J-B, Rijsdijk F, Schoeler T, Choi SW, Selzam S, Krapohl E, et al. Estimating the sensitivity of associations between risk factors and outcomes to shared genetic effects. bioRxiv. 2019:592352.
46. Plomin R, Bergeman CS. The nature of nurture: Genetic influence on "environmental" measures. Behavioural and Brain Sciences. 2011;14(3):373-86.
47. Wertz J, Moffitt TE, Agnew-Blais J, Arseneault L, Belsky DW, Corcoran DL, et al. Using DNA from mothers and children to study parental investment in children's educational attainment. bioRxiv. 2018:489781.
48. Sulc J, Mounier N, Felix G, Winkler T, Wood AR, Frayling TM, et al. Maximum likelihood method quantifies the overall contribution of gene-environment interaction to complex traits: an application to obesity traits. bioRxiv. 2019:632380.

49. Rimfeld K, Malanchini M, Spargo T, Spickernell G, Selzam S, McMillan A, et al. Twins Early Development Study: A Genetically Sensitive Investigation into Behavioural and Cognitive Development from Infancy to Emerging Adulthood. Twin Research and Human Genetics. 2019:1-6.
50. Rimfeld K, Malanchini M, Hannigan LJ, Dale PS, Allen R, Hart SA, et al. Teacher assessments during compulsory education are as reliable, stable and heritable as standardized test scores. Journal of Child Psychology and Psychiatry. 2019.
51. Surveys OoPaC. Standard occupational classification. Her Majesty's Stationery Office 1991;volume 3.
52. Matheny Jr AP, Wachs TD, Ludwig JL, Phillips K. Bringing order out of chaos: Psychometric characteristics of the confusion, hubbub, and order scale. Journal of Applied Developmental Psychology. 1995;16(3):429-44.
53. Coddington RD. The significance of life events as etiologic factors in the diseases of children: II. A study of a normal population. Journal of psychosomatic research. 1972.
54. Tibshirani R. Regression shrinkage and selection via the Lasso. J R Stat Soc Ser B Methodol. 1996(58):267–88.
55. Berisa T, Pickrell JK. Approximately independent linkage disequilibrium blocks in human populations. Bioinformatics. 2016;32(2):283-5.
56. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, de Leeuw CA, et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nature genetics. 2018;50(7):912-9.
57. Hill WD, Davies NM, Ritchie SJ, Skene NG, Bryois J, Bell S, et al. Genome-wide analysis identifies molecular systems and 149 genetic loci associated with income. Nature communications. 2019;10(1):1-16.
58. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nature genetics. 2019;51(3):431.
59. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nature genetics. 2018;50(5):668.
60. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. Nature genetics. 2019;51(5):793.
61. Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nature genetics. 2018;50(3):381.

62. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nature genetics. 2019;51(1):63.
63. Genetics IOCDF, Arnold PD, Askland KD, Barlassina C, Bellodi L, Bienvenu O, et al. Revealing the complex genetic architecture of obsessive–compulsive disorder using meta-analysis. Molecular psychiatry. 2018;23(5):1181.
64. Watson HJ, Yilmaz Z, Thornton LM, Hübel C, Coleman JR, Gaspar HA, et al. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. Nature genetics. 2019;51(8):1207-14.
65. Duncan LE, Ratanatharathorn A, Aiello AE, Almli LM, Amstadter AB, Ashley-Koch AE, et al. Largest GWAS of PTSD (N= 20 070) yields genetic overlap with schizophrenia and sex differences in heritability. Molecular psychiatry. 2018;23(3):666.
66. Howard DM, Adams MJ, Clarke T-K, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nature neuroscience. 2019;22(3):343.
67. Hail: An Open-Source Framework for Scalable Genetic Data [Internet]. 2017.
68. Okbay A, Baselmans BM, De Neve J-E, Turley P, Nivard MG, Fontana MA, et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. Nature genetics. 2016;48(6):624.
69. Luciano M, Hagenaars SP, Davies G, Hill WD, Clarke T-K, Shirali M, et al. Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. Nature genetics. 2018;50(1):6.
70. Jansen PR, Watanabe K, Stringer S, Skene N, Bryois J, Hammerschlag AR, et al. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. Nature genetics. 2019;51(3):394.
71. Linnér RK, Biroli P, Kong E, Meddens SFW, Wedow R, Fontana MA, et al. Genome-wide association analyses of risk tolerance and risky behaviours in over 1 million individuals identify hundreds of loci and shared genetic influences. Nature genetics. 2019;51(2):245.
72. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in∼ 700000 individuals of European ancestry. Human molecular genetics. 2018;27(20):3641-9.
73. Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology). 2005;67(2):301-20.
74. Liu K, Markovic J, Tibshirani R. More powerful post-selection inference, with application to the lasso. arXiv preprint arXiv:180109037. 2018.

75. Loftus JR. Selective inference after cross-validation. arXiv preprint arXiv:151108866. 2015.
76. Rosseel Y. Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). Journal of statistical software. 2012;48(2):1-36.
77. MacKinnon DP, Krull JL, Lockwood CM. Equivalence of the mediation, confounding and suppression effect. Prevention science. 2000;1(4):173-81.

Chapter 4 – The p factor: genetic analyses support a general dimension of psychopathology in childhood and adolescence

This chapter is presented as a published paper. It is an exact copy of the following publication:

Allegrini, A. G.*, Cheesman, R.*, Rimfeld, K., Selzam, S., Pingault, J. B., Eley, T. C., & Plomin, R. (2019). The p factor: genetic analyses support a general dimension of psychopathology in childhood and adolescence. *Journal of Child Psychology and Psychiatry*, *61*(1), 30-39.

*joint authorship

Supplementary materials are included in Appendix 3.

Journal of Child Psychology and Psychiatry 61:1 (2020), pp 30–39 doi:10.1111/jcpp.1311

The p factor: genetic analyses support a general dimension of psychopathology in childhood and adolescence

Andrea G. Allegrini,[1]* Rosa Cheesman,[1]* Kaili Rimfeld,[1] Saskia Selzam,[1] Jean-Baptiste Pingault,[1,2] Thalia C. Eley,[1] and Robert Plomin[1]

[1]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; [2]Division of Psychology and Language Sciences, University College London, London, UK

Background: Diverse behaviour problems in childhood correlate phenotypically, suggesting a general dimension of psychopathology that has been called the p factor. The shared genetic architecture between childhood psychopathology traits also supports a genetic p. This study systematically investigates the manifestation of this common dimension across self-, parent- and teacher-rated measures in childhood and adolescence. **Methods:** The sample included 7,026 twin pairs from the Twins Early Development Study (TEDS). First, we employed multivariate twin models to estimate common genetic and environmental influences on p based on diverse measures of behaviour problems rated by children, parents and teachers at ages 7, 9, 12 and 16 (depressive traits, emotional problems, peer problems, autism traits, hyperactivity, antisocial behaviour, conduct problems and psychopathic tendencies). Second, to assess the stability of genetic and environmental influences on p across time, we conducted longitudinal twin modelling of the first phenotypic principal components of childhood psychopathological measures across each of the four ages. Third, we created a genetic p factor in 7,026 unrelated genotyped individuals based on eight polygenic scores for psychiatric disorders to estimate how a general polygenic predisposition to mostly adult psychiatric disorders relates to childhood p. **Results:** Behaviour problems were consistently correlated phenotypically and genetically across ages and raters. The p factor is substantially heritable (50%–60%) and manifests consistently across diverse ages and raters. However, residual variation in the common factor models indicates unique contributions as well. Genetic correlations of p components across childhood and adolescence suggest stability over time (49%–78%). A polygenic general psychopathology factor derived from studies of psychiatric disorders consistently predicted a general phenotypic p factor across development (0.3%–0.9%). **Conclusions:** Diverse forms of psychopathology generally load on a common p factor, which is highly heritable. There are substantial genetic influences on the stability of p across childhood. Our analyses indicate genetic overlap between general risk for psychiatric disorders in adulthood and p in childhood, even as young as age 7. The p factor has far-reaching implications for genomic research and, eventually, for diagnosis and treatment of behaviour problems. **Keywords:** Childhood psychopathology; behavioural genetics; genomics.

Introduction

The p factor, analogous to the concept of general intelligence ('g'), reflects the observation that individuals who score highly on certain psychopathological traits also score highly on others (Caspi et al., 2014). Recent research suggests that this single continuous dimension can, in part, summarise and explain liability to a wide range of psychopathologies in childhood.

Interest in the p factor stemmed initially from high levels of psychopathological comorbidity in adults. The co-occurrence of psychiatric disorders is strikingly high, with up to 50% of individuals diagnosed with a mental illness going on to develop two or more comorbidities in a 12-month period (Kessler et al., 2005). Already during childhood and adolescence, forms of psychopathology are often comorbid. A recent report found that 1 in 20 British young people under 20 years of age met criteria for 2 or more mental disorders (NHS Digital 2017).

Quantitative genetic research suggests that shared genetic factors contribute substantially to the observed co-occurrence of psychopathological traits (Plomin, DeFries, Knopik, & Neiderhiser, 2016). Several multivariate twin and family studies have replicated the finding that a common genetic factor influences a wide range of emotional and behavioural problems in childhood (Lahey, Van Hulle, Singh, Waldman, & Rathouz, 2011; Pettersson, Larsson, & Lichtenstein, 2016; Tackett et al., 2013; Waldman, Poore, van Hulle, Rathouz, & Lahey, 2016). Many studies have investigated developmental genetic effects on specific psychopathological traits in childhood (e.g. Pingault et al., 2015), yet little is known about the genetic and environmental architecture of general psychopathology across development. Stability and change in p across time and the extent to which genetic influences drive age-related patterns remain largely unknown. Here, for the first time, we systematically investigate p across diverse ages, raters and measures in childhood and adolescence.

Conflict of interest statement: No conflicts declared.
*These authors contributed equally to this work.

It is also unknown to what extent a general p factor across earlier development relates to adult psychopathology. In addition to genetic analyses using the twin and family designs, polygenic scores are a new genomic tool that can be used to test for shared genetic effects across traits. Polygenic scores are constructed by aggregating genetic risk across thousands of genetic variants, thus indexing the genetic liability that each individual carries for a specific trait. A landmark study in the field of psychiatric genetics (International Schizophrenia Consortium et al. 2009) first showed that a polygenic score for schizophrenia was also associated with bipolar disorder, suggesting a shared genetic component underlying these two disorders, which has been substantiated further more recently (Cross-Disorder Group of the Psychiatric Genomics Consortium et al. 2019). Several studies have used polygenic scores for schizophrenia, ADHD and other psychiatric disorders to predict general psychopathology in childhood. An increasing amount of evidence converges on the finding that few polygenic effects specific to individual aspects of psychopathology remains after conditioning on the p factor (Brikell et al., 2017; Jones et al., 2016, 2018; Riglin et al., 2018). These studies also suggest that genetic risk for psychiatric disorders emerges in childhood, in the form of continuously measured behaviour problems. More recently, a study using different genomic methods provided evidence for a 'polygenic p' factor (Selzam, Coleman, Caspi, Moffitt, & Plomin, 2018). However, no studies to date have empirically related 'polygenic p' to 'phenotypic p' or systematically tested the architecture of p across development and across different raters.

Here, we investigated the structure of general psychopathology across childhood and adolescence. Our study has three aims:

1. Investigate the genetic architecture of p in childhood through common pathway twin models across ages and raters.
2. Test the stability of p across childhood and adolescence through longitudinal quantitative genetic analysis of first principal components of psychopathology across ages (7, 9, 12 and 16) and raters (parent, teacher and self-ratings).
3. Estimate associations between childhood phenotypic p and adult polygenic p. The latter can be constructed by principal component analysis of polygenic scores for mostly adult psychiatric disorders created for each TEDS participant.

Methods

Sample

The sampling frame is the Twins Early Development Study (TEDS), a multivariate, longitudinal study of >10,000 twin pairs representative of England and Wales, recruited from 1994 to 1996 births (Haworth, Davis, & Plomin, 2013). The following exclusions were applied: extreme perinatal conditions, severe medical conditions, uncertain zygosity and unknown gender. Analyses were conducted on a subsample of unrelated individuals with available genotype data and their cotwins (N = 7,026). Genomic analyses were limited to unrelated individuals (one twin from each pair).

Genotyping

Data were available for 3,057 individuals genotyped on the Affymetrix GeneChip 6.0 and 3,969 individuals genotyped on HumanOmniExpressExome-8v1.2 arrays. Typical quality control procedures were followed (e.g. samples were removed based on call rate <0.98, MAF < 0.5%). Genotypes from the two platforms were separately imputed and then harmonised (for detail see Selzam et al., 2018).

Measures

Twins Early Development Study (TEDS) measures have been described previously (Haworth et al., 2013). Measures administered at ages 7, 9, 12 and 16 were included in our analyses. Some of these measures (e.g. peer problems, prosocial behaviour (reversed), autism traits) have not previously been used in other studies of general psychopathology, but we adopted a hypothesis-free approach in an attempt to capture a general trait that is pervasive across diverse domains. For similar reasons, we included all measures available at each age, even though some measures (e.g. aggression) were available only at one age. Table 1 summarises the measures included in this study. Due to the wide range of ages, raters and measures used in the analyses; for information on response rates, please see Haworth et al. (2013).

For all phenotypes, z-standardised residuals were derived for each scale regressed on sex and age. Standardised scores for each scale were calculated as mean scores, with the requirement of complete data for more than half of the items (i.e. 3 of 4 or 2 of 3). All procedures were executed using RStudio (version 1.1.419; Rstudio 2019).

Age 7 measures. We used both parent and teacher ratings of all subscales of the Strengths and Difficulties Questionnaire (SDQ; Hyperactivity, Conduct Problems, Peer Problems, Emotional Problems and Prosocial (reversed; Goodman, 1997), as well as the Antisocial Process Screening Device (APSD; Frick & Hare, 2001) and a measure of autism traits.

Age 9 measures. The five subscales of the SDQ and the Childhood Autism Spectrum Test (CAST; Scott, Baron-Cohen, Bolton, & Brayne, 2002; Williams et al., 2005) were included in the set of self-, parent- and teacher-reported measures. In addition, we used parent- and teacher-rated APSD and aggression (a mean of proactive and reactive scales) measures (Dodge & Coie, 1987).

Age 12 measures. The five subscales of the SDQ, the APSD and the CAST were included in the set of self-, parent- and teacher-reported measures. Parent reports of the Moods and Feelings Questionnaire (MFQ) assessing depressive traits (Angold, Costello, Messer, & Pickles, 1995) and the Conners' ADHD behaviour measure (Conners, 2003) were also available.

Age 16 measures. The five subscales of the SDQ, MFQ and The Autism Quotient (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) were available from self-, parent and teacher reports. Parent-rated data on Conners

Table 1 Summary of psychopathology measures available in the Twins Early Development Study (TEDS)

Construct	Measure	Age/reporter	Reference
ADHD behaviours	Conners Parent Rating Scale	P12, P16	Conners (2003)
Aggression	Reactive-Proactive Aggression Questionnaire	P9, T9	Dodge and Coie (1987)
Anxiety-related Behaviours	ARBQ	P16	Eley et al. (2003)
Autism traits	Autism traits	P7, T7	DSM-IV criteria items
	Childhood Autism Spectrum Test (CAST)	C9, P9, T9, C12, P12, T12	Scott et al. (2002) and Williams et al. (2005)
	Autism Quotient (AQ)	C16, T16	Baron-Cohen et al. (2001)
Callous Unemotional Traits	Callous Unemotional Scale	P16	Kimonis et al. (2008)
Conduct Problems	Strengths and Difficulties Questionnaire (SDQ)	P7, T7, C9, P9, T9, C12, P12, T12, C16, T16	Goodman (1997)
Depressive Symptoms	Moods and Feelings Questionnaire (MFQ)	P12, C16, T16	Angold et al. (1995)
Emotional Problems	Strengths and Difficulties Questionnaire (SDQ)	P7, T7, C9, P9, T9, C12, P12, T12, C16, T16	Goodman (1997)
Hyperactivity	Strengths and Difficulties Questionnaire (SDQ)	P7, T7, C9, P9, T9, C12, P12, T12, C16, T16	Goodman (1997)
Peer Problems	Strengths and Difficulties Questionnaire (SDQ)	P7, T7, C9, P9, T9, C12, P12, T12, C16, T16	Goodman (1997)
Prosocial Behaviour	Strengths and Difficulties Questionnaire (SDQ)	P7, T7, C9, P9, T9, C12, P12, T12, C16, T16	Goodman (1997)
Psychopathic Tendencies	Antisocial Process Screening Device (APSD)	P7, T7, P9, T9, C12, P12, T12	Frick and Hare (2001)

Letters refer to the reporter (C = child self-report, P = parent report, T = teacher report), and numbers refer to the ages at which measures were available.

ADHD measure, the inventory for the Callous Unemotional scale (Kimonis et al., 2008) and the Anxiety-related Behaviors Questionnaire (Eley et al., 2003) were also included.

Statistical analyses

Common pathway twin models of behaviour problem measures for each rater at each age. To estimate the genetic and environmental influence on phenotypic variance in general psychopathology and to examine loadings of individual psychopathology measures on p, we conducted multivariate twin model-fitting analyses. In the twin design, differences in within-pair trait correlations for monozygotic (MZ) and dizygotic (DZ) twins are used to estimate genetic, shared environmental and nonshared environmental effects on traits. Greater MZ than DZ similarity indicates additive genetic influence (A). Within-pair similarity that is not due to genetic factors is attributed to shared environmental influences (C). Nonshared environment (E) accounts for individual-specific factors that influence differences among siblings from the same family, plus measurement error. We considered genetic and environmental associations between all psychopathology measures at each age and separately for each rater. Specifically, we fit the data to the common pathway model (Rijsdijk, 2014). This is a multivariate twin model, in which common genetic and environmental variation influence all measures via a single common latent (p) factor. The model allows the estimation of genetic and environmental influences on a common factor (p) and of the factor loadings of each measure of psychopathology on the latent liability (p). The common pathway model also allows the estimation of genetic and environmental variance in each trait that is independent of the common factor.

Longitudinal twin analysis: Cholesky decomposition of phenotypic principal components. We performed a Cholesky decomposition of the parent-rated phenotypic p principal components, allowing for the investigation of stability and innovation in the genetic and environmental influences on our measures of p across the four ages. We focused on parent-rated data since measures were much more consistent across time than for self-report and teacher report. The first genetic factor (A1) represents genetic influences on p at age 7. The extent to which these same genes also influence p at ages 9, 12 and 16 is also estimated, and is represented by the diagonal pathways from A1 to the other variables. The second genetic factor (A2) represents genetic influences on p at age 9 that are independent of those influencing age 7. The extent to which these genes also influence p at ages 12 and 16 is also estimated. The third genetic factor (A3) indexes genetic influences on p at age 12 that are independent of genetic influences shared with the previous ages. The impact of these genes on age 16 general psychopathology is also estimated. Finally, the fourth genetic factor (A4) represents residual genetic influences on age 16 general psychopathology. The same decomposition is done for the shared environmental and nonshared environmental influences (C1–4 and E1–4, respectively). All twin model fitting analyses using full-information maximum likelihood were carried out with structural equation modelling software OpenMx (Neale et al., 2016).

Extracting p: Principal Component Analyses (PCA). In preparation for longitudinal analyses and genomic prediction analyses, we obtained the first principal component (1st PC) of behaviour problem phenotypes at each age separately for child, parent and teacher ratings. Only individuals with complete data were used to generate PCs, as PCA does not allow for missing data. We report full results from PCA, which in themselves give insights into the phenotypic architecture of p in childhood. The variance explained by the first PC suggests how much the p factor underpins diverse forms of psychopathology, and loadings of each measure on the first PC

indicate the extent to which variables reflect general psychopathology.

We also obtained the first PC from polygenic scores for psychiatric disorders (polygenic p). We used publicly available genome-wide association summary statistics for eight major psychiatric traits: autism spectrum disorder (Grove et al., 2019), major depressive disorder (MDD; Wray et al., 2018), bipolar disorder (BIP), schizophrenia (SCZ; Pardiñas et al., 2018), attention deficit hyperactivity disorder (ADHD; Demontis et al., 2019), obsessive–compulsive disorder (OCD; International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OCGAS), 2018), anorexia nervosa (AN; Duncan et al., 2017) and posttraumatic stress disorder (PTSD; Duncan et al., 2018). For each psychiatric disorder, polygenic scores for each TEDS participant were created in LDpred (Vilhjálmsson et al., 2015), assuming a fraction of causal markers of 1 (analysis steps were similar to Selzam et al., 2018).

Assessing the association between the polygenic 1st PC and the phenotypic 1st PC across childhood and adolescence. To assess the extent to which the genetic predisposition for a general psychopathology factor relates to p in childhood, we performed ordinary least square regression analyses of phenotypic p on polygenic p at each age separately by each rater. Age, sex and the first 10 genomic principal components were regressed from all dependent and independent variables, and standardised residuals were used in all linear models.

Results

Common pathway twin models

Common pathway twin models showed substantial heritability for the p factor at each age for all raters (50%–60%). See Figure 1 for parent-rated measures and Figure S1 for teacher-rated and child-rated measures. Figure S2 summarises the heritability estimates for the common factor at each age and for each rater. Shared environmental effects were moderate for the parent-rated common factors (~30%; Figure S1), absent for the teacher-rated common factors (~0%; Figure S1) and weak for the self-rated common factors (~15%, declining with age; Figure S1). Autism traits, conduct problems, antisocial behaviour and psychopathic tendencies loaded the highest on the parent-rated and teacher-rated common factors, while emotional problems, depression and anxiety loaded the highest for the child-rated p factor. We also found substantial specific genetic and environmental variance for all measures suggesting unique influences on psychopathological measures beyond the p factor. See Table S1 for full estimates of common and specific genetic and environmental influences. Table S1 also contains full model-fitting results, and sample sizes of measures, which ranged from 2,216 to 5,592 twin pairs who also had genotype data. See Table S2 for model fit statistics.

Cholesky decomposition of p across development

The Cholesky decomposition of principal components suggests stability of genetic effects on general psychopathology across childhood and adolescence, in addition to new genetic components at each age, as shown in Figure 2 for parent ratings. Figure S5 shows genetic correlations derived from a correlated factor solution. Age-to-age genetic correlations derived from these results are high, ranging from 0.49 to 0.78 (see Figure S5). Figures S3 and S4 present the Cholesky model-fitting results for shared and nonshared environmental variance components, respectively. Figures S6–S15 indicate phenotypic correlations among psychopathology measures at all ages and for all raters. These correlations are notably similar to genetic correlations from the Cholesky model. Figure S16 shows phenotypic correlations between principal components across age and raters. Parent-rated correlations were the strongest, ranging from .47 to .68, while child and teacher-rated correlations were somewhat weaker, but still substantial (i.e. ~.3 to ~.4 for both). Cross-rater correlations were strongest for child- and parent-rated p factors, ranging between ~.3 and ~.4, and weakest between teacher-rated and child-rated p factors (~.1 to ~.2). Table S3 lists loadings of observed measures on first principal components, which shows that loadings are consistently substantial for all measures, ages and raters. The first unrotated principal component of phenotypic measures accounted for 40% to 50% of the variance across ages and raters (see Table S4, which also shows the sample sizes for each 1st PC, which ranged from 1,391 to 4,490).

Prediction of phenotypic p with polygenic p

A polygenic p score defined as the first unrotated principal component of polygenic scores for mostly adult psychiatric disorders was significantly associated with phenotypic p scores in childhood, predicting 0.3%–0.9% of the variance across ages and raters. See Table S5 for full polygenic prediction results. Prediction was generally consistent across ages and raters, although standard errors are largely overlapping (see Figure 3). Figure S17 shows correlations between the polygenic scores in TEDS used to derive polygenic p. Although these correlations are modest (0.01–0.32), the first principal component of polygenic scores from psychiatric traits explained up to 20% of the polygenic score variability. The loadings on polygenic p (shown in Figure S18) were all above 0.3, apart from obsessive–compulsive disorder (0.13) and posttraumatic stress disorder (0.18). This could be because the GWA summary statistics for these disorders were derived from smaller samples than the others. Analyses applying to differentially powered summary statistics for the same traits to TEDS data have demonstrated that, as GWA study sample sizes increase, factor loadings on a polygenic p factor are likely to approach those derived from family studies (Selzam et al., 2018).

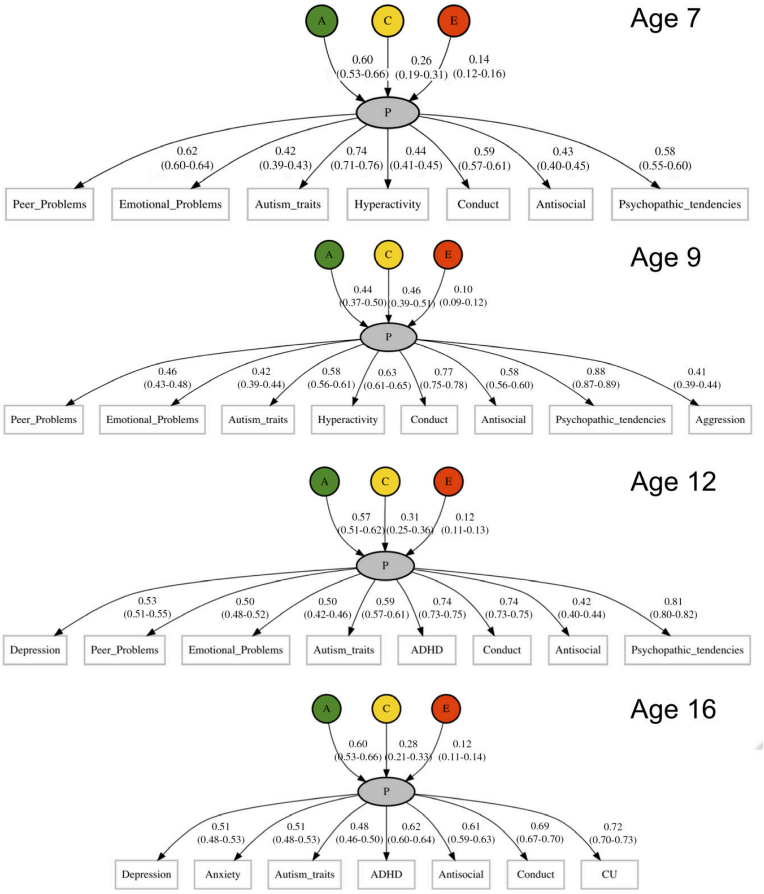


Figure 1 Common pathway twin models of p (parent rated) at ages 7, 9, 12 and 16

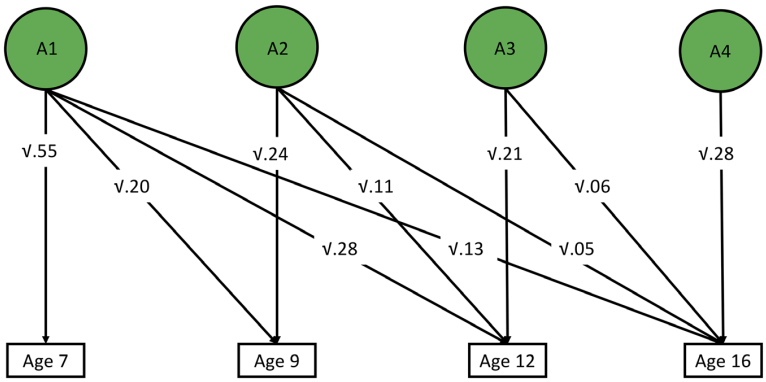


Figure 2 Additive genetic influences on parent-rated p across age, derived from longitudinal twin model-fitting (Cholesky decomposition)

Discussion

For the first time, we systematically quantified the extent to which a single common factor relates to diverse forms of psychopathology across childhood and adolescence using phenotypic, genetic and genomic methods. Phenotypically, our results parallel previous findings, suggesting a common psychopathology factor. We show that p emerges consistently across different measures at different ages and raters. Our genetic results support three main conclusions. First, multivariate twin analyses revealed that 48%–80% of the variance in the

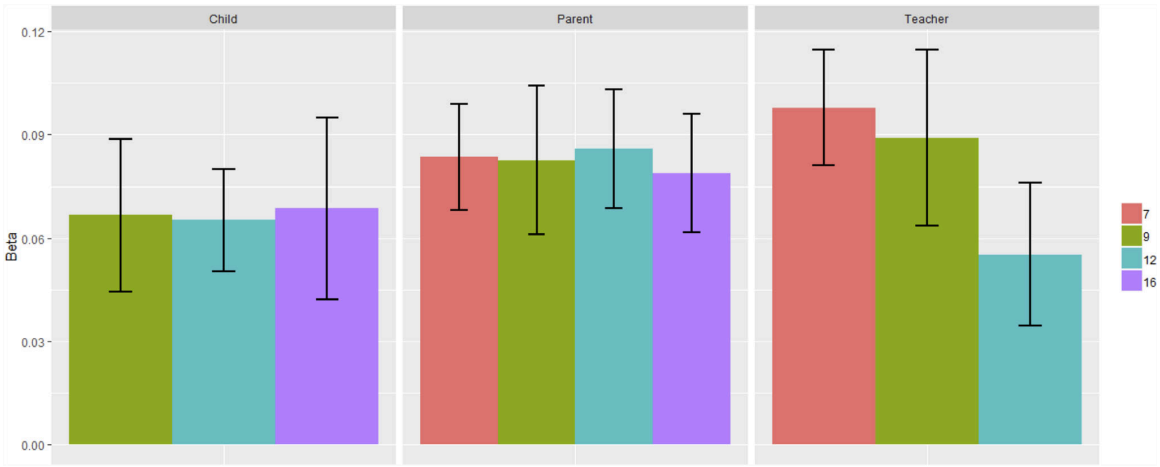


Figure 3 Prediction of phenotypic p with polygenic p by ages and raters. *Note*: Error bars represent ± 1 standard error

common factor was due to genetic influences, depending on age and raters considered. It is important to note, however, that although we found a consistent and stable genetic p factor across childhood and adolescence, substantial unique genetic and environmental influences indicate that there are also genetic components specific to each trait and each age beyond p. Second, longitudinal twin model fitting showed that this genetic p factor was stable across time. Third, polygenic prediction analyses demonstrate that there are shared genetic influences connecting childhood psychopathology to general risk for (mostly) adult psychiatric disorders. Even though variance predicted is low (i.e. ~1%), effect sizes are within the expected range considering previous research in this area (e.g. Riglin et al., 2018; Grotzinger et al., 2019; see below). In sum, these analyses provide further evidence that a common genetic substrate permeates the landscape of psychopathology, across measures, ages and raters.

Our common pathway twin modelling analyses, for which we adopted a hypothesis-free approach to the inclusion of measures, show that diverse psychopathological traits contribute to p. Furthermore, it is commonly acknowledged that all psychopathological traits are dimensional traits both at the phenotypic and genetic levels (Plomin, Haworth, & Davis, 2009). Future research might investigate the extent to which p extends to other behavioural domains. For example, suggestive evidence of links between p and personality has begun to emerge (Rosenström et al., 2018). In addition, instead of testing competing factor structures, we focused on the common pathway model, since the present study aimed to investigate the most parsimonious highest order part of the hierarchy that we call p. This is further justified by evidence for correlations and heterotypic sequential comorbidity across the internalising and externalising domains (Caspi & Moffitt, 2018).

Differences between raters in our common pathway twin analyses suggested some additional insights. First, inspection of the loadings of psychopathology measures revealed that 'externalising' problems relating to conduct and antisocial behaviour contributed most to parent- and teacher-rated common factors, whereas 'internalising' problems such as depression and anxiety loaded highest for the child-rated p factor. This could suggest that parents report on overt behaviours, which might stem from worry and sadness from the child's perspective. Second, we observed that shared environmental influences were moderate for the parent-report-based p factor, but negligible for self- and teacher-rated p, respectively. This pattern of results is most likely due to rater bias in that parent ratings are based on a single informant rating both twins, whereas for teacher and self-ratings different informants rate each twin (Bartels et al., 2004).

Our longitudinal twin model fitting and polygenic scoring revealed substantial genetic influences on stability of general psychopathology across childhood. Our polygenic score results suggest that these stable genetic influences overlap with those underlying adult psychiatric disorders.

In terms of predictive value, effect sizes of our polygenic p score in association with phenotypic p are weak (~1%). However, these are within the expected range for polygenic prediction of psychiatric traits, and consistent with previous literature on polygenic risk and general psychopathology, whereby current polygenic scores for adult psychiatric traits often explain < 1% of the variance in general psychopathology (Riglin et al., 2018), similar to a polygenic score created from a p factor GWAS (Grotzinger et al., 2019). The predictive accuracy of a polygenic p score will increase as the power of single GWAS of psychiatric traits grows, especially when GWAS go beyond DNA arrays consisting of common

SNPs to include all DNA variants as assessed by whole-genome sequencing. In addition, there is increasing evidence that joint multivariate analyses of traits are likely to increase the predictive power of polygenic scores (e.g. Grotzinger et al., 2019; Maier et al., 2018).

Future research could assess influences on different temporal trajectories of p across childhood and adolescence. One study recently showed that polygenic scores for neurodevelopmental disorders (schizophrenia, ADHD) and depression were associated with early adolescent onset depression, whereas later onset depression was only predicted significantly by depression polygenic scores (Rice et al., 2018). This could be repeated with more powerful polygenic p scores.

Notably, some interesting results also emerge about the environment. There are some known general 'environmental' risks for psychopathology such as birthweight, birth complications and childhood maltreatment that are associated with diverse neurodevelopmental outcomes (Caspi & Moffitt, 2018; Lim et al., 2018). However, we find that nonshared environmental effects contribute less than genetic effects to the general psychopathology factor and its temporal stability. As has been demonstrated in previous studies of specific psychopathology, nonshared environment is largely time-specific, and genetic effects clearly contribute more to stability.

Naturally, through the course of multivariate longitudinal studies like TEDS, there are changes in available measures and informants, which in turn can introduce variability in the pattern of results. That is, our measures of p are not perfect indices of general liability to psychopathology, but reflect the specific measures and raters available at each age. This is problematic when estimating genetic and environmental influences on stability and change in p across time. Specifically, any innovation cannot solely be attributed to p, as it will reflect new influences on new measures that were not available at the previous age. This criticism is difficult to overcome even with the availability of consistent data: exactly the same measure at different time points does not necessarily reflect the same thing. We consider that the availability of varied measures is a strength rather than a limitation of the present study because this means that our strong evidence for genetic p and genetic stability for p emerges despite the use of different measures. In the cognitive literature on g, this phenomenon is known as the *indifference of the indicator* – any set of diverse cognitive measures yields a strong g factor (Spearman, 1904). Factor loadings were consistently substantial, not only across measures but also across ages and raters. Importantly, the phenotypic correlations between first principal components across time (ranging between ~0.5 and ~0.7) suggest that p indexes a consistent core psychopathology trait.

The fact that we can predict childhood p using polygenic p derived from typically adult case–control genome-wide association studies has several interesting implications. First, it suggests that in young children there are already continuous manifestations of genetic risk for adult case–control psychiatric disorders that are unmeasured in our population-based, developmental sample. Therefore, this extends the insight from twin analyses within our sample that genetic risk for psychopathology at age 7 correlates about 0.50 with genetic risk for psychopathology at age 16. In other words, early onset behavioural and emotional problems are early signs of psychiatric genetic risk. This supports other evidence for the usefulness of early intervention for psychiatric problems. The second implication of the genetic overlap between p in childhood and adulthood relates to research design. Specifically, researchers could increase the power of genome-wide association studies to detect DNA variation associated with general risk for psychopathology by aggregating diverse traits across wide age ranges. One way to implement this is a common factor genome-wide association analysis using Genomic SEM (Grotzinger et al., 2019). Similarly, the modest power of psychiatric polygenic scores to predict traits in childhood could be enhanced using multitrait frameworks to generate predictors that leverage the shared genetic risk between traits (e.g. SMTpred; Maier et al., 2018).

The current clinical zeitgeist focuses on specificity. The recognition that a common factor transcends diverse aspects of psychopathology in childhood is of primary importance, as this knowledge can inform early detection of children at risk in the general population.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Figure S1. Common pathway twin models for child-rated and teacher-rated psychopathology measures by age.
Figure S2. Comparison of twin heritability estimates from common pathway models.
Figures S3–S4. Shared environmental and non-shared environmental influences on p (parent-rated) across age, derived from longitudinal twin model-fitting (Cholesky decomposition).
Figure S5. Correlated factor solution of the longitudinal Cholesky decomposition.
Figures S6–S15. Phenotypic correlations among psychopathology measures used to construct phenotypic p factors.
Figure S16. Correlations of 1st PCs across ages.
Figure S17. Correlations between polygenic scores for psychiatric traits used to construct polygenic p.
Figure S18. PCA results for polygenic p-factor.

Table S1. Additional parameters derived from common pathway twin models of childhood psychopathology in TEDS.
Table S2. Model fit statistics for common pathway twin models of childhood psychopathology in TEDS.
Table S3. Loadings on first principal components of psychopathology measures for each age and rater.
Table S4. Variance explained by 1st PCs for each age and rater.
Table S5. Association statistics for polygenic p across phenotypic p measures.

Acknowledgements
The authors gratefully acknowledge the ongoing contribution of the participants in the Twins Early Development Study (TEDS) and their families. TEDS is supported by a programme grant to R.P. from the UK Medical Research Council (MR/M021475/1 and previously G0901245), with additional support from the US National Institutes of Health (AG046938). The research leading to these results has also received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013)/grant agreement no. 602768 and ERC grant agreement no. 295366). R.P. is supported by a Medical Research Council Professorship award (G19/2). T.C.E. is part funded by the above programme grant from the UK Medical Research Council (MR/M021475/1). This study represents independent research part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. High-performance computing facilities were funded with capital equipment grants from the GSTT Charity (TR130505) and Maudsley Charity (980). R.C. is supported by an ESRC studentship. A.G.A. has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 721567. J-B.P. is a fellow of MQ: Transforming Mental Health (MQ16IP16). The authors have declared that they have no competing or potential conflicts of interest.

Correspondence
Andrea G. Allegrini, Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, 16 de Crespigny Park, Denmark Hill, London SE5 8AF, UK; Email: andrea.allegrini@kcl.ac.uk

Key points

- We investigated the underlying structure of p across diverse measures, ages and raters, and consistently found a substantial genetic component, in line with previous theory.
- We showed that this genetic component is stable across time, with influences in childhood being pervasive across development through to adolescence.
- Genomic analyses revealed shared genetic risk between p in children as young as 7 and general risk for adult psychiatric disorders.
- We provide further evidence that, in addition to residual variation specific to each trait, a common genetic substrate permeates the landscape of psychopathology.

References
Angold, A., Costello, E.J., Messer, S.C., & Pickles, A. (1995). The development of a Questionnaire for use in epidemiological studies of depression in children and adolescents. *International Journal of Methods in Psychiatric Research*, *5*, 237–249.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*, 5–17. https://doi.org/10.1023/A:1005653411471.

Bartels, M., Boomsma, D.I., Hudziak, J.J., Rietveld, M.J.H., van Beijsterveldt, T.C.E.M., & van den Oord, E.J.C.G. (2004). Disentangling genetic, environmental, and rater effects on internalizing and externalizing problem behavior in 10-year-old twins. *Twin Research*, *7*, 162–175.

Brikell, I., Larsson, H., Lu, Y., Pettersson, E., Chen, Q., Kuja-Halkola, R., … & Martin, J. (2017). The contribution of common genetic risk variants for ADHD to a general factor of childhood psychopathology. *BioRxiv*.

Caspi, A., Houts, R.M., Belsky, D.W., Goldman-Mellor, S.J., Harrington, H., Israel, S., … & Moffitt, T.E. (2014). The p factor: One General psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, *2*, 119–137.

Caspi, A., & Moffitt, T.E. (2018). All for one and one for all: Mental disorders in one dimension. *The American Journal of Psychiatry*, *175*, 831–844.

Conners, C.K. (2003). *Conners' Rating Scales-revised: Technical manual*. New York: Multi-Health System Inc.

Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee, P.H., Anttila, V., Won, H., Feng, Y.-C.A., Rosenthal, J., … & Smoller, J.W. (2019). Genome wide meta-analysis identifies genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders: Supplemental Tables 1-18. *BioRxiv*.

Demontis, D., Walters, R.K., Martin, J., Mattheisen, M., Als, T.D., Agerbo, E., … & Neale, B.M. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature Genetics*, *51*, 63–75.

Dodge, K.A., & Coie, J.D. (1987). Social-information-processing factors in reactive and proactive aggression in children's

peer groups. *Journal of Personality and Social Psychology*, *53*, 1146–1158.
Duncan, L.E., Ratanatharathorn, A., Aiello, A.E., Almli, L.M., Amstadter, A.B., Ashley-Koch, A.E., … & Koenen, K.C. (2018). Largest GWAS of PTSD (N = 20,070) yields genetic overlap with schizophrenia and sex differences in heritability. *Molecular Psychiatry*, *23*, 666–673.
Duncan, L., Yilmaz, Z., Gaspar, H., Walters, R., Goldstein, J., Anttila, V., … & Bulik, C.M. (2017). Significant locus and metabolic genetic correlations revealed in genome-wide association study of anorexia nervosa. *The American Journal of Psychiatry*, *174*, 850–858.
Eley, T.C., Bolton, D., O'Connor, T.G., Perrin, S., Smith, P., & Plomin, R. (2003). A twin study of anxiety related behaviours in pre-school children. *Journal of Child Psychology and Psychiatry*, *44*, 945–960.
Frick, P., & Hare, R.D. (2001). *The antisocial process screening device*. Toronto, ON: Multi-Health Systems.
Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, *38*, 581–586.
Grotzinger, A.D., Rhemtulla, M., de Vlaming, R., Ritchie, S.J., Mallard, T.T., Hill, W.D., … & Tucker-Drob, E.M. (2019). Genomic SEM Provides insights into the multivariate genetic architecture of complex traits. *Nature Human Behaviour*, *3*, 513–525.
Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R.K., Won, H., … & Børglum, A.D. (2019). Identification of common genetic risk variants for autism spectrum disorder. *Nature Genetics*, *51*, 431–444.
Haworth, C.M.A., Davis, O.S.P., & Plomin, R. (2013). Twins Early Development Study (TEDS): A genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin Research and Human Genetics*, *16*, 117–125.
International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC), & OCD Collaborative Genetics Association Studies (OCGAS) (2018). Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Molecular Psychiatry*, *23*, 1181–1188.
International Schizophrenia Consortium, Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., … & Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, *460*, 748–752.
Jones, H.J., Heron, J., Hammerton, G., Stochl, J., Jones, P.B., Cannon, M., … & Me Research Team (2018). Investigating the genetic architecture of general and specific psychopathology in adolescence. *Translational Psychiatry*, *8*, 145.
Jones, H.J., Stergiakouli, E., Tansey, K.E., Hubbard, L., Heron, J., Cannon, M., … & Zammit, S. (2016). Phenotypic manifestation of genetic risk for schizophrenia during adolescence in the general population. *JAMA Psychiatry*, *73*, 221–228.
Kessler, R.C., Berglund, P., Demler, O., Jin, R., Merikangas, K.R., & Walters, E.E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*, 593–602.
Kimonis, E.R., Frick, P.J., Skeem, J.L., Marsee, M.A., Cruise, K., Munoz, L.C., … & Morris, A.S. (2008). Assessing callous-unemotional traits in adolescent offenders: Validation of the Inventory of Callous Unemotional Traits. *Journal of the International Association of Psychiatry and Law*, *31*, 241–252.
Lahey, B.B., Van Hulle, C.A., Singh, A.L., Waldman, I.D., & Rathouz, P.J. (2011). Higher-order genetic and environmental structure of prevalent forms of child and adolescent psychopathology. *Archives of General Psychiatry*, *68*, 181–189.
Lim, K.X., Liu, C.-Y., Schoeler, T., Cecil, C.A.M., Barker, E.D., Viding, E., … & Pingault, J.-B. (2018). The role of birth weight on the causal pathway to child and adolescent ADHD symptomatology: A population-based twin differences longitudinal design. *Journal of Child Psychology and Psychiatry*, *59*, 1036–1043.
Maier, R.M., Zhu, Z., Lee, S.H., Trzaskowski, M., Ruderfer, D.M., Stahl, E.A., … & Robinson, M.R. (2018). Improving genetic prediction by leveraging genetic correlations among human diseases and traits. *Nature Communications*, *9*, 989.
Neale, M.C., Hunter, M.D., Pritikin, J.N., Zahery, M., Brick, T.R., Kirkpatrick, R.M., … & Boker, S.M. (2016). Openmx 2.0: Extended structural equation and statistical modeling. *Psychometrika*, *81*, 535–549.
NHS Digital (2017). Mental Health of Children and Young People in England, 2017 [PAS] – NHS Digital [Online]. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2017/2017 [last accessed 11 December 2018].
Pardiñas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., Carrera, N., … & CRESTAR Consortium (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nature Genetics*, *50*, 381–389.
Pettersson, E., Larsson, H., & Lichtenstein, P. (2016). Common psychiatric disorders share the same genetic origin: A multivariate sibling study of the Swedish population. *Molecular Psychiatry*, *21*, 717–721.
Pingault, J.-B., Viding, E., Galéra, C., Greven, C.U., Zheng, Y., Plomin, R., & Rijsdijk, F. (2015). Genetic and environmental influences on the developmental course of attention-deficit/hyperactivity disorder symptoms from childhood to adolescence. *JAMA Psychiatry*, *72*, 651–658.
Plomin, R., DeFries, J.C., Knopik, V.S., & Neiderhiser, J.M. (2016). Top 10 replicated findings from behavioral genetics. *Perspectives on Psychological Science*, *11*, 3–23.
Plomin, R., Haworth, C.M.A., & Davis, O.S.P. (2009). Common disorders are quantitative traits. *Nature Reviews. Genetics*, *10*, 872–878.
Rice, F., Riglin, L., Thapar, A.K., Heron, J., Anney, R., O'Donovan, M.C., & Thapar, A. (2018). Characterizing Developmental trajectories and the role of neuropsychiatric genetic risk variants in early-onset depression. *JAMA Psychiatry*, *76*, 306.
Riglin, L., Thapar, A.K., Leppert, B., Martin, J., Richards, A., Anney, R., … & Thapar, A. (2018). The contribution of psychiatric risk alleles to a general liability to psychopathology in early life. *BioRxiv*.
Rijsdijk, F. (2014). Common pathway model. In *Wiley StatsRef: Statistics reference online*. Chichester, UK: John Wiley & Sons Ltd.
Rosenström, T., Gjerde, L.C., Krueger, R.F., Aggen, S.H., Czajkowski, N.O., Gillespie, N.A., … & Ystrom, E. (2018). Joint factorial structure of psychopathology and personality. *Psychological Medicine*, 1–10.
Rstudio (2019). Open source and enterprise-ready professional software for data science – RStudio [Online]. Available from: https://www.rstudio.com/ [last accessed 15 March 2019].
Scott, F.J., Baron-Cohen, S., Bolton, P., & Brayne, C. (2002). The CAST (Childhood Asperger Syndrome Test): Preliminary development of a UK screen for mainstream primary-school-age children. *Autism*, *6*, 9–31.
Selzam, S., Coleman, J.R.I., Caspi, A., Moffitt, T.E., & Plomin, R. (2018). A polygenic p factor for major psychiatric disorders. *Translational Psychiatry*, *8*, 205.
Spearman, C. (1904). "General Intelligence", Objectively determined and measured. *The American Journal of Psychology*, *15*, 201.
Tackett, J.L., Lahey, B.B., van Hulle, C., Waldman, I., Krueger, R.F., & Rathouz, P.J. (2013). Common genetic influences on

negative emotionality and a general psychopathology factor in childhood and adolescence. *Journal of Abnormal Psychology*, *122*, 1142–1153.

Vilhjálmsson, B.J., Yang, J., Finucane, H.K., Gusev, A., Lindström, S., Ripke, S., … & Price, A.L. (2015). Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *American Journal of Human Genetics*, *97*, 576–592.

Waldman, I.D., Poore, H.E., van Hulle, C., Rathouz, P.J., & Lahey, B.B. (2016). External validity of a hierarchical dimensional model of child and adolescent psychopathology: Tests using confirmatory factor analyses and multivariate behavior genetic analyses. *Journal of Abnormal Psychology*, *125*, 1053–1066.

Williams, J., Scott, F., Stott, C., Allison, C., Bolton, P., Baron-Cohen, S., & Brayne, C. (2005). The CAST (Childhood Asperger Syndrome Test): Test accuracy. *Autism*, *9*, 45–68.

Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., … & Sullivan, P.F. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*, *50*, 668–681.

Accepted for publication: 9 July 2019
First published online: 20 September 2019

Chapter 5 – Directional relationships between childhood psychopathology dimensions across development

This chapter is adapted from a manuscript in preparation for peer-review.

Supplementary materials are included in Appendix 5.

Andrea G Allegrini[1], Toos van Beijsterveldt[2], Dorret Boomsma[2], Kaili Rimfeld[1], Jean-Baptiste Pingault[1,3], Robert Plomin[1], Meike Bartels[2]*, Michel G Nivard[2]*

*joint authorship

[1]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK

[2]Biological Psychology department, Vrije University Amsterdam, NL

[3]Division of Psychology and Language Sciences, University College London, UK

Abstract

Disentangling between-person (co)variance from within-person (co)variance between psychiatric measures across childhood is vital to understand causes of (adult) comorbidity, and may reveal developmental pathways underlying mental health problems. We present results of a preregistered study conducted in two large population-based cohorts, the Twin Early Developmental Study (n=8,549) and the Netherlands Twin Register (n=16,677). We investigated the longitudinal associations between measures of common psychopathologies from childhood to early adolescence (age 7 to 12), jointly estimating trait-like between-person and state-like within-person processes across time to determine whether and to what extent directional relationships between psychiatric traits within-person, and between individuals within families, play a role in multivariate comorbidity. We conducted random intercepts cross-lagged panel model (RI-CLPM) analyses to unravel the longitudinal co-occurrence of child psychopathology dimensions, and developed an extension of the model to estimate sibling effects within-family (wf-RI-CLPM). Analyses were separately conducted in two large population-based cohorts, the Twin Early Developmental Study (n=8,549) and the Netherlands Twin Register (n=16,677), including measures of child problem behaviours based on the SDQ and CBCL scales respectively. We found evidence for strong between-person effects underlying the positive intercorrelation between problem behaviours across time. We further identify within-person positive directed relationships between measures of different psychopathologies that partly overlapped between cohorts and accounted up to 15% of the variance over time depending on measure. Lastly, by accommodating family-level data, we found evidence for reciprocal directional influences within sib-pairs over time, after accounting for similarities that arise through shared (genetic or environmental) influences. Our results indicate that directed relationships between psychopathology dimensions within-person and between siblings, within-family, partly explain the cooccurrence of psychopathologies in childhood and should be taken into account in developmental models of comorbidity.

Introduction

For many psychiatric disorders, disease onset can be traced back to childhood or adolescence (Akingbuwa et al., 2020; Jansen et al., 2018; Kessler, Chiu, Demler, & Walters, 2005; Riglin et al., 2018). Psychopathological comorbidity – the co-occurrence of psychiatric disorders – is pervasive, half of individuals with a mental health diagnosis is likely to be diagnosed with one or multiple other disorders within a year time (Bartels et al., 2018; Kessler et al., 2005; Kessler et al., 1994). Converging evidence from the quantitative genetics literature suggests that a common genetic predisposition may underlie psychopathology throughout development (Allegrini et al., 2019; Caspi & Moffitt, 2018; Grotzinger et al., 2019). However, residual trait variation in (longitudinal) common factor models of psychopathology, suggests that there are specific genetic contributions to separate psychopathological traits across time (Bartels et al., 2004). Probing the nature of co-morbidity as a population phenomenon, but perhaps more importantly across development within an individual, is critical for our understanding of the development of psychopathology.

Analogous to the concept of general intelligence (Spearman, 1961), the observation that psychopathology traits are positively intercorrelated across the lifespan has led to the conclusion that a common cause might underly such positive covariation (also called positive manifold; Borg, 2018). A number of studies have looked at the common aetiology between traits, both in childhood and adulthood, modelling the correlated factor structure of psychiatric traits (Lahey et al., 2012; Pettersson, Anckarsäter, Gillberg, & Lichtenstein, 2013). Considering several psychopathology traits, a seminal study in the field (Caspi et al., 2014) showed that a hierarchical structure produces the best fit to the data, pinnacle of which is a common psychopathology factor (called the p-factor; Caspi et al., 2014).

However, the positive intercorrelation observed across psychopathology measures can arise from different mechanisms than a common cause. In fact, it has been noted that a hierarchical solution is a mathematical necessity when observing a positive manifold (van Bork, Epskamp, Rhemtulla, Borsboom, & van der Maas, 2017). That is, at one point of the hierarchical structure there is always a common latent component that can be extracted from positive intercorrelations (van Bork et al., 2017). There is a variety of data generating processes that can give rise to the positive manifold between measures of psychopathology (van Bork et al., 2017). For example, a compelling alternative for the p-factor model is that proposed by the network approach of psychopathology, which poses that the observed correlation structure between psychopathology traits can arise from an underlying pattern of causal effects at the

symptom level (Borsboom, 2017). In this regard the temporal causal relationships between symptoms can induce a positive correlation between psychopathology traits, which appears consistent with the presence of a common cause. The different models, though, are not easily distinguished based on cross sectional data.

In summary, there is evidence that psychopathologies are the result of a developmental process, and tend to co-occur. This co-occurrence has been attributed to a common genetic aetiology (Caspi et al., 2014; Pettersson et al., 2013). However, it remains unclear whether and to what extent the correlation between psychopathologies is the product of correlated fundamental individual differences *between* people on stable traits (stable across the life course, with their correlation attributable to a general "p-factor"), or the product of a causal process *within* people where the temporal state on one variable (e.g. mood) causally influences the state of another variable (e.g. attention), inducing correlation. These two processes are not mutually exclusive and it is conceivable, perhaps even likely, that both correlated stable traits as well as direct influences of within-person states on other within-person states play a role in the development of (comorbid) psychopathology.

Disentangling trait like between-person from state like within-person processes of psychiatric traits across childhood is vital to understand causes of comorbidity, and to gain insights on developmental pathways underlying mental health problems. Here we set out to investigate the longitudinal directional relationships between psychopathology related traits from childhood to early adolescence (age 7 to 12), jointly estimating between-person and within-person processes. Traditionally, trait associations in panel data are modelled within a cross-lagged panel model (CLPM) to infer causal predominance of one variable over another across time (in the sense of Granger causality; Granger, 1969). This model cannot, however, distinguish a between-person from a within-person level of variation. Here we employ an alternative approach, the random intercept cross-lagged panel model (RI-CLPM), that formally models between-person trait-like stability across time by the addition of a random intercept (Hamaker, Kuiper, & Grasman, 2015). This random intercept can be thought of as a latent factor accounting for between-person stability over time.

Individuals, and their psychopathologies, do not develop in a vacuum, but rather are nested in social structures, often, but not always, primarily in a nuclear family. We can extend the contrast between trait like individual differences and state like within person processes to the family. Members of a family, siblings in particular, are known to behave alike and this is

attributed to similarities in their upbringing, the means their parents had available to support their development, their cultural capital, and shared heritable influences. Similarities between siblings in terms of their symptoms of psychopathology and co-morbidities can be attributed to state like individual differences which, in twin family models, can be further specified to be heritable stable influences or influences of the shared environment. However, there are obvious direct interactions between siblings, where age specific symptoms in one sibling, could precipitate mental symptoms in the other sibling at a later age.

Sibling interactions can result in cooperation and contrast effects, which respectively index the extent to which siblings will tend to imitate/stimulate each other increasing their phenotypic similarity, or contrast each other, decreasing their resemblance (Carey, 1986). Siblings interactions have been investigated in the quantitative genetics literature especially in regard to externalizing traits such as hyperactivity and conduct disorder (Boomsma, 2014; Rebollo & Boomsma, 2006; Simonoff et al., 1998; Thapar, Hervas, & McGuffin, 1995). A typical observation for traits such as these is the low dizygotic twin correlations compared to monozygotic twin correlations, which can be attributed to genetic dominance (interactions of two alleles at the same locus) or contrast effects. Results of such studies are mixed, finding that contrast effects are either due to siblings interactions (Thapar et al., 1995) or due to parental bias (Simonoff et al., 1998), where the difference in, say, hyperactivity ratings within sibling pairs index parental perception rather than a true difference between siblings. Importantly contrast effects can mask dominance (Simonoff et al., 1998). This picture is complicated by issues of power, which makes it difficult to estimate contrast versus dominant effects concurrently (Rebollo & Boomsma, 2006). With few notable exceptions (Carey, 1986; Dolan, de Kort, van Beijsterveldt, Bartels, & Boomsma, 2014; M. Rietveld, Posthuma, Dolan, & Boomsma, 2003; M. J. Rietveld, Hudziak, Bartels, Van Beijsterveldt, & Boomsma, 2004) these studies infer contrast effects rather than directly quantifying the reciprocal effects of sibling phenotypes within families. Here we investigate within-family reciprocal directed influences between siblings across several behavioural problems over time.

We extend the RICLPM to accommodate trait like similarities between siblings, leveraging monozygotic and dizygotic twin pairs, while concurrently allowing for siblings direct influences on each other's behaviours. The aim of this extension is to estimate reciprocal directional influences between siblings over time from similarities between siblings that arise through shared (genetic or environmental) influences that exist in a family. The pre-registered aim of the present study (See

(https://osf.io/dtvc8/?view_only=fe8e8eff7df1417da399a05625a5df8e) is exploratory in nature, as we do not have an assumption on the underlying directional relationships between the measures of psychopathology. We jointly estimate the between- and within-person contributions to the correlations between conduct problems, hyperactivity/inattention, emotional problems, and peer problems across development. Here the presence of within-person correlations between psychopathologies over time is viewed as a pre-requisite for, but not definitive evidence for, within-person direct causal effects between features of psychopathologies. In additional analyses, that were not pre-registered, we test whether similarities in symptoms of psychopathology between siblings are a function of correlated stable traits, due to heritable or environmental factors, or direct mutual influences of the behaviour of one sibling on the other.

Methods

Samples

We tested our primary analyses using data from the Twin Early Development Study (Rimfeld et al., 2019), a large longitudinal population-based study involving 16,810 pairs of twins born in England and Wales between 1994-1996. Here we focused on parent rated (mainly maternal) psychopathology measures administered when the twins were aged 7, 9 and 12 (total sample used in analyses n = 8,549). We then conducted a replication of these analyses in the Netherlands Twin register (NTR) (Ligthart et al., 2019), a longitudinal population-based sample with new twins data added every year. The Young-NTR (Bartels et al., 2007) contains data on twins from birth onwards. Twins are categorized by birth cohort and data collection is cohort driven. Here we also focused on maternal rated psychopathological measures administered when the twins were aged 7, 10, and 12 (total sample used in analyses n = 16,677).

Measures

In TEDS we employed the parent rated version of the Strength and Difficulties Questionnaire (SDQ) (Goodman, 1997) comprising four scales indexing emotional symptoms (anxiety and depression), conduct problems, hyperactivity/inattention, peer relationship problems. In NTR we applied the same analysis pipeline using the mother rated Child Behaviour Checklist (Achenbach, Ivanova, & Rescorla, 2017) at similar age intervals as in TEDS (mean age 7, 9,

and 12, vs mean age 7, 10 and 12). The CBCL comprises eight syndrome scales including aggressive behaviour, anxious/depressed, attention problems, rule-breaking behaviour, somatic complaints, social problems, thought problems, and withdrawn/depressed. For consistency we use equivalent scales between the SDQ (available in TEDS) and the CBCL (available in NTR), i.e. conduct problems, hyperactivity/inattention, emotional problems, and peer problems in SDQ, and externalizing, inattention, internalizing and social problems in CBCL. These are henceforth called CND, HYP, EMO and PER for both TEDS and NTR, with subscript 1, 2, or 3 to indicate the different time points (age 7, 9-10, and 12 respectively). SDQ and CBCL scales have been shown to correlate highly with each other (r =.59 to .84) in childhood (Goodman & Scott, 1999) and to be highly comparable across cohorts. For example, scales from the SDQ and CBCL show very similar patterns of co-occurrence in TEDS and NTR with aggression (Bartels et al., 2018) and similar genetic architecture (Porsch et al., 2016).

As age ranges partly overlapped across time lags in both cohorts, first, we excluded individuals with overlapping information across age ranges, to create non-overlapping age bins. Then, we derived z-standardised residuals for each SDQ and CBCL subscale regressed on sex and age (at each age bin). We then used standardized residuals in structural equation models. Supplementary Tables S1a and S1b report descriptive statistics of all measures used in the study by cohort.

Analyses

We modelled psychopathology measures longitudinally fitting random intercept cross-lagged panel models (RI-CLPM; Hamaker et al., 2015; Mund & Nestler, 2019) to test for the presence of within-person directional influences of psychopathological traits over time. The RI-CLPM (Figure 1, left and right models) is an alternative to the widely used Cross-Lagged panel model (CLPM). A detailed critique of the CLPM is available elsewhere (Hamaker et al., 2015). Briefly, the CLPM cannot disentangle within-person effects over time from between-person stable effects. However, the inclusion of a random intercept capturing between-person individual differences over time allows for this separation. There are several generalizations and alternatives to the RI-CLPM described elsewhere (Epskamp, 2020; Mund & Nestler, 2019; Zyphur et al., 2019).

Figure 1 is a schematic representation of a RI-CLPM for two traits and three measurements. The random intercepts (ω or κ) represent the stable between-person trait like influences across time, while the fixed intercepts estimate the group mean level at each measurement for

a trait. The individual level deviation at each measurement is estimated by the latent variables p_{it} and q_{it}, while α_2 and α_3 (and δ_2, δ_3) estimate the within-person carry-over effects between measurement x_{it} and measurement x_{it+1}. Finally, the cross-lagged paths, β_2 and β_3 (and γ_2, γ_3) estimate the effects of the within-person deviation over the expected mean (μ_t or π_t) for a trait at one measurement, on the deviation from the expected mean for another trait at another measurement. These latter effects are reciprocal directed influences between traits from one time point to another, over and above (controlled for) group-level mean differences, between-person trait-level stability over time, and within-person carry-over effects. In other words, any effect found in this regard is indicative of processes acting at the individual level, indexing the extent to which a person's unusually high (or low) levels for a trait at one time point are predictive of the person's unusually high (or low) levels for another trait later in time.

We present results for TEDS followed by replication of analyses in NTR. For both TEDS and NTR we conducted the same procedures as described below. First, we fit a model in which we constrain variances, autoregressive, and cross-lagged paths to be the same across time and compare it to an unconstrained model in which we let these parameters vary freely overtime. An unconstrained model may be warranted as our observations are taken further apart in time. In this case we might expect to be measuring a latent overarching trait rather than simply random intercepts (Hamaker et al., 2015), such as a common (heritable) predisposition underlying psychopathological stability over time (e.g. the p factor).

The best fitting model is then carried forward in analyses. Within this model we specifically test whether the longitudinal (cross-lagged) paths (β_2, β_3 and γ_2, γ_3) in the example diagram in Figure 1) are significantly different from 0. We remove paths for which we do not find evidence against the null-hypothesis of no effect (using the Benjamini-Hocberg FDR procedure for multiple testing) (Cribbie, 2007), and refit the model. Finally, we fit a 'null' model with all (cross-lagged) longitudinal paths removed. We then evaluate the difference in model fit between these nested models. To study sex specific developmental processes, we perform multi-group analyses by sex, using parameter constraints to test whether regressions differ between males and females. For all analyses we used a Maximum Likelihood estimator with robust standard errors (MLR), and full information maximum likelihood to treat missing data (FIML). Analyses were performed in Rstudio (v1.2.1335), structural models were specified using Lavaan (v0.6-5). See https://osf.io/dtvc8/?view_only=fe8e8eff7df1417da399a05625a5df8e for the preregistered protocol.

Within family RI-CLPM

In analyses that were not pre-registered, we extended the RI-CLPM to family data, by considering siblings pairs instead of unrelated individuals. Specifically, we extended the model to include monozygotic (MZ) and dizygotic (DZ) twin pairs, employing different model specifications depending on zygosity. We call this model within-family RI-CLPM (wfRI-CLPM). Aim of this extension is to further separate reciprocal directional influences between siblings (within families) from similarities between siblings that arise through shared (genetic or environmental) influences that exist in a family. Furthermore, this application can be used to parse out genetic and environmental components of variance at the level of time-invariant overarching stable traits, and age specific effects. In practice this extension takes the network model from an individual to a family level, while controlling for the fact that family members are related to each other.

First, we run a multi-group RI-CLPM on MZ and DZ twins. We fix random intercepts, variances, covariances, and regressions within individuals to be equal across zygosity and for twin 1 and twin 2. Conversely, we let between person covariances of random intercepts and latent factors to vary between zygosity groups (MZ vs DZ). Finally, between-sibling regressions are constrained to be equal across zygosity.

Similar to the RI-CLPM, the measurement model can be expressed as follows for Twin 1 (i) and Twin 2 (j):

$$x_{it} = \mu_t + \kappa_i + p_{it} \qquad (1)$$

$$y_{it} = \pi_t + \omega_i + q_{it}$$

$$x_{jt} = \mu_t + \kappa_j + p_{jt}$$

$$y_{jt} = \pi_t + \omega_j + q_{jt}$$

Where $\mu \; and \; \pi$ are the group means at measurement t for trait x_t and trait y_t, κ and ω are the between-person latent factors (random intercepts) for the two traits respectively measured over time. p_t and q_t are the deviations from a person expected score (i.e. $\kappa + \mu$). In the wfRI-CLPM the within-person variance for the random intercepts is the same for both MZ and DZ, and twin 1 (i), twin 2 (j):

$$\text{var}(\omega_i) = \text{var}(\omega_j) \qquad (2)$$

$\text{var}(\kappa_i) = \text{var}(\kappa_j)$

While covariance between two members of a twin pair is freely estimated across zygosity groups:

$\text{cov}_{MZ}(\omega_i, \omega_j)$ vs $\text{cov}_{DZ}(\omega_i, \omega_j)$ (3)

$\text{cov}_{MZ}(\omega_i, \kappa_j)$ vs $\text{cov}_{DZ}(\omega_i, \kappa_j)$

$\text{cov}_{MZ}(p_{it}, p_{jt})$ vs $\text{cov}_{DZ}(p_{it}, p_{jt})$

$\text{cov}_{MZ}(p_{it}, q_{jt})$ vs $\text{cov}_{DZ}(p_{it}, q_{jt})$

Expanding on the equation for change of the RICLPM using the original notation from (Hamaker et al., 2015) the model for the longitudinal deviations can be expressed as follows for a given MZ or DZ pair (example for two traits *p* and *q*):

Trait *p*:

$$p_{it} = \alpha_t p_{i,t-1} + \beta_t q_{i,t-1} + \boldsymbol{\alpha_t p_{ji,t-1} + \beta_t q_{ji,t-1}} + \upsilon_{it} \quad (4)$$

$$p_{jt} = \alpha_t p_{j,t-1} + \beta_t q_{j,t-1} + \boldsymbol{\alpha_t p_{ij,t-1} + \beta_t q_{ij,t-1}} + \upsilon_{jt}$$

Trait q:

$$q_{it} = \delta_t q_{i,t-1} + \gamma_t p_{i,t-1} + \boldsymbol{\delta_t q_{ji,t-1} + \gamma_t p_{ji,t-1}} + \nu_{it} \quad (5)$$

$$q_{jt} = \delta_t q_{j,t-1} + \gamma_t p_{j,t-1} + \boldsymbol{\delta_t q_{ij,t-1} + \gamma_t p_{ij,t-1}} + \nu_{jt}$$

The first part of the equation $p_{it} = \alpha_t p_{i,t-1} + \beta_t q_{i,t-1} + \upsilon_{it}$ is akin to the equation for the time deviations of RI-CLPM in (Hamaker et al., 2015). Where $\alpha_t p_{i,t-1}$ and $\beta_t q_{i,t-1}$ are the within-person regressions within-trait and cross-trait (respectively) for *trait p* in twin 1 (subscript i), while $\delta_t q_{j,t-1}$ and $\gamma_t p_{j,t-1}$ are the within-person regressions within-trait and cross-trait (respectively) for *trait q* in twin 2 (subscript j). υ_{it} is the residual trait variation for *trait p* in twin 1 and ν_{jt} is the residual trait variance of *trait q* in twin 2. The second part of the equation (bold) differs in the following ways: $\alpha_t p_{ji,t-1}$ and $\beta_t q_{ji,t-1}$ are the between-sibling regressions within-trait and cross-trait (respectively) of twin 1 on twin 2 (subscript ji), for trait *p*. While $\delta_t q_{ij,t-1}$ and $\gamma_t p_{ij,t-1}$ are the between-sibling regressions within-trait and cross-trait (respectively) of twin 2 on twin 1 (subscript ij), for trait *q*. Note that in the model specification $\alpha_t p_{ji,t-1} =$

$\alpha_t p_{ij,t-1}$ and $\gamma_t p_{ij,t-1} = \gamma_t p_{ji,t-1}$ (similarly $\delta_t q_{i,t-1} = \delta_t q_{j,t-1}$ and $\beta_t q_{ji,t-1} = \beta_t q_{ij,t-1}$) within and across zygosity. In this context α_t and δ_t represents the between sibling effects within-trait and γ_t and β_t index between sibling effects cross-trait, after accounting for within-person changes from one time point to the next, group mean level at each time point, and between person differences overtime. Here the main interest is in both within-trait cross-twin and between-trait cross-twin effects (conversely in the standard RICLPM the main interest is in cross-lagged effects). In Appendix (see **Appendix 4**) we report model specification of the within-family RI-CLPM. Figure 1 is a schematic depiction of the wfRI-CLPM for two traits and three measurements in a sib pair.

For the wf-RI-CLPM we take a three-step procedure akin to the RI-CLPM analyses in singletons. First, we fit a full (unconstrained) model estimating all within-person within and cross-trait effects, and all between-sibling within and cross-trait effects. We then formally compare this model (Model 1) with nested models: 1) a pruned model in which we drop all non-significant between-sibling paths (using a nominal significance threshold of $\alpha < 0.05$); 2) and a 'null' model in which we drop all between-sibling paths. We use the "qgraph" R package (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012) to plot associations between random intercepts (between-person networks), and regression estimates within-person (within-person networks) and between networks within families (sibling network).

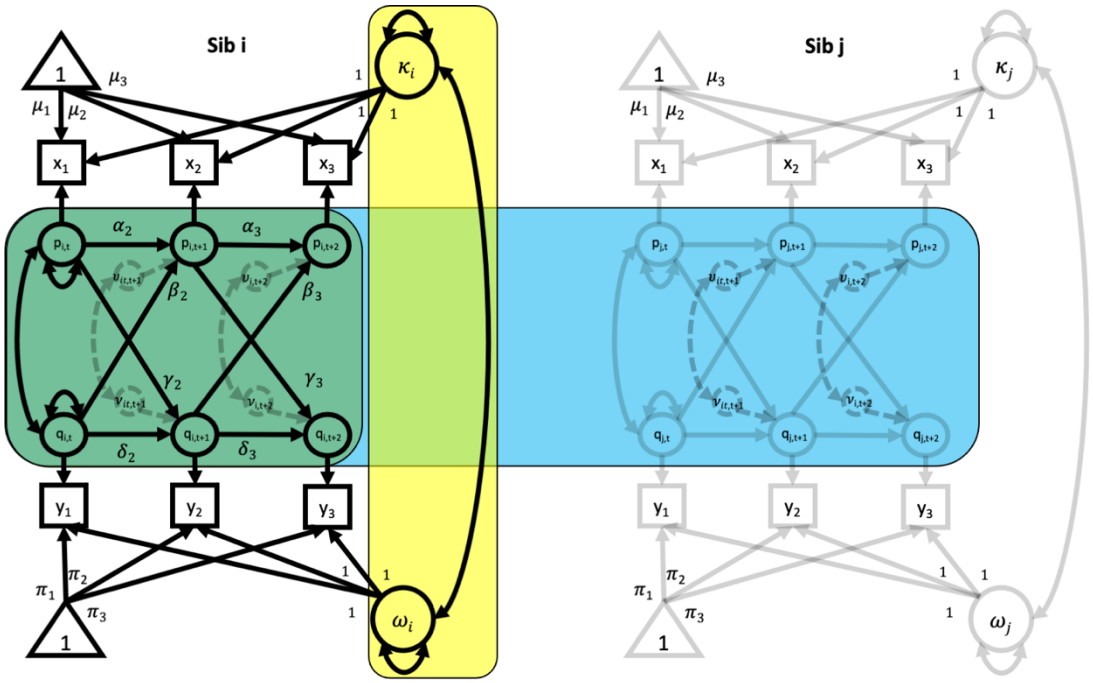


Figure 1. Schematic depiction of the random-intercept cross-lagged panel model extended to sibling pairs (i.e. within-family; wfRI-CLPM) for two traits and three measurements. The green and yellow shades respectively outline within-person and between-person effects (for sibling 'i') modelled in the RI-CLPM. The blue shade outlines between-sibling (i and j) effects (regressions overtime) within a family modelled in the wfRI-CLPM, on top of within-person and between-person effects, and after accounting for similarities due to family influences. **Note.** This model has depiction purposes only, actual models tested included 4 traits and 3 measurement occasions for both the singletons model, and the within-family model including regressions of one sibling's deviations over the other's across time lags.

Genetic and environmental variance components

Every twin phenotypic value at a given time point is a function of the person genetic contributions at that time (h^2, additive or dominance effects), shared environmental (c^2) and unique environmental (plus error; e^2) variance. In the wfRI-CLPM we further separate genetic and environmental effects ($h^2 + c^2 + e^2$) of the stable between-person component, and time specific ($h^2_t + c^2_t + e^2_t$) effects, which can be parsed out into genetic and environmental component of variance on top of within-person effects from one time point to the next (independent of between-person effects and time specific cross-twin cross-trait covariance):

For latent phenotype 'p' of person *i* at time point *t*:

$$p_{it} = h^2_t + c^2_t + e^2_t + \alpha_t p_{i,t-1} + \beta_t q_{i,t-1} + \alpha_t p_{ji,t-1} + \beta_t q_{ji,t-1} \qquad (6)$$

that is, the phenotypic value of each twin at a certain time point is given by the within-person effects and the sibling effects from the previous time point (within and cross trait), and an age specific residual that is separated in shared genetic, shared environmental, and unique environmental influences, after controlling for between-person stable effects overtime. Modelling the between-person covariances separately for MZ and DZ twins as a multi-group analysis, we can thus estimate genetic and environmental relative contributions to the variation in latent stable traits (the random intercepts) and age specific effects (the time specific residuals for each trait). Based on Falconer's formula, we can estimate genetic (h^2 and d^2) and environmental (c^2 and e^2) contributions by comparing MZ twins correlations (rMZ) to DZ twins correlations (rDZ) as follows:

$$h^2 = 2*(r_{MZ} - r_{DZ}) \quad (7)$$

$$c^2 = 2*r_{DZ} - r_{MZ} \quad (8)$$

For traits where non additive effects are evident (negative c^2 represented by r_{mz} being more than twice r_{DZ}) we estimate additive vs dominance effects as follows:

$$a^2 = 4*r_{DZ} - r_{MZ} \quad (9)$$

$$d^2 = 2*r_{MZ} - 4*r_{DZ} \quad (10)$$

Unique environmental contributions (+ error) can be estimated as follows:

$$e^2 = 1 = r_{MZ} \quad (11)$$

An ADE or ACE model can be specified a priori upon inspection of the covariance structure for each trait.

Results

Supplementary Tables S2a and S2b show bivariate correlations between all study variables. In TEDS, both the full constrained and unconstrained models showed an excellent fit (table S3a). A chi-square test showed that the unconstrained model was favoured over the constrained model: $\Delta\chi^2(26) = 32.688$, $p = 0.170$. However, CFI and RMSEA favoured the unconstrained model. Upon inspection of cross-lagged regressions it was evident that the pattern of relationships differed between the two time lags (age 7-9 and age 9-12; see below) indicating a developmental change in the within-person process overtime. As such we carried forward

the unconstrained model: χ^2 (6) = 10.550, p = .103, RMSEA= 0.009, SRMR = 0.004, CFI = 0.999. In NTR the full unconstrained model had an excellent fit too and was favoured over the constrained model based on the chi-square test $\Delta\chi^2$ (26) = 64.238, p = 4.403E-5 and standard fit indices (table S3c and S3d): χ^2 (6) = 10.550, p = .103, RMSEA= 0.009, SRMR = 0.004, CFI = 0.999.

Between-person stable effects

In TEDS, between-person individual differences as indexed by the random intercepts accounted for a substantial proportion the variation of the constructs under study over time (CND = 48%, HYP = 56%, EMP = 45%, PER = 42%). There were positive correlations (ranging from 0.36 to 0.60, see between-person network in Figure 2) between all traits indicating that higher rating for a particular child problem behaviour across time also tended to be higher for other problem behaviours across the three measurements waves. These between-person correlations were twice the magnitude within-person correlations at any given time point (Table S4a).

Consistent with TEDS findings, in NTR we observed substantial between-person effects across time for all traits (CND = 57%, HYP = 54%, EMO = 44%, PER = 43%) and strong positive correlations (Figure 2). Again, these were more than twice in extent the within-person simultaneous correlations at any given time (Table S4b). We then considered evidence for within-person associations.

Within-person time-varying effects

We detected several positive directed effects between problem behaviours across time that were significant after FDR correction, indicating the extent to which deviations from a person expected score in one problem behaviour at one time point (say age 7) predicted deviations in the person's other problem behaviour at a subsequent time point (say age 9), after accounting for stable between-person differences and time-varying carry-over effects. Table S5a and S5b report coefficients for all regression in TEDS and NTR, Figure 2 shows network plots of directed within-person relationships.

Within-person networks evidenced a reciprocal pattern of relationships between several dimensions over time. Of note, for example, were reciprocal effects of conduct and hyperactivity/inattention at both time lags in TEDS and between emotional and peer problems at both time lags in NTR. In both cases directed effects were of the same magnitude ($\beta \sim 0.1$)

indicating a positive loop overtime, rather than causal predominance of one variable over the other across time.

The pattern of relationship emerging in the network plots partly overlapped between TEDS and NTR, specifically conduct problems were predictive of emotional problems (β = 0.140, se = 0.037, p = 4.486E-5; and β = 0.100, se = 0.042, p = 5.39E-3) and hyperreactivity (β = 0.099, se = 0.034, p = 4.465E-3; and β = 0.089, se = 0.037, p = 1.14E-2) from age 7 to 9 (age 7 to 10 in NTR); while peer problems predicted emotional problems (β = 0.170, se = 0.034, p = 6.000E-4; and β = 0.080, se = 0.032, p = 1.26E-2) from age 7 to 9-10; in turn emotional problems were predictive of peer problems (β = 0.115, se = 0.040, p = 3.70E-3; and β = 0.097, se = 0.033, p = 3.17E-3) from age 9-10 to age 12.

Within-person directed effects accounted for 4% to 11% of the variance in TEDS and 10% to 18% of the variance in NTR in the full models, depending on measure and wave. Nested model comparisons between the full model, a pruned model in which only FDR adjusted parameters (as shown in Figure 2 and S5 tables) were included in the model, and a null model in which all cross-lagged paths were dropped, favoured the pruned model in TEDS and the full unconstrained model in NTR (Tables S3b, S3d).

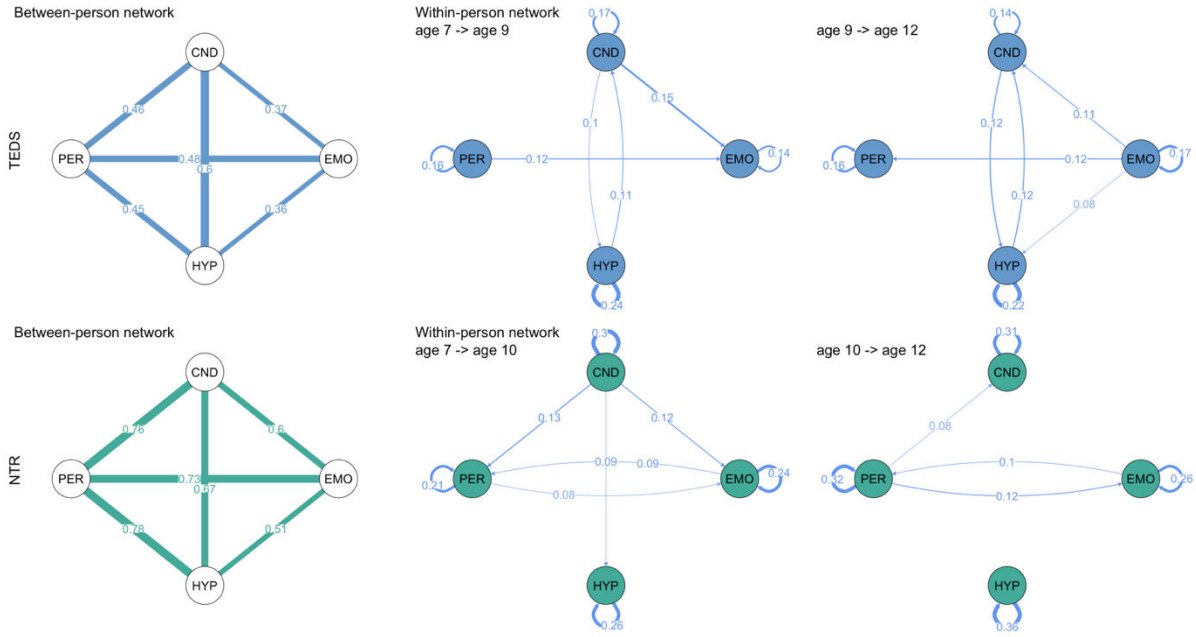


Figure 2. Between-person and within-person (directed) networks of relationships in TEDS (blue) and NTR (green) obtained from the RI-CLPM. Nodes represent the measure of interest (the random intercept in the case of between-person networks, and residual deviation of measurement occasion for the within-person network). Edges width and labels indicate and quantify the strength of relationships between nodes, and in the case of within-person networks also the temporal direction of the effect. For every time lag (7->9 and 9->12 for TEDS; 7-10 and 10->12 for NTR) edges represent directional effects within-trait (self-pointing arrows; αs and δs in Schematic figure 1) or cross-trait (βs and γs in Schematic Figure 1). **Note.** TEDS/NTR acronyms: CND = conduct/externalizing, HYP = hyperactivity/ hyperactivity-inattention, EMO = emotional problems/internalizing, PER = peer problems/social problems. All edges survived FDR correction for multiple testing.

Sex differences

While in TEDS we did not find evidence for sex differences as indicated by multigroup comparisons, in NTR developmental differences between groups were evident (Tables S6a-S6d). On one hand, in males, autoregressive effects were stronger for emotional and conduct problems than in females. The network of cross-lag relationship involved effects of conduct problems on emotional problems across time lags (age 7-10 and 10-12) and on hyperactivity from age 10 to 12. In turn emotional problems and peer problems were reciprocally predictive form age 10 to 12. On the other hand, in females we find only weaker evidence (nominal significance) for within-person effects, which suggested a role of conduct problems only in the first time lag (on peer problems and hyperactivity), with a more predominant role of peer problems later in life (on conduct and emotional problems; Tables 7a-7b and Figure S1).

Within family extension of the RI-CLPM

Both in TEDS and NTR the full unconstrained models including sibling effects had an excellent fit excellent fit: $\chi^2(428) = 848.883$, p < 1E-8, RMSEA= 0.015, SRMR = 0.033, CFI = 0.992 and $\chi^2(428) = 848.883$, p < 1E-8, RMSEA= 0.012, SRMR = 0.031, CFI = 0.997 respectively (Table S9a/S9c). Tables S9a to S9d show correlations for observed variables by zygosity. Figure S2 shows MZ vs DZ twin correlations for random intercepts and age specific residual variances for TEDS and NTR.

Overall, these were consistent with an additive model of genetic variance, and little shared environmental effects. For inattention/hyperreactivity, genetic dominance effects were evident with the exception of age 9 and age 12 in TEDS. In TEDS for the HYP random intercept we found evidence of overdominance (*d2* effects exceeding 1) possibly due to a mix of true dominance effects and contrast effects. Figure 3 depicts variance component estimates for latent stable traits, and age specific residual variances. The pattern of variance components estimates was not always consistent between TEDS and NTR. Inspection of the observed covariance matrix for DZs and MZs twins (Tables S9a to S9d; figure S3 for twin correlations for random intercepts and age specific residuals) showed that the overdominance evident in TEDS was underlay by DZs correlations of ~0 for hyperactivity/inattention at age 7. On the other hand, the differing ADE vs ACE pattern for HYP at ages 9 and 12 for NTR vs TEDS might be attributable to differences in measures employed (see discussion).

Figure 3. Variance components for random intercepts and age specific residual variances estimated from MZ and DZ correlations, corresponding to shared additive (a2), dominance (d2), environmental effects (c2), and unique environmental effects (e2). Note the over-dominance for hyperactivity is consistent with DZ correlations < .25* the MZ correlation.

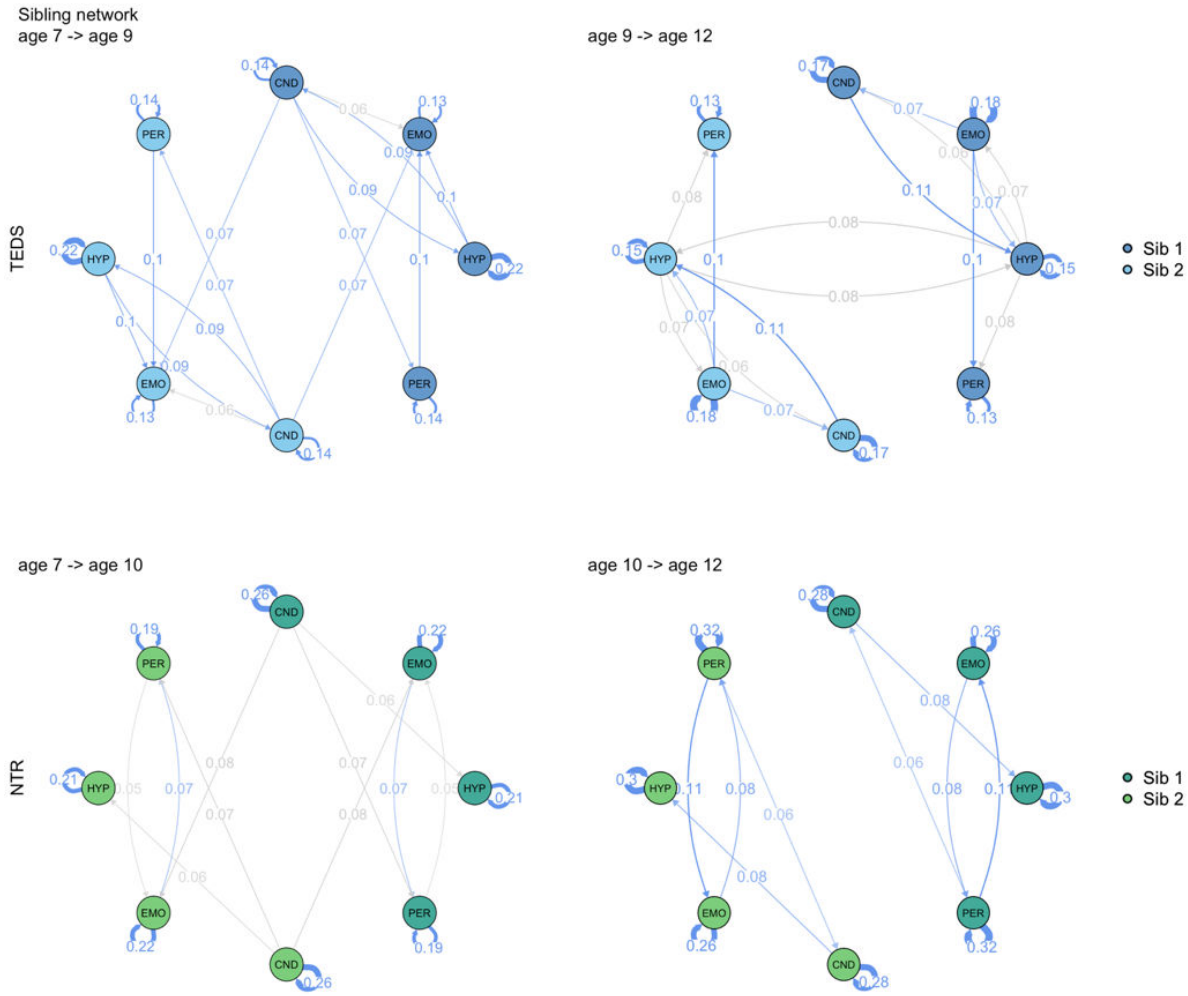


Figure 4. Sibling network plots in TEDS (blue) and NTR (green) obtained from the w-fRI-CLPM. Figure shows within-person networks for sibling pairs within-families (light and dark blue for TEDS; light and dark green for NTR), with between-sibling relationship represented by directed edges form one within-person network to the other. Blue edges survived FDR correction for multiple testing, while gray edges correspond to nominal significance $\alpha < 0.05$. **Note.** TEDS/NTR acronyms: CND = conduct/externalizing, HYP = hyperactivity/ hyperactivity-inattention, EMO = emotional problems/internalizing, PER = peer problems/social problems. All edges survived FDR correction for multiple testing.

Within family RI-CLPM: sibling effects

Nested comparisons indicated that the pruned model where only significant between-sibling regression paths were retained was to be favoured (Tables S9b/S9d). Figure 4 shows siblings network plots where the concept of Figure 2 is extended to include regressions from one within-sibling network to another's, within a family (Tables S10a/S10b).

In both TEDS and NTR we detected a positive cross-trait between-sibling effect, at the first time lag, of conduct on emotional problems (β = 0.066, se = 0.028, p = 8.190E-3 for TEDS,

and β = 0.064, se = 0.03, p = 1.180E-2 for NTR). This indicated that unexpectedly high conduct problems for one sibling at age 7 longitudinally predicted unexpectedly high emotional problems for their sibling at a later time point (age 9 for TEDS and age 10 for NTR), after controlling for within-person time-varying effects and between-person individual differences, such as stable genetic or environmental confounds. We further detected a within-trait between-sibling effects of hyperactivity on subsequent hyperactivity in TEDS (β = 0.084, se = 0.038, p = 2.800E-2), this effect was however not replicated in NTR (Table S10a/S10b for all regression estimates). In total, sibling effects accounted for less than 3% of the variance in psychopathology measures on top of within-person effects at age 9 and 12 in both TEDS and NTR.

Discussion

In the present study we investigated directional relationships between problem behaviours in two population based twin cohorts (TEDS and NTR), separating within- vs between-person effects over time. We found that modelling cross-trait relationships over time provides the best fit to the data, indicating that within-person effects are an important source of covariation between problem behaviours in childhood. These state-like relationships partly account for the observed correlation between psychopathology dimensions over time after controlling for between-person effects. This finding provides evidence towards the hypothesis that the positive intercorrelation between psychiatric traits arises as a function of both common underlying predispositions of a trait-like nature, accounting for between-person individual differences across time, and an underlying network of state-like directional effects between psychopathology dimensions.

Several within-person directed relationships replicated across cohorts. First, we found directed within-person relationships of conduct problems age 7 on emotional problems at age 9-10 in both TEDS and NTR. This is consistent with recent evidence (Hannigan et al., 2018) pointing to shared genetic causes accounting for child co-development of conduct and emotional problems trajectories over time. This in turn suggests that part of the reason these traits co-develop are within-person state-like directed effects of conduct on emotional problems over time. Peer problems at age 7 were also predictive of emotional problems at age 9-10. Problematic relationships with peers in early childhood have been known to precipitate emotional and behavioural problems later in childhood (Menting, Koot, & van Lier, 2015; van Lier & Koot, 2010). In turn emotional problems predicted peer problems from age 9-10 to age 12 within-person, indicating a positive feedback over time. Finally, within-person

associations of conduct problems on hyperactivity from age 7 to age 9-10 were evident in both cohorts

We also found several cohort specific effects. For example, in TEDS we found a reciprocal relationship between conduct and hyperactivity from age 9 to 12, consistent with previous findings (Thapar, Harrington, & McGuffin, 2001; Waschbusch, 2002), but with no evidence of causal predominance in either direction. Findings in NTR on the other hand indicated reciprocal relationships of emotional problems and peer problems across time, again with no evidence for causal predominance. In NTR sex differences were also evident: the reciprocal influence between emotional problems and peer problems from age 10 to 12 only hold for males, while for females there was only (weaker) evidence for a directed relationship of peer problems on emotional problems. Finally, perhaps the most salient difference was that conduct problems seemed to play a more prominent role in males than in females, predicting all traits across time, while in females we only find weak evidence for directed within-person influences of conduct problems from age 7 to age 10.

We further modelled sibling interactions to account for within family reciprocal relationships between siblings over time. Overall, we found that accounting for such relationships within-trait and cross-trait provided the best fit to the data, suggesting that reciprocal relationships between siblings should be take into account in developmental models of comorbidity. These longitudinal between-sibling effects from one time point to the next were detected on top of stable and age specific genetic and environmental effects that make children more similar (or different) within families. The effects detected were in a positive direction indicating that one twin behaviour reinforced the other twin's behaviour or the perception of parents thereof.

Sibling interactions can be separated into cooperative and contrast effects, indicated in our models by positive or negative predictions of one twin behaviour at one time point on the other twin behaviour at another time point. Within age specific traits, however, sibling interactions would load on either d^2 (dominance effects) or c^2 (shared environmental factors), depending on whether these are contrast or cooperative effects respectively. There is evidence, at least for childhood hyperactivity, that maternal ratings may suffer from rater contrast (Meike Bartels, Boomsma, Hudziak, van Beijsterveldt, & van den Oord, 2007; Simonoff et al., 1998), as such contrast and cooperative effects between siblings are rather an index of parental perception of siblings interacting rather than an index of true behavioural interaction between siblings. Again, these effects will inflate d^2 and c^2 estimates, decreasing or increasing DZ similarities respectively. Contrast effects can also mimic dominance

(Simonoff et al., 1998) and this would partly explain the out of bounds estimates we find in TEDS for the latent time invariant hyperactivity trait. We observe overdominance (d^2 above 1) in TEDS for hyperactivity which is likely a combination of true dominance effects (that we also observe in NTR for this trait) and competition effects (which might in part be inflated by parent bias).

We did not attempt to formally distinguish between contrast and cooperative effects when separating genetic and environmental influences at each age. However, no dominance effects were observed for emotional problems or conduct problems, which were the two problem behaviours involved in the between sibling effects that replicated across cohorts. In this regard, if rater contrast accounted for the underlying pattern of relationships, we should have observed negative directed relationships between siblings. That is, the more one twin is rated high on a construct the less the other twin will be perceived to be problematic and thus rated accordingly. However, this effect is in the opposite direction than what we generally observed, with higher ratings on one trait corresponding to higher ratings on the other trait. We could have the reverse where globally we perceive siblings to have same levels of problem behaviours, this would inflate c^2 estimates as observed elsewhere (Allegrini et al., 2019). However, we do not observe consistent c^2 estimates for any of the replicated findings across cohorts, both within and between sibling effects.

The approach employed in the current study can be extended to other within family designs to help separating relative familial contributions to trait variation. Evidence in this regard can in turn reinforce our confidence in a putative causal relationship. For example, it is intriguing that the effects of conduct/externalizing on emotional problems/internalizing we observed is not only a function of within-person effects, but it is partly accounted for by between sibling relationships. Further triangulation is however warranted before any conclusion can be drawn. Another exciting avenue for future studies is to extend this model further by the incorporation of polygenic scores. This may allow extending already existing approaches to estimate gene-environment interplay and causality in family-based designs (Dolan, Huijskens, Minica, Neale, & Boomsma, 2019; Minică, Dolan, Boomsma, de Geus, & Neale, 2018). In principle implementing the current approach to extended family-based designs, may help separating measured direct and indirect genetic effects of siblings, on top of parent-child relationships.

We should note that similar methods have been developed that leverage longitudinal sibling based designs to infer gene-environment covariance via phenotypic transmission (De Kort et

al., 2014; Dolan et al., 2014). Alternative methods also exist that consider difference scoring instead of sibling relationships within a cross-lagged framework (Moscati, Verhulst, McKee, Silberg, & Eaves, 2018; Ritchie, Bates, & Plomin, 2015). However, out method emphasizes between siblings phenotypic effects, rather than gene-environment interplay, and could be further extended to include parental effects as well as polygenic scores effects.

It is important to highlight that the between-person genetic effects captured by the current model specification are of a stable nature. However, if developmental genetic changes were present these will likely be pushed into within-person effects affecting directed relationships within-trait (autoregressive effects) and cross-trait (cross-lag effects). Including random-slopes effects in this model might shed light on whether this is the case. However, this model would necessitate of at least four measurement occasions across time to be specified. In addition, siblings directed effects as estimated in the within-family extension of the RICLPM, could also capture such genetic developmental effects, although in a diluted form as DZ sibling are only 50% genetically similar on average.

A few other limitations should be highlighted. First, the use of different scales in TEDS and NTR might have influenced our results. For example, although as noted earlier SDQ and CBCL scales have been found to be highly comparable across cohorts (Meike Bartels et al., 2018; Goodman & Scott, 1999; Porsch et al., 2016), differences between constructs might still account for some of the cohort specific effects we observed in our study. A second limitation is that our measurements were taken relatively far apart in time. Arguably closer in time measurement are needed to more robustly detect state-like effects as those presently investigated. By the same token however it is all the more surprising that we find that such effects hold across time, and partly replicate across cohorts, suggesting that we are tapping into something worth investigating further. Application of the wfRI-CLPM to panel data with measurements closer in time might yield even more interesting results. A last limitation is that while we talk about directed influences, these need not be causal influences, there might still be third variable confounding at play. In this regard triangulation with other genetically sensitive approaches, and integration with extended family-based designs might help increasing confidence in our findings.

Notwithstanding these limitations, to our knowledge the present study is the first systematically investigating directional within-person and between-sibling effects between psychopathology dimensions across childhood. In conclusion, our analyses provided

substantive results on trait associations between behavioural problems within an open science framework replicating our analyses in two well powered longitudinal child cohorts. We found that correlations between psychopathology dimensions are partly attributable to a network of direct within-person relationships between traits acting against a background of between-person trait-like differences. Finally extending this approach to family-level data we provided a framework that can be extended to other within-family, genetically sensitive designs in the future.

References

Achenbach, T. M., Ivanova, M. Y., & Rescorla, L. A. (2017). Empirically based assessment and taxonomy of psychopathology for ages 1½–90+ years: Developmental, multi-informant, and multicultural findings. *Comprehensive Psychiatry, 79*, 4-18.

Akingbuwa, W. A., Hammerschlag, A. R., Jami, E. S., Allegrini, A. G., Karhunen, V., Sallis, H., ... Middledorp, C. M. (2020). Genetic Associations Between Childhood Psychopathology and Adult Depression and Associated Traits in 42 998 Individuals: A Meta-Analysis. *JAMA psychiatry*.

Allegrini, A. G., Cheesman, R., Rimfeld, K., Selzam, S., Pingault, J.-B., Eley, T., & Plomin, R. (2019). The p factor: Genetic analyses support a general dimension of psychopathology in childhood and adolescence. *bioRxiv*, 591354.

Bartels, M., Boomsma, D. I., Hudziak, J. J., van Beijsterveldt, T. C., & van den Oord, E. J. (2007). Twins and the study of rater (dis) agreement. *Psychological methods, 12*(4), 451.

Bartels, M., Hendriks, A., Mauri, M., Krapohl, E., Whipp, A., Bolhuis, K., . . . Hagenbeek, F. (2018). Childhood aggression and the co-occurrence of behavioural and emotional problems: results across ages 3–16 years from multiple raters in six cohorts in the EU-ACTION project. *European child & adolescent psychiatry, 27*(9), 1105-1121.

Bartels, M., van Beijsterveldt, C. T., Derks, E. M., Stroet, T. M., Polderman, T. J., Hudziak, J. J., & Boomsma, D. I. (2007). Young Netherlands Twin Register (Y-NTR): a longitudinal multiple informant study of problem behaviour. *Twin Research and Human Genetics, 10*(1), 3-11.

Bartels, M., Van den Oord, E., Hudziak, J., Rietveld, M., Van Beijsterveldt, C., & Boomsma, D. (2004). Genetic and environmental mechanisms underlying stability and change in problem behaviours at ages 3, 7, 10, and 12. *Developmental psychology, 40*(5), 852.

Boomsma, D. I. (2014). Sibling interaction effects. *Wiley StatsRef: Statistics Reference Online*, 1-3.

Borg, I. (2018). A note on the positive manifold hypothesis. *Personality and Individual Differences, 134*, 13-15.

Borsboom, D. (2017). A network theory of mental disorders. *World psychiatry, 16*(1), 5-13.

Carey, G. (1986). Sibling imitation and contrast effects. *Behaviour Genetics, 16*(3), 319-341.

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., . . . Poulton, R. (2014). The p factor: one general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science, 2*(2), 119-137.

Caspi, A., & Moffitt, T. E. (2018). All for one and one for all: Mental disorders in one dimension. *American Journal of Psychiatry, 175*(9), 831-844.

Cribbie, R. A. (2007). Multiplicity control in structural equation modeling. *Structural Equation Modeling*, 14(1), 98-112.

De Kort, J. M., Dolan, C. V., Kan, K.-J., Van Beijsterveldt, C. E., Bartels, M., & Boomsma, D. I. (2014). Can GE-covariance originating in phenotype to environment transmission account for the Flynn Effect? *Journal of Intelligence, 2*(3), 82-105.

Dolan, C. V., de Kort, J. M., van Beijsterveldt, T. C., Bartels, M., & Boomsma, D. I. (2014). GE covariance through phenotype to environment transmission: an assessment in longitudinal twin data and application to childhood anxiety. *Behaviour Genetics, 44*(3), 240-253.

Dolan, C. V., Huijskens, R. C., Minica, C. C., Neale, M. C., & Boomsma, D. I. (2019). Incorporating polygenic scores in the twin model to estimate genotype-environment covariance: exploration of statistical power. *bioRxiv*, 702738.

Epskamp, S. (2020). Psychometric network models from time-series and panel data. *psychometrika*, 1-26.

Epskamp, S., Cramer, A. O., Wadorp, L.J. Schmittmann, V.D. & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. Journal of statistical software, 48, 1-18.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry, 38*(5), 581-586.

Goodman, R., & Scott, S. (1999). Comparing the Strengths and Difficulties Questionnaire and the Child Behaviour Checklist: is small beautiful? *Journal of abnormal child psychology, 27*(1), 17-24.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424-438.

Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S. J., Mallard, T. T., Hill, W. D., . . . Deary, I. J. (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature human behaviour, 3*(5), 513.

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological methods, 20*(1), 102.

Hannigan, L. J., Pingault, J.-B., Krapohl, E., McAdams, T. A., Rijsdijk, F. V., & Eley, T. C. (2018). Genetics of co-developing conduct and emotional problems during childhood and adolescence. *Nature human behaviour, 2*(7), 514-521.

Jansen, P. R., Polderman, T. J., Bolhuis, K., van der Ende, J., Jaddoe, V. W., Verhulst, F. C., . . . Tiemeier, H. (2018). Polygenic scores for schizophrenia and educational attainment are associated with behavioural problems in early childhood in the general population. *Journal of Child Psychology and Psychiatry, 59*(1), 39-47.

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of general psychiatry, 62*(6), 617-627.

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., . . . Kendler, K. S. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey. *Archives of general psychiatry, 51*(1), 8-19.

Lahey, B. B., Applegate, B., Hakes, J. K., Zald, D. H., Hariri, A. R., & Rathouz, P. J. (2012). Is there a general factor of prevalent psychopathology during adulthood? *Journal of abnormal psychology, 121*(4), 971.

Ligthart, L., van Beijsterveldt, C. E., Kevenaar, S. T., de Zeeuw, E., van Bergen, E., Bruins, S., . . . Hottenga, J.-J. (2019). The Netherlands Twin Register: Longitudinal Research Based on Twin and Twin-Family Designs. *Twin Research and Human Genetics*, 1-14.

Menting, B., Koot, H., & van Lier, P. (2015). Peer acceptance and the development of emotional and behavioural problems: Results from a preventive intervention study. *International Journal of Behavioural Development, 39*(6), 530-540.

Minică, C. C., Dolan, C. V., Boomsma, D. I., de Geus, E., & Neale, M. C. (2018). Extending causality tests with genetic instruments: An integration of Mendelian randomization with the classical twin design. *Behaviour Genetics, 48*(4), 337-349.

Moscati, A., Verhulst, B., McKee, K., Silberg, J., & Eaves, L. (2018). Cross-lagged analysis of interplay between differential traits in sibling pairs: validation and application to parenting behaviour and ADHD symptomatology. *Behaviour Genetics, 48*(1), 22-33.

Mund, M., & Nestler, S. (2019). Beyond the cross-lagged panel model: Next-generation statistical tools for analyzing interdependencies across the life course. *Advances in Life Course Research, 41*, 100249.

Pettersson, E., Anckarsäter, H., Gillberg, C., & Lichtenstein, P. (2013). Different neurodevelopmental symptoms have a common genetic etiology. *Journal of Child Psychology and Psychiatry, 54*(12), 1356-1365.

Porsch, R. M., Middeldorp, C. M., Cherny, S. S., Krapohl, E., Van Beijsterveldt, C. E., Loukola, A., . . . Rhee, S. (2016). Longitudinal heritability of childhood aggression.

American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 171(5), 697-707.

Rebollo, I., & Boomsma, D. I. (2006). Genetic analysis of anger: genetic dominance or competitive sibling interaction. *Behaviour Genetics, 36*(2), 216-228.

Rietveld, M., Posthuma, D., Dolan, C., & Boomsma, D. (2003). ADHD: sibling interaction or dominance: an evaluation of statistical power. *Behaviour Genetics, 33*(3), 247-255.

Rietveld, M. J., Hudziak, J. J., Bartels, M., Van Beijsterveldt, C., & Boomsma, D. I. (2004). Heritability of attention problems in children: longitudinal results from a study of twins, age 3 to 12. *Journal of Child Psychology and Psychiatry, 45*(3), 577-588.

Riglin, L., Thapar, A. K., Leppert, B., Martin, J., Richards, A., Anney, R., . . . Lahey, B. B. (2018). Using genetics to examine a general liability to childhood psychopathology. *bioRxiv*, 409540.

Rimfeld, K., Malanchini, M., Spargo, T., Spickernell, G., Selzam, S., McMillan, A., . . . Plomin, R. (2019). Twins Early Development Study: A Genetically Sensitive Investigation into Behavioural and Cognitive Development from Infancy to Emerging Adulthood. *Twin Research and Human Genetics*, 1-6.

Ritchie, S. J., Bates, T. C., & Plomin, R. (2015). Does learning to read improve intelligence? A longitudinal multivariate analysis in identical twins from age 7 to 16. *Child development, 86*(1), 23-36.

Simonoff, E., Pickles, A., Hervas, A., Silberg, J., Rutter, M., & Eaves, L. (1998). Genetic influences on childhood hyperactivity: Contrast effects imply parental rating bias, not sibling interaction. *Psychological Medicine, 28*(4), 825-837.

Spearman, C. (1961). " General Intelligence" Objectively Determined and Measured.

Thapar, A., Harrington, R., & McGuffin, P. (2001). Examining the comorbidity of ADHD-related behaviours and conduct problems using a twin study design. *The British Journal of Psychiatry, 179*(3), 224-229.

Thapar, A., Hervas, A., & McGuffin, P. (1995). Childhood hyperactivity scores are highly heritable and show sibling competition effects: Twin study evidence. *Behaviour Genetics, 25*(6), 537-544.

van Bork, R., Epskamp, S., Rhemtulla, M., Borsboom, D., & van der Maas, H. L. (2017). What is the p-factor of psychopathology? Some risks of general factor modeling. *Theory & Psychology, 27*(6), 759-773.

van Lier, P. A., & Koot, H. M. (2010). Developmental cascades of peer relations and symptoms of externalizing and internalizing problems from kindergarten to fourth-grade elementary school. *Development and psychopathology, 22*(3), 569-582.

Waschbusch, D. A. (2002). A meta-analytic examination of comorbid hyperactive-impulsive-attention problems and conduct problems. *Psychological bulletin, 128*(1), 118.

Zyphur, M. J., Allison, P. D., Tay, L., Voelkle, M. C., Preacher, K. J., Zhang, Z., . . . Koval, P. (2019). From data to causes I: Building a general cross-lagged panel model (GCLM). *Organizational Research Methods*, 1094428119847278.

Chapter 6 – Discussion

This thesis explored questions of prediction and statistical inference of complex traits in childhood within a genetics-based multivariate framework. Throughout the thesis I employed different multivariate genetic and genomic approaches applied to the prediction of complex cognitive traits and the co-occurrence of child psychopathology dimensions across development. This chapter summarises salient findings from the previous chapters of the thesis, highlights limitations and outlines possible future research directions.

Summary of key findings, and limitations

With increasing availability of large genome-wide summary statistics accuracy of polygenic scores is increasing. However, while advances in genomic methods afford powerful methods to harness the information from GWAS to predict complex traits, there is no consensus on which polygenic score method performs best. Chapter 1 discussed genomic prediction in the context of child and adolescence cognitive related traits. First, a comparison of the predictive accuracy of three state-of-the-art polygenic score approaches (PRSice, LDpred and Lassosum) employing the most powerful GWAS summary statistics of intelligence (IQ3) and educational attainment (EA3) showed that, at the univariate level, we can now predict up to 9% of the variance in general intelligence and 14% in educational achievement during childhood and adolescence. This polygenic prediction substantially increases from age 12 to 16 and does not differ significantly between boys and girls. A formal comparison of approaches showed that depending on how the correlated information between SNPs across the genome is handled (i.e. linkage disequilibrium; LD) polygenic score prediction can be improved over a more standard (p-value clumping + thresholding; C+T) approach, by discarding less information genome-wide. However, it should be noted that the data QC procedure employed, described in Appendix 1, while allowing to compare methods on a same set of SNPs and reducing computational burden, might have affected accuracy of the C+T approach by limiting the available information to the set of genotyped SNPs.

By extending these models to include information about genetically correlated traits, emerging multi-trait genomic approaches can be used to boost power of GWAS summary statistics and therefore prediction accuracy of polygenic scores. By comparing several multi-trait approaches (Genomic SEM, MTAG, SMTpred) I showed that polygenic score prediction of cognitive related traits can be increased up to 3% of the variance depending on polygenic score approach and target phenotype. Notably, within polygenic score method, multi-trait approaches had comparable performance with respect to predictive power suggesting that

improving accuracy of GWAS SNP effect sizes with a multivariate approach is equally effective regardless of the specific method employed. Finally, the difference between top and bottom polygenic scores deciles for the best model implied a difference of one full standard deviation on average for general intelligence, and two full grades difference for educational achievement.

Importantly, if our ultimate goal is prediction (and not, say, causal or statistical inference of trait associations) more comprehensive models should be considered. For example, improved prediction accuracy could be achieved by integrating polygenic score and measured environmental effects in a multivariable framework. As large biobank data become increasingly available prediction models integrating both genomic based and environmental (risk) factors will become more common. In chapter 3, employing a different multi-trait framework, I consider the joint modelling of several polygenic scores and environments measured early in life as they predict educational achievement (EA) at age 16. By employing a regularization technique, elastic net, I show that the joint predictive performance of several polygenic scores is substantial (R^2 = .18) and that jointly modelling polygenic scores and measured environments significantly improves hold-out set prediction of EA (R^2 = .34) over either nested model (multiple polygenic scores or multiple environmental measures) considered alone.

Furthermore, I considered the role of gene-environment correlation and interaction on hold-out set prediction, showing that two-way gene-environment interactions do not significantly contribute to overall prediction over main effects. These results indicated that if GxE interactions will be of any use in predictive modelling, it will only be by considering higher order interactions (three way or more) at least for EA. It should be noted, however, that this conclusion is conditional on the variables and predictive modelling framework employed. In contrast, gene-environment correlation as measured by the information overlap between the two predictive models was substantial (r = .40). Furthermore, I showed that polygenic scores effects on EA are partly accounted for by their correlation with environmental effects, similarly environmental effects on EA are linked to polygenic scores effects (genetic confounding). However, despite ubiquitous GE correlation jointly modelling multiple polygenic scores and correlated environmental effects significantly improved overall prediction of EA, showing that beyond their correlation, polygenic scores and environmental measures still provide unique information.

An important multivariate theme tackled in chapter 4 and 5 of the thesis regards the mechanisms underlying covariation of psychopathology related traits across development. In chapter 4, I explored the manifestation of a common cause of psychopathology accounting for the covariation of problem behaviours throughout development. First, I showed that this common psychopathology component consistently arises across development and is robust across different raters (parent, self, teacher) and psychopathology measures employed, at the phenotypic level. This common dimension is consistently heritable ~60% and genetically stable across time.

It must be highlighted, however, that the spectrum of traits was not always captured by the common (p) factor to the same degree in each instance. Some traits had low common genetic factor loadings depending on the rater considered, for example teacher rated emotional problems had lower loadings (.17 to .24) on the common factor across all ages. While for the child rated p-factor antisocial behaviour showed lower loading compared to other psychopathology measures (.12 to .25) across time. As noted in chapter 4, this partly reflect differences in the composition of this common latent factor across raters. In addition, substantial specific ACE factors (reported in supplementary table S1, Appendix 3) were evident across models considered especially in regard to teacher and child measures. For example, substantial specific A factor variance (As) was evident across most measures in teacher reports at age 7. Particularly, the most extreme case at age 7 was represented by teacher rated emotional problems, for which specific A contributions were sizeable, As = .48, while virtually no variance in this trait was shared, Ac = .02. Another such example was child rated antisocial behaviour at age 12, with substantial specific A contributions (As = .42), but virtually no shared A contributions (Ac =.02). This suggests the potential for specificity in psychopathology dimensions, perhaps depending on rater and age considered, and future studies might consider testing independent pathway in contrast to common pathway models (information on specific ACE component is not reported in Figure 1 of Chapter 4, supplementary Figure S1 in Appendix 6 reports these estimates in an improved version of that diagram).

Lastly, I demonstrated that childhood phenotypic p (as soon as age 7) is consistently associated with adult polygenic p (constructed from polygenic scores reflecting adult psychiatric disorders). These associations (modest in extent: R^2~.01) are consistent with other reports of polygenic scores for adult psychiatric traits on childhood psychopathology (see below) and suggests a (genetic) link between the two. However, given the modest effect size estimates, conclusions that can be drawn from these specific analyses remain limited,

especially when considered in light of potential confounding factors such as those described in Chapter 3.

One pitfall of longitudinal studies is represented by attrition. This is especially an issue when considering complete cases on many traits concurrently in multivariate analyses such as those carried out in Chapter 4. Specifically, in principal component analyses individuals were dropped listwise. This can lead to participation bias as the subset of individuals included in analyses with all available data might not be representative of the originally intended population (in this case of UK-based youth), in other words this will limit generalizability of findings. There is evidence that TEDS remained fairly representative throughout collection waves with respect to demographics collected at first recruitment (Rimfeld et al., 2019). For the analysed measures investigated in Chapter 4 with respect to all complete cases (i.e. complete cases across all measures considered across age by rater bins) I provide measures of associations with socio economic variables available at recruitment (Table S1 and S2, Appendix 6), showing that overall individuals with complete data were generally coming from a higher socio economic status background, families with higher education and occupational status, compared to dropouts since inclusion. Although, this is to some extent expected, it should be taken into account when considering findings from Chapter 4.

In Chapter 5, I further investigated the idea of developmental co-occurrence of psychopathology. A common heritable cause is only one of the plausible mechanisms underlying covariation between psychopathology traits in childhood (the positive manifold). Other data generating mechanisms can account for the positive intercorrelation between psychopathology dimensions observed across development. In fact, another plausible reason for this positive manifold is that psychopathology dimensions temporally cause each other inducing correlation. That is a causal network might (at least in part) underly psychopathological comorbidity in childhood. In chapter 5, I explored to what extent the longitudinal covariation between child problem behaviours results from between-person stable individual differences, attributable to heritable trait-like stable effects, or from state-like temporal directed influences of one variable over the other. With replication in two large twin cohorts (the Twin Early Development Study and the Netherlands Twin Register) I show that beyond between-person differences, a temporal network of directed effects accounts for up to 15% of trait variation (depending on specific trait) from age 7 to age 12. This provides promising evidence for the existence of a network of causal influences, and once the links between psychopathology dimensions are accurately mapped and further empirically

replicated, interventions along the causal chain may be implemented to eventually ameliorate the outcome of child problem behaviours.

I further developed an extension of this network model to family-level data towards the goal of quantifying direct reciprocal sibling effects over time. Again, with replication in TEDS and NTR, I show that sibling effects are an important component of variation and covariation of psychopathology traits within-family accounting for up to 3% more of the variance on top of within-person effects after controlling for shared genetic and environmental effects that make siblings alike. Burgeoning availability of large family datasets with genotypic data, this model could be extended to include parental effects as well as genomic based predictors in the future (see *Integrating family and genomic methods* section below).

Challenges and prospects for developmental quantitative genetics

Prediction of complex developmental traits using genomic data

As shown in chapter 2 and 3, polygenic scores are becoming a powerful tool for prediction of cognitive relevant traits in childhood. However, in the context of genomic prediction educational attainment is a special phenotype because it is typically available as standard demographic variable in most studies, it's relatively straightforward to measure (years of education), and therefore it is easier to collect very large GWAS samples compared to other behavioural phenotypes. Prediction of behavioural problems in childhood from GWAS of adult psychiatric traits, on the other hand, has a longer way to go. An example is a recent effort to meta-analyse polygenic score associations with problem behaviours in childhood (Akingbuwa et al., 2020), in ~50k individuals where the strongest effects detected for any problem behaviour were predicted by EA3 and fell short of 1% of the variance explained. In this regard the polygenic-p effects on child problem behaviour discussed in chapter 4 are at the upper bound of this association effects.

Thus, at present, polygenic scores have limited value for prediction of child psychopathology. However, a promising avenue for future polygenic score work in this field might lie in identifying strata of individuals at increased predisposition to develop mental health disorders, to then target screening interventions accordingly (Lewis & Vassos, 2020), akin to the medical field where this is already implementable for example in cancer screening or coronary heart disease (Jia et al., 2020; Khera et al., 2018).

The low polygenic score heritability of childhood psychopathology can be in partly attributed to the low SNP heritability of child psychopathology (Cheesman et al., 2017) roughly around 10%. In this regard it has been shown that focusing on variation in common across traits can improve estimation of SNP h^2 (Cheesman et al., 2017). Part of the problem, however, can be ascribed to the fact that power of polygenic scores (R^2) is a function of GWAS sample size (Daetwyler, Villanueva, & Woolliams, 2008; Dudbridge, 2013; Pasaniuc & Price, 2017; Wray, Kemper, Hayes, Goddard, & Visscher, 2019; Wray et al., 2013). The equation that governs polygenic score predictive power can be defined as follows:

(1)

$$R^2 = \frac{h_M^2}{1 + M/(Nh_M^2)}$$

Where M is the effective number of independent SNPs contributing to the trait (this can be approximated to 50,000 for complex behavioural phenotypes; Wray et al., 2013), N is the sample size of the discovery GWAS sample and h^2 is the heritability (hence h_M^2 correspond to the SNP heritability). Thus, as N increases to infinity the denominator becomes 1 and $R^2 = h_M^2$. Only relatively recently have we started accumulating large enough samples for adult psychiatric disorders (e.g. Wray et al., 2018), while, for most traits, child samples efforts are still currently under way (with two notable exceptions, such as autism spectrum disorder, Grove et al., 2019, and ADHD, Demontis et al., 2019). Some studies have accrued samples comparable in magnitude to EA (e.g. Howard et al., 2019). However, this comes at the inevitable expense of the phenotypic definition (also known as shallow, or minimal, phenotyping; Cai et al., 2020), and therefore an increase in heterogeneity at the expense of predictive power (see below). For example there is evidence that lowering heterogeneity improves detection and estimation of variants effect sizes (Manchia et al., 2013) this in turn would increase power of polygenic scores. Summary statistics obtained from GWAS of minimal phenotypic definitions also produce less powerful polygenic scores predictors for specific disorders, compared to more strict phenotypic definitions at equal sample size (Cai et al., 2020).

A solution to this problem might be afforded by large international efforts to accrue child samples large enough to directly conduct GWAS of child phenotypes. While several efforts in this regard are under way, it is typically difficult when pulling many child cohorts together to be able to keep a homogenous phenotypic definition across samples. One example is the

recent GWAS of (broad) aggression (Ip et al., 2019) where notwithstanding the sheer size of the discovery set (~500k) the polygenic score derived from the GWAS only achieved an out-of-sample prediction of .2% and .4% depending on the target set. This might be partly attributable to the low SNP h^2 for the broad aggression phenotype (SNP h^2 = 3.3%; which in turn might be due to measurement error across the many different phenotypic definitions employed by the various datasets), in part to the difference between out-of-sample and discovery set phenotypes (e.g. self-reported retrospective conduct disorder and mother reported childhood aggression). That is, the discrepancy between these phenotypic definitions might compromise polygenic prediction.

In fact, we can quantify heterogeneity in the context of polygenic prediction as the imperfect genetic overlap within trait, between the cohort where we estimate the SNP effect sizes (the discovery set) and the hold-out set, where we test polygenic score performance (de Vlaming et al., 2017). Hold-out sample predictive power for polygenic scores can be written as follows (de Vlaming et al., 2017):

(2)

$$R^2 = \frac{r_g h_M^2 h_M^{2^H}}{h_M^2 + M/N}$$

Where $h_M^{2^H}$ is the hold-out set trait SNP h^2, h_M^2 is the training set SNP h^2, and r_g is the genetic correlation between the two (this equation is also generalizable to cross-trait prediction). When genetic correlation equals unity, and assuming that we are measuring the same trait with equal SNP heritability in the discovery and hold-out set, this equation equals Equation 1. Otherwise R^2 will grow as a function of r_g and the upper limit of R^2 will be given by $r_g h_M^2$ as sample size in the discovery set increases (de Vlaming et al., 2017; Figure 1).
This implies that any deviation from the phenotypic definition between the two cohorts introduce heterogeneity at the expense of out-of-sample prediction, governed by discovery set sample size and r_g between the two traits. Tangentially this is part of the reason why depending on genetic architecture we can achieve higher cross-trait predictive power than within-trait hold-out set prediction (depending on the relative combination of discovery and hold-out set SNP heritability).

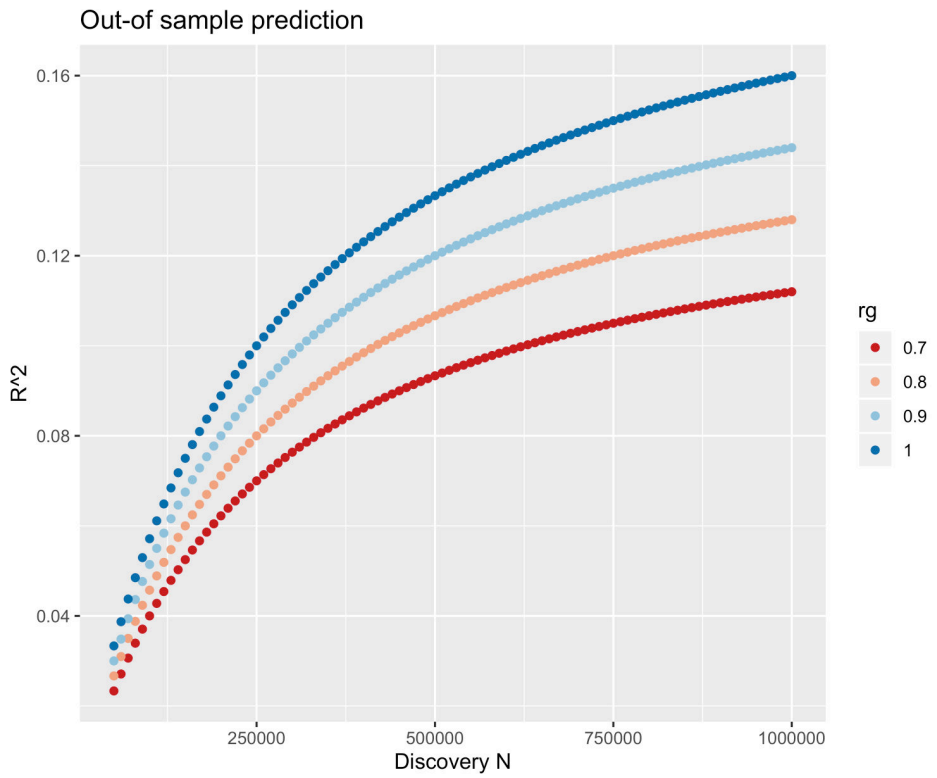


Figure 1. Heterogeneity and out-of-sample prediction: Out-of-sample prediction of polygenic scores as a function of genetic correlations (r_g) between discovery and target set according to Equation 2, assuming both traits have SNP h^2 = .2 and M = 50,000.

This suggest that before we can achieve large GWAS sample sizes of childhood psychopathology traits with homogenous phenotypic definitions across cohorts, multivariate approaches to trait prediction will lead the way in terms of polygenic score prediction. For example, as we have seen in chapter 2 multi-trait genomic approaches can be used to boost power for discovery by jointly analysing correlated traits. Another, perhaps more easily implemented, way is to use a multivariable approach (as seen in chapter 3) where combining multiple polygenic predictors in the same model yielded a powerful hold-out-set prediction.

Improving prediction by solving the still missing heritability problem

Bridging the gap between SNP h^2 and polygenic score heritability will, however, remain an active area of research. One way forward is afforded by whole genome sequence (WGS) data. As WGS data will become cheaper we will start bridging the gap between SNP and narrow sense heritability, as we will start mapping rare variants in low LD regions across the population (Wainschtein et al., 2019). For example, using WGS data recovered almost all the 'lost' heritability for height and BMI. By increasing the theoretical ceiling of polygenic prediction with recovery of the so called still missing heritability (Wainschtein et al., 2019),

WGS data will in turn lead to improve genomic prediction by integrating rare variants in polygenic scores.

However, by the same token, this will also imply a substantial increase of independent variants involved in the trait of interests, according to early estimations approximately a 10-fold increase (~500k, Wray et al., 2019). In turn, this will lead to a lower R^2, despite the increase in estimated SNP heritability, because the amount of noise along with more SNP estimates will increase (Figure 2; de Vlaming et al., 2017; Wray et al., 2019). Future polygenic scores methods will need to tackle this problem by trying to balance more effectively M vs SNP h^2 (Wray et al., 2019; Wray et al., 2013). One way to do that might be incorporating prior information such as genomic annotations to bin variants, and regularization methods that introduce sparsity in the data (akin to Lassosum employed in both chapter 2 and 3; or Bayesian shrinkage such as LDpred; chapter 2). In practice it has been shown that binning variants according to functional annotations to then introduce sparsity via a regularization method improves prediction over existing methods by effectively reducing the number of estimated effect sizes (Marquez-Luna et al., 2020; Wray et al., 2019). Another option is to include information on close relatives in the discovery set. This in practice reduces M by decreasing effective sample size, in turn minimizing the number of parameters to be estimated by having discovery and target samples sharing recent common ancestors (Lee, Weerasinghe, Wray, Goddard, & Van Der Werf, 2017; Truong et al., 2020). The usefulness of this approach increases as large biobanks datasets and registries include both unrelated as well as closely and distantly related individuals, bringing personalized medicine from unrelated samples to a family-type level.

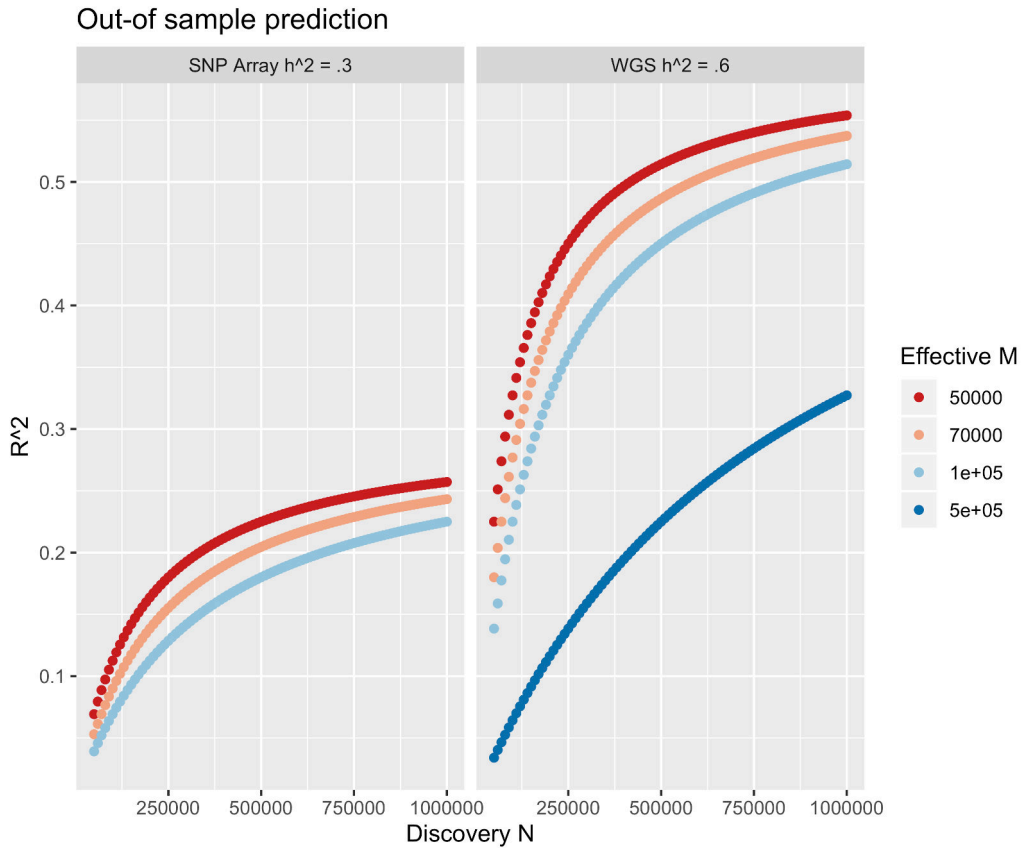


Figure 2. Effective M and out-of-sample prediction: Predictive power of polygenic scores according to Equation 1 as a function of effective M (number of independent SNPs involved in the trait). Left: example for a trait with SNP h^2 = .3 based on genotype array data. Right: example for the same trait with WGS data increasing estimation to SNP h^2 = .6. In blue is the number of independent SNPs from WGS involved in the trait (500k), based on Wray et al., 2019.

This also implies that at low sample sizes, which can be currently obtained until costs can be reduced, the value of WGS for genomic prediction will be limited (Figure 2), that is WGS based prediction will perform similarly to genotyped data for the same trait (with lower h^2), for which much larger samples can be more easily and cheaply obtained. This is further complicated by heterogeneity, in the sense that even when WGS data will become available, an important role will remain collecting high quality data with homogenous definitions across child cohorts. In addition, it remains to be established whether the theoretical expectation for recovery of missing heritability will be the same for (developmental) behavioural phenotypes.

Until we can collect large discovery (and hold-out) WGS samples, multi-trait approaches will be powerful tools to increase predictive capacity of genomic based predictors. As sample size of GWAS grows, joint modelling of multiple polygenic score predictors from genotype (non

WGS) data will remain most effective, straightforward and easily implemented. Notwithstanding these challenges, large collaborations between child (twin) cohorts will be an essential tool to map the genetic architecture of (psychopathology) related complex traits and aiming to collect ever growing sample sizes will remain the most effective way to improve predictive accuracy of polygenic scores, with important implications for downstream analyses (see section below).

Multi-trait genomic analyses of developmental phenotypes

As large international GWAS efforts accumulate one important game changer in the GWAS landscape will be employing results from GWAS of childhood phenotypes using multivariate data from large longitudinal cohorts. With increasingly large collaborative efforts 'stratified' GWAS for several developmental phenotypes are under way. One example is the GWAS of broad aggression mentioned above (Ip et al., 2019), where separate GWAS were run for different strata of rater by age bins (which were then meta-analysed together). Stratified GWAS such as this are an important resource for future studies investigating child development, because the genetic architectures and correlations between GWAS traits thus calculated can then be formally modelled to investigate developmental questions at the genomic level. Several similar efforts are under way for a number of complex traits across childhood. At the time of writing this thesis, TEDS has participated in several such collaborations including cognitive-related and psychopathology traits. This is one very exciting avenue for future research.

For example, in chapter 2 a multi-trait method, Genomic SEM (Grotzinger et al., 2019), was employed to fit a common factor model using GWAS summary statistics for several cognitive-related traits. In turn, a common factor GWAS was run to obtain summary statistics for this common factor. In a similar vein, we could explore other (longitudinal) factor structures using stratified summary statistics for GWASs of childhood traits. Akin to twin methods, several factor structures and competing models can be tested, for example, fitting a Cholesky model to infer genetic stability or innovation of a particular trait over time (similar to what was done in chapter 4 with a phenotypically defined parent-rated p-factor); or fitting a common factor (chapter 4) vs independent pathway model of childhood psychopathology to then run a GWAS of this factor structure to infer the mechanisms through which single variants affect variation and covariation across traits.

Other multivariate approaches exist that employ SEM to model (individual-level) genomic data (St Pourcain et al., 2018; Verhulst, Maes, & Neale, 2017). Multivariate methods that model the (genetic) relationships between traits are an important resource for studying multivariate developmental data and are becoming increasingly important with the availability of large longitudinal family data such as the Norwegian Mother father and Child cohort (MOBA). However, an advantage of methods such as Genomic SEM is the use of summary-level data instead of relying on measured genotypes from individual-level data, thereby achieving powerful samples without issues of data sharing, and in addition being able to study traits at the multivariate level that might be difficult to obtain within the same sample.

Summary statistics generated from different factor structures can be used to construct 'bespoke' polygenic scores, for example capturing common vs trait specific polygenic effects, which in turn could be employed for trait prediction or in SEM models for more nuanced hypothesis tests. For example, an interesting avenue could be to test gene-environment interactions using polygenic scores for trait-specific effects vs common polygenic liability to psychiatric traits. The integration of more powerful and nuanced polygenic scores with family-level data will help illuminate several important developmental questions in the near future. This type of approaches which analyse genomic factor structures will likely lead the way to tackle important questions related to developmental complex traits. By employing powerful modelling techniques widely applied in the twin literature modern quantitative genetics is bringing together new genomic methods and family-level data. As large longitudinal family cohorts invest in genotyping the integration between the two worlds is becoming more common.

Integrating family and genomic methods

The integration between family-level data and genomic methods is an important avenue for future research in child development. Studies that leverage the availability of genotyped cohort with family data such as DZ twin pairs and family trios (parents and offspring) are emerging to investigate questions related to gene-environment correlation and genetic nurture using polygenic scores (Bates et al., 2018; Kong et al., 2018). As discussed in chapter 3 gene-environment correlation refers to the ways in which an individual's genotype covary with the environment; we can typically distinguish three main type of GE correlation: passive, active and evocative. Genetic nurture can be thought of as a special type of GE correlation in which parental genotypes (partly shared with the offspring) influence child phenotype via the rearing environment. In practice this opens a 'backdoor path' from child genotypes to child

phenotype mediated by the environment. In chapter 3 I discussed how we can think about GE correlation within a prediction modelling framework in terms of genetic confounding and environmentally mediated polygenic scores effects on hold-out-set prediction.

As explained, this concept is not novel to the genomic era, but historically it has been investigated by the quantitative genetics literature, with extensive evidence showing that measures of the environment are themselves heritable (Plomin, 2014; Plomin, DeFries, Knopik, & Neiderhiser, 2016). An advantage of genomic-based methods in this regard, however, is that they allow us to directly quantify this environmentally mediated effects using measured genetics. In turn this can help us better understand the mechanisms through which genetic predisposition links to the eventual phenotype and separate so called 'direct' from 'indirect' genetic effects.

For example, emerging methods using family trios infer direct and indirect genetic effects by testing associations of polygenic scores created using transmitted and non-transmitted alleles from parent to offspring (Kong et al., 2018). This phenomenon of environmental mediation, has been shown also at the level of SNP heritability estimation by leveraging family data (Eilertsen et al., 2020; Young et al., 2018) to partition paternal and maternal indirect genetic effects from offspring genetic effects. Data such as MOBA, which includes information of parent and offspring genotypes, can be leveraged to separate 'direct' genetic effects from indirect mediated effects via the rearing environment. However, recent advances in genomic methods will make this type of research generally accessible to (twin) cohorts via imputation of parental genotypes from genotyped sibling pairs (Young et al., 2020; Hwang et al., 2020). These data in turn could be leveraged in large collaborative efforts to reach larger sample sizes.

More broadly, the integration of polygenic scores in family-based (twin) models have been already employed to investigate several important questions related to gene-environment interplay and causality (Dolan, Huijskens, Minica, Neale, & Boomsma, 2019; Minică, Dolan, Boomsma, de Geus, & Neale, 2018). Particularly, it seems clear that there is a lot of potential for developing statistical models to strengthen causal inference with family data. For example, in chapter 5 I discussed an extension of a network approach to sibling pairs that can be used to assess reciprocal sibling effects longitudinally while controlling for shared genetic and environmental effects that make siblings alike. A model such as this could be further extended to include parental effects as well as polygenic scores effects. If phenotypic information from

parents is available a model could be fit to test for reciprocal longitudinal effects between parents and siblings, on top of reciprocal sibling effects and within-person phenotypic effects, after controlling for shared genetic and environmental effects that make people in families alike.

On top of these effects we could look at measured genetic differences, and test how parental polygenic scores relate to child residual variance at each measurement over time. A study design such as this could provide a direct test of genetic nurturing effects, testing for phenotypic mediation of parental (non-shared) polygenic scores effects on child outcomes longitudinally. Of course, this would be a complex model to fit requiring a large sample size to account for subtle effect sizes, and polygenic scores more powerful than the ones that are currently available. This integration will also be more challenging for psychopathology related traits than for other complex traits for which we possess powerful polygenic scores (e.g. BMI or educational attainment) because the variance explained by the polygenic scores will be underpowered in longitudinal models including familial effects and reciprocal sibling effects. Until more powerful polygenic scores become available, one possible solution will be to employ multivariable polygenic score approaches to boost predictive power (chapter 3). However, as discussed, as we keep accumulating data on large longitudinal child cohorts, advances in genomic methods (for example, imputing parents' genotypes when these are not available) will make models such as this feasible in the near future.

Conclusion

It seems likely that a triangulation between multivariate methods will be at the forefront of research discovery for child development in the years to come. On one hand, with increasingly powerful GWAS summary statistics, genomic multi-trait methods will be able to be leveraged to further boost predictive power of polygenic scores. Powerful multivariable models integrating several such scores jointly could be implemented along with environmental effects. This will be key for further polygenic score work especially with respect to child psychopathology. Furthermore, modelling of genomic structures in multivariate SEM might be leveraged to create more nuanced polygenic scores, as well as to infer mechanisms underpinning covariance between traits, which will in turn help illuminating important developmental questions. Concurrently efforts to bring together high-quality phenotypic data from several child cohorts, as well as genotyping family-level data,

will afford the possibility to implement genomic methods at the family level by exploiting powerful quantitative genetics approaches on an unprecedented scale.

References

Akingbuwa, W. A., Hammerschlag, A. R., Jami, E. S., Allegrini, A. G., Karhunen, V., Sallis, H., ... Middledorp, C. M. (2020). Genetic Associations Between Childhood Psychopathology and Adult Depression and Associated Traits in 42 998 Individuals: A Meta-Analysis. *JAMA psychiatry*.

Bates, T. C., Maher, B. S., Medland, S. E., McAloney, K., Wright, M. J., Hansell, N. K., . . . Gillespie, N. A. (2018). The nature of nurture: Using a virtual-parent design to test parenting effects on children's educational attainment in genotyped families. *Twin Research and Human Genetics, 21*(2), 73-83.

Cai, N., Revez, J. A., Adams, M. J., Andlauer, T. F., Breen, G., Byrne, E. M., . . . Flint, J. (2020). Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nature genetics, 52*(4), 437-447.

Cheesman, R., Selzam, S., Ronald, A., Dale, P. S., McAdams, T. A., Eley, T. C., & Plomin, R. (2017). Childhood behaviour problems show the greatest gap between DNA-based and twin heritability. *Translational psychiatry, 7*(12), 1284.

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.

Demontis, D., Walters, R. K., Martin, J., Mattheisen, M., Als, T. D., Agerbo, E., ... & Neale, B. M. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nature genetics, 51(1), 63-75.

de Vlaming, R., Okbay, A., Rietveld, C. A., Johannesson, M., Magnusson, P. K., Uitterlinden, A. G., . . . Koellinger, P. D. (2017). Meta-GWAS Accuracy and Power (MetaGAP) calculator shows that hiding heritability is partially due to imperfect genetic correlations across studies. *PLoS genetics, 13*(1).

Dolan, C. V., Huijskens, R. C., Minica, C. C., Neale, M. C., & Boomsma, D. I. (2019). Incorporating polygenic scores in the twin model to estimate genotype-environment covariance: exploration of statistical power. *bioRxiv*, 702738.

Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLoS genetics, 9*(3), e1003348.

Eilertsen, E. M., Jami, E. S., McAdams, T. A., Hannigan, L. J., Havdahl, A. S., Magnus, P. M., . . . Ystrom, E. (2020). Direct and indirect effects of maternal, paternal, and offspring genotypes: Trio-GCTA. *bioRxiv*.

Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S. J., Mallard, T. T., Hill, W. D., . . . Tucker-Drob, E. M. (2019). Genomic structural equation modelling provides

insights into the multivariate genetic architecture of complex traits. *Nature human behaviour, 3*(5), 513.

Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R. K., Won, H., ... & Børglum, A. D. (2019). Identification of common genetic risk variants for autism spectrum disorder. Nature genetics, 51(3), 431-444.

Howard, D. M., Adams, M. J., Clarke, T.-K., Hafferty, J. D., Gibson, J., Shirali, M., . . . McIntosh, A. M. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature neuroscience, 22*(3), 343.

Hwang, L.-D., Tubbs, J. D., Luong, J., Moen, G.-H., Sham, P. C., Partida, G. C., & Evans, D. M. (2020). Estimating indirect parental genetic effects on offspring phenotypes using virtual parental genotypes derived from sibling and half sibling pairs. bioRxiv.

Ip, H. F., van der Laan, C. M., Brikell, I., Sánchez-Mora, C., Nolte, I. M., St Pourcain, B., . . . Boomsma, D.I. (2019). Genetic Association Study of Childhood Aggression across raters, instruments and age. *bioRxiv*, 854927.

Jia, G., Lu, Y., Wen, W., Long, J., Liu, Y., Tao, R., . . . Zheng, W. (2020). Evaluating the utility of polygenic risk scores in identifying high-risk individuals for eight common cancers. *JNCI Cancer Spectrum*.

Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., . . . Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature genetics, 50*(9), 1219-1224.

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.

Lee, S. H., Weerasinghe, W. S. P., Wray, N. R., Goddard, M. E., & Van Der Werf, J. H. (2017). Using information of relatives in genomic prediction to apply effective stratified medicine. *Scientific reports, 7*, 42091.

Lewis, C. M., & Vassos, E. (2020). Polygenic risk scores: from research tools to clinical instruments. *Genome Medicine, 12*, 1-11.

Manchia, M., Cullis, J., Turecki, G., Rouleau, G. A., Uher, R., & Alda, M. (2013). The impact of phenotypic and genetic heterogeneity on results of genome wide association studies of complex diseases. *PloS one, 8*(10), e76295.

Marquez-Luna, C., Gazal, S., Loh, P.-R., Kim, S. S., Furlotte, N., Auton, A., . . . Price, A. L. (2020). LDpred-funct: incorporating functional priors improves polygenic prediction accuracy in UK Biobank and 23andMe data sets. *bioRxiv*, 375337.

Minică, C. C., Dolan, C. V., Boomsma, D. I., de Geus, E., & Neale, M. C. (2018). Extending causality tests with genetic instruments: An integration of Mendelian randomization with the classical twin design. *Behaviour Genetics, 48*(4), 337-349.

Pasaniuc, B., & Price, A. L. (2017). Dissecting the genetics of complex traits using summary association statistics. *Nature reviews genetics, 18*(2), 117-127.

Plomin, R. (2014). Genotype-environment correlation in the era of DNA. *Behaviour Genetics, 44*(6), 629-638.

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 replicated findings from behavioural genetics. *Perspectives on psychological science, 11*(1), 3-23.

Rimfeld, K., Malanchini, M., Spargo, T., Spickernell, G., Selzam, S., McMillan, A., . . . Plomin, R. (2019). Twins Early Development Study: A Genetically Sensitive Investigation into Behavioural and Cognitive Development from Infancy to Emerging Adulthood. *Twin Research and Human Genetics*, 1-6.

St Pourcain, B., Eaves, L. J., Ring, S. M., Fisher, S. E., Medland, S., Evans, D. M., & Smith, G. D. (2018). Developmental changes within the genetic architecture of social communication behaviour: a multivariate study of genetic variance in unrelated individuals. *Biological psychiatry, 83*(7), 598-606.

Truong, B., Zhou, X., Shin, J., Li, J., van der Werf, J. H., Le, T. D., & Lee, S. H. (2020). Efficient polygenic risk scores for biobank scale data by exploiting phenotypes from inferred relatives. *Nature communications, 11*(1), 1-11.

Verhulst, B., Maes, H. H., & Neale, M. C. (2017). GW-SEM: a statistical package to conduct genome-wide structural equation modeling. *Behaviour Genetics, 47*(3), 345-359.

Wainschtein, P., Jain, D. P., Yengo, L., Zheng, Z., Cupples, L. A., Shadyab, A. H., . . . Visscher, P.M. (2019). Recovery of trait heritability from whole genome sequence data. *bioRxiv*, 588020.

Wray, N. R., Kemper, K. E., Hayes, B. J., Goddard, M. E., & Visscher, P. M. (2019). Complex trait prediction from genome data: contrasting EBV in livestock to PRS in humans: genomic prediction. *Genetics, 211*(4), 1131-1141.

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., . . . Sullivan, P. F. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature genetics, 50*(5), 668-681.

Wray, N. R., Yang, J., Hayes, B. J., Price, A. L., Goddard, M. E., & Visscher, P. M. (2013). Pitfalls of predicting complex traits from SNPs. *Nature reviews genetics, 14*(7), 507-515.

Young, A. I., Frigge, M. L., Gudbjartsson, D. F., Thorleifsson, G., Bjornsdottir, G., Sulem, P., . . . Kong, A. (2018). Relatedness disequilibrium regression estimates heritability without environmental bias. *Nature genetics, 50*(9), 1304-1310.

Young, A. I., Nehzati, S. M., Lee, C., Benonisdottir, S., Cesarini, D., Benjamin, D. J., . . . Kong, A. (2020). Mendelian imputation of parental genotypes for genome-wide estimation of direct and indirect genetic effects. *bioRxiv*.

Appendix 1– supplementary material for Chapter 2

Contents:

Supplementary methods

Methods S1: Genotyping protocol and quality control

DNA for 8,743 individuals (including 3,722 dizygotic co-twin samples) was extracted from saliva and buccal cheek swab samples and hybridized to HumanOmniExpressExome-8v1.2 genotyping arrays at the Institute of Psychiatry, Psychology and Neuroscience Genomics & Biomarker Core Facility. The raw image data from the array were normalized, pre-processed, and filtered in GenomeStudio according to Illumina Exome Chip SOP v1.4. (http://confluence.brc.iop.kcl.ac.uk:8090/display/PUB/Production+Version%3A+Illumina+Exome+Chip+SOP+v1.4). In addition, prior to genotype calling, 919 multi-mapping SNPs and 501 samples with callrate <0.95 were removed. The ZCALL program was used to augment the genotype calling for samples and SNPs that passed the initial QC.
DNA from 3,747 samples was extracted from buccal cheek swabs and genotyped at Affymetrix, Santa Clara, California, USA. From this sample, 3,665 samples were successfully hybridized to AffymetrixGeneChip 6.0 SNP genotyping arrays (http://www.affymetrix.com/support/technical/datasheets/genomewide_snp6_datasheet.pdf) using experimental protocols recommended by the manufacturer (Affymetrix Inc., Santa Clara, CA). The raw image data from the arrays were normalized and pre-processed at the Wellcome Trust Sanger Institute, Hinxton, UK for genotyping as part of the Wellcome Trust Case Control Consortium 2 (https://www.wtccc.org.uk/ccc2/) according to the manufacturer's guidelines (http://www.affymetrix.com/support/downloads/manuals/genomewidesnp6_manual.pdf). Genotypes for the Affymetrix arrays were called using CHIAMO (https://mathgen.stats.ox.ac.uk/genetics_software/chiamo/chiamo.html).
After initial quality control and genotype calling, the same quality control was performed on the samples genotyped on the Illumina and Affymetrix platforms separately using PLINK(Chang et al., 2015; Purcell et al., 2007), R(R Core Team, 2018), BCFtools(Li, 2011), and EIGENSOFT(Patterson, Price, & Reich, 2006; Price et al., 2006).
Samples were removed from subsequent analyses on the basis of call rate (<0.98), suspected non-European ancestry, heterozygosity, and relatedness other than dizygotic twin status. SNPs were excluded if the minor allele frequency was smaller than 0.5%, if more than 2% of genotype data were missing, or if the Hardy Weinberg *p*-value was lower than 10^{-5}. Non-autosomal markers and indels were removed. Association between SNP and the platform, batch, plate or well on which samples were genotyped was calculated; SNPs with an effect *p*-

value < 10^{-4} were excluded. A total sample of 10,346 samples (including 3,320 dizygotic twin pairs and 7,026 unrelated individuals), with 7,289 individuals and 559,772 SNPs genotyped on Illumina and 3,057 individuals and 635,269 SNPs genotyped on Affymetrix remained after quality control.

Genotypes from the two platforms were separately phased using EAGLE2 (Loh et al., 2016), and imputed into the Haplotype Reference Consortium (release 1.1) using the Positional Burrows-Wheeler Transform method (Durbin, 2014) through the Sanger Imputation Service(McCarthy et al., 2016). Prior to merging, we excluded variants with info <0.75 and removed non-overlapping SNPs between platforms. After merging, we tested for minor allele frequency differences between platforms and removed SNPs with an effect p-value < 10^{-4}, and Hardy Weinberg p-value > 10^{-5}. Using these criteria, 7,363,646 genotyped and well-imputed SNPs were retained for the analyses.

To generate principal components to be used as covariates, we performed principal component analysis on a subset of 39,353 common (MAF > 5%), perfectly imputed (info = 1) autosomal SNPs, after stringent pruning to remove markers in linkage disequilibrium (r2 > 0.1) and excluding high linkage disequilibrium genomic regions so as to ensure that only genome-wide effects were detected.

Methods S2: Description of phenotypes

Outcome Variables

Intelligence. Intelligence was operationalized as general cognitive ability. At age 12 and 16 twins participated in web-based testing, assessing verbal and non-verbal cognitive ability. At age 12 we administered two verbal ability test: WISC-III-PI Multiple choice Information (general knowledge)(Wechsler D, 1992); and WISC-III-PI Vocabulary Multiple-Choice; and two non-verbal tests: Raven's Standard Progressive Matrices (Raven, Court, & Raven, 1996) and WISC-III-UK Picture Completion (Wechsler & Golombok, 1992). At age 16, web testing consisted of one verbal: 'Mill Hill Vocabulary test'; and one non-verbal: 'Raven's Progressive Matrices' tests. We computed scales as the mean of the standardized web-test scores. Intelligence was defined as a mean composite of the verbal and non-verbal cognitive standardized scores.

Educational achievement. At age 12, educational achievement was based on teacher-ratings and test grades obtained through the UK National Pupil Database (NPD) (https://www.gov.uk/government/collections/national-pupil-database). Test grades are based on the National Curriculum, which is a set of standardized subjects taught in all primary and secondary schools in the UK. It consists of national tests and teacher assessments in English, mathematics and science (https://www.gov.uk/national-curriculum). At age 12 teachers rated NPD scores on a 9-point Likert scale, with higher scores indicating higher educational achievement.

At age 16, educational achievement was indexed as performance on standardized UK General Certificate of Secondary Education (GCSE) exams obtained via the NPD. This is the examination for educational achievement at the end of compulsory education. Individuals take around 8-10 subjects (Shakeshaft et al., 2013), with three of them being compulsory ('core subjects'): English, mathematics, and science. Families of TEDS twins were contacted following GCSE examinations by email and twins' parents completed test results forms and reported on results qualifications (Shakeshaft et al., 2013). For individuals for whom self-report data were not available, families of TEDS' twins were contacted and asked for consent to access the National Pupil Database. In the present study educational achievement was operationalized as the mean grade of the three compulsory subjects, with results coded from 4 (G, or lowest grade) to 11 (A+, or highest grade).

All phenotypes were corrected for age, sex and 10 genetic principal components. The obtained residuals were used in all subsequent analyses. Figure S1 shows correlations between the four outcome variables, and Table S1 reports descriptive statistics.

Methods S3: Polygenic Score Approaches

After stringent quality control, we restricted the sample to include only unrelated individuals (N=7,026). In addition, in order to ease computational burden across polygenic scores approaches, and to be able to efficiently compare results across them, we further limited analyses to variants genotyped or imputed at Info = 1 (with minor allele frequency > 0.5% and Hardy-Weinberg equilibrium p-value $>1x10^{-5}$), leaving a total of 515,100 SNPs.

PRSice2

Summary statistics used in analyses were coordinated with our genotyped sample and p-value clumping was performed with a $R^2 = 0.25$ cutoff within a 250-kb window. After coordination and clumping, 106,000 SNPs remained for analyses in PRSice (Euesden, Lewis, & O'Reilly, 2015). In PRSice we constructed polygenic scores as the sum of individuals' genotypes across SNPs, weighted by the betas of IQ3 and EA3 GWAS, as well as the summary statistics derived from MTAG (for IQ3 and EA3) and Genomic SEM (the common factor GWAS). Next, a best-fit polygenic score was obtained for intelligence and educational achievement at ages 12 and 16 through the high-resolution scoring option in PRSice. This consisted in regressing each of our phenotypes in the validation set on GPS calculated at a high number of p-value thresholds (from 0.001 to 1 with increments of 0.001) for each of our base datasets (IQ3, EA3, the summary statistics for MTAG IQ3, MTAG EA3 and the one obtained from the common factor GWAS in Genomic SEM), until the best-fit GPS was identified as the GPS threshold associated with the phenotype at the lowest p-value (and the highest R^2).

LDpred

LDpred (Vilhjalmsson et al., 2015) is a Bayesian method that calculates GPS based on GWAS summary statistics, by taking into account the trait-heritability, assuming a prior for the polygenicity of a trait (i.e. fraction of associated markers) and adjusting for linkage disequilibrium (LD) from a reference panel. In LDpred no LD clumping or pruning is performed on the individual-level genotyped data, which avoids losing information across the genome. Rather, the prior is used to stabilize the betas by re-weighing the coefficients. Genotypes were coordinated with the summary statistics, leaving 496,633 SNPs for the EA3 GWAS, 497,059 for IQ3 GWAS, 475,767 for EA3 MTAG, 411,947 for IQ3 MTAG and

476,321 for the common factor GWAS (Genomic SEM). LD adjustment was performed using the target sample (TEDS) genotype data as LD reference panel. The weights were then estimated based on the heritability explained by the markers in the GWAS summary statistics and the assumed fraction of markers with non-zero effects. For each summary statistic we created *LDpred* GPS based on 9 fractions (i.e. 1,0.3,0.1,0.03,0.01,0.003,0.001,0.0003,0.0001) which we then optimized with respect to prediction accuracy (highest R^2) in our validation set for all our phenotypes of interest.

Lassosum

Lassosum (Mak, Porsch, Choi, Zhou, & Sham, 2017) is a machine-learning approach which uses penalized regression in the context of GWAS summary statistics. The LASSO, or L1 regularization, approach is particularly useful within highly dimensional data where *m* (number of SNPs) > *N* (individuals), and most betas weights are assumed to be 0, because it allows variable selection (i.e. the removal of SNPs).
In LASSO regression an L1 penalty, or L1 regularization term, ($\lambda||\beta||_1$) is applied to the least square estimator. That is minimize

$$\beta||y-X\beta||^2$$

subject to a constraint parameter, *s:*

$$||\beta||_1 \leq s$$

The notation $||\beta||_1$ describes the L1 norm of a coefficient vector β, defined as

$$||\beta||_1 = \sum|\beta_j|.$$

This L1 norm depends on a specific parameter λ, which controls the amount of shrinkage applied to the estimator (Tibshirani, 1996), while *s* is the boundary for how large $\sum|\beta_j|$ can be. λ and *s* have thus an inverse relationship: as *s* approaches infinity λ becomes 0, no shrinkage is applied on the least square estimator and the problem becomes unconstrained (such as ordinary least squares). On the other hand as *s* becomes 0 λ approaches infinity thus betas are shrunk to 0. A simpler way to look at this is that depending on λ, some SNP betas are shrunk to 0, that is, the higher the penalty the more sparse the model will be. In the extreme, when λ = 0, all betas are retained in the model (no betas are set to 0 and therefore eliminated from the

model); when λ approaches infinity all betas are shrunk to 0. LASSO optimizes the tuning parameter λ to perform shrinkage and variable selection, thereby dealing with overfit and multicollinearity.
λ and s are hyperparameters which are tuned in a cross-validation step (see below).
In the context of Lassosum, the equation for the LASSO is rewritten to include a matrix correlation between SNPs and phenotype, and LD information from a reference panel is directly integrated in the formula as a matrix correlation between SNPs: see (Tibshirani, 1996) for more details.

Tuning and constraint parameters, λ and s, are chosen in the validation step (this is akin to optimization that can be performed in p-value thresholding methods). As with other approaches we used our validation set to perform parameter tuning and we retained the best polygenic score (with respect to R^2) to assess model performance in the test set. LD was accounted for via a reference panel, here the same as the test sample, and estimation of LD blocks was performed using LD regions defined in (Berisa & Pickrell, 2016), as recommended (Mak et al., 2017).

Methods S4: Multi-trait approaches.

Genomic SEM

Genomic SEM (Grotzinger et al., 2018) is a structural equation modeling approach, which can be used to jointly analyze GWA summary statistics to boost power for discovery and polygenic prediction. In the form used here, the common factor GWA analysis, we specifically test SNP effects at the level of a latent common factor constructed from GWA summary statistics for several traits. We used the most recent GWA summary statistics for intelligence (IQ3) and educational attainment (EA3) along with three UKbiobank traits: 'Age when completed full time education', and 'Time spent using computer', and 'Household Income'. Table S2 provides more information about these GWA summary statistics.
First, we estimated the genetic covariance and corresponding sampling covariance matrices by using multivariate LD-score regression (LDSC) a multivariate extension of LDSC, which is used in Genomic SEM to populate the off-diagonal elements of the sampling covariance matrix, accounting for unknown sample overlap (Grotzinger et al., 2018). Secondly, we fitted a common factor model to the data using a Diagonally Weighted Least Square (DWLS) estimator (default in Genomic SEM). Table S3 shows results for the common factor model in terms of standardized and unstandardized factor loadings and residual variance of the indicators after removing the effects of the common factor. The model fit indices showed a reasonably good fit for our specified model: $\chi^2(5) = 45.899$, AIC = 65.899, CFI = .99, SRMR = .032.

We then proceeded to run the common factor GWAS. Here, the genetic and sampling covariance matrices are expanded to include effects of individual SNPs, and a common factor model is fitted to the data for each SNP in common between the indicators (in our case ~ 7 million SNPs in common across all summary statistics).

After the common factor GWAS analysis we calculated an effective sample size for the common factor GWAS summary statistics(Nivard, 2018), in order to be able to run further analyses (e.g. polygenic scores in LDpred):

First, we computed the dot product of the square root of N (where N is a matrix containing GWAS sample sizes):
sqrt(N)%*%t(sqrt(N))

Second, we extended the dot product with an inner matrix, which is the correlation matrix between the square root of the sample sizes:

```
I <- diag(5)

sqrt(N)%*%solve(I) %*% t(sqrt(N))
```

Lastly, we replaced the inner matrix with the matrix holding the cross-trait intercepts from the multivariable LDSC function. This yielded an effective sample size of N = 1,387,848.
The analyses yielded a common factor trait with SNP h^2 of 8.4%. It has been noted (Grotzinger et al., 2018) that the WLS estimator will tend to produce a solution mainly reflecting the loadings of the most powered traits and will boost power for prediction accuracy accordingly. This property was desirable in our case as EA3 had the highest loading on the common factor (Table S4).

This was also reflected in the pattern of genetic correlations from Genomic SEM (r_g estimates presented in Supplementary Figure S3), where the common factor trait had a genetic correlation of r_g = .96 with EA3 and of r_g = .86 with IQ3. This finding suggest that the common factor trait captured more of the genetic architecture of educational attainment than of that of IQ, which was also suggested by standardized factor loadings on the common factor of .92 and .79 for EA3 and IQ3, respectively."

MTAG

MTAG (Turley et al., 2018) is a meta-analytic approach that jointly analyzes univariate GWAS summary statistics from several potentially different traits, and outputs trait-specific, power-boosted summary statistics, which can be used for variant discovery or polygenic prediction. In a similar fashion as Genomic SEM, bivariate LDSC is used to account for unknown sample overlap between GWA summary statistics. As for Genomic SEM, here we jointly analyzed IQ3 and EA3 (our GPS of interest) along with Income, 'Age completed full time education' and 'time spent using computer'. Table S5 reports mean χ^2 of summary statistics before and after analysis in MTAG. Both in the case of IQ3 and EA3 the analysis yielded substantial gains in power, with a mean χ^2 increase of 19.1% for IQ3 which corresponds to an increase in power from a GWA sample size of N = 266,453 to N = 383,743 (i.e. 44% increase in GWAS sample size), and a mean χ^2 increase of 8.9% for EA3, corresponding to an increase in GWA sample size from N= 766,345 to N = 883,280 (i.e. 15% increase in GWAS sample size). Table S4 shows MTAG results for all summary statistics. SNP h^2 of intelligence increased from 18.4% to 28.2% and SNP h^2 of educational achievement from 10.7% to 12.6%.

MTAG makes the assumption that SNP effects share the same variance-covariance matrix across traits (Turley et al., 2018). This assumption is however likely to be violated because effects across traits are heterogeneous, which could lead to inflated type I error. Nevertheless, as outlined in the original manuscript (Turley et al., 2018), even if this assumption is not met, polygenic scores constructed from MTAG summary statistics are expected to yield more accurate prediction estimates (R^2) than polygenic scores derived from GWA summary statistics for individual traits. However, this may still cause problems at the level of interpretability, because increases in false positive rates could bias the genetic architecture of MTAG analyses towards the traits with the most powered GWAS (Hill, 2018).

To keep track of Type I error inflations, we ran the recommended max False Discovery Rate (maxFDR) (Supplementary table S5) analyses on our traits, which showed that Type I error did not seem inflated for IQ3 (FDR <. 0001) or EA3 (FDR <.0001). However, maxFDR calculations only provide information regarding contamination of genome-wide significant hits. As such, the insights provided by these analyses regarding the level of possible genetic confound are limited.

We therefore estimated genetic correlations between our main variables of interest before and after MTAG as well as with a schizophrenia GWA (Schizophrenia Working Group of the Psychiatric Genomics, 2014), used here as a control variable. As shown elsewhere (Hill, 2018), since schizophrenia has a negative moderate genetic correlation with IQ and a very weak and positive, or non-significant, association with EA, this analysis provided insights about the degree to which MTAG summary statistics for IQ3 and EA3 were leaning more towards one trait or the other. Figure S4 shows genetic correlations between these traits. Before being meta-analyzed in MTAG, IQ3 had a moderate negative genetic correlation with schizophrenia, r_g = - 0.20 (p = 2.95e-22), while EA3 had a negligible positive genetic correlation with SCZ, r_g = 0.063 (p = 0.0006). After the analysis in MTAG the genetic correlation between IQ3 and SCZ was somewhat attenuated r_g = -0.10 (p = 3.0475e-08), but still consistent and in the direction expected. The genetic correlation between EA3 and SCZ was even closer to 0 and was not significant, $r_g < 0.003$ (p = .86). These analyses, along with FDR results, suggested that confounding due to violation of MTAG assumptions was not likely for IQ3 and EA3.

SMTpred

SMTpred **(**Maier et al., 2018**)** is a multi-trait method that combines beta weights from GWAS summary statistics or, more directly, already calculated polygenic scores, based on sample size, SNP h2 and genetic correlations between traits, in order to boost prediction accuracy of polygenic scores **(**Maier et al., 2018**)**. In the form used here, SMTpred is an extension of the genomic best linear unbiased predictor (GBLUP) method (whereby in a linear mixed model framework SNP effects are jointly fit in the same model; (Yang et al., 2012) to the extent that we apply multi-trait weighting to BLUP predictors based on summary statistics (wMT-SBLUP). First, we converted beta (OLS) estimates from GWA summary statistics of 'IQ', 'EA', 'Income', 'Age completed full time education' and 'Time spent using computed', to summary statistics based BLUP estimates using GCTA-Cojo (Yang et al., 2012). As recommended (Maier et al., 2018) we used a shrinkage λ parameter $M^*(1-h^2/h^2)$, where M is the number of SNPs and h^2 is the SNP h^2 of the trait, and an LD window of 2000 Kb, using our target sample (TEDS) as the reference panel.

We then calculated SNP h2 and genetic correlation estimates between traits in LDSC (Bulik-Sullivan et al., 2015) to obtain SNP h2 and genetic correlation estimates between traits. These were input in SMTpred, along with SBLUP estimates, to derive wMT-SBLUP estimates using an optimal index weighting for each trait (Maier et al., 2018), which takes into account contributions of jointly analysed traits depending on GWAS sample size, SNP h2 and genetic correlation with the target trait. Polygenic scores were then calculated for the multi-trait weighted IQ3 and EA3 summary statistics using PLINK [2].

Supplementary figures

Figure S1. Phenotypic correlations between outcome variables.

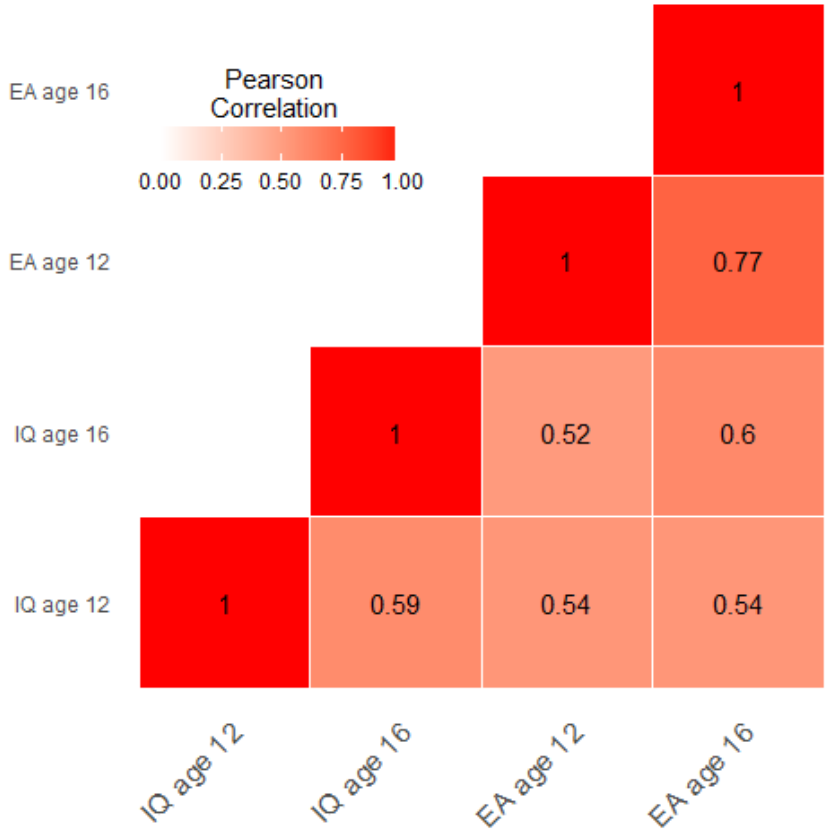
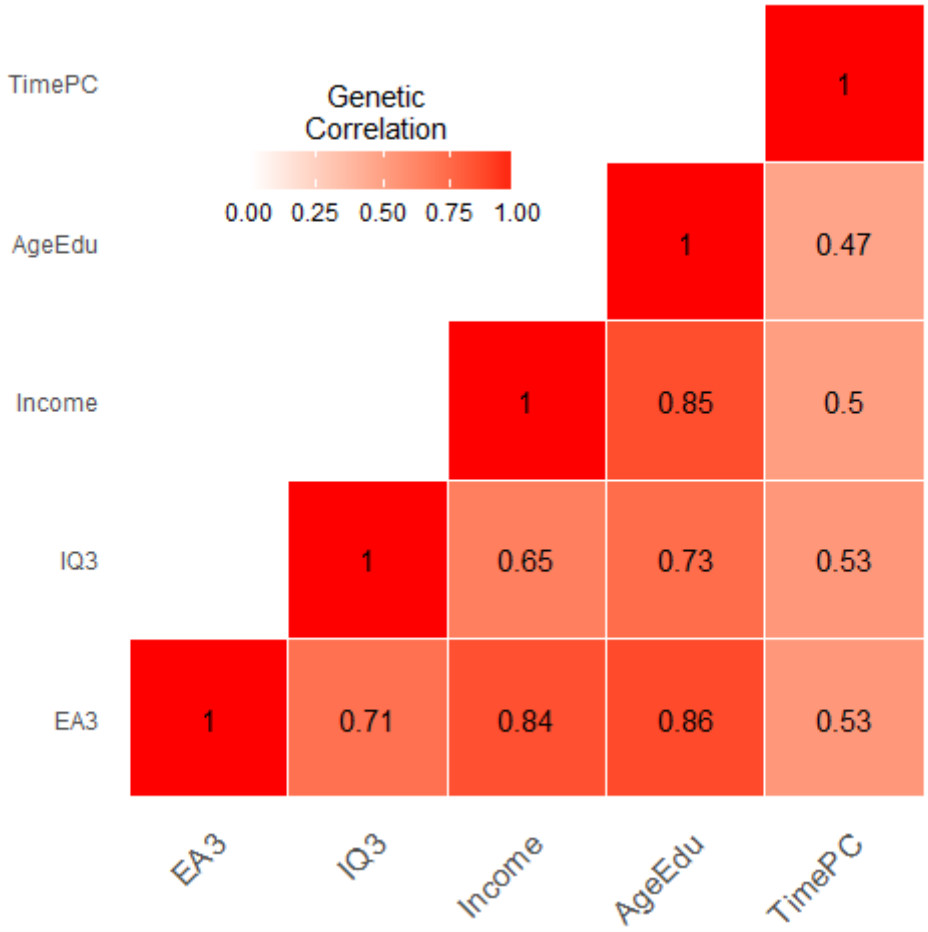
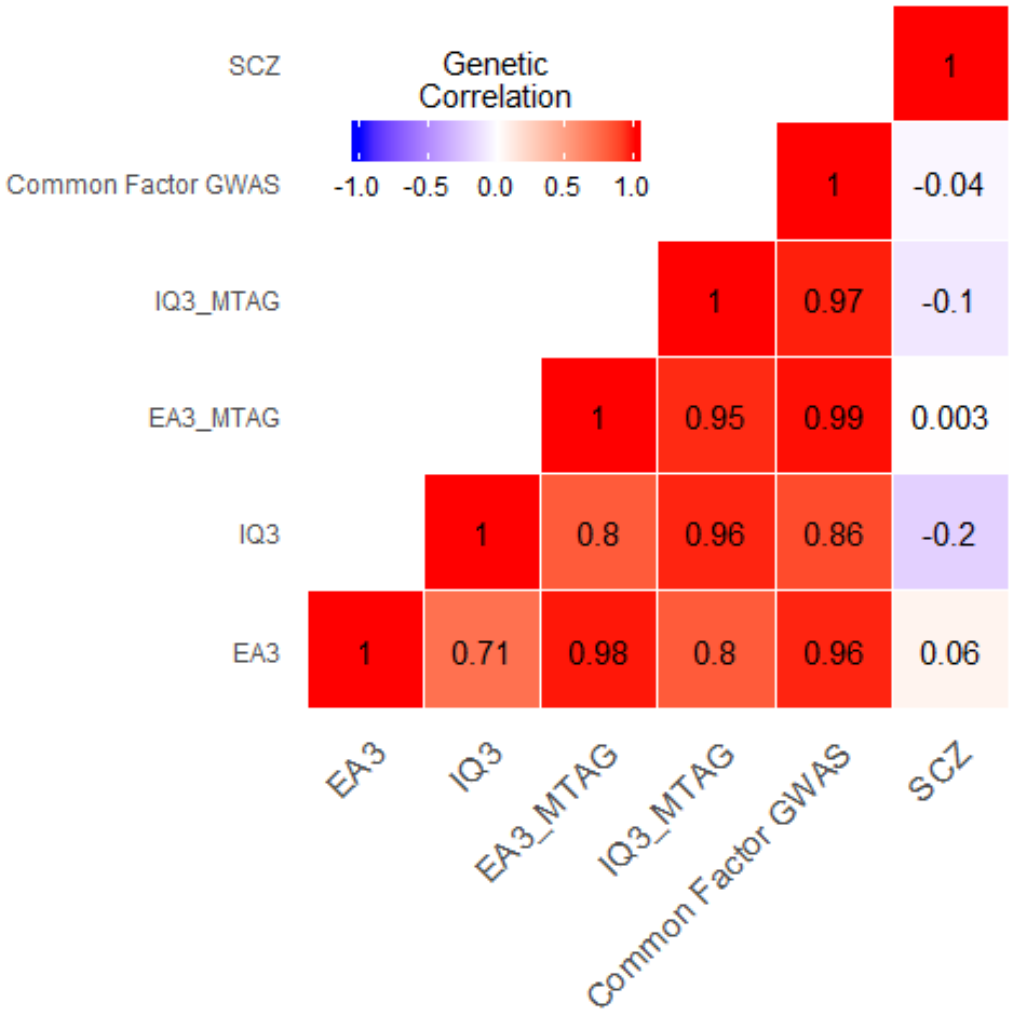


Figure S2. Genetic correlations between GWAS summary statistics.



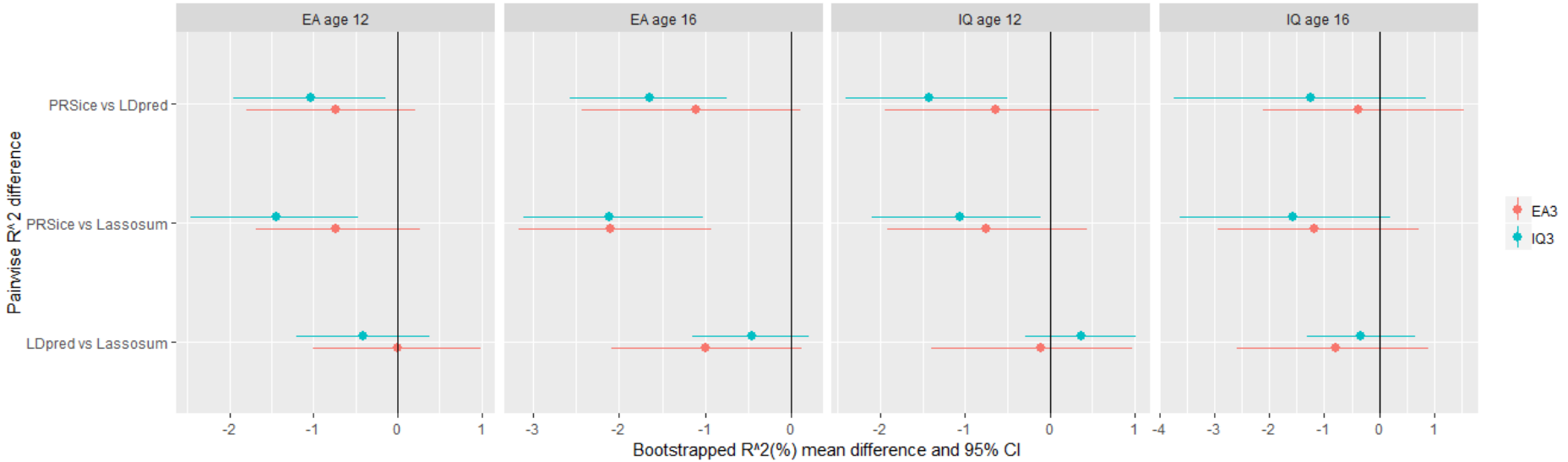
Note. EA3 = educational attainment GWAS, IQ3 = intelligence GWAS, Income = Income GWAS, AgeEdu = Age completed full time education GWAS, TimePC = Time spent at computer GWAS.

Figure S3. Genetic correlations between pre and post multi-trait analyses.



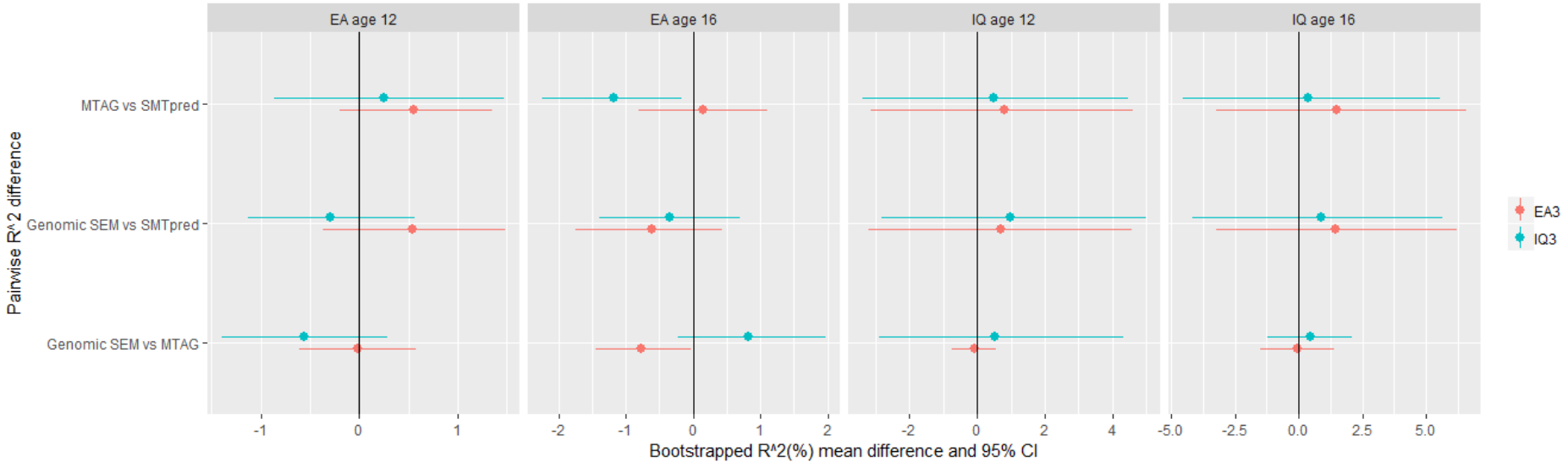
Note. EA3 = educational attainment GWAS, IQ3 = intelligence GWAS, SCZ = Schizophrenia GWAS.

Figure S4. Test of difference in predictive power between polygenic score approaches.



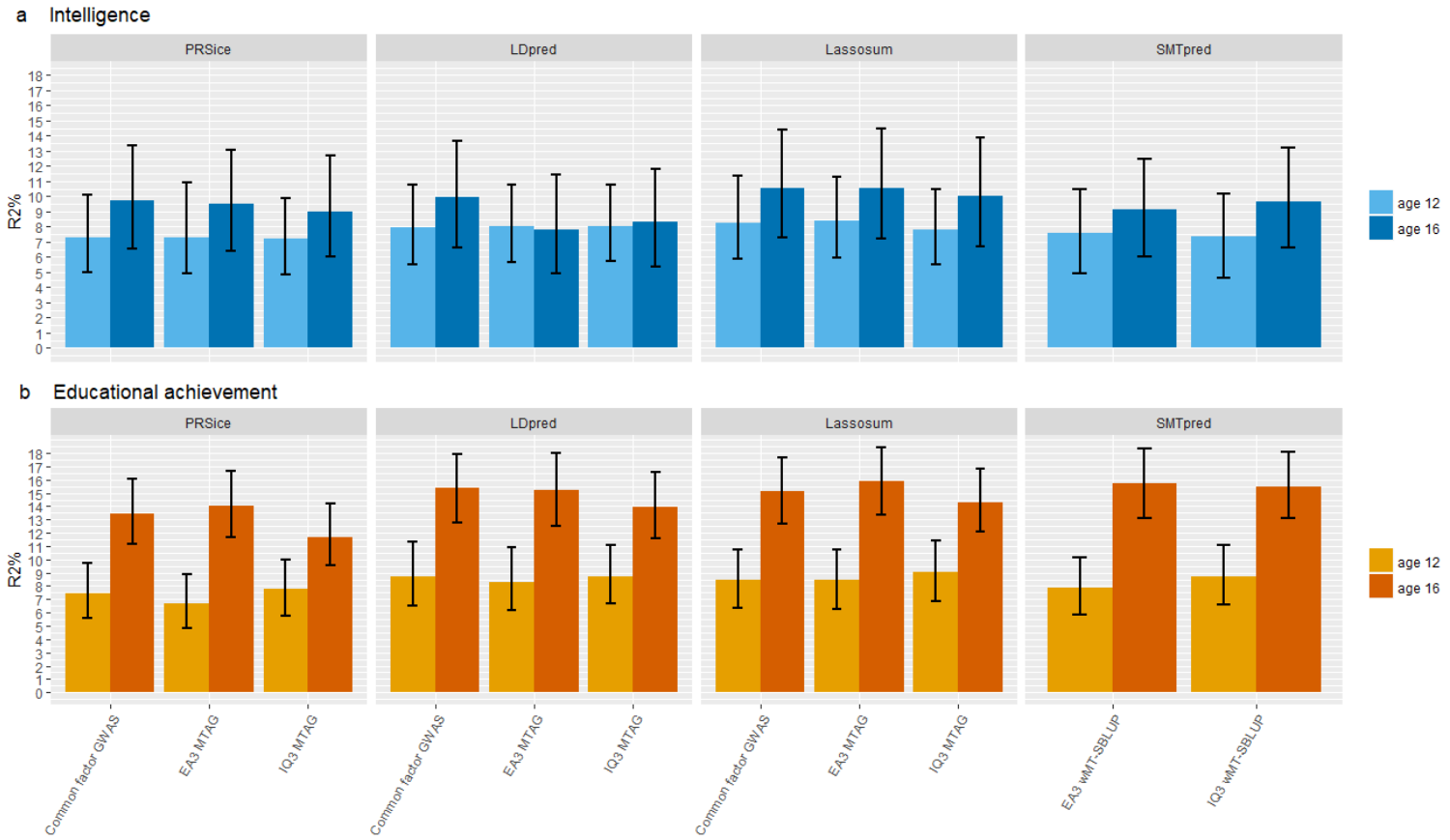
Note. Figure shows pairwise bootstrapped R^2 differences (%) and 95% confidence intervals calculated as the 2.5th and 97.5th percentiles of the distribution of R^2 differences for the polygenic score methods tested.

Figure S5. Test of difference in predictive power between multi-trait approaches.



Note. Figure shows pairwise bootstrapped R^2 differences (%) and 95% confidence intervals calculated as the 2.5th and 97.5th percentiles of the distribution of R^2 differences, for the multi-trait methods tested. Polygenic scores were calculated using Lassosum.

Figure S6. Variance predicted in intelligence and educational achievement at age 12 and 16 across multi-trait and polygenic scores methods.



Note. Figure shows variance predicted in intelligence (panel a) and educational achievement (panel b) at age 12 and 16 using polygenic scores constructed from IQ3 MTAG, EA3 MTAG, common factor GWAS (summary statistics from genomic SEM), and multi-trait weighted IQ3 and EA3 SBLUP predictors (wMT-SBLUP).

Error bars are bootstrapped 95% confidence intervals based on 1,000 replications

Supplementary References

Achenbach, T. M., Ivanova, M. Y., & Rescorla, L. A. (2017). Empirically based assessment and taxonomy of psychopathology for ages 1½–90+ years: Developmental, multi-informant, and multicultural findings. *Comprehensive Psychiatry, 79*, 4-18.

Bartels, M., Hendriks, A., Mauri, M., Krapohl, E., Whipp, A., Bolhuis, K., . . . Hagenbeek, F. (2018). Childhood aggression and the co-occurrence of behavioural and emotional problems: results across ages 3–16 years from multiple raters in six cohorts in the EU-ACTION project. *European child & adolescent psychiatry, 27*(9), 1105-1121.

Bartels, M., van Beijsterveldt, C. T., Derks, E. M., Stroet, T. M., Polderman, T. J., Hudziak, J. J., & Boomsma, D. I. (2007). Young Netherlands Twin Register (Y-NTR): a longitudinal multiple informant study of problem behaviour. *Twin Research and Human Genetics, 10*(1), 3-11.

Bergen, S. E., Gardner, C. O., & Kendler, K. S. (2007). Age-related changes in heritability of behavioural phenotypes over adolescence and young adulthood: a meta-analysis. *Twin Research and Human Genetics, 10*(3), 423-433.

Berisa, T., & Pickrell, J. K. (2016). Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics, 32*(2), 283-285. doi:10.1093/bioinformatics/btv546

Bulik-Sullivan, B. K., Loh, P. R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., . . . Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet, 47*(3), 291-295. doi:10.1038/ng.3211

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience, 4*, 7. doi:10.1186/s13742-015-0047-8

Cribbie, R. A. (2007). Multiplicity control in structural equation modeling. *Structural Equation Modeling, 14*(1), 98-112.

Durbin, R. (2014). Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). *Bioinformatics, 30*(9), 1266-1272. doi:10.1093/bioinformatics/btu014

Epskamp, S., Cramer, A. O., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of statistical software, 48*(4), 1-18.

Euesden, J., Lewis, C. M., & O'Reilly, P. F. (2015). PRSice: Polygenic Risk Score software. *Bioinformatics, 31*(9), 1466-1468. doi:10.1093/bioinformatics/btu848

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry, 38*(5), 581-586.

Goodman, R., & Scott, S. (1999). Comparing the Strengths and Difficulties Questionnaire and the Child Behaviour Checklist: is small beautiful? *Journal of abnormal child psychology, 27*(1), 17-24.

Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S. J., Mallard, T. T., Hill, W. D., . . . Tucker-Drob, E. M. (2018). Genomic SEM Provides Insights into the Multivariate Genetic Architecture of Complex Traits. *bioRxiv*. Retrieved from http://biorxiv.org/content/early/2018/04/21/305029.abstract

Hill, W. D. (2018). Comment on 'Large-Scale Cognitive GWAS Meta-Analysis Reveals Tissue-Specific Neural Expression and Potential Nootropic Drug Targets' by Lam et al. *Twin Res Hum Genet, 21*(2), 84-88. doi:10.1017/thg.2018.12

Knopik, V. S., Neiderhiser, J. M., DeFries, J. C., & Plomin, R. (2016). *Behavioural genetics*: Macmillan Higher Education.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics, 27*(21), 2987-2993. doi:10.1093/bioinformatics/btr509

Ligthart, L., van Beijsterveldt, C. E., Kevenaar, S. T., de Zeeuw, E., van Bergen, E., Bruins, S., . . . Hottenga, J.-J. (2019). The Netherlands Twin Register: Longitudinal Research Based on Twin and Twin-Family Designs. *Twin Research and Human Genetics*, 1-14.

Loh, P. R., Danecek, P., Palamara, P. F., Fuchsberger, C., Y, A. R., H, K. F., . . . A, L. P. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet, 48*(11), 1443-1448. doi:10.1038/ng.3679

Maier, R. M., Zhu, Z., Lee, S. H., Trzaskowski, M., Ruderfer, D. M., Stahl, E. A., . . . Robinson, M. R. (2018). Improving genetic prediction by leveraging genetic correlations among human diseases and traits. *Nature Communications, 9*(1), 989. doi:10.1038/s41467-017-02769-6

Mak, T. S. H., Porsch, R. M., Choi, S. W., Zhou, X., & Sham, P. C. (2017). Polygenic scores via penalized regression on summary statistics. *Genet Epidemiol, 41*(6), 469-480. doi:10.1002/gepi.22050

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., . . . Durbin, R. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet, 48*(10), 1279-1283. doi:10.1038/ng.3643

Moore, D. S., & Shenk, D. (2017). The heritability fallacy. *Wiley Interdisciplinary Reviews: Cognitive Science, 8*(1-2), e1400.

Nivard, M. G. (2018). [Personal communication].

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet, 2*(12), e190. doi:10.1371/journal.pgen.0020190

Plomin, R., & Deary, I. J. (2015). Genetics and intelligence differences: five special findings. *Molecular psychiatry, 20*(1), 98-108.

Porsch, R. M., Middeldorp, C. M., Cherny, S. S., Krapohl, E., Van Beijsterveldt, C. E., Loukola, A., . . . Rhee, S. (2016). Longitudinal heritability of childhood aggression. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 171*(5), 697-707.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet, 38*(8), 904-909. doi:10.1038/ng1847

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet, 81*(3), 559-575. doi:10.1086/519795

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Raven, J. C., Court, J., & Raven, J. (1996). *Raven Manual: Section 3 Standard ProgressiveMatrices With Adult US Norms by JC Raven, JH Court AndJ. Raven*: Oxford Psychologist Press.

Rimfeld, K., Malanchini, M., Spargo, T., Spickernell, G., Selzam, S., McMillan, A., . . . Plomin, R. (2019). Twins Early Development Study: A Genetically Sensitive Investigation into Behavioural and Cognitive Development from Infancy to Emerging Adulthood. *Twin Research and Human Genetics*, 1-6.

Schizophrenia Working Group of the Psychiatric Genomics, C. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature, 511*, 421. doi:10.1038/nature13595

https://www.nature.com/articles/nature13595#supplementary-information

Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Rimfeld, K., Krapohl, E., Haworth, C. M. A., . . . Plomin, R. (2013). Strong Genetic Influence on a UK Nationwide Test of Educational Achievement at the End of Compulsory Education at Age 16. *PLOS ONE, 8*(12), e80341. doi:10.1371/journal.pone.0080341

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *J R Stat Soc Ser B Methodol*(58), 267–288.

Turley, P., Walters, R. K., Maghzian, O., Okbay, A., Lee, J. J., Fontana, M. A., . . . Benjamin, D. J. (2018). Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet, 50*(2), 229-237. doi:10.1038/s41588-017-0009-4

Vilhjalmsson, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindstrom, S., Ripke, S., . . . Price, A. L. (2015). Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am J Hum Genet, 97*(4), 576-592. doi:10.1016/j.ajhg.2015.09.001

Wechsler, D., & Golombok, S. (1992). WISC-III UK. *Sidcup, Kent: The Psychological Corporation*.

Wechsler D, G. S., Rust J. (1992). WISC-III UK Wechsler Intelligence Scale for Children: UK manual. *Sidcup UK Psychol Corp*.

Yang, J., Ferreira, T., Morris, A. P., Medland, S. E., Madden, P. A., Heath, A. C., . . . Visscher, P. M. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet, 44*(4), 369-375, s361-363. doi:10.1038/ng.2213

Supplementary Tables

Supplementary Table S1. Descriptives statistics of phenotypes.

Outcome	N	Mean	SD
Intelligence age 12	3273	0.003	0.982
Educational achievement age 12	4537	4.428	0.577
Intelligence age 16	1925	-0.016	0.983
Educational achievement age 16	4925	8.754	1.275
Training set			
Intelligence age 12	1734	0	0.99
Educational achievement age 12	2331	4.42	0.57
Intelligence age 16	994	0	0.96
Educational achievement age 16	2516	8.74	1.24
Test set			
Intelligence age 12	1539	0.01	0.97
Educational achievement age 12	2206	4.44	0.58
Intelligence age 16	931	-0.03	1
Educational achievement age 16	2409	8.77	1.31

Supplementary Table S2. GWA summary statistics.

Name	N	h2	SE h2	Repository
EA3 (without 23andMe)	766,345	0.1066	0.0026	https://www.dropbox.com/s/ho58e9jmytmpaf8/GWAS_EA_excl23andMe.txt?dl=0
IQ3 (without TEDS)	266,450	0.1841	0.0063	https://ctg.cncr.nl/documents/p1651/SavageJansen_IntMeta_sumstats.zip
Income (UK Biobank)	96,900	0.0703	0.0065	Hill2016_UKB_Income_summary_results_21112016.txt.gz
Age completed full time education (UKB)	226,899	0.0831	0.0043	https://www.dropbox.com/s/ia7o57jpr4so3y7/189.assoc.tsv.gz?dl=0
Time spent using computer (UKB)	261,987	0.0959	0.0041	https://www.dropbox.com/s/cxp82nmxfxi1fwz/2178.assoc.tsv.gz?dl=0

Name	References
EA3 (without 23andMe)	Lee et al., 2018. Nat. Genet. 2018; 50:1112–1121.
IQ3 (without TEDS)	Savage et al., 2018. Nat Genet. 2018;50:912–9.
Income (UKB)	Hill et al., 2016. Curr Biol. 2016;26:3083–9.
Age completed full time education (UKB)	Seed C Hail: An Open-Source Framework for Scalable Genetic Data. 2017.
Time spent using computer (UKB)	Seed C Hail: An Open-Source Framework for Scalable Genetic Data. 2017.

Supplementary Table S3. Model selection and model performance in training and test sets.

		Training sets							
		Intelligence age 12				Educational achievement age 12			
Polygenic score approach	Summary statistics	Beta (SE)	R2%	tuning parameter	N SNPs	Beta (SE)	R2%	tuning parameter	N SNPs
PRSice	GWAS IQ3	0.203 (0.024)	3.9	p-value threshold = 0.3	72,614	0.22 (0.020)	5.18	p-value threshold = 0.3	72,614
	GWAS EA3	0.230 (0.023)	5.1	p-value threshold = 0.3	75,860	0.270 (0.019)	7.89	p-value threshold = 0.3	75,860
LDpred	GWAS IQ3	0.210 (0.023)	4.3	Fraction = 1	497,059	0.257 (0.019)	6.77	Fraction = 1	497,059
	GWAS EA3	0.221 (0.023)	4.8	Fraction = 1	496,633	0.271 (0.019)	7.5	Fraction = 1	496,633
Lassosum	GWAS IQ3	0.217 (0.023)	4.6	$\lambda = 0.001; s = 0.9$	149,922	0.265 (0.019)	7.19	$\lambda = 0.00162; s = 0.9$	107,322
	GWAS EA3	0.237 (0.023)	5.4	$\lambda = 0.001; s = 0.5$	74,009	0.290 (0.019)	8.55	$\lambda = 0.001; s = 0.9$	110,698
		Intelligence age 16				Educational achievement age 16			
Polygenic score approach	Summary statistics	Beta (SE)	R2%	tuning parameter	N SNPs	Beta (SE)	R2%	tuning parameter	N SNPs
PRSice	GWAS IQ3	0.253 (0.029)	6.8	p-value threshold = 0.1	10,399	0.25 (0.019)	6.46	p-value threshold = 1	141,805
	GWAS EA3	0.274 (0.029)	7.9	p-value threshold = 0.1	41,905	0.365(0.018)	13.8	p-value threshold = 0.1	41,905
LDpred	GWAS IQ3	0.280 (0.029)	8.4	Fraction = 1	497,059	0.274 (0.018)	7.76	Fraction = 1	497,059
	GWAS EA3	0.270 (0.029)	8.2	Fraction = 1	496,633	0.367 (0.018)	14	Fraction = 1	496,633
Lassosum	GWAS IQ3	0.266 (0.032)	8.8	$\lambda = 0.001; s = 0.9$	149,922	0.284 (0.018)	8.32	$\lambda = 0.001; s = 0.9$	149,922
	GWAS EA3	0.285 (0.029)	8.6	$\lambda = 0.001; s = 0.9$	110,698	0.391 (0.018)	15.9	$\lambda = 0.001; s = 0.5$	74,009
		Test sets							
		Intelligence age 12				Educational achievement age 12			
Polygenic score approach	Summary statistics	Beta (SE)	R2%	95% CI (lower/upper bound)		Beta (SE)	R2%	95% CI (lower/upper bound)	
PRSice	GWAS IQ3	0.189 (0.024)	3.9	2.22	5.95	0.230 (0.020)	5.64	3.03	7.48
	GWAS EA3	0.250 (0.024)	6.4	4.05	9.09	0.240 (0.020)	5.9	4.1	7.99
LDpred	GWAS IQ3	0.221 (0.023)	5.3	3.42	7.7	0.259 (0.020)	6.65	4.85	8.75
	GWAS EA3	0.263 (0.024)	7.1	4.93	9.58	0.259 (0.020)	6.66	4.62	9.1
Lassosum	GWAS IQ3	0.215 (0.024)	4.9	3.1	7.37	0.267 (0.020)	7.06	5.19	9.22
	GWAS EA3	0.265 (0.024)	7.2	4.83	9.97	0.259 (0.020)	6.64	4.83	8.7
		Intelligence age 16				Educational achievement age 16			
Polygenic score approach	Summary statistics	Beta (SE)	R2%	95% CI (lower/upper bound)		Beta (SE)	R2%	95% CI (lower/upper bound)	
PRSice	GWAS IQ3	0.232 (0.032)	5.2	2.82	8.09	0.273 (0.019)	7.25	5.55	9.49
	GWAS EA3	0.304 (0.032)	8.7	5.62	12.12	0.360 (0.019)	12.7	10.28	15.41
LDpred	GWAS IQ3	0.260 (0.032)	6.4	3.92	9.71	0.302 (0.019)	8.91	7.05	11.18
	GWAS EA3	0.308 (0.031)	9.1	6.06	12.71	0.376 (0.019)	13.8	11.16	16.23
Lassosum	GWAS IQ3	0.266 (0.032)	6.7	4.01	10.03	0.310 (0.019)	9.4	7.24	11.63
	GWAS EA3	0.321 (0.031)	9.9	6.7	13.7	0.389 (0.019)	14.8	12.27	17.28

Note: significance test across all: p < 2e-16

Supplementary Table S3a. Bootstrapped R2 difference and 95% confidence intervals for polygenic score methods tested.

		Intelligence age 12			Intelligence age 16		
Polygenic score approach comparison	Summary statistics	Bootstrapped R2 difference	95% CI (lower/upper bound)		Bootstrapped R2 difference	95% CI (lower/upper bound)	
PRSice vs LDpred	GWAS IQ3	-0.014281321	-0.023996651	-0.00492503	-0.012479971	-0.03739398	0.0084961
	GWAS EA3	-0.006440552	-0.019373234	0.00580973	-0.003773637	-0.02116681	0.0154004
LDpred vs Lassosum	GWAS IQ3	0.003639075	-0.002952799	0.01017529	-0.003304759	-0.01321172	0.0066538
	GWAS EA3	-0.001096851	-0.013921895	0.00971285	-0.007979123	-0.02598166	0.0089394
PRSice vs Lassosum	GWAS IQ3	-0.010642246	-0.020898629	-0.00107778	-0.01578473	-0.03638437	0.0020232
	GWAS EA3	-0.007537404	-0.019097489	0.00443493	-0.01175276	-0.02939713	0.0072051
		Educational achievement age 12			Educational achievement age 16		
Polygenic score approach comparison	Summary statistics	Bootstrapped R2 difference	95% CI (lower/upper bound)		Bootstrapped R2 difference	95% CI (lower/upper bound)	
PRSice vs LDpred	GWAS IQ3	-1.03E-02	-0.01956885	-0.00146299	-0.016504655	-0.02579594	-0.0075171
	GWAS EA3	-7.33E-03	-0.01791592	0.0021386	-0.01101581	-0.02435639	0.0011394
LDpred vs Lassosum	GWAS IQ3	-4.08E-03	-0.01200445	0.0038833	-0.00464733	-0.01151893	0.001991
	GWAS EA3	-2.41E-05	-0.01007874	0.0099258	-0.009969713	-0.02088693	0.0011764
PRSice vs Lassosum	GWAS IQ3	-1.44E-02	-0.0246109	-0.00463441	-0.021151985	-0.03111475	-0.0103181
	GWAS EA3	-7.35E-03	-0.01691077	0.00273215	-0.020985522	-0.03170075	-0.009227

Supplementary Table S4. Model selection and model performance in training and test sets.

		Training sets							
		intelligence age 12				educational achievement age 12			
Polygenic score approach	**Summary statistics**	**Beta (SE)**	**R2%**	**tuning parameter**	**N SNPs**	**Beta (SE)**	**R2%**	**tuning parameter**	**N SNPs**
PRSice	Common factor GWAS	0.025 (0.023)	6.23	p-value threshold = 0.1	43,889	0.301 (0.019)	9.27	p-value threshold = 0.1	43,889
	MTAG IQ3	0.247 (0.023)	5.99	p-value threshold = 0.1	37,912	0.297 (0.019)	9.04	p-value threshold = 0.1	37,912
	MTAG EA3	0.253 (0.023)	6.21	p-value threshold = 0.3	77,465	0.299 (0.019)	9.14	p-value threshold = 1	141,470
Ldpred	Common factor GWAS	0.278 (0.024)	5.74	Fraction = 1	476,321	0.303 (0.019)	9.38	Fraction = 1	476,321
	MTAG IQ3	0.247 (0.023)	5.91	Fraction = 1	411,947	0.302 (0.019)	9.32	Fraction = 1	411,947
	MTAG EA3	0.234 (0.023)	5.3	Fraction = 1	475,767	0.292 (0.019)	8.7	Fraction = 1	475,767
Lassosum	Common factor GWAS	0.253 (0.023)	6.21	$\lambda = 0.001$; s = 0.5	58,988	0.317 (0.019)	10.28	$\lambda = 0.001$; s = 0.5	58,988
	MTAG IQ3	0.260 (0.023)	6.6	$\lambda = 0.001$; s = 0.5	103,554	0.316 (0.019)	10.21	$\lambda = 0.00264$; s = 0.2	28,276
	MTAG EA3	0.252 (0.023)	6.14	$\lambda = 0.001$; s = 0.5	76,252	0.309 (0.019)	9.76	$\lambda = 0.001$; s = 0.5	76,252
		intelligence age 16				**educational achievement age 16**			
		Beta (SE)	**R2%**	**tuning parameter**	**N SNPs**	**Beta (SE)**	**R2%**	**tuning parameter**	**N SNPs**
PRSice	Common factor GWAS	0.311 (0.029)	9.96	p-value threshold = 0.1	43,889	0.370 (0.018)	14.17	p-value threshold = 0.1	43,889
	MTAG IQ3	0.315(0.029)	10.4	p-value threshold = 0.3	71,903	0.352 (0.018)	12.76	p-value threshold = 0.03	20,156
	MTAG EA3	0.299 (0.029)	9.39	p-value threshold = 0.3	77,465	0.378 (0.018)	14.83	p-value threshold = 0.1	44,159
Ldpred	Common factor GWAS	0.316 (0.029)	10.23	Fraction = 1	476,321	0.396 (0.017)	16.19	Fraction = 1	476,321
	MTAG IQ3	0.316 (0.029)	10.23	Fraction = 1	411,947	0.363 (0.018)	13.58	Fraction = 1	411,947
	MTAG EA3	0.326 (0.029)	10.88	Fraction = 1	475,767	0.382 (0.018)	15.06	Fraction = 1	475,767
Lassosum	Common factor GWAS	0.327 (0.029)	10.84	$\lambda = 0.00127$; s = 0.2	32,257	0.408 (0.017)	17.2	$\lambda = 0.001$; s = 0.5	58,988
	MTAG IQ3	0.336 (0.020)	11.56	$\lambda = 0.001$; s = 0.5	103,554	0.386 (0.018)	15.42	$\lambda = 0.001$; s = 0.5	103,554
	MTAG EA3	0.310 (0.029)	9.97	$\lambda = 0.001$; s = 0.9	112,655	0.410 (0.017)	17.39	$\lambda = 0.001$; s = 0.5	76,252
		Test sets							
		intelligence age 12				**educational achievement age 12**			
Polygenic score approach	**Summary statistics**	**Beta (SE)**	**R2%**	**95% CI (lower/upper bound)**		**Beta (SE)**	**R2%**	**95% CI (lower/upper bound)**	
PRSice	Common factor GWAS	0.267 (0.024)	7.32	4.99	10.1	0.274 (0.020)	7.42	5.57	9.7
	MTAG IQ3	0.263 (0.024)	7.2	4.87	9.91	0.281 (0.020)	7.79	5.79	10.01
	MTAG EA3	0.266 (0.024)	7.26	4.9	10.88	0.261 (0.020)	6.72	4.87	8.92
Ldpred	Common factor GWAS	0.278 (0.024)	7.95	5.51	10.78	0.296 (0.020)	8.69	6.56	11.35
	MTAG IQ3	0.270 (0.023)	8.02	5.72	10.76	0.297 (0.020)	8.72	6.69	11.06
	MTAG EA3	0.280 (0.024)	8.06	5.66	10.8	0.288 (0.020)	8.27	6.2	10.93
Lassosum	Common factor GWAS	0.284 (0.024)	8.28	5.9	11.39	0.292 (0.020)	8.43	6.38	10.77
	MTAG IQ3	0.274 (0.024)	7.79	5.48	10.47	0.302 (0.020)	9.03	6.89	11.43
	MTAG EA3	0.287 (0.024)	8.38	5.96	11.3	0.292 (0.020)	8.43	6.27	10.76
SMTpred	IQ3	0.261 (0.025)	7.33	4.65	10.21	0.297 (0.020)	8.74	6.63	11.12
	EA3	0.267 (0.025)	7.59	4.95	10.49	0.283 (0.020)	7.89	5.88	10.13
		intelligence age 16				**educational achievement age 16**			
		Beta (SE)	**R2%**	**95% CI (lower/upper bound)**		**Beta (SE)**	**R2%**	**95% CI (lower/upper bound)**	
PRSice	Common factor GWAS	0.320 (0.032)	9.75	6.58	13.39	0.371 (0.019)	13.45	11.14	16.07
	MTAG IQ3	0.309 (0.032)	9.02	6.04	12.7	0.345 (0.019)	11.65	9.56	14.21
	MTAG EA3	0.317 (0.032)	9.5	6.4	13.05	0.379 (0.019)	14.06	11.66	16.67
Ldpred	Common factor GWAS	0.322 (0.031)	9.97	6.62	13.63	0.396 (0.018)	15.36	12.81	17.93
	MTAG IQ3	0.296 (0.032)	8.36	5.34	11.81	0.378 (0.019)	13.98	11.61	16.6
	MTAG EA3	0.280 (0.032)	7.77	4.88	11.4	0.395 (0.019)	15.25	12.54	18.02
Lassosum	Common factor GWAS	0.334 (0.032)	10.53	7.25	14.41	0.393 (0.019)	15.11	12.68	17.68
	MTAG IQ3	0.327 (0.032)	10.04	6.7	13.9	0.382 (0.019)	14.28	12.1	16.81
	MTAG EA3	0.332 (0.031)	10.58	7.19	14.5	0.403 (0.018)	15.91	13.38	18.42
SMTpred	IQ3	0.318 (0.031)	9.67	6.59	13.22	0.0397 (0.018)	15.45	13.09	18.1
	EA3	0.308 (0.031)	9.12	6.03	12.49	0.401 (0.018)	15.75	13.13	18.37

Note: significance test across all: p < 2e-16

Supplementary Table S4a. Bootstrapped R2 difference and 95% confidence intervals for multi-trait methods tested.

		Intelligence age 12			Intelligence age 16		
Multi-trait approach comparisor	**Summary statistics**	**Bootstrapped R2 difference**	**95% CI (lower/upper bound)**		**Bootstrapped R2 difference**	**95% CI (lower/upper bound)**	
Genomic SEM vs MTAG	IQ	0.005195566	-0.028898336	0.043116661	0.004721998	-0.012023	0.02106937
	EA	-0.000932435	-0.007487815	0.005488146	-0.000536255	-0.015058	0.01376225
Genomic SEM vs SMTpred	IQ	0.009886851	-0.028121942	0.049892766	0.008583246	-0.041507	0.05659638
	EA	0.00711602	-0.032066286	0.045629472	0.014250987	-0.032128	0.0618653
MTAG vs SMTpred	IQ	0.004691286	-0.033835197	0.044525301	0.003861247	-0.045346	0.05543739
	EA	0.008048454	-0.031312676	0.045999744	0.014787242	-0.032516	0.06579134

		Educational achievement age 12			Educational achievement age 16		
		Bootstrapped R2 difference	**95% CI (lower/upper bound)**		**Bootstrapped R2 difference**	**95% CI (lower/upper bound)**	
Genomic SEM vs MTAG	IQ	-0.005536917	-0.013945717	0.002838683	0.00823242	-0.002308	0.01970279
	EA	-0.000126251	-0.006123323	0.005733628	-0.007720841	-0.014386	-0.000281249
Genomic SEM vs SMTpred	IQ	-0.002970655	-0.01132872	0.005650657	-0.003527456	-0.013971	0.007059419
	EA	0.005402023	-0.003684722	0.014830101	-0.006145427	-0.017474	0.004328821
MTAG vs SMTpred	IQ	0.002566261	-0.008591518	0.014724984	-0.011759876	-0.022392	-0.001644046
	EA	0.005528274	-0.001956338	0.013552312	0.001575414	-0.00799	0.01109316

Note. Estimates refer to the Lassosum-based polygenic scores assosciations depicted in Figure 3 of the manuscirpt.

Supplementary Table S5. MTAG results.

	pre-MTAG		post-MTAG		
Trait	**Sample size**	**mean χ2**	**Sample size (equivalent)**	**mean χ2**	**maxFDR**
1. Intelligence (IQ3)	266,450	1.67	392,833	1.989	0.000607
2. Educational attainment (EA3)	766,345	2.235	887,057	2.434	2.06E-05
3. Age when completed full time education	226,899	1.258	874,284	1.992	0.036225
4. Income	96,900	1.089	948,643	1.874	0.157761
5. Time spent using computer	261,987	1.354	388,744	1.525	0.003028

Supplementary Table S6. Genomic SEM: model fit and indicator loadings.

Modelfit:

χ2	df	χ2 p-value	AIC	CFI	SRMR
45.899	5	9.52E-09	65.899	0.995	0.032

Factor Loadings and Residual Variance

	Unstandardized		Standardized	
	Estimate	SE	Estimate	SE
F1 =~ EA3	0.31017126	0.006074982	0.9260441	0.0181374
F1 =~ IQ3	0.35044795	0.008897467	0.7969648	0.02150712
F1 =~ Age Edu	0.26558754	0.006239188	0.9155075	0.01467975
F1 =~ Time PC	0.23375656	0.007166691	0.8819124	0.02703836
F1 =~ Income	0.18364141	0.006360183	0.5913828	0.02048178
EA ~~ EA3	0.01598009	0.002449829	0.1424423	0.02183715
IQ ~~ IQ3	0.07054708	0.005003863	0.3648469	0.02587836
Age Edu ~~ Age Edu	0.01362054	0.002863664	0.1618458	0.03402753
Time PC ~~ Time PC	0.01561277	0.005021845	0.2222307	0.07148034
Income ~~ Income	0.06270398	0.003160402	0.6502664	0.03277469

Note. F1 =~ trait : factor loading; trait ~~ trait : residual variance.
AgeEdu = Age when completed full time education; Time PC = Time spent using computer.

Supplementary Table S7. Multiple regression of univariate GPS calculated in LDp

	Intelligence at age 12				
	Beta	**SE**	**Pvalue**	**R2(%)**	**adj R2 (%)**
Polygenic score				8.9	8.6
EA3	0.194	0.023	**<.0001**		
IQ3	0.128	0.029	**<.0001**		
Income	-0.027	0.026	0.289		
AgePC	0.006	0.025	0.795		
AgeEdu	0.045	0.028	0.107		
	Intelligence at age 16				
	Beta	**SE**	**Pvalue**	**R2(%)**	**adj R2 (%)**
				10.49	9.92
EA3	0.224	0.042	**<.0001**		
IQ3	0.155	0.04	**<.0001**		
Income	-0.009	0.037	0.803		
AgePC	0.03	0.037	0.409		
AgeEdu	0.004	0.039	0.917		
	Educational achievement at age 12				
	Beta	**SE**	**Pvalue**	**R2(%)**	**adj R2 (%)**
				9.85	9.58
EA3	0.186	0.022	**<.0001**		
IQ3	0.174	0.027	**<.0001**		
Income	0.001	0.025	0.993		
AgePC	-0.003	0.024	0.874		
AgeEdu	-0.003	0.026	0.894		
	Educational achievement at age 16				
	Beta	**SE**	**Pvalue**	**R2 (%)**	**adj R2 (%)**
				16.98	16.76
EA3	0.3	0.026	**<.0001**		
IQ3	0.167	0.024	**<.0001**		
Income	-0.006	0.023	0.783		
AgePC	-0.052	0.023	**0.023**		
AgeEdu	0.053	0.025	**0.037**		

Note. AgeEdu = Age when completed full time education; Time PC = Time spent using computer.

Appendix 2– supplementary material for Chapter 3

Contents:

Supplemental figures

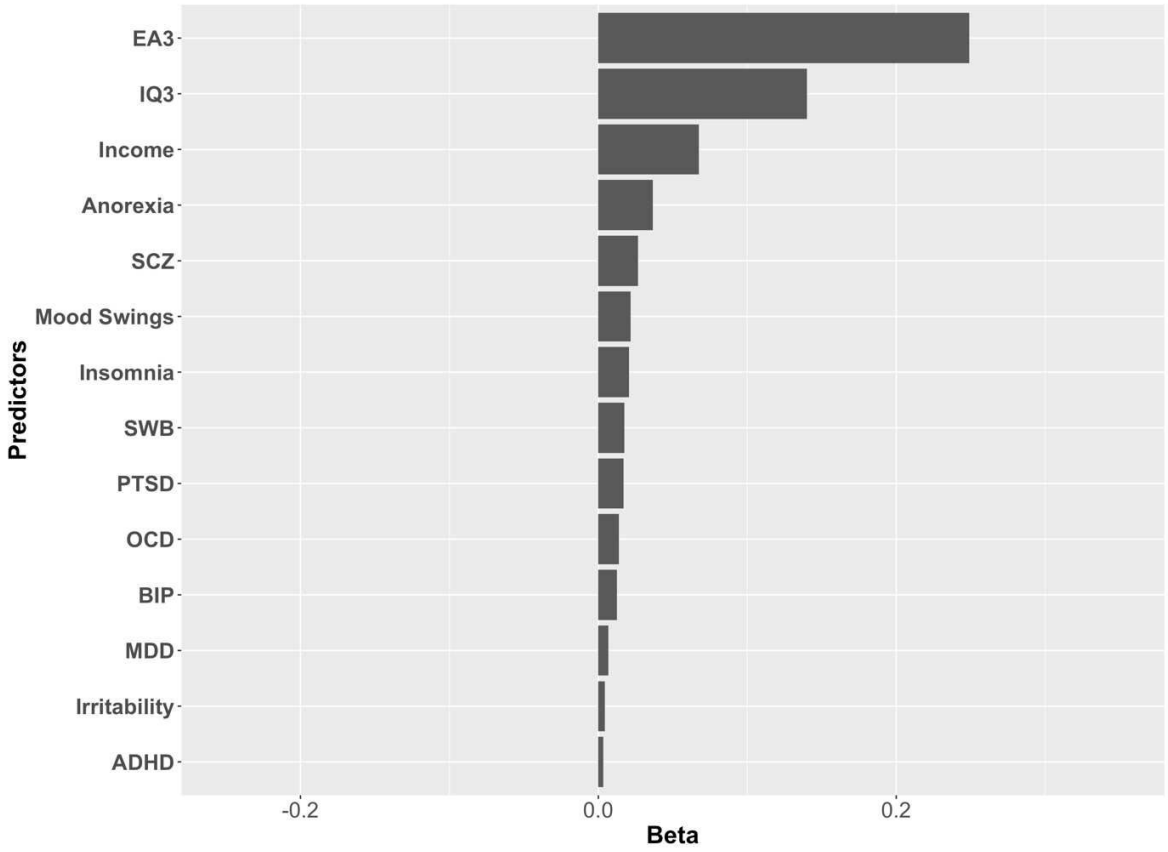


Figure S1. Polygenic score (G) model used in hold-out set prediction. Figure shows variables importance for the best G model selected via repeated cross-validation in the training set.
Note. ASD = Autism Spectrum Disorder, ADHD = Attention-Deficit Hyperactivity Disorder, BIP = Bipolar Disorder, EA3 = educational attainment, IQ3 = intelligence, MDD = Major Depressive Disorder, SWB = Subjective Well-Being, OCD = Obsessive Compulsive Disorder, PTSD = Post-Traumatic Stress Disorder, SCZ = Schizophrenia.

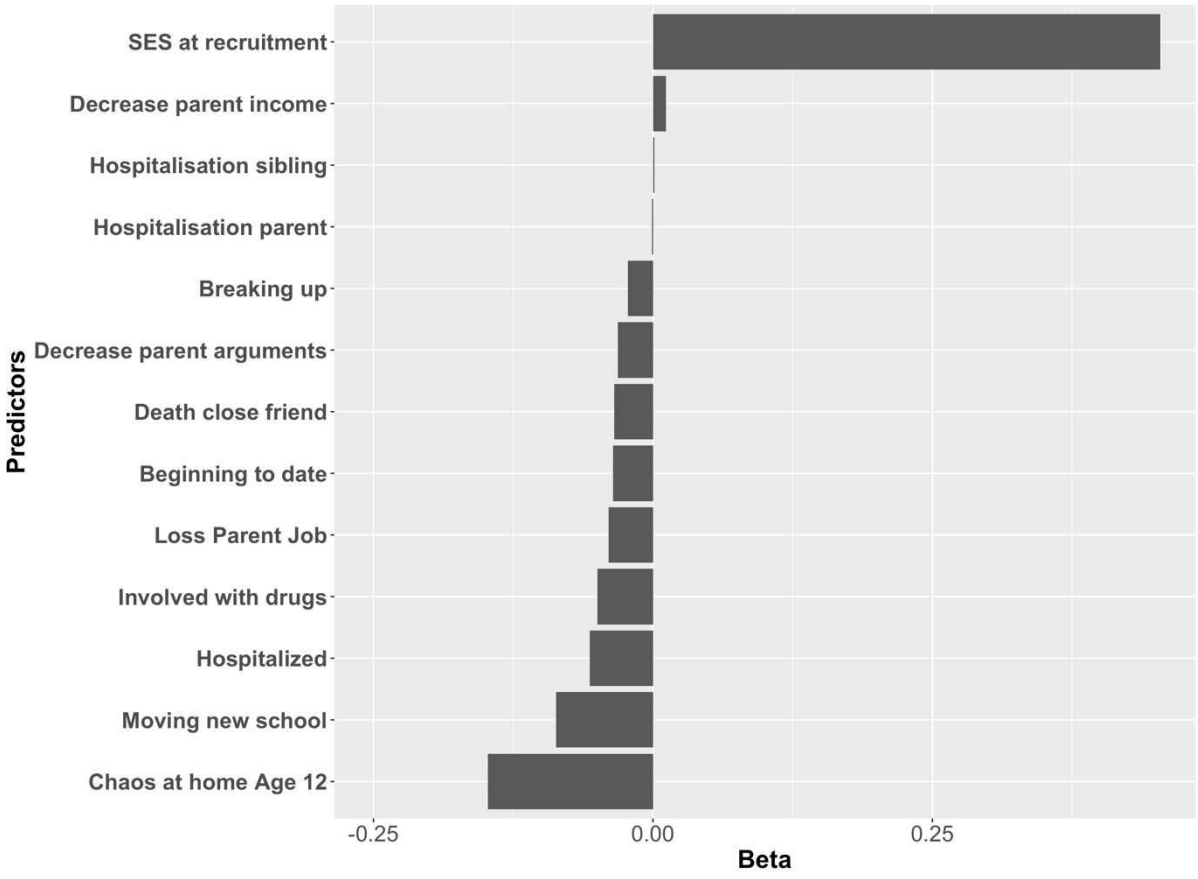


Figure S2. Environmental predictors (E) model used in hold-out set prediction. Figure shows variables importance for the best E model selected via repeated cross-validation in the training set.

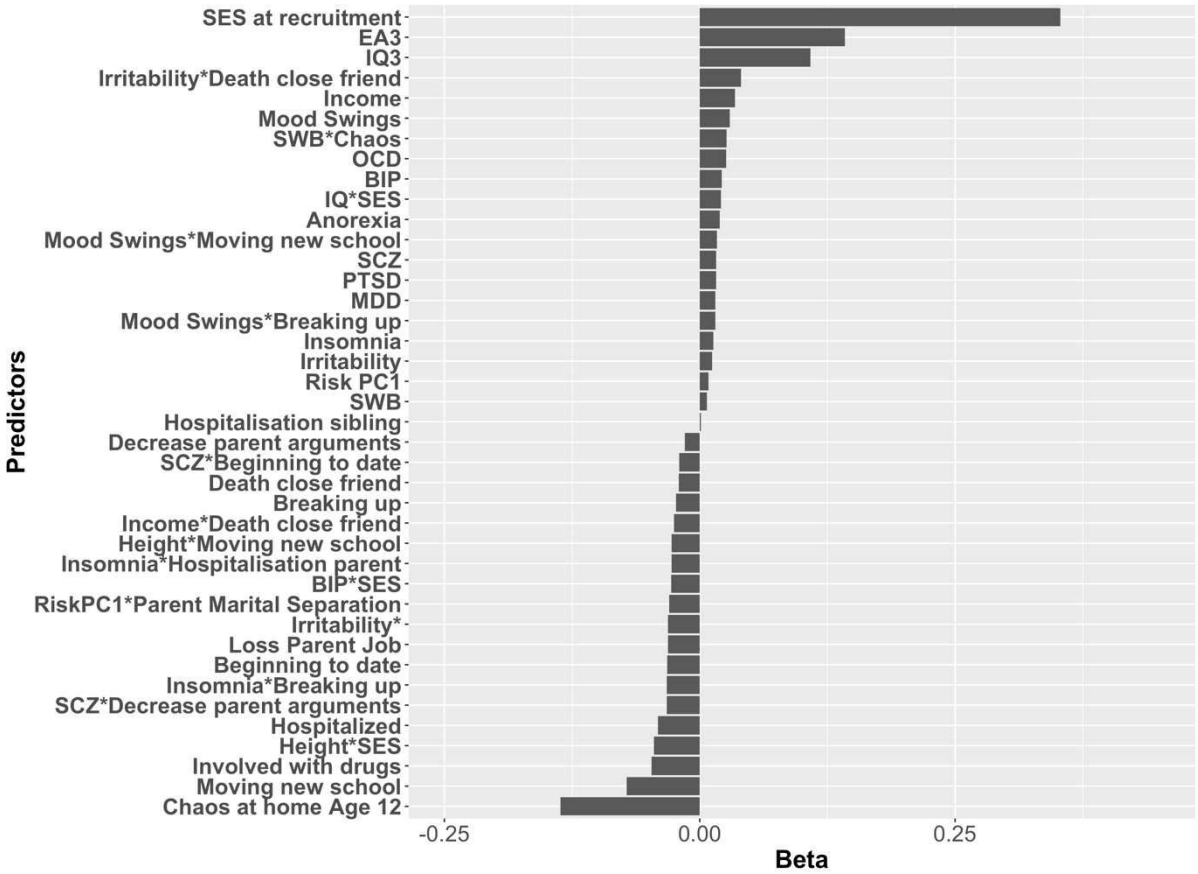


Figure S3. G*E model used in hold-out set prediction. Figure shows variables importance for the best G*E model selected via repeated cross-validation in the training set. **Note**. For interactions the first name refers to polygenic scores, the second name refers to environmental predictors. ASD = Autism Spectrum Disorder, ADHD = Attention-Deficit Hyperactivity Disorder, BIP = Bipolar Disorder, EA3 = educational attainment, IQ3 = intelligence, MDD = Major Depressive Disorder, SWB = Subjective Well-Being, OCD = Obsessive Compulsive Disorder, PTSD = Post-Traumatic Stress Disorder, SCZ = Schizophrenia.

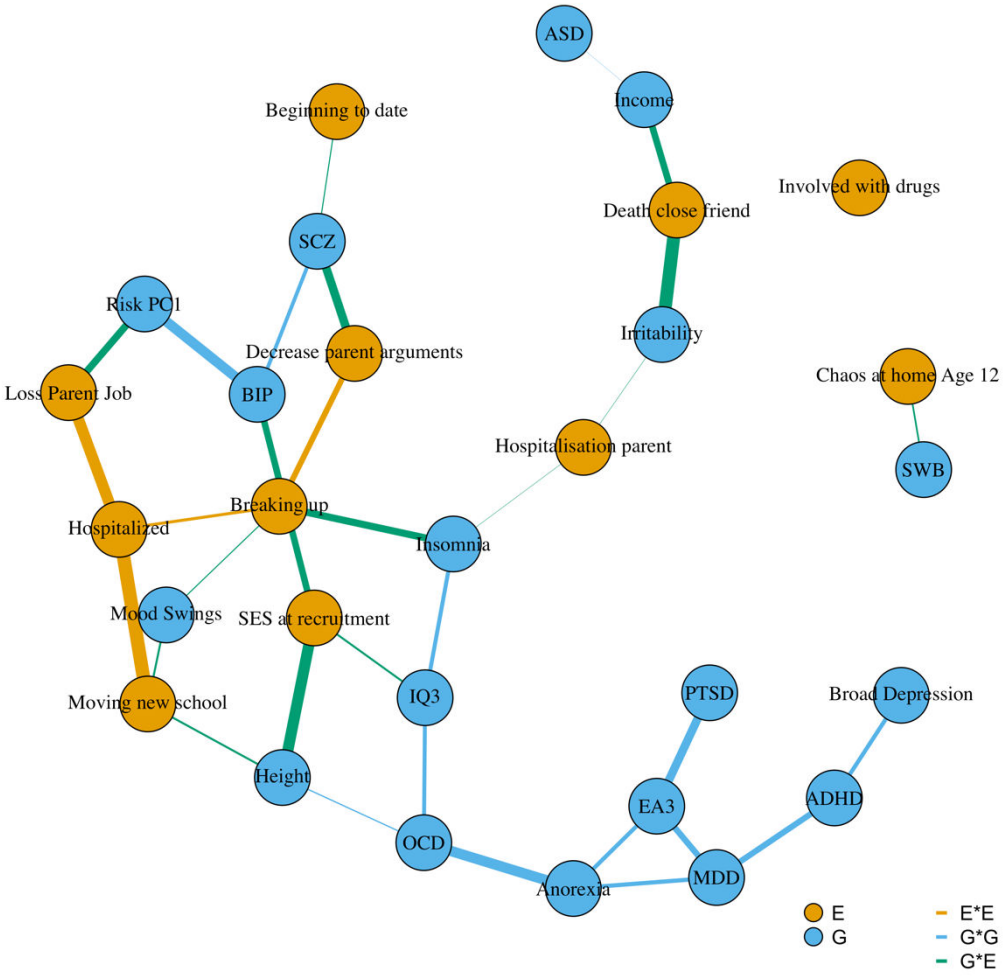


Figure S4. Interaction network of glinternet model. **Note.** E = Environmental measure, G = Genome-wide polygenic score. ASD = Autism Spectrum Disorder, ADHD = Attention-Deficit Hyperactivity Disorder, BIP = Bipolar Disorder, EA3 = educational attainment, IQ3 = intelligence, MDD = Major Depressive Disorder, SWB = Subjective Well-Being, OCD = Obsessive Compulsive Disorder, PTSD = Post-Traumatic Stress Disorder, Risk PC1 = first principal component of risky behaviours, SCZ = Schizophrenia.

Supplemental information

QC and genotyping protocol

DNA for 8,743 individuals (including 3,722 dizygotic co-twin samples) was extracted from saliva and buccal cheek swab samples and hybridized to HumanOmniExpressExome-8v1.2 genotyping arrays at the Institute of Psychiatry, Psychology and Neuroscience Genomics & Biomarker Core Facility, London, UK. The raw image data from the array were normalized, pre-processed, and filtered in GenomeStudio according to Illumina Exome Chip SOP v1.4. (http://confluence.brc.iop.kcl.ac.uk:8090/display/PUB/Production+Version%3A+Illumina+Exome+Chip+SOP+v1.4). In addition, prior to genotype calling, 919 multi-mapping SNPs and 501 samples with callrate <0.95 were removed. The ZCALL program was used to augment the genotype calling for samples and SNPs that passed the initial QC.

DNA from 3,747 samples was extracted from buccal cheek swabs and genotyped at Affymetrix, Santa Clara, California, USA. From this sample, 3,665 samples were successfully hybridized to AffymetrixGeneChip 6.0 SNP genotyping arrays (http://www.affymetrix.com/support/technical/datasheets/genomewide_snp6_datasheet.pdf) using experimental protocols recommended by the manufacturer (Affymetrix Inc., Santa Clara, CA). The raw image data from the arrays were normalized and pre-processed at the Wellcome Trust Sanger Institute, Hinxton, UK for genotyping as part of the Wellcome Trust Case Control Consortium 2 (https://www.wtccc.org.uk/ccc2/) according to the manufacturer's guidelines (http://www.affymetrix.com/support/downloads/manuals/genomewidesnp6_manual.pdf). Genotypes for the Affymetrix arrays were called using CHIAMO (https://mathgen.stats.ox.ac.uk/genetics_software/chiamo/chiamo.html).

After initial quality control and genotype calling, the same quality control was performed on the samples genotyped on the Illumina and Affymetrix platforms separately using PLINK (Chang et al., 2015; Purcell et al., 2007), R (R Core Team, n.d.), BCFtools (Li, 2011), and EIGENSOFT (Patterson, Price, & Reich, 2006; Price et al., 2006).

Samples were removed from subsequent analyses on the basis of call rate (<0.98), suspected non-European ancestry, heterozygosity, and relatedness other than dizygotic twin status. SNPs were excluded if the minor allele frequency was smaller than 0.5%, if more than 2% of genotype data were missing, or if the Hardy Weinberg *p*-value was lower than 10^{-5}. Non-autosomal markers and indels were removed. Association between SNP and the platform, batch, plate or well on which samples were genotyped was calculated; SNPs with an effect *p*-value < 10^{-4} were excluded. A total sample of 10,346 samples (including 3,320 dizygotic twin pairs and 7,026 unrelated individuals), with 7,289 individuals and 559,772 SNPs genotyped on Illumina and 3,057 individuals and 635,269 SNPs genotyped on Affymetrix remained after quality control.

Genotypes from the two platforms were separately phased using EAGLE2 (Loh et al., 2016), and imputed into the Haplotype Reference Consortium (release 1.1) using the Positional Burrows-Wheeler Transform method (Durbin, 2014) through the Sanger Imputation Service (McCarthy et al., 2016). Prior to merging, we excluded variants with info <0.75 and removed non-overlapping SNPs between platforms. After merging, we tested for minor allele frequency differences between platforms and removed SNPs with an effect p-value < 10^{-4}, and Hardy Weinberg p-value > 10^{-5}. Using these criteria, 7,363,646 genotyped and well-imputed SNPs were retained for the analyses.

We performed principal component analysis on a subset of 39,353 common (MAF > 5%), perfectly imputed (info = 1) autosomal SNPs, after stringent pruning to remove markers in linkage disequilibrium ($r^2 > 0.1$) and excluding high linkage disequilibrium genomic regions so as to ensure that only genome-wide effects were detected.

Supplemental Tables

Table S1. Descriptive statistics

Training set	**N = 2170**						**Frequency**	
	mean	**sd**	**min**	**max**	**skew**	**kurtosis**	**level = 0**	**level = 1**
sex							0.56	0.44
Age chaos at home	11.21	0.68	9.80	13.31	-0.29	-0.92		
Age life events were reported	16.32	0.67	14.92	18.76	-0.33	-1.03		
Age GCSE scores taken	16.31	0.29	15.83	16.75	-0.05	-1.25		
SES at recruitment	0.30	0.95	-2.18	2.40	-0.05	-0.82		
Chaos at home Age 12	3.88	2.05	0.00	12.00	0.53	0.32		
Moving new school			0.00	1.00			0.74	0.26
Loss Parent Job			0.00	1.00			0.91	0.09
Involved with drugs			0.00	1.00			0.95	0.05
Death close friend			0.00	1.00			0.78	0.22
Hospitalized			0.00	1.00			0.92	0.08
Breaking up			0.00	1.00			0.77	0.23
Hospitalisation sibling			0.00	1.00			0.93	0.07
Hospitalisation parent			0.00	1.00			0.92	0.08
Decrease parent income			0.00	1.00			0.93	0.07
Decrease parent arguments			0.00	1.00			0.77	0.23
Beginning to date			0.00	1.00			0.77	0.23
Educational Achievement	0.00	1.00	-3.80	1.80	-0.50	0.06		
Educational Achievement transformed	0.00	1.00	-3.26	3.41	0.00	-0.07		
ADHD	0.00	0.00	0.00	0.00	-0.57	-0.67		
Anorexia	31.91	2.92	21.58	41.66	0.05	0.04		
ASD	12.32	2.64	4.73	23.34	0.11	0.01		
BIP	9.17	3.34	-2.65	20.25	0.02	0.10		
BMI	0.54	0.59	-1.49	2.45	0.06	0.04		
Broad Depression	0.00	0.00	-0.02	0.00	-0.75	0.25		
EA3	-0.30	0.33	-1.26	0.88	0.15	-0.02		
Height	-0.21	0.41	-1.76	1.22	-0.02	0.24		
Income	-1.36	0.64	-3.33	0.87	-0.02	-0.14		
Insomnia	0.00	0.00	0.00	0.01	0.31	-0.42		
IQ3	-1.08	1.33	-6.46	3.56	0.04	0.13		
Irritability	-0.15	0.39	-1.34	1.17	0.03	-0.03		
MDD	-0.01	0.01	-0.04	0.01	-0.87	0.64		
Mood Swings	0.00	0.00	0.00	0.01	0.37	-0.30		
Neuroticism	-0.01	0.01	-0.03	0.01	-0.09	-0.79		
OCD	0.07	0.05	-0.09	0.25	0.12	-0.21		
PTSD	0.02	0.03	-0.07	0.11	-0.08	-0.02		
Risk PC1	0.00	0.00	0.00	0.00	-0.36	-0.25		
SCZ	5.50	1.91	-1.59	11.34	-0.11	0.01		
SWB	-0.54	0.29	-1.49	0.37	0.05	-0.03		
PC1	0.00	0.01	-0.04	0.05	0.03	1.50		
PC2	0.00	0.01	-0.03	0.06	0.65	2.47		
PC3	0.00	0.01	-0.04	0.04	-0.03	0.77		
PC4	0.00	0.01	-0.04	0.03	-0.04	0.65		
PC5	0.00	0.01	-0.04	0.04	-0.07	0.90		
PC6	0.00	0.01	-0.03	0.04	0.04	0.70		
PC7	0.00	0.01	-0.04	0.04	-0.07	0.56		
PC8	0.00	0.01	-0.04	0.04	-0.02	0.94		
PC9	0.00	0.01	-0.05	0.04	-0.15	1.06		
PC10	0.00	0.01	-0.04	0.04	-0.01	0.52		
Chip	1.56	0.50	1.00	2.00	-0.23	-1.95	0.44	0.56

Hold-out set	N = 542						Frequency	
	mean	sd	min	max	skew	kurtosis	level = 0	level = 1
sex	1.44	0.50	1.00	2.00	0.24	-1.94	0.56	0.44
Age chaos at home	11.23	0.72	9.84	12.76	-0.42	-1.00		
Age life events were reported	16.33	0.72	14.97	18.58	-0.34	-0.97		
Age GCSE scores taken	16.29	0.28	15.83	16.75	-0.02	-1.21		
SES at recruitment	0.30	0.96	-2.18	2.40	-0.06	-0.83		
Chaos at home Age 12	3.75	1.92	0.00	11.00	0.45	0.29		
Moving new school			0.00	1.00			0.73	0.27
Loss Parent Job			0.00	1.00			0.88	0.12
Involved with drugs			0.00	1.00			0.94	0.06
Death close friend			0.00	1.00			0.78	0.22
Hospitalized			0.00	1.00			0.91	0.09
Breaking up			0.00	1.00			0.76	0.24
Hospitalisation sibling			0.00	1.00			0.93	0.07
Hospitalisation parent			0.00	1.00			0.92	0.08
Decrease parent income			0.00	1.00			0.91	0.09
Decrease parent arguments			0.00	1.00			0.77	0.23
Beginning to date			0.00	1.00			0.75	0.25
Educational Achievement	0.00	1.00	-3.20	1.97	-0.55	-0.09		
Educational Achievement transformed	0.00	1.00	-2.78	2.81	-0.02	-0.24		
ADHD	0.00	0.00	0.00	0.00	-0.61	-0.74		
Anorexia	32.07	2.82	23.18	42.37	0.12	0.37		
ASD	12.43	2.57	4.27	19.21	0.10	-0.13		
BIP	9.24	3.31	-0.95	22.23	0.25	0.20		
BMI	0.55	0.63	-1.16	2.65	-0.02	-0.10		
Broad Depression	0.00	0.00	-0.02	0.00	-0.81	0.80		
EA3	-0.31	0.33	-1.39	0.67	-0.05	-0.06		
Height	-0.21	0.41	-1.36	1.15	0.11	0.16		
Income	-1.35	0.67	-3.45	0.57	-0.05	-0.23		
Insomnia	0.00	0.00	0.00	0.01	0.34	-0.53		
IQ3	-1.11	1.43	-4.99	3.68	0.14	-0.15		
Irritability	-0.15	0.41	-1.39	1.26	0.07	0.21		
MDD	0.00	0.01	-0.04	0.01	-0.93	1.04		
Mood Swings	0.00	0.00	0.00	0.01	0.29	-0.32		
Neuroticism	-0.01	0.01	-0.02	0.01	-0.04	-0.82		
OCD	0.07	0.05	-0.07	0.28	0.37	0.31		
PTSD	0.02	0.03	-0.05	0.09	0.00	-0.41		
Risk PC1	0.00	0.00	0.00	0.00	-0.17	-0.39		
SCZ	5.47	1.87	-0.75	11.56	0.08	0.10		
SWB	-0.54	0.27	-1.47	0.23	-0.14	-0.11		
PC1	0.00	0.01	-0.04	0.04	-0.01	1.43		
PC2	0.00	0.01	-0.03	0.03	0.08	0.44		
PC3	0.00	0.01	-0.03	0.03	-0.15	0.64		
PC4	0.00	0.01	-0.03	0.03	-0.04	0.86		
PC5	0.00	0.01	-0.03	0.04	0.10	1.26		
PC6	0.00	0.01	-0.03	0.03	-0.02	0.39		
PC7	0.00	0.01	-0.03	0.03	-0.16	0.36		
PC8	0.00	0.01	-0.03	0.03	0.22	0.85		
PC9	0.00	0.01	-0.03	0.03	0.11	0.27		
PC10	0.00	0.01	-0.03	0.04	0.23	1.08		
Chip	1.57	0.50	1.00	2.00	-0.27	-1.93	0.43	0.57

Table S2. Train and hold-out set fit indices

								95% Cis		Best tuning parameters	
	model	**median R^2 cv Train**	**sd R^2 cv Train**	**median RMSE cv Train**	**sd RMSE cv Train**	**RMSE test**	**R^2 test**	**0.025**	**0.975**	**Alpha**	**Lambda**
Elastic Net	G	0.167	0.042	0.912	0.037	0.901	0.183	0.127	0.236	0.200	0.062
	E	0.280	0.049	0.846	0.036	0.834	0.301	0.243	0.356	0.800	0.006
	E + G	0.331	0.050	0.819	0.037	0.798	0.361	0.304	0.416	0.300	0.034
	E + G + E*G	0.339	0.050	0.813	0.037	0.798	0.361	0.305	0.418	0.200	0.034
Glinternet	E + G + E*G + E*E + G*G	0.354	0.037	0.804	0.024	0.799	0.364	0.297	0.411	NA	0.001

Note: cv = cross-validation

Table S2b. Bootstrapped R^2 difference and 95% confidence intervals for prediction models tested.

Prediction model comparison	**median R^2 difference**	**95% Cis (lower/upper bound)**	
G+E - E	0.059	0.028	0.091
G+E - G	0.177	0.132	0.223
G*E - G+E	0.001	-0.012	0.013

Table S3. Statistical inference

names	Coef Naive	Pval Naive	Lci Naive	Uci Naive	Coef Test	Pval Test	Lci Test	Uci Test	Coef Cond	Pval Cond	Lci Cond	Uci Cond
chaos at home	-0.144	0.000	-0.179	-0.109	-0.200	0.000	-0.269	-0.131	-0.145	0.000	-0.175	-0.116
moving to new school	-0.076	0.000	-0.111	-0.041	-0.066	0.062	-0.135	0.003	-0.076	0.000	-0.105	-0.046
involved with drugs	-0.054	0.002	-0.089	-0.019	-0.065	0.069	-0.136	0.005	-0.055	0.002	-0.085	-0.025
hospitalized	-0.048	0.006	-0.083	-0.014	-0.001	0.980	-0.071	0.069	-0.049	0.112	-0.078	0.019
beginning to date	-0.035	0.056	-0.071	0.001	0.013	0.725	-0.060	0.087	-0.034	0.063	-0.065	0.003
loss of parent job	-0.036	0.042	-0.070	-0.001	-0.037	0.298	-0.108	0.033	-0.040	0.039	-0.070	-0.003
breaking up	-0.025	0.171	-0.061	0.011	-0.078	0.036	-0.151	-0.005	-0.025	0.168	-0.056	0.020
death of close friend	-0.028	0.120	-0.062	0.007	-0.028	0.432	-0.098	0.042	-0.029	0.107	-0.058	0.011
decrease parent arguments	-0.024	0.183	-0.058	0.011	-0.016	0.656	-0.086	0.054	-0.025	0.163	-0.054	0.019
Risk PC1	0.029	0.190	-0.014	0.071	0.013	0.768	-0.074	0.101	0.029	0.181	-0.026	0.066
SWB	0.016	0.356	-0.018	0.051	-0.001	0.977	-0.071	0.069	0.016	0.366	-0.046	0.045
Irritability	0.019	0.284	-0.016	0.054	-0.044	0.229	-0.116	0.028	0.018	0.310	-0.039	0.049
Insomnia	0.021	0.228	-0.013	0.056	0.072	0.042	0.003	0.142	0.021	0.233	-0.028	0.052
MDD	0.022	0.248	-0.015	0.058	-0.035	0.337	-0.107	0.037	0.037	0.208	-0.040	0.080
PTSD	0.025	0.161	-0.010	0.059	0.017	0.621	-0.052	0.086	0.025	0.163	-0.018	0.053
SCZ	0.024	0.230	-0.015	0.063	0.055	0.164	-0.023	0.133	0.025	0.217	-0.030	0.059
Anorexia	0.023	0.209	-0.013	0.058	-0.054	0.125	-0.124	0.015	0.023	0.204	-0.025	0.066
OCD	0.030	0.092	-0.005	0.064	0.051	0.148	-0.018	0.121	0.030	0.092	-0.008	0.059
BIP	0.024	0.210	-0.014	0.062	0.043	0.272	-0.034	0.119	0.024	0.227	-0.030	0.056
Mood Swings	0.047	0.032	0.004	0.089	0.024	0.592	-0.065	0.113	0.046	0.037	0.004	0.082
Income	0.040	0.057	-0.001	0.081	0.028	0.498	-0.054	0.111	0.042	0.048	0.000	0.076
IQ3	0.119	0.000	0.078	0.159	0.132	0.003	0.045	0.219	0.120	0.000	0.085	0.154
EA3	0.137	0.000	0.092	0.181	0.169	0.001	0.072	0.265	0.137	0.000	0.097	0.177
SES at recruitment	0.368	0.000	0.331	0.406	0.331	0.000	0.255	0.408	0.369	0.000	0.337	0.405

Note. Naïve = coefficients from multiple regression of selected variables in training set; Test = coefficients from multiple regression of selected variables in test set; Cond = coefficients estimated with a conditional probability from truncated distribution.

Table S4. Mediation models, bootstrapped estimates and 95% Confidence Intervals.

	Environmentally mediated effects			**Genetically mediated effects (genetic confounding)**		
	Beta	**lower CI**	**Upper CI**	**Beta**	**lower CI**	**Upper CI**
Y~X	0.26	0.189	0.333	0.451	0.387	0.511
Y~M	0.452	0.385	0.51	0.262	0.185	0.335
M~X	0.38	0.306	0.451	0.382	0.316	0.455
a*b	0.172	0.135	0.212	0.1	0.065	0.137
c+(a*b)	0.432	0.363	0.503	0.551	0.499	0.6
ab/total	0.399	0.31	0.497	0.182	0.117	0.247

Note. Y = outcome; X = predictor; M = mediator; Y~X = effects of X on Y while controlling for M; Y ~ M = effects of M on Y while controlling for X; M~X effects of X on M; a*b = indirect path; c+(a*b) = total effects; ab/total = percentage of effect mediated.

Table S5. List of interactions between significant main effects identified through Glinternet

Main effect 1	Main effect 2	Weights
SES at recruitment	BIP	-0.007
SES at recruitment	Height	-0.013
SES at recruitment	IQ3	0.002
Chaos at home Age 12	SWB	0.002
Loss Parent Job	Hospitalized	-0.013
Loss Parent Job	Risk PC1	-0.008
Death close friend	Income	-0.008
Death close friend	Irritability	0.015
Hospitalized	Breaking up	-0.003
Hospitalized	Moving new school	-0.015
Breaking up	Decrease parent arguments	-0.007
Breaking up	Insomnia	-0.008
Breaking up	Mood Swings	0.001
Hospitalisation parent	Insomnia	0.000
Hospitalisation parent	Irritability	0.000
Decrease parent arguments	SCZ	-0.010
Beginning to date	SCZ	-0.001
Moving new school	Height	-0.002
Moving new school	Mood Swings	0.002
ADHD	Broad Depression	-0.005
ADHD	MDD	-0.007
Anorexia	EA3	0.005
Anorexia	MDD	0.005
Anorexia	OCD	0.012
ASD	Income	0.000
BIP	Risk PC1	0.011
BIP	SCZ	-0.004
EA3	MDD	0.007
EA3	PTSD	-0.010
Height	OCD	0.001
Insomnia	IQ3	0.005
IQ3	OCD	-0.004

Note. Interactions are listed in order of discovery.

Table S5b. List of GxE interactions introduced in Elastic Net model

Environment	Polygenic score	weight
SES at recruitment	BIP	-0.007
SES at recruitment	Height	-0.013
SES at recruitment	IQ3	0.002
Chaos at home Age 12	SWB	0.002
Loss Parent Job	Risk PC1	-0.008
Death close friend	Income	-0.008
Death close friend	Irritability	0.015
Breaking up	Insomnia	-0.008
Breaking up	Mood Swings	0.001
Hospitalisation parent	Insomnia	0.000
Hospitalisation parent	Irritability	0.000
Decrease parent arguments	SCZ	-0.010
Beginning to date	SCZ	-0.001
Moving new school	Height	-0.002
Moving new school	Mood Swings	0.002

Table S6. GWAS Summary statistics

Trait	Cases	Controls	Sample Size	h2	se	Download link for summary stats
ADHD	20183	35191	55374	0.220	0.01	http://www.med.unc.edu/pgc/results-and-downloads/data-use-agreement-forms/copy_of_ADHDEA2017_data_download_agreement
Anorexia Nervosa	16992	55525	72517	0.11-.17	0.01	https://www.med.unc.edu/pgc/download-results/ed/
Autism Spectrum Disorder	18381	27969	46350	0.118	0.01	https://www.med.unc.edu/pgc/results-and-downloads
Bipolar Disorder	20352	31358	51710	0.200	0.3	https://www.dropbox.com/s/kzg4hhttihva95d/2395_4.gwas.imputed_v3.male.tsv.bgz?dl=0 -O 2395_4.gwas.imputed_v3.male.tsv.bgz
BMI			681275	0.246	0.037	https://portals.broadinstitute.org/collaboration/giant/images/c/c8/Meta-analysis_Locke_et_al%2BUKBiobank_2018_UPDATED.txt.gz
Broad depression			500199	0.089	0.003	https://datashare.is.ed.ac.uk/handle/10283/3203
Years of Education (EA3) excluding 23andMe			766345	0.106	0.0026	https://www.dropbox.com/s/y3tkhblvbrzilzl/GWAS_EduYears_excl23andMe.txt?dl=0
Standing Height			693529	0.483	0.037	https://portals.broadinstitute.org/collaboration/giant/images/6/63/Meta-analysis_Wood_et_al%2BUKBiobank_2018.txt.gz
Household Income			286301	0.074	0.033	https://www.lothianbirthcohort.ed.ac.uk/content/gwas-summary-data
Insomnia			113006	0.110	0.009	http://ctg.cncr.nl/software/summary_statistics
Intelligence (IQ3) excluding TEDS			266453	0.190	0.01	correspondence
Irritability	90282	232386	322668	0.117	0.0122	https://www.dropbox.com/s/r8tcj6ppl7lpq28/1940.assoc.tsv.gz?dl=0%20-O%201940.assoc.tsv.gz
MDD no 23andme	59851	113154	173005	0.087	0.004	https://www.med.unc.edu/pgc/results-and-downloads/downloads
Mood swings	148601	180827	329428	0.108	0.00637	https://www.dropbox.com/s/4on2uuwif6dg5y5/1920.assoc.tsv.gz?dl=0%20-O%201920.assoc.tsv.gz
Neuroticism			329821	0.088	0.0068	https://www.thessgac.org/data
OCD	2688	7037	9725	0.280	0.04	http://www.med.unc.edu/pgc/results-and-downloads/data-use-agreement-forms/pgc-ocd_aug2017.zip%20data_download_agreement
PTSD	2424	7113	9537	0.180	0.06	https://www.med.unc.edu/pgc/results-and-downloads/data-use-agreement-forms/PTSD%20EA_data_download_agreement
Risk - First PC from automobile speading, drinks per week, ever smoker, sexual partners			315894	0.156	0.004	https://www.thessgac.org/data
Schizophrenia	40675	64643	105318	0.240	0.0068	http://walters.psycm.cf.ac.uk
Well-being			298420	0.047	0.004	http://ssgac.org/documents/SWB_Full.txt.gz

Table S7. Parameter tuning for lassosum GPS

	Tuning parameters			
GPS names	**s**	**lambda**	**N best beta**	**R^2% validation**
ADHD	0.5	0.029763514	1	0.0052
Anorexia	0.9	0.004281332	233163	0.7482
ASD	0.9	0.001	386670	0.1524
BIP	1	0.004281332	197683	0.3387
BMI	1	0.006951928	3672	0.0741
Broad Depression	0.9	0.011288379	20	0.0655
EA3	0.5	0.001	149650	14.4394
Height	0.2	0.001	135386	0.3581
Income	0.9	0.002069138	143019	6.328
Insomnia	0.2	0.018329807	4	0.2378
IQ3	0.9	0.001	304122	9.1025
Irritability	0.5	0.001	193689	0.0133
MDD	0.2	0.011288379	49	0.0561
Mood Swings	0.2	0.011288379	19	0.0159
Neuroticism	0.2	0.014384499	35	0.0028
OCD	1	0.037926902	159	0.0907
PTSD	0.9	0.037926902	117	0.1974
Risk PC1	0.2	0.018329807	26	0.0078
SCZ	0.9	0.001623777	239742	0.0863
SWB	0.2	0.001274275	109447	0.2172

Supplementary References

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience, 4(1), 7. http://doi.org/10.1186/s13742-015-0047-8

Durbin, R. (2014). Efficient haplotype matching and storage using the positional Burrows–Wheeler transform (PBWT). Bioinformatics, 30(9), 1266–1272. http://doi.org/10.1093/bioinformatics/btu014

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics, 27(21), 2987–2993. http://doi.org/10.1093/bioinformatics/btr509

Loh, P.-R., Danecek, P., Palamara, P. F., Fuchsberger, C., A Reshef, Y., K Finucane, H., et al. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. Nature Genetics, 48(11), 1443–1448. http://doi.org/10.1038/ng.3679

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nature Genetics, 48(10), 1279–1283. http://doi.org/10.1038/ng.3643

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. PLOS Genetics, 2(12), e190. http://doi.org/10.1371/journal.pgen.0020190

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics, 38(8), 904–909. http://doi.org/10.1038/ng1847

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. American Journal of Human Genetics, 81(3), 559–575. http://doi.org/10.1086/519795

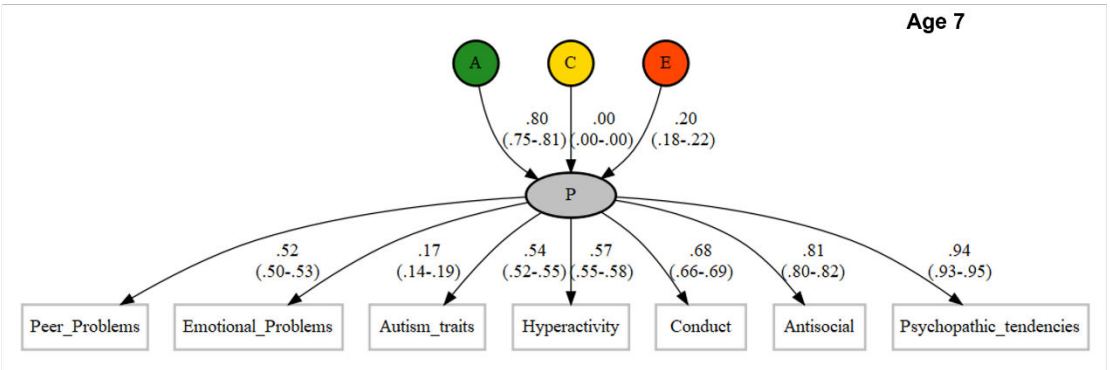
R Core Team. (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Retrieved from https://www.r-project.org

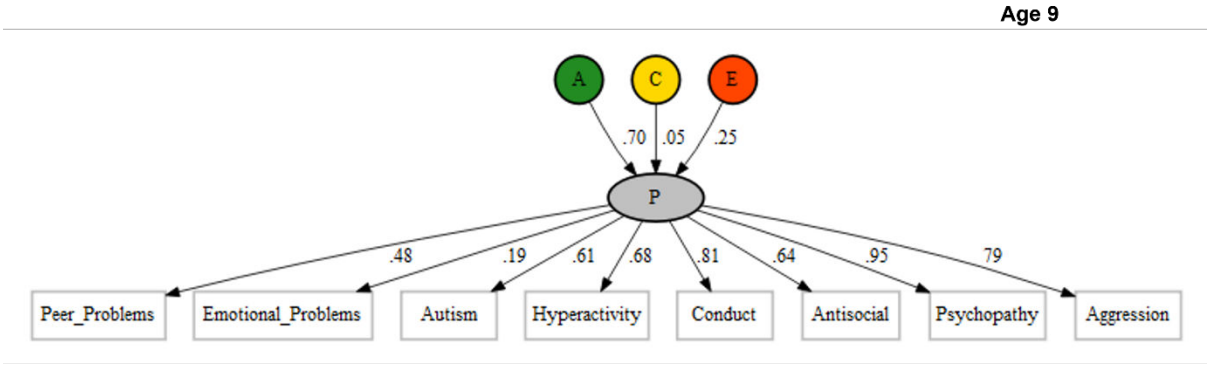
Appendix 3 – Supplementary material for Chapter 4

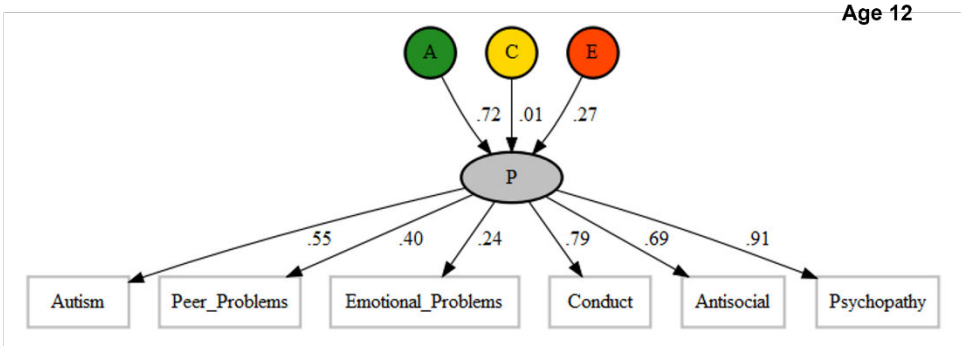
Contents:

Supplementary Figure S1 Common pathway twin models for teacher-rated and child-rated measures by age. *Note:* The ACE variance decomposition results for the common factor are presented in the top half of each figure, and the factor loadings of observed psychopathology variables on p are presented in the bottom half. See Supplementary Tables 1 and 2 for additional model parameters and model fit statistics, respectively. Also note that Antisocial = prosocial SDQ scale reversed; and psychopathy = APSD scale.

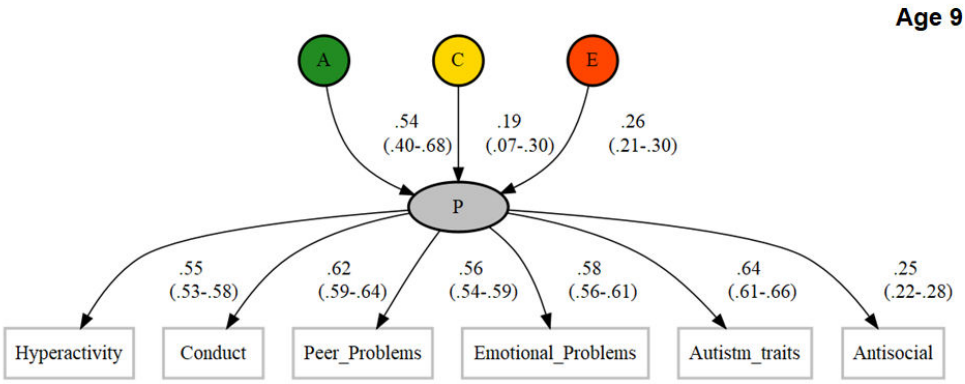
Teacher-rated p factor.

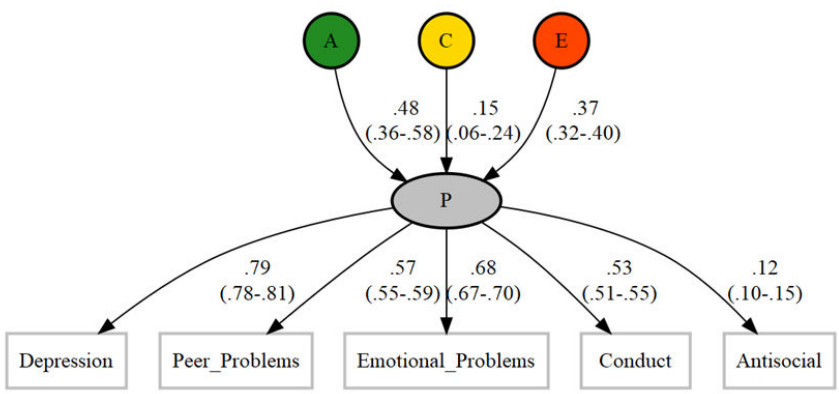


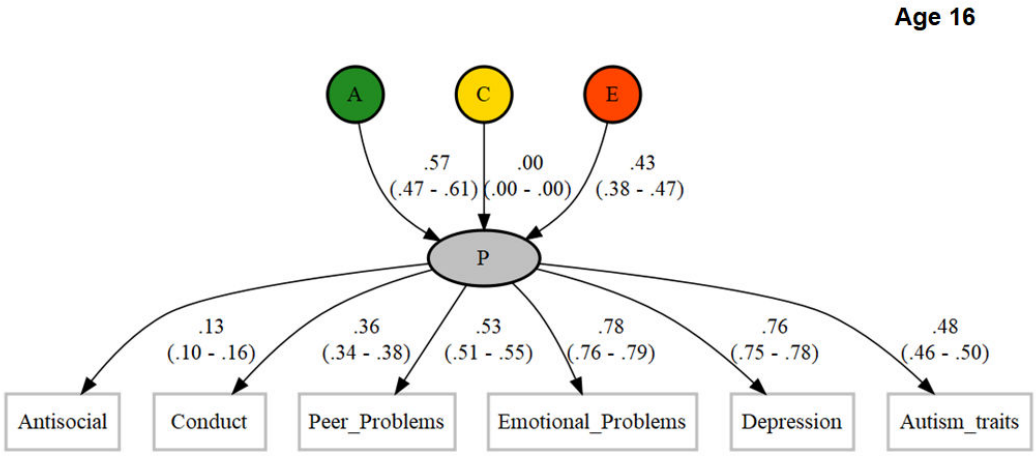




Child-rated p factor.







Supplementary Figure S2. Comparison of twin heritability estimates from common pathway models. Error bars are 95% CIs.

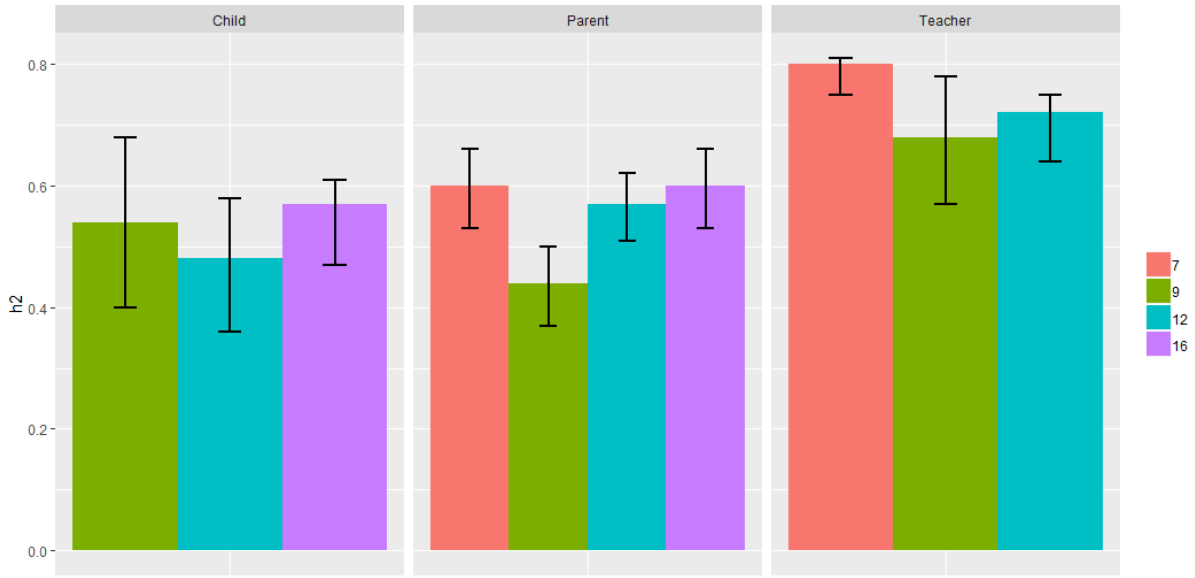


Figure S3. Shared environmental influences on p (parent-rated) across age, derived from longitudinal twin model-fitting (Cholesky decomposition)

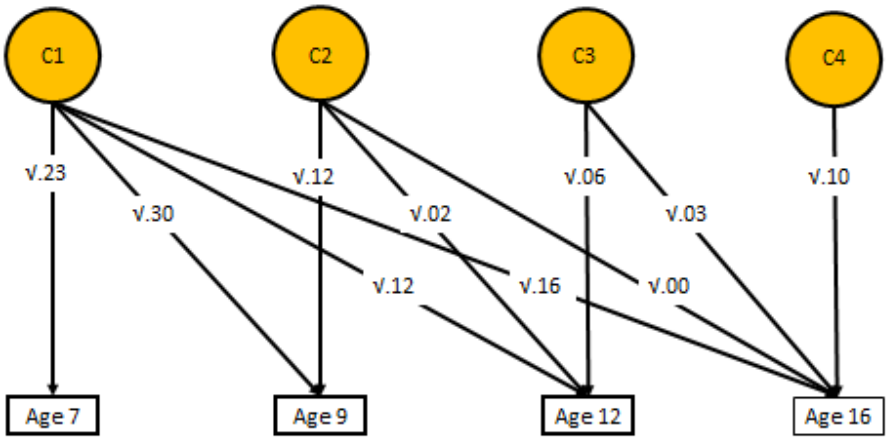


Figure S4. Unique environmental influences on p (parent-rated) across age, derived from longitudinal twin model-fitting (Cholesky decomposition)

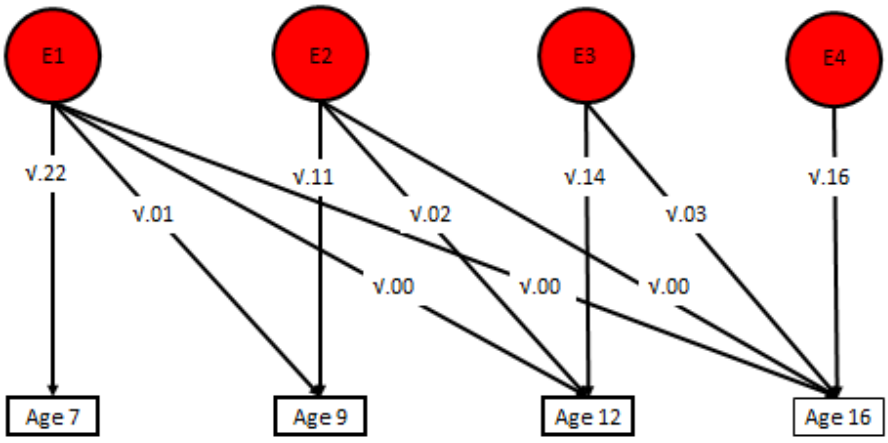
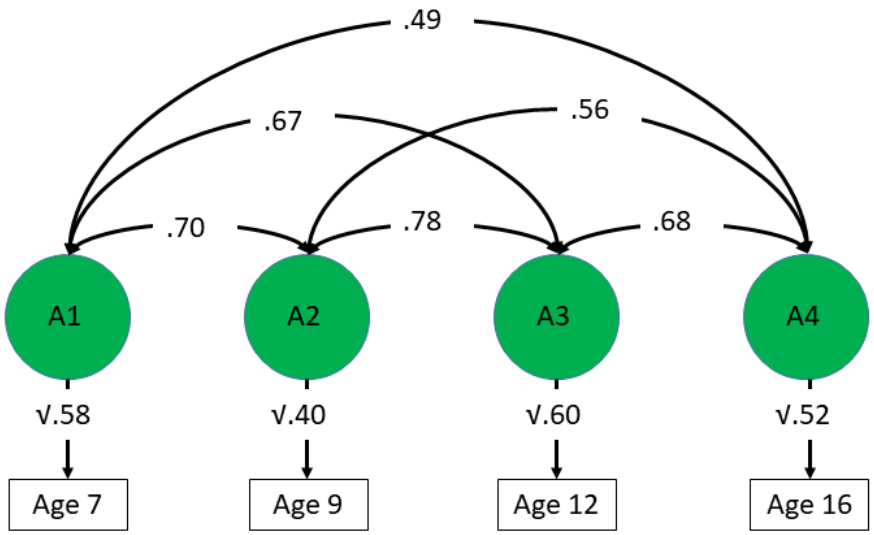


Figure S5. Correlated factor solution of the longitudinal Cholesky decomposition. The figure shows genetic correlations between and univariate heritability of phenotypic p, defined by the first unrotated principal component of psychopathology measures, using parent-reported data at age 7, 9, 12 and 16.



Supplementary Figures S6- to S15: Phenotypic correlations among psychopathology measures used to construct phenotypic p factors. *Darker blue indicates stronger positive correlation.*

Figure S6. Phenotypic correlations between parent-rated traits at age 7. Note prosocial was reversed.

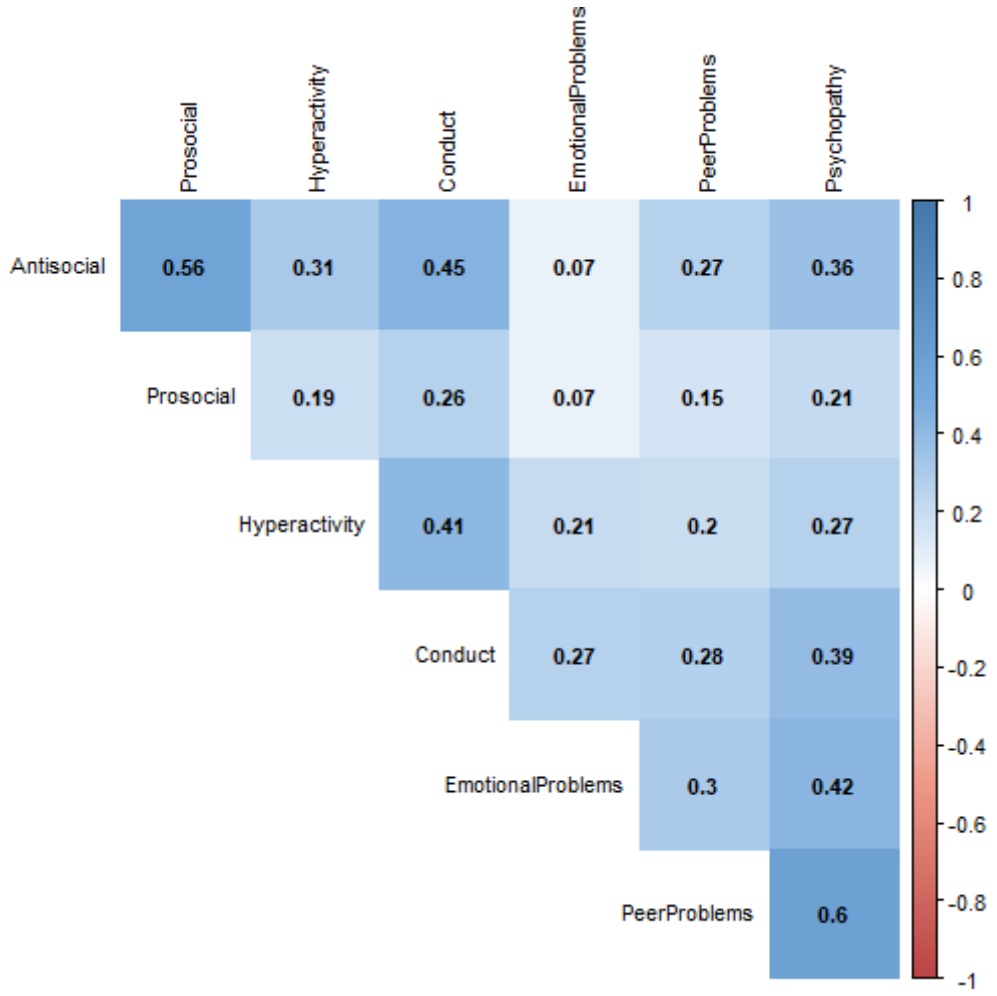


Figure S7. Phenotypic correlations between Teacher rated psychiatric traits at age 7. Note prosocial was reversed.

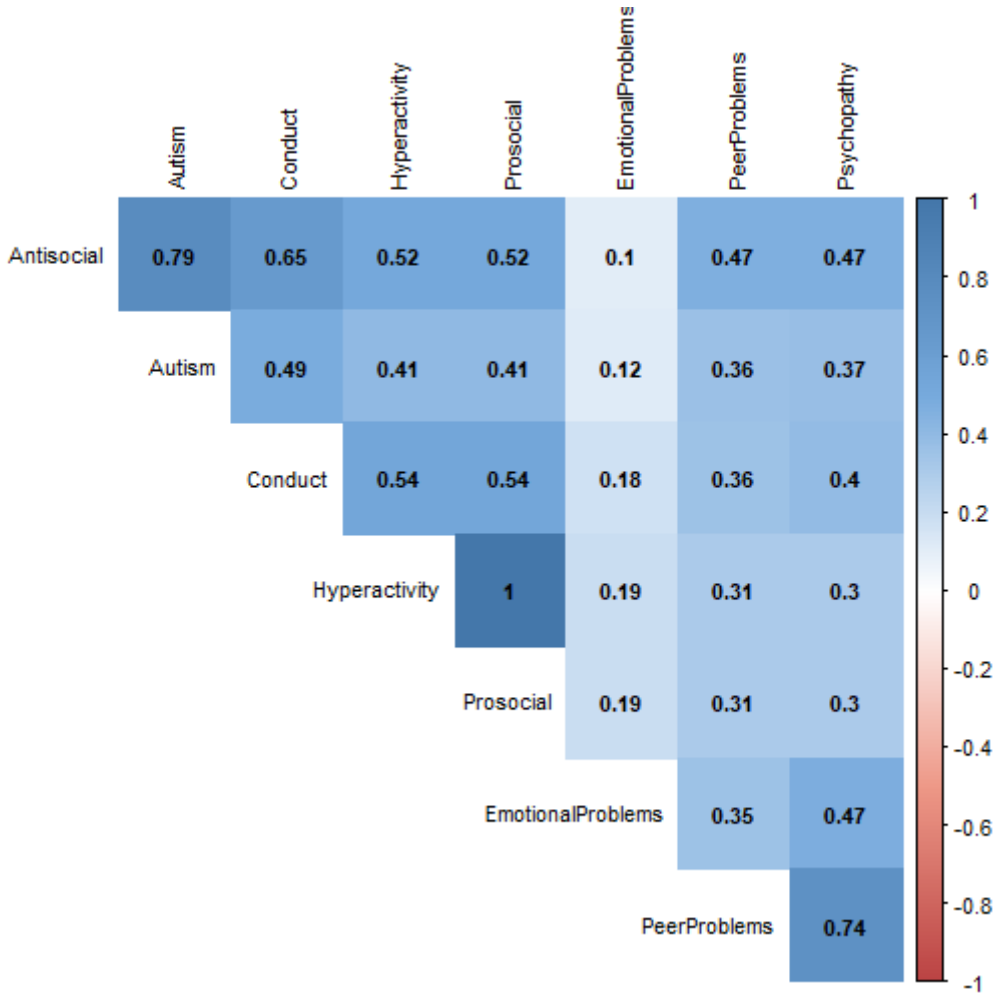


Figure S8. Phenotypic correlations between Child rated psychiatric traits at age 9.

Figure S9. Phenotypic correlations between Parent rated psychiatric traits at age 9.

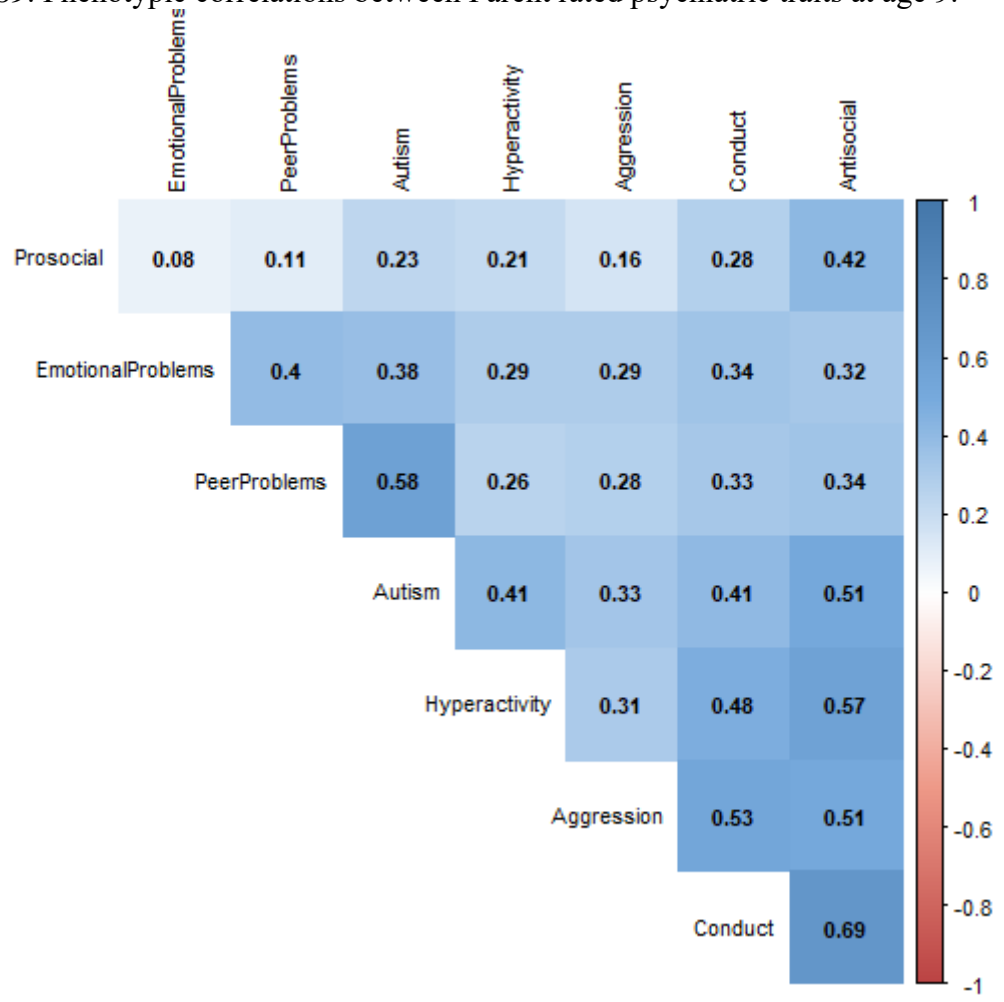


Figure S10. Phenotypic correlations between Teacher rated psychiatric traits at age 9.

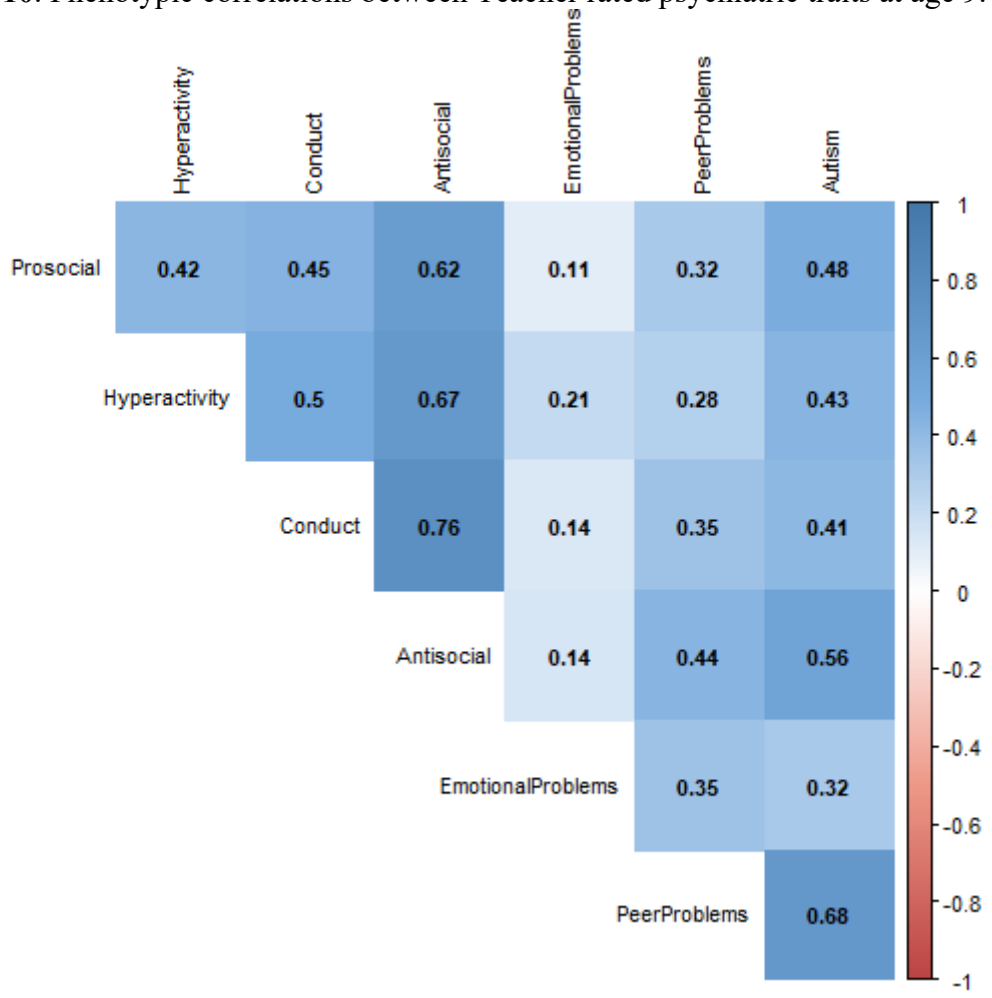


Figure S11. Phenotypic correlations between Parent rated psychiatric traits at age 12.

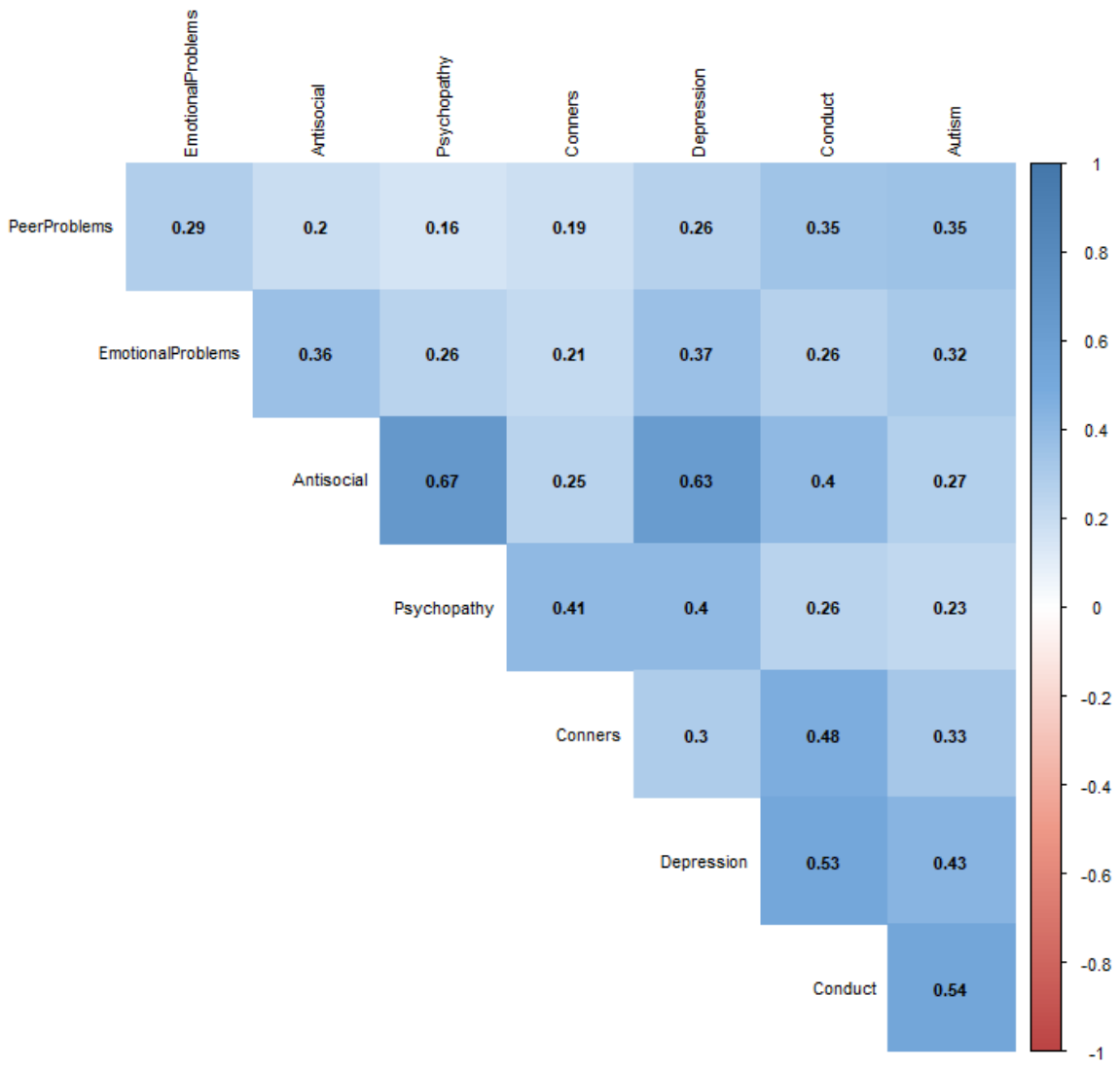


Figure S12. Phenotypic correlations between Teacher rated psychiatric traits at age 12.

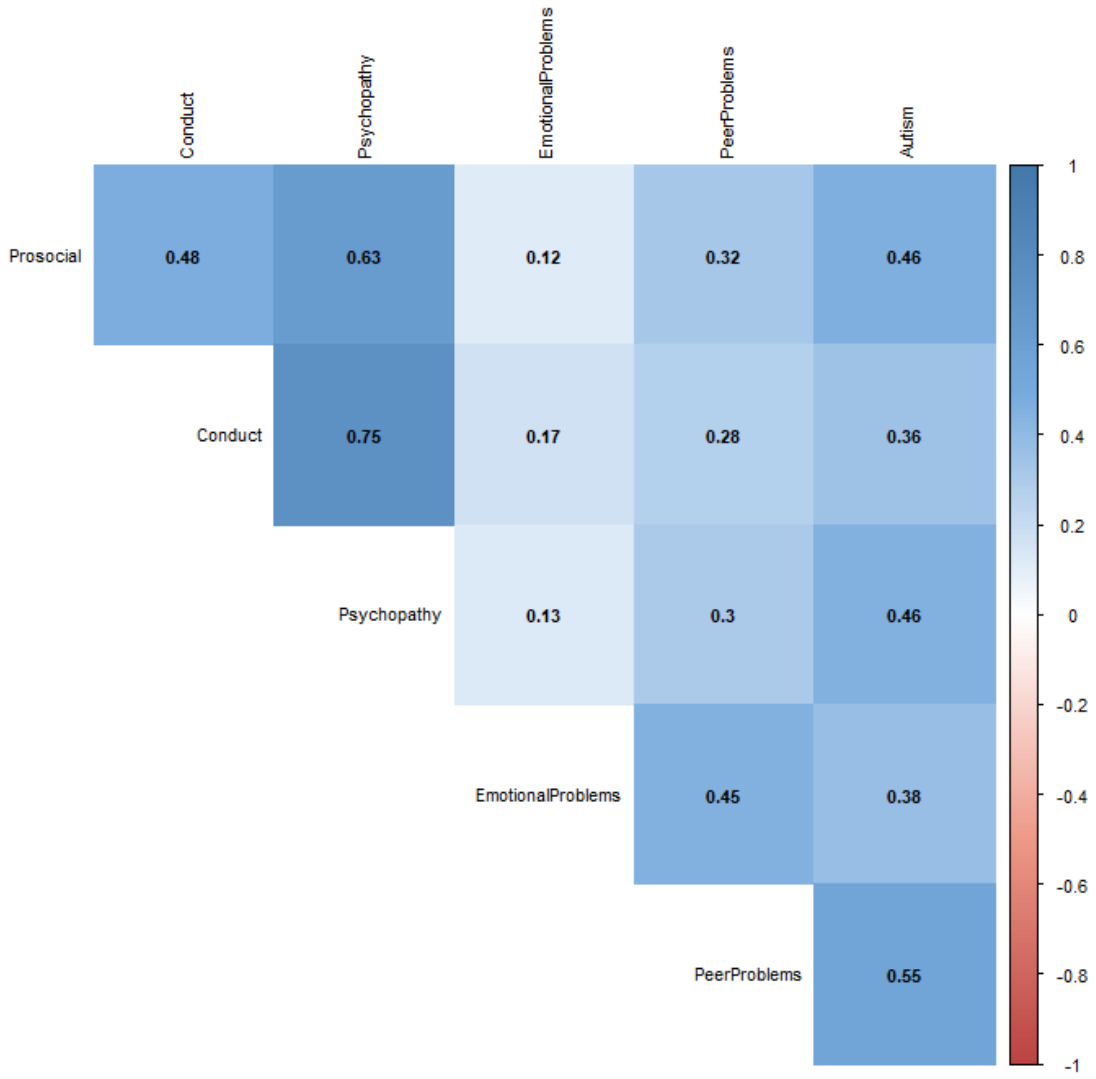


Figure S13. Phenotypic correlations between Child rated psychiatric traits at age 12.

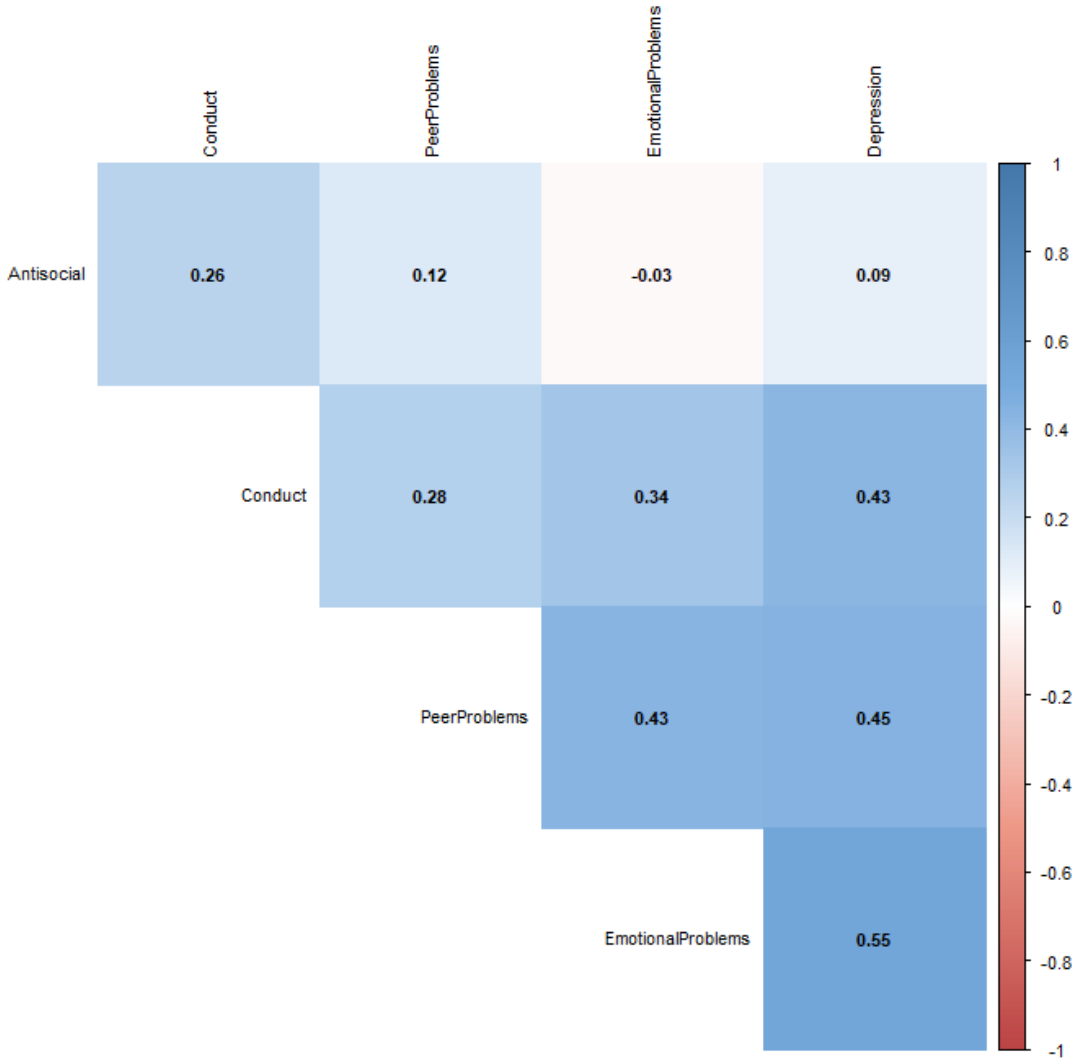


Figure S14. Phenotypic correlations between Parent rated psychiatric traits at age 16.

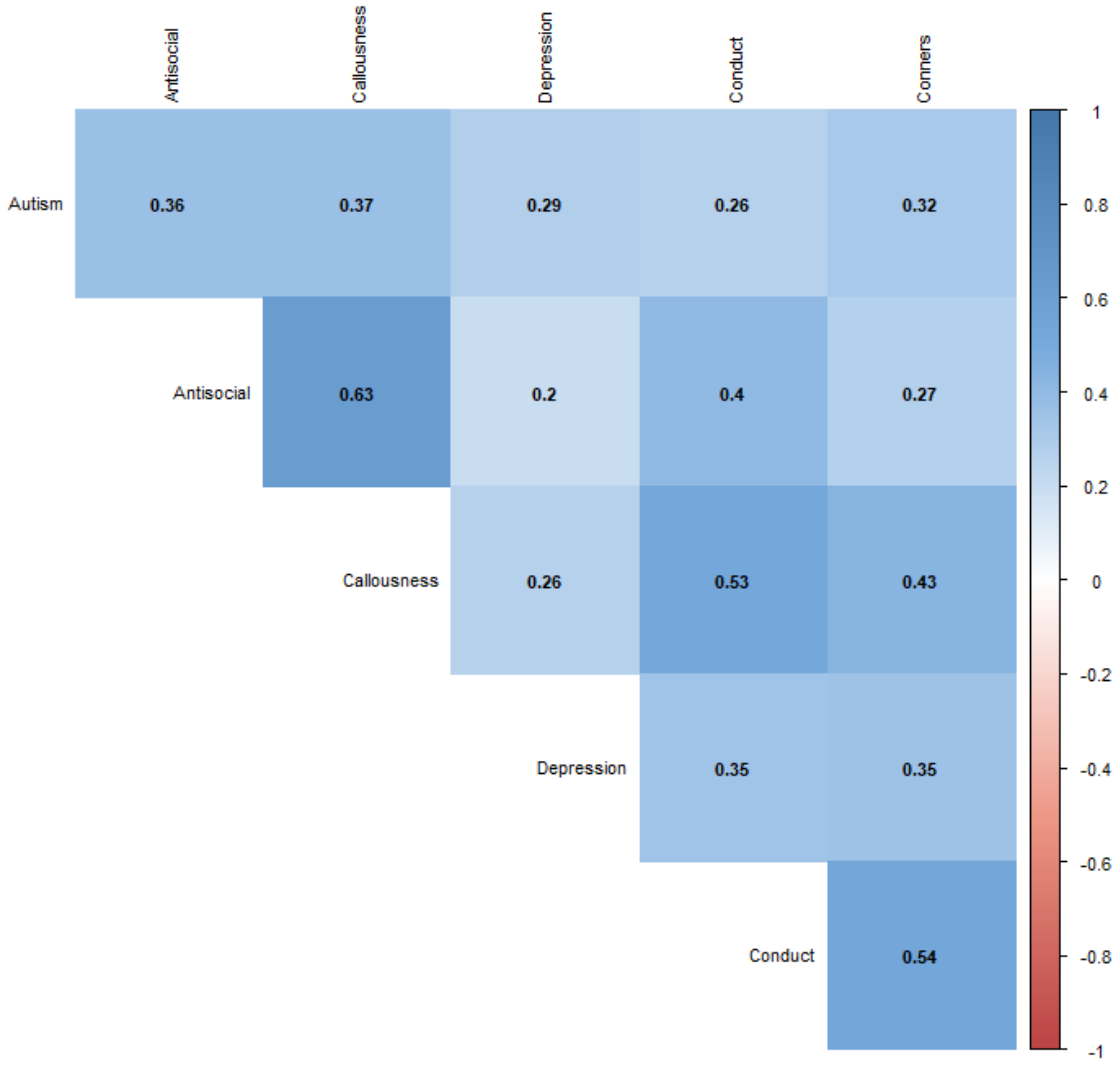


Figure S15. Phenotypic correlations between Child rated psychiatric traits at age 16.

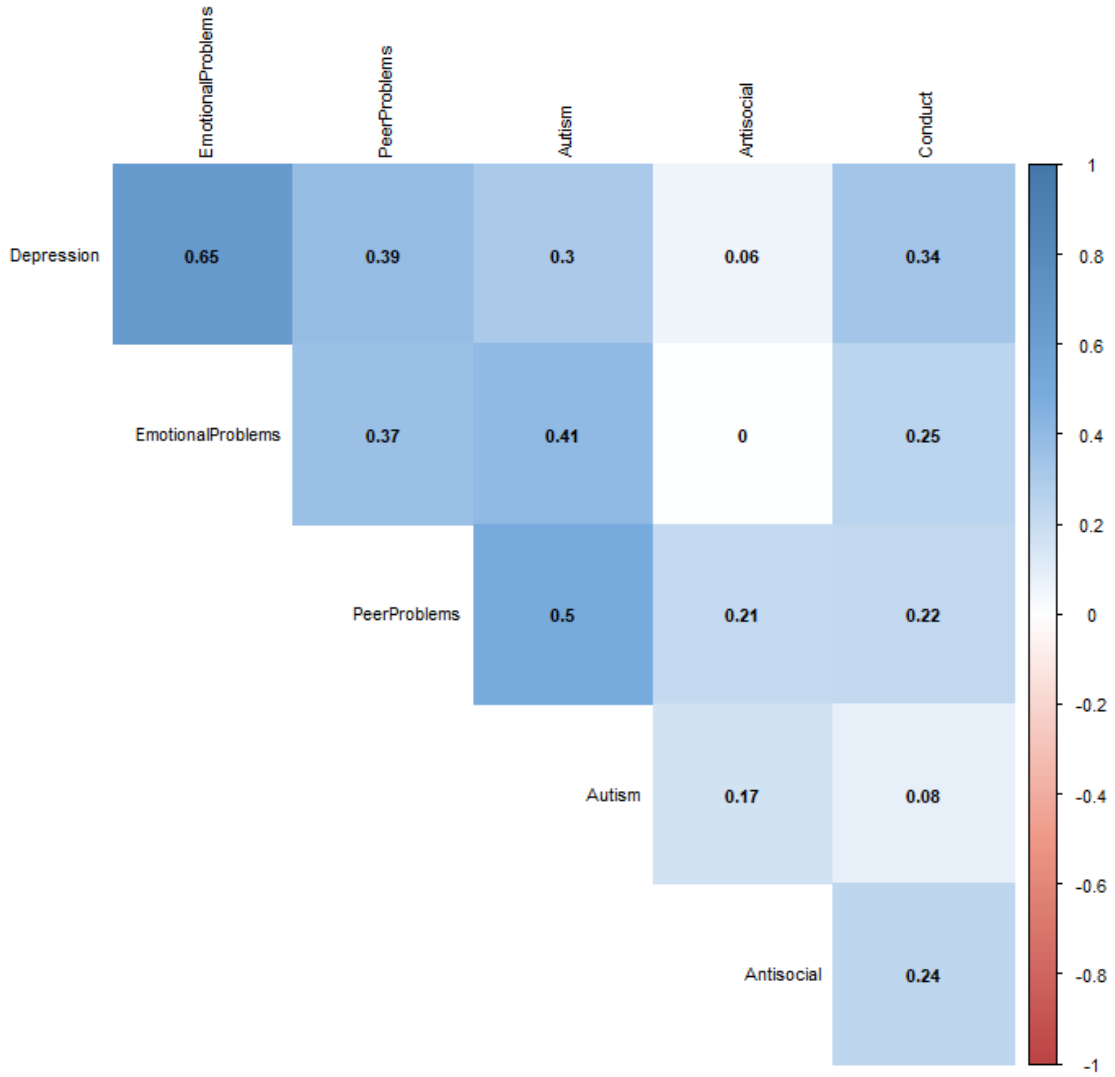
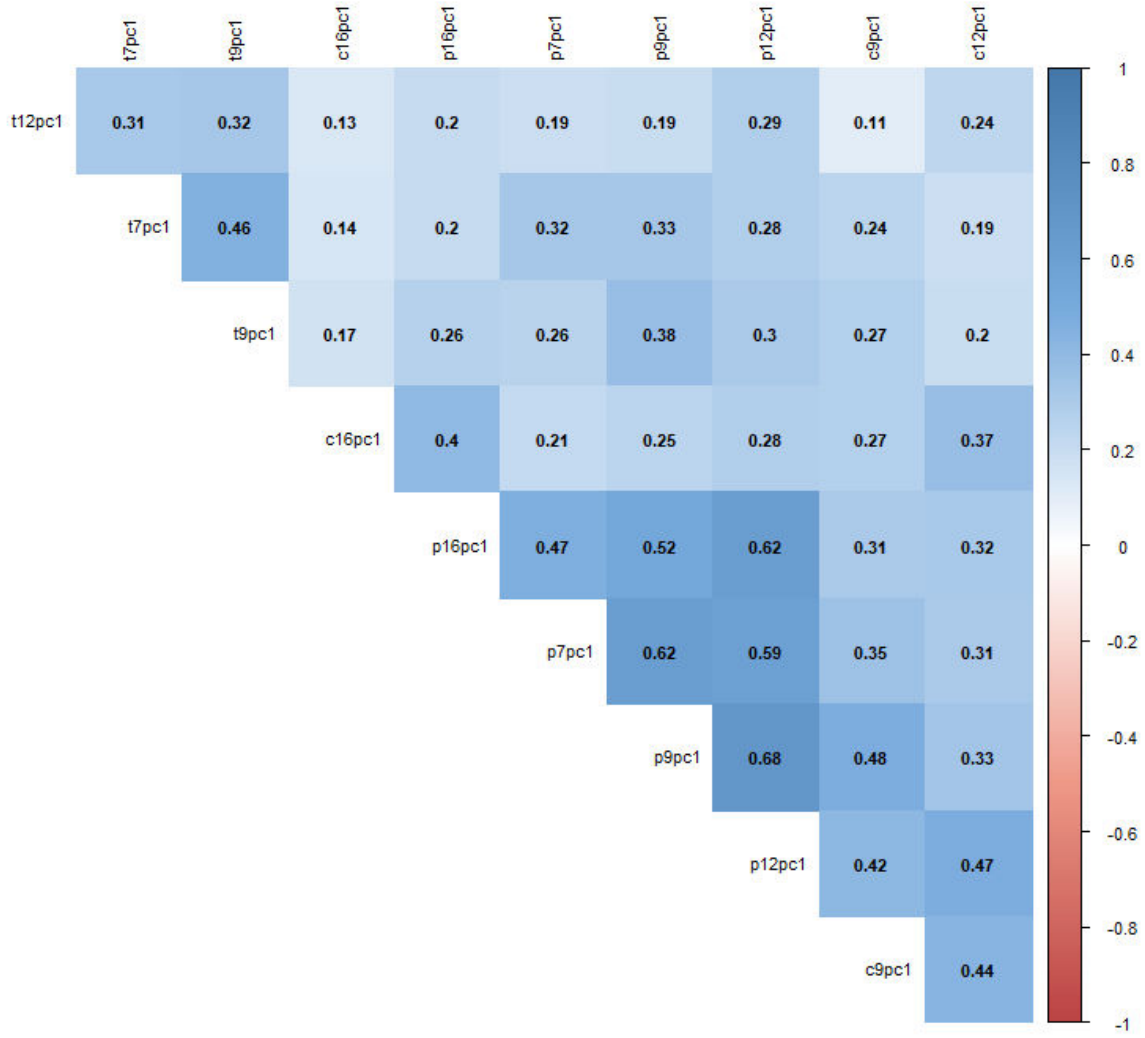
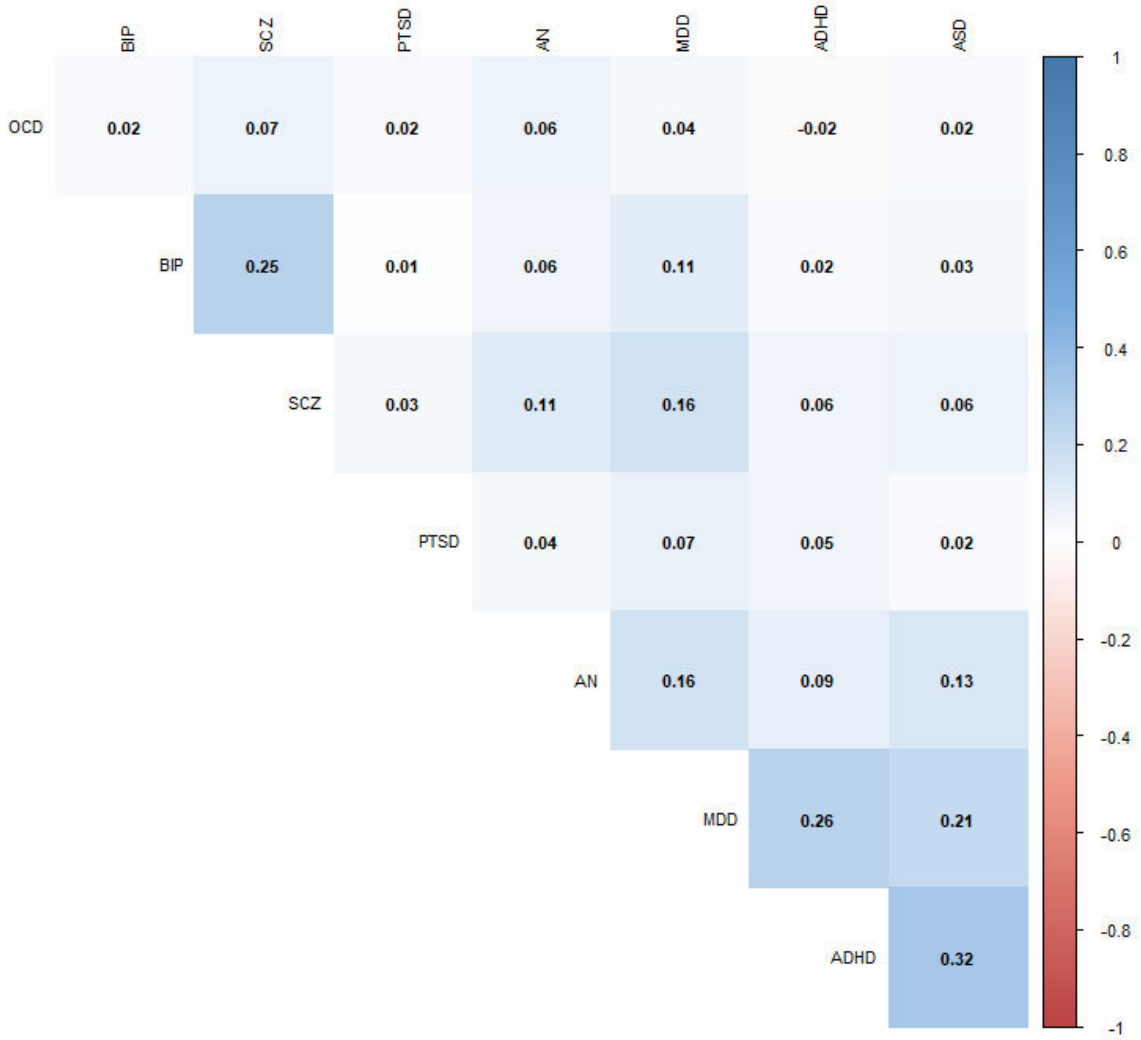


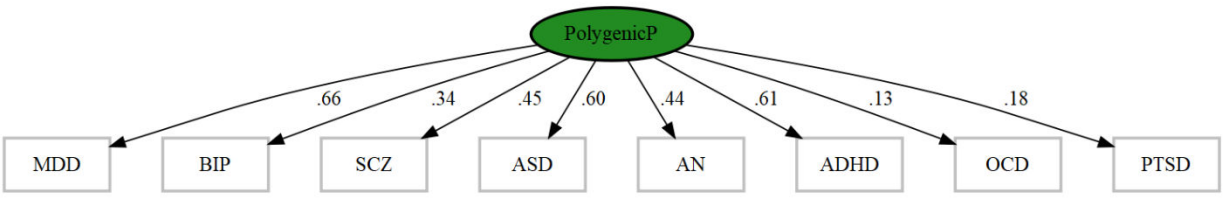
Figure S16. Correlations of 1st PCs of parent teacher and child rated measures across ages. Note variable names denote first principal component for parent reported data at age 7 to 16. E.g. P7pc1 = First principal component for parent rated age 7 data.



Supplementary Figure S17. Correlations between polygenic scores for psychiatric traits used to construct polygenic p. *OCD =obsessive compulsive disorder; BIP =bipolar disorder; SCZ = schizophrenia; PTSD = Post-traumatic stress disorder; AN = anorexia nervosa; MDD = major depressive disorder; ADHD = attention deficit hyperactivity disorder; AUT = Autism. Darker blue indicates stronger positive correlation.*



Supplementary Figure S18. PCA results for polygenic p factor. Note labels for polygenic scores: MDD= major depressive disorder, BIP= bipolar disorder, SCZ= schizophrenia, ASD= autism, AN= anorexia nervosa, ADHD=attention-deficit hyperactivity disorder, OCD= obsessive compulsive disorder, PTSD= post-traumatic stress disorder.



Supplementary Table S1: Additional parameters derived from common pathway twin models of childhood psychopathology. Note that c= common; s=specific; Tot=total.

Parent Report Age 7

	Ac	As	Cc	Cs	Ec	Es	Tot-H2	Tot-C2	Tot-E2	N
Hyperactivity SDQ	0.11	0.23	0.05	0	0.03	0.58	0.35	0.05	0.61	5588
Conduct SDQ	0.21	0.45	0.09	0	0.05	0.2	0.66	0.09	0.25	5592
Peer Problems SDQ	0.23	0.29	0.1	0	0.06	0.32	0.53	0.1	0.37	5591
Emotional Problems SDQ	0.1	0.44	0.04	0	0.02	0.39	0.54	0.04	0.41	5591
APSD Psychopathy= Narcissism + CU-Traits	0.2	0.36	0.09	0	0.05	0.3	0.56	0.09	0.35	5587
Autism Traits = Social + Nonsocial	0.33	0.29	0.14	0	0.08	0.16	0.62	0.14	0.24	5591
Prosocial SDQ	0.11	0.52	0.05	0.02	0.03	0.27	0.63	0.07	0.3	5591

Teacher Report Age 7

	Ac	As	Cc	Cs	Ec	Es	Tot-H2	Tot-C2	Tot-E2	N
Hyperactivity SDQ	0.26	0.44	0	0	0.07	0.23	0.7	0	0.3	4608
Conduct SDQ	0.37	0.31	0	0	0.09	0.22	0.69	0	0.31	4610
Peer Problems SDQ	0.21	0.43	0	0	0.05	0.3	0.64	0	0.36	4602
Emotional Problems SDQ	0.02	0.48	0	0	0.01	0.49	0.5	0	0.5	4591
APSD Psychopathy= Narcissism + CU-Traits	0.7	0.04	0	0	0.18	0.08	0.74	0	0.26	4591
Autism Traits = Social + Nonsocial	0.23	0.47	0	0	0.06	0.24	0.7	0	0.3	4601
Prosocial SDQ	0.53	0.21	0	0	0.13	0.13	0.73	0	0.27	4603

Child Report Age 9

	Ac	As	Cc	Cs	Ec	Es	Tot-H2	Tot-C2	Tot-E2	N
Hyperactivity SDQ	0.17	0.17	0.06	0	0.08	0.52	0.34	0.06	0.6	2635
Conduct SDQ	0.21	0.17	0.08	0	0.1	0.44	0.38	0.08	0.54	2631
Peer Problems SDQ	0.17	0.17	0.06	0	0.08	0.51	0.34	0.06	0.6	2621
Emotional Problems SDQ	0.19	0.19	0.07	0	0.09	0.46	0.38	0.07	0.56	2633
CAST autism traits	0.22	0.16	0.08	0	0.11	0.43	0.38	0.08	0.54	2556
Prosocial SDQ	0.04	0.33	0.01	0.02	0.02	0.58	0.37	0.04	0.59	2637

Parent Report Age 9

	Ac	As	Cc	Cs	Ec	Es	Tot-H2	Tot-C2	Tot-E2	N
Hyperactivity SDQ	0.18	0.32	0.18	0.00	0.04	0.28	0.50	0.18	0.32	2671
Conduct SDQ	0.26	0.26	0.27	0.02	0.06	0.13	0.52	0.29	0.19	2673
Peer Problems SDQ	0.09	0.51	0.10	0.00	0.02	0.27	0.61	0.10	0.30	2673
Emotional Problems SDQ	0.08	0.44	0.08	0.02	0.02	0.36	0.52	0.10	0.38	2673
CAST autism traits	0.15	0.52	0.16	0.00	0.04	0.13	0.67	0.16	0.17	2675

APSD Psychopathy	0.34	0.13	0.35	0.00	0.08	0.09	0.48	0.35	0.17	2676
Aggression	0.15	0.09	0.15	0.02	0.04	0.55	0.24	0.18	0.58	2666
Prosocial SDQ	0.08	0.53	0.08	0.14	0.02	0.16	0.61	0.22	0.18	2675

Teacher Report Age 9

	Ac	As	Cc	Cs	Ec	Es	Tot-H2	Tot-C2	Tot-E2	N
Hyperactivity SDQ	0.32	0.26	0.05	0	0.11	0.26	0.58	0.05	0.37	2217
Conduct SDQ	0.42	0.14	0.06	0	0.14	0.23	0.56	0.06	0.37	2221
Peer Problems SDQ	0.16	0.35	0.02	0.06	0.05	0.35	0.51	0.08	0.41	2229
Emotional Problems SDQ	0.03	0.46	0.004	0	0.01	0.50	0.49	0.004	0.51	2225
CAST autism traits	0.27	0.42	0.04	0	0.09	0.18	0.69	0.02	0.27	2228
APSD Psychopathy	0.61	0.00	0.09	0	0.20	0.10	0.61	0.09	0.30	2229
Aggression	0.24	0.20	0.04	0	0.08	0.44	0.44	0.04	0.52	2216
Prosocial SDQ	0.29	0.23	0.04	0.1	0.09	0.25	0.52	0.14	0.34	2226

Parent Report Age 12

	Ac	As	Cc	Cs	Ec	Es	Tot-H2	Tot-C2	Tot-E2	N
Prosocial	0.1	0.6	0.05	0.02	0.02	0.2	0.7	0.07	0.22	4654
Conduct	0.31	0.27	0.17	0.04	0.07	0.14	0.58	0.21	0.21	4645
Depression	0.16	0.33	0.09	0.04	0.04	0.34	0.49	0.13	0.38	4644
Peer Problems	0.15	0.48	0.08	0	0.03	0.27	0.62	0.08	0.3	4644
Emotional Problems	0.11	0.38	0.06	0	0.02	0.42	0.49	0.06	0.45	4644
CAST autism traits	0.2	0.44	0.11	0	0.04	0.2	0.65	0.11	0.24	4743
Conners	0.31	0.34	0.17	0	0.07	0.11	0.66	0.17	0.17	4652
APSD	0.37	0.26	0.2	0	0.08	0.09	0.63	0.2	0.17	4643

Teacher Report Age 12

	Ac	As	Cc	Cs	Ec	Es	Tot-H2	Tot-C2	Tot-E2	N
Prosocial	0.35	0.22	0	0.02	0.13	0.28	0.57	0.02	0.41	3812
Conduct	0.46	0.13	0	0	0.18	0.23	0.59	0	0.41	3842
Peer Problems	0.12	0.44	0	0	0.04	0.4	0.56	0	0.44	3842
Emotional Problems	0.04	0.39	0	0	0.02	0.55	0.44	0	0.56	3836
CAST autism traits	0.22	0.35	0	0	0.08	0.34	0.57	0	0.43	3811
APSD	0.61	0.06	0	0	0.23	0.1	0.67	0	0.33	3828

Child Report Age 12

	Ac	As	Cc	Cs	Ec	Es	Tot-H2	Tot-C2	Tot-E2	N
Prosocial	0.01	0.41	0	0	0.01	0.57	0.42	0	0.58	4631

Conduct	0.14	0.24	0.04	0	0.1	0.48	0.37	0.04	0.59	4631
Peer Problems	0.16	0.2	0.05	0	0.12	0.47	0.36	0.05	0.59	4631
Emotional Problems	0.23	0.1	0.07	0	0.17	0.42	0.33	0.07	0.6	4631
Depression	0.3	0.01	0.1	0.02	0.23	0.33	0.32	0.12	0.56	4649

Parent Report Age 16

	Ac	As	Cc	Cs	Ec	Es	Tot-H2	Tot-C2	Tot-E2	N
Prosocial	0.22	0.33	0.1	0.18	0.05	0.1	0.56	0.28	0.15	4011
Conduct	0.28	0.31	0.13	0	0.6	0.2	0.6	0.13	0.26	4011
Depression	0.15	0.3	0.07	0	0.03	0.43	0.46	0.7	0.12	4011
CAST autism traits	0.14	0.66	0.06	0	0.02	0.09	0.81	0.6	0.12	4010
Cu	0.31	0.27	0.14	0.07	0.6	0.12	0.58	0.22	0.18	4010
Conners	0.23	0.44	0.1	0	0.04	0.16	0.68	0.1	0.21	4010
Arbq	0.15	0.34	0.07	0.07	0.03	0.32	0.49	0.14	0.35	3995

Child Report Age 16

	Ac	As	Cc	Cs	Ec	Es	Tot-H2	Tot-C2	Tot-E2	N
Prosocial	0.01	0.37	0.00	0.00	0.01	0.62	0.38	0.00	0.62	3999
Conduct	0.08	0.26	0.00	0.00	0.06	0.61	0.33	0.00	0.67	3998
Peer Problems	0.17	0.25	0.00	0.00	0.12	0.46	0.41	0.00	0.59	3999
Emotional Problems	0.35	0.06	0.00	0.00	0.26	0.33	0.41	0.00	0.59	3999
Depression	0.34	0.06	0.00	0.00	0.25	0.35	0.40	0.00	0.60	4000
CAST autism traits	0.13	0.35	0.00	0.00	0.10	0.42	0.48	0.00	0.52	4002

Table S2: Model fit statistics for common pathway twin models of childhood psychopathology. Note that Sat= saturated model; CPACE = the common pathway ACE model

Model	base	comparison	ep	minus2LL	df	AIC	diffLL	diffdf	p
Age 7 parent report	Sat	CPACE	38	198398.6	77905	42588.63	8872.522	201	0.0000
Age 7 teacher report	Sat	CPACE	38	153092	64219	24654	8562.541	201	0.0000
Age 9 self report	Sat	CPACE	33	83002.4	31375	20252.4	1032.731	148	0.0000
Age 9 parent report	Sat	CPACE	43	101865.3	42701	16463.31	3960.66	262	0.0000
Age 9 teacher report	Sat	CPACE	43	83562.71	35208	13146.70	4933.27	262	0.0000
Age 12 self report	Sat	CPACE	28	121208.1	46270	28668.05	966.53	103	0.0000
Age 12 parent report	Sat	CPACE	43	178367.5	74408	29551.5	5872.72	262	0.0000
Age 12 teacher report	Sat	CPACE	33	113280.2	45806	21668.24	4135.12	148	0.0000
Age 16 self report	Sat	CPACE	48	170085.1	65525	39035.12	4212.57	331	0.0000
Age 16 parent report	Sat	CPACE	38	134663.4	55982	22699.37	6624.52	201	0.0000

Supplementary Table S3: Loadings on first principal components of psychopathology measures for each age and rater.

Parent report age 7	
Hyperactivity SDQ	0.34
Conduct SDQ	0.42
Peer problems SDQ	0.38
Emotional problems SDQ	0.29
APSD psychopathy = Narcissism + CU-traits	0.41
Autism = Social + Nonsocial	0.45
Prosocial SDQ (reversed)	0.32

Teacher report age 7	
Hyperactivity SDQ	0.35
Conduct SDQ	0.40
Peer problems SDQ	0.39
Emotional problems SDQ	0.22
APSD psychopathy= Narcissism + CU-traits	0.45
Autism = Social + Nonsocial	0.40
Prosocial SDQ (reversed)	0.40

Child report age 9	
Hyperactivity SDQ	0.42
Conduct SDQ	0.45
Peer problems SDQ	0.42
Emotional problems SDQ	0.43
Autism CAST= Social + Nonsocial	0.46
Prosocial SDQ (reversed)	0.21

Parent report age 9	
Hyperactivity SDQ	0.36
Conduct SDQ	0.41
Peer problems SDQ	0.32
Emotional problems SDQ	0.30

Autism CAST	0.38
APSD psychopathy	0.44
Aggression	0.34
Prosocial SDQ (reversed)	0.22

Teacher report age 9	
Hyperactivity SDQ	0.36
Conduct SDQ	0.40
Peer problems SDQ	0.32
Emotional problems SDQ	0.17
CAST	0.38
APSD psychopathy	0.46
Aggression	0.31
Prosocial SDQ (reversed)	0.35

Child report age 12	
Prosocial (reversed)	0.16
Peer Problems SDQ	0.47
Emotional Problems SDQ	0.5
Conduct	0.48
Depression	0.54

Parent report age 12	
Conners	0.5
Conduct SDQ	0.5
CU-traits	0.32
Emotional problems SDQ	0.43
Depression	0.47

Teacher report age 12	
Conduct SDQ	0.44
Peer problems SDQ	0.38
Emotional problems SDQ	0.26
CAST	0.44
APSD psychopathy	0.47
Prosocial SDQ	0.43

Child report age 16	

Prosocial	0.07
Peer Problems SDQ	0.32
Emotional Problems SDQ	0.49
Conduct	0.4
Depression	0.47
CAST	0.39
CASI	0.41

Parent report age 16	
Conduct SDQ	0.4
Conners	0.38
CAST	0.36
ARBQ	0.36
Prosocial SDQ	0.36
Depression	0.34
CU	0.4

Supplementary Table S4. Variance explained by first principal components (phenotypic p factors) for each age and rater, plus the sample size for each 1st PC. P/T/C= parent/teacher/child ratings; 7/9/12/16 = age.

Age/rater	R^2 by 1st PC	N
P7	0.40	4109
T7	0.50	3435
C9	0.42	2074
P9	0.45	2157
T9	0.48	1594
C12	0.46	4490
P12	0.45	3227
T12	0.50	2146
C16	0.42	1391
P16	0.46	3258

Supplementary Table S5. Association statistics for polygenic p across phenotypic p measures.

Outcome	beta	se	p-value	r squared	df residual
P age 7	0.083554	0.015365	**5.70E-08**	0.007149	4107
P age 9	0.082633	0.021656	**0.00014**	0.006711	2155
P age 12	0.085961	0.017255	**6.63E-07**	0.007637	3225
P age 16	0.078936	0.017206	**4.65E-06**	0.006423	3256
T age 7	0.097842	0.016814	**6.47E-09**	0.009767	3433
T age 9	0.089132	0.025625	**0.000518**	0.007542	1592
T age 12	0.055183	0.020751	0.007889	0.003288	2144
C age 9	0.066584	0.022225	**0.002769**	0.004313	2072
C age 12	0.065225	0.014811	**1.09E-05**	0.004303	4488
C age 16	0.068554	0.026306	0.009258	0.005750	1389

Bold. Significant after Bonferroni correction (alpha 0.05/10).
Note. P = parent rated , T = teacher rated, C = child-rated.

Appendix 4 – Model specification of the wfRI-CLPM example for two siblings measured on two traits at two waves.

Measurement model:

$$\begin{bmatrix} x_{i1} \\ y_{i1} \\ x_{j1} \\ y_{j1} \\ x_{i2} \\ y_{i2} \\ x_{j2} \\ y_{j2} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \pi_1 \\ \mu_1 \\ \pi_1 \\ \mu_2 \\ \pi_2 \\ \mu_2 \\ \pi_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_{i1} \\ q_{i1} \\ p_{j1} \\ q_{j1} \\ p_{i2} \\ q_{i2} \\ p_{j2} \\ p_{j2} \\ \kappa_i \\ \omega_i \\ \kappa_j \\ \omega_j \end{bmatrix}$$

Structural model:

$$\begin{bmatrix} p_{i1} \\ q_{i1} \\ p_{j1} \\ q_{j1} \\ p_{i2} \\ q_{i2} \\ p_{j2} \\ p_{j2} \\ \kappa_i \\ \omega_i \\ \kappa_j \\ \omega_j \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha_{i2} & \beta_{i2} & \alpha_{j2} & \beta_{j2} & 0 & 0 & 0 & 0 \\ \gamma_{i2} & \delta_{i2} & \gamma_{j2} & \delta_{j2} & 0 & 0 & 0 & 0 \\ \alpha_{j2} & \beta_{j2} & \alpha_{i2} & \beta_{i2} & 0 & 0 & 0 & 0 \\ \gamma_{j2} & \delta_{j2} & \gamma_{i2} & \delta_{i2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} p_{i1} \\ q_{i1} \\ p_{j1} \\ q_{j1} \\ p_{i2} \\ q_{i2} \\ p_{j2} \\ q_{j2} \\ \kappa_i \\ \omega_i \\ \kappa_j \\ \omega_j \end{bmatrix} + \begin{bmatrix} p_{i1} \\ q_{i1} \\ p_{j1} \\ q_{j1} \\ \upsilon_{i2} \\ \nu_{i2} \\ \upsilon_{j2} \\ \nu_{j2} \\ \kappa_i \\ \omega_i \\ \kappa_j \\ \omega_j \end{bmatrix}$$

Variance/covariance matrix of residuals:

$$\psi = \begin{bmatrix} \sigma^2_{p_{i1}} & & & & & & & & & & & \\ \sigma_{p_{i1}q_{i1}} & \sigma^2_{q_{i1}} & & & & & & & & & & \\ \sigma_{p_{i1}p_{j1}} & \sigma_{q_{i1}p_{j1}} & \sigma^2_{p_{j1}} & & & & & & & & & \\ \sigma_{p_{i1}q_{j1}} & \sigma_{q_{i1}q_{j1}} & \sigma_{p_{j1}q_{j1}} & \sigma^2_{q_{j1}} & & & & & & & & \\ 0 & 0 & 0 & 0 & \sigma^2_{\upsilon_{i2}} & & & & & & & \\ 0 & 0 & 0 & 0 & \sigma_{\upsilon_{i2}\nu_{i2}} & \sigma^2_{\nu_{i2}} & & & & & & \\ 0 & 0 & 0 & 0 & \sigma_{\upsilon_{i2}\upsilon_{j2}} & \sigma_{\nu_{i2}\upsilon_{j2}} & \sigma^2_{\upsilon_{j2}} & & & & & \\ 0 & 0 & 0 & 0 & \sigma_{\upsilon_{i2}\nu_{j2}} & \sigma_{\nu_{i2}\nu_{j2}} & \sigma_{\upsilon_{j2}\nu_{j2}} & \sigma^2_{\nu_{j2}} & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{\kappa_i} & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{\kappa_i\omega_i} & \sigma^2_{\omega_i} & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{\kappa_i\kappa_j} & \sigma_{\omega_i\kappa_j} & \sigma^2_{\kappa_j} & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{\kappa_i\omega_j} & \sigma_{\omega_i\omega_j} & \sigma_{\kappa_j\omega_j} & \sigma^2_{\omega_j} \end{bmatrix}$$

Appendix 5 – Supplementary material for Chapter 5

Contents:

- **Table 9c.** MZ Cross-twin cross-trait correlations of variables under study - NTR
- **Table 9d.** DZ Cross-twin cross-trait correlations of variables under study - NTR
- **Table 10a.** Within-person and between-sibling regression estimates from the wfRICLPM - TEDS
- **Table 10b.** Within-person and between-sibling regression estimates from the wfRICLPM - NTR

Supplementary Figures

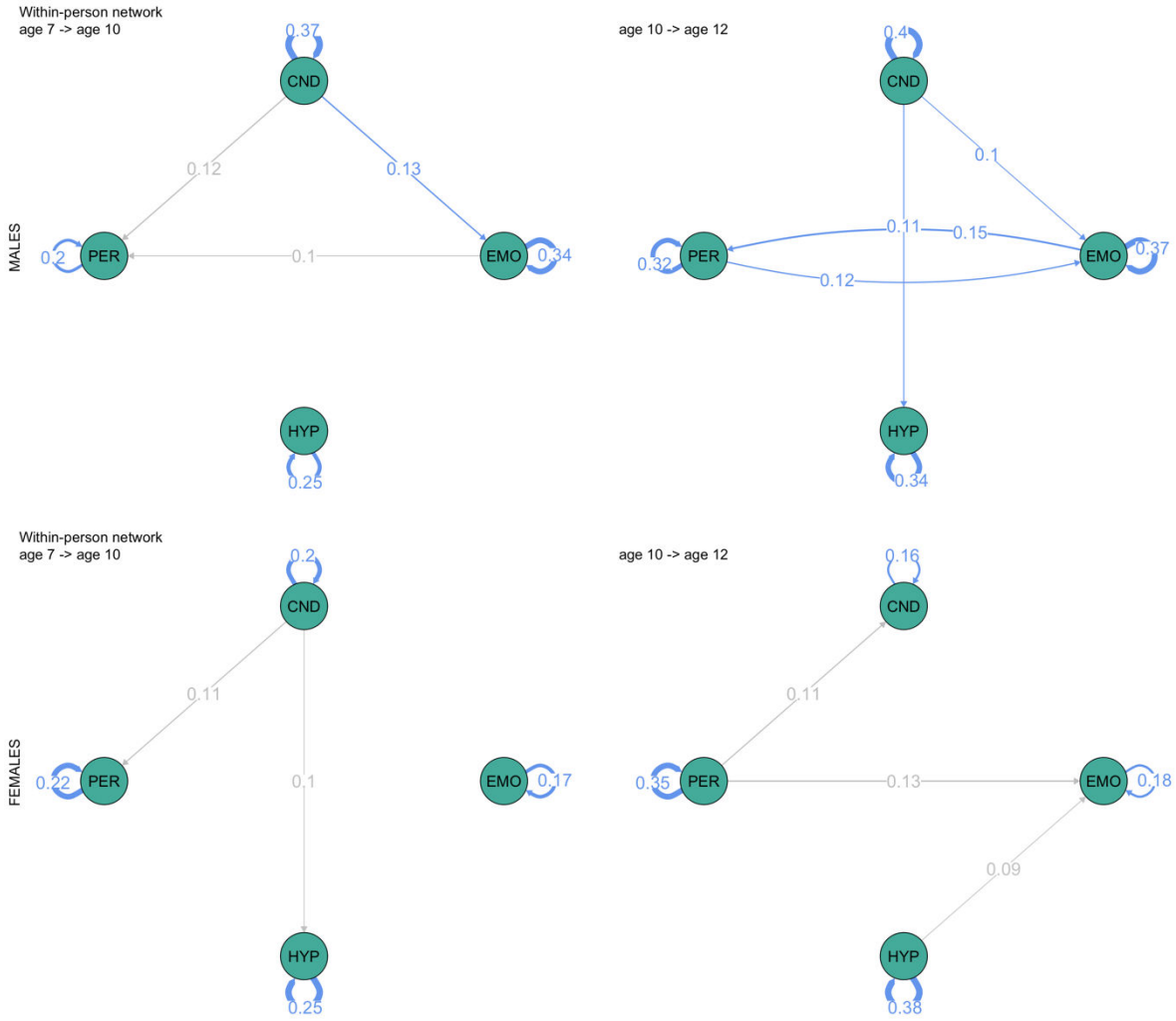


Figure S1. Between-person and within-person (directed) networks of relationships in Males (top) and Females (bottom) obtained from the RI-CLPM in NTR. Nodes represent the measure of interest (the random intercept in the case of between-person networks, and residual deviation of measurement occasion for the within-person network). Edges width and labels indicate and quantify the strength of relationships between nodes, and in the case of within-person networks also the temporal direction of the effect. For every time lag (7-10 and 10->12) edges represent directional effects within-trait (self-pointing arrow) or cross-trait. **Note.** Acronyms: CND = conduct/externalizing, HYP = hyperactivity/ hyperactivity-inattention, EMO = emotional problems/internalizing, PER = peer problems/social problems. All edges survived FDR correction for multiple testing.

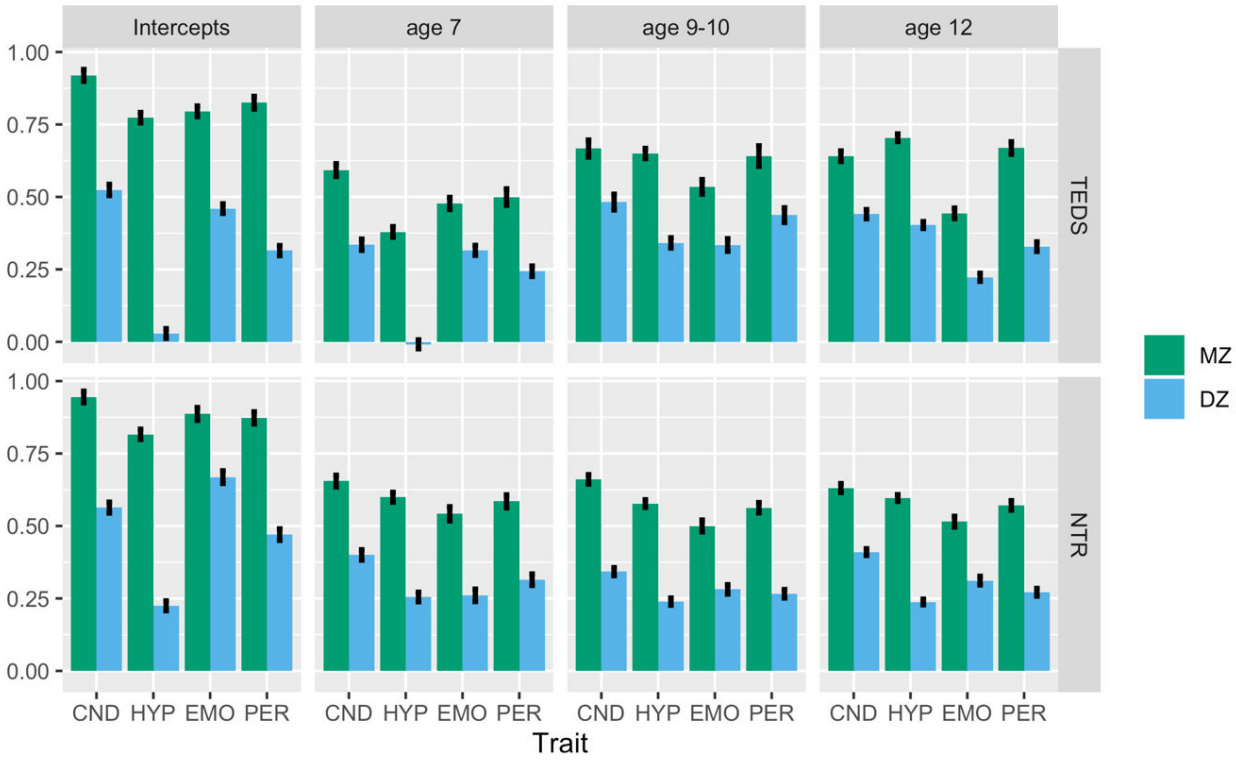


Figure S2. Twin correlations for random intercepts and residual deviations in the wfRICLPM in TEDS and NTR.

Supplementary Tables

Table S1a. Descriptive Statistics TEDS

Measure	n	mean	sd	median	min	max	skew	kurtosis	se	% females
CON_t1a	7758	1.709	1.644	1	0	10	1.186	1.616	0.019	0.51
HYP_t1a	7753	3.611	2.556	3	0	10	0.551	-0.444	0.029	0.51
PEp_t1a	7754	1.007	1.429	0	0	10	1.897	4.262	0.016	0.51
EMp_t1a	7754	2.232	1.868	2	0	10	0.972	0.74	0.021	0.51
CON_t2a	3388	1.278	1.439	1	0	10	1.525	3.253	0.025	0.525
HYP_t2a	3386	3.26	2.384	3	0	10	0.814	0.114	0.041	0.525
PEp_t2a	3388	1.07	1.545	0	0	9	1.884	3.873	0.027	0.525
EMp_t2a	3388	1.761	1.917	1	0	10	1.297	1.451	0.033	0.526
CON_t3a	5829	1.337	1.483	1	0	10	1.471	3.074	0.019	0.525
HYP_t3a	5829	2.848	2.272	2	0	10	0.857	0.347	0.03	0.525
PEp_t3a	5828	1.106	1.536	1	0	9	1.88	4.088	0.02	0.525
EMp_t3a	5829	1.843	1.946	1	0	10	1.252	1.394	0.025	0.525
CON_t1b	7758	1.725	1.661	1	0	10	1.203	1.726	0.019	0.51
HYP_t1b	7752	3.634	2.573	3	0	10	0.565	-0.448	0.029	0.51
PEp_t1b	7754	1.017	1.455	0	0	10	2.006	5.021	0.017	0.51
EMp_t1b	7754	2.202	1.843	2	0	10	0.989	0.815	0.021	0.51
CON_t2b	3387	1.308	1.466	1	0	10	1.531	3.274	0.025	0.525
HYP_t2b	3385	3.274	2.407	3	0	10	0.839	0.194	0.041	0.525
PEp_t2b	3388	1.073	1.57	0	0	10	2.02	4.738	0.027	0.525
EMp_t2b	3387	1.717	1.9	1	0	10	1.386	1.823	0.033	0.526
CON_t3b	5829	1.322	1.462	1	0	10	1.427	2.694	0.019	0.525
HYP_t3b	5829	2.843	2.281	2	0	10	0.884	0.442	0.03	0.525
PEp_t3b	5829	1.122	1.535	1	0	10	1.858	4.145	0.02	0.525
EMp_t3b	5830	1.811	1.927	1	0	10	1.266	1.395	0.025	0.526
Age7	7732	7.065	0.248	7.044	5.574	7.984	0.259	-0.011	0.003	0.511
Age9	3411	9.019	0.275	9.01	8.082	9.966	0.268	-0.484	0.005	0.527
Age12	5989	11.36	0.669	11.48	10.001	13.495	-0.157	-0.491	0.009	0.519

Note: a = Sib 1; b = Sib 2; t1 = Age 7; t2 = Age 9; t3 = Age 12

Table S1b. Descriptive Statistics NTR

Measure	n	mean	sd	median	min	max	skew	kurtosis	se	% females
CON_t1a	12674	6.446	6.324	5	0	50	1.631	3.664	0.056	0.509
HYP_t1a	12710	3.118	3.1	3	0	20	1.239	1.629	0.027	0.509
PEp_t1a	12700	2.133	2.443	1	0	19	1.86	4.496	0.022	0.509
EMp_t1a	12474	4.603	4.639	3	0	45	1.857	4.929	0.042	0.509
CON_t2a	11344	5.654	6.212	4	0	53	1.887	4.977	0.058	0.508
HYP_t2a	11418	3.024	3.125	2	0	20	1.29	1.728	0.029	0.509
PEp_t2a	11412	2.023	2.547	1	0	18	1.917	4.498	0.024	0.509
EMp_t2a	11260	4.732	5.075	3	0	45	1.961	5.379	0.048	0.509
CON_t3a	8815	4.824	5.631	3	0	50	2.001	5.398	0.06	0.513
HYP_t3a	8854	2.719	3.017	2	0	20	1.409	2.045	0.032	0.513
PEp_t3a	8842	1.674	2.395	1	0	20	2.197	5.799	0.025	0.513
EMp_t3a	8748	4.267	4.895	3	0	48	2.121	6.546	0.052	0.513
CON_t1b	12667	6.482	6.366	5	0	52	1.607	3.439	0.057	0.509
HYP_t1b	12717	3.084	3.092	3	0	20	1.285	1.867	0.027	0.509
PEp_t1b	12700	2.11	2.437	1	0	20	1.907	5.014	0.022	0.508
EMp_t1b	12488	4.619	4.693	3	0	41	1.921	5.329	0.042	0.508
CON_t2b	11362	5.655	6.189	4	0	47	1.847	4.575	0.058	0.509
HYP_t2b	11416	3.033	3.15	2	0	20	1.244	1.389	0.029	0.509
PEp_t2b	11422	2.017	2.525	1	0	19	1.91	4.505	0.024	0.509
EMp_t2b	11250	4.719	5.108	3	0	44	1.987	5.294	0.048	0.508
CON_t3b	8806	4.751	5.675	3	0	49	2.09	6.058	0.06	0.512
HYP_t3b	8852	2.704	3.001	2	0	18	1.455	2.295	0.032	0.513
PEp_t3b	8838	1.669	2.406	1	0	20	2.277	6.67	0.026	0.513
EMp_t3b	8738	4.316	4.994	3	0	48	2.181	6.879	0.053	0.512
Age7	12751	7.454	0.386	7.4	6.2	9	0.996	0.973	0.003	0.508
Age10	10878	9.985	0.388	10	9.1	11	0.107	-0.293	0.004	0.51
Age12	8909	12.265	0.393	12.2	11.1	13.5	0.54	-0.131	0.004	0.513

Note: a = Sib 1; b = Sib 2; t1 = Age 7; t2 = Age 10; t3 = Age 12

Table S2a. Correlation matrix of variables under study - TEDS

	CON_t1	CON_t2	CON_t3	HYP_t1	HYP_t2	HYP_t3	EMp_t1	EMp_t2	EMp_t3	PEp_t1	PEp_t2	PEp_t3
CON_t1	1	0	0	0	0	0	0	0	0	0	0	0
CON_t2	0.577	1	0	0	0	0	0	0	0	0	0	0
CON_t3	0.514	0.584	1	0	0	0	0	0	0	0	0	0
HYP_t1	0.422	0.375	0.334	1	0	0	0	0	0	0	0	0
HYP_t2	0.389	0.48	0.388	0.669	1	0	0	0	0	0	0	0
HYP_t3	0.34	0.402	0.483	0.589	0.674	1	0	0	0	0	0	0
EMp_t1	0.252	0.221	0.193	0.204	0.21	0.195	1	0	0	0	0	0
EMp_t2	0.237	0.323	0.241	0.205	0.295	0.237	0.539	1	0	0	0	0
EMp_t3	0.199	0.25	0.327	0.195	0.226	0.301	0.468	0.549	1	0	0	0
PEp_t1	0.28	0.228	0.216	0.221	0.22	0.217	0.321	0.263	0.227	1	0	0
PEp_t2	0.247	0.327	0.247	0.241	0.301	0.245	0.255	0.409	0.28	0.518	1	0
PEp_t3	0.225	0.258	0.331	0.226	0.257	0.306	0.226	0.274	0.412	0.441	0.525	1

Note: t1 = Age 7; t2 = Age 9; t3 = Age 12

Table S2b. Correlation matrix of variables under study - NTR

	CON_t1	CON_t2	CON_t3	HYP_t1	HYP_t2	HYP_t3	EMp_t1	EMp_t2	EMp_t3	PEp_t1	PEp_t2	PEp_t3
CON_t1	1	0	0	0	0	0	0	0	0	0	0	0
CON_t2	0.71	1	0	0	0	0	0	0	0	0	0	0
CON_t3	0.624	0.73	1	0	0	0	0	0	0	0	0	0
HYP_t1	0.573	0.444	0.402	1	0	0	0	0	0	0	0	0
HYP_t2	0.478	0.584	0.462	0.69	1	0	0	0	0	0	0	0
HYP_t3	0.424	0.488	0.59	0.604	0.729	1	0	0	0	0	0	0
EMp_t1	0.515	0.372	0.336	0.428	0.345	0.297	1	0	0	0	0	0
EMp_t2	0.415	0.545	0.417	0.324	0.452	0.359	0.616	1	0	0	0	0
EMp_t3	0.358	0.432	0.554	0.293	0.366	0.462	0.512	0.651	1	0	0	0
PEp_t1	0.655	0.485	0.428	0.608	0.492	0.432	0.622	0.462	0.391	1	0	0
PEp_t2	0.508	0.659	0.517	0.466	0.609	0.497	0.461	0.671	0.506	0.613	1	0
PEp_t3	0.439	0.512	0.646	0.421	0.495	0.614	0.388	0.502	0.68	0.512	0.665	1

Note: t1 = Age 7; t2 = Age 9; t3 = Age 12

Table S3a. Model fit Indices - TEDS

	chisq	df	chisqPval	rmsea	srmr	cfi
TEDS_Constrained	54.6854	32	0.0075	0.0091	0.014	0.9988
TEDS_Unconstrained	10.5506	6	0.1033	0.0094	0.0049	0.9998
TEDS_Pruned	40.9425	22	0.0084	0.01	0.0123	0.999
TEDS_NoLags	91.2658	30	0	0.0155	0.0216	0.9968

Table S3b. Nested models comparison - TEDS

	Df	AIC	BIC	Chisq	Chisq diff	Df diff	Pr(>Chisq)
TEDS_Unconstrained	6	170165.783	170758.283	10.5506	NA	NA	NA
TEDS_Pruned	22	170164.175	170643.817	40.9425	25.5521	16	0.0607
TEDS_NoLags	30	170198.498	170621.712	91.2658	43.3825	8	0

Table S3c. Model fit Indices - NTR

	chisq	df	chisqPval	rmsea	srmr	cfi
NTR_constrained	172.5375	32	0	0.0162	0.0125	0.9983
NTR_unconstrained	54.9716	6	0	0.0221	0.007	0.9994
NTR_pruned	96.4998	22	0	0.0142	0.0115	0.9991
NTR_noLags	172.4929	30	0	0.0169	0.0177	0.9982

Table S3d. Nested models comparison - NTR

	Df	AIC	BIC	Chisq	Chisq diff	Df diff	Pr(>Chisq)
NTR_unconstrained	6	283997.669	284646.299	54.9716	NA	NA	NA
NTR_pruned	22	284007.197	284532.278	96.4998	26.6536	16	0.0455
NTR_noLags	30	284067.19	284530.497	172.4929	44.9768	8	0

Table S4a. Simultaneous residual correlations and residual variances - TEDS

correlations	beta	lower CI	upper CI	stdBeta	robust SE	p.value	p_value adjusted	index
CND1<->CND1	0.511	0.463	0.559	1	0.025	2.22E-16	2.22E-16	*
CND2<->CND2	0.533	0.466	0.6	0.956	0.034	2.22E-16	2.22E-16	*
CND3<->CND3	0.494	0.455	0.534	0.923	0.02	2.22E-16	2.22E-16	*
EMO1<->CND1	0.077	0.046	0.109	0.145	0.016	2.22E-16	2.22E-16	*
EMO1<->EMO1	0.556	0.513	0.6	1	0.022	2.22E-16	2.22E-16	*
EMO1<->HYP1	0.017	-0.013	0.046	0.034	0.015	2.65E-01	2.74E-01	
EMO2<->CND2	0.151	0.113	0.189	0.28	0.019	7.33E-15	2.22E-16	*
EMO2<->EMO2	0.546	0.49	0.602	0.928	0.029	2.22E-16	2.22E-16	*
EMO2<->HYP2	0.111	0.076	0.147	0.226	0.018	5.49E-10	6.87E-10	*
EMO3<->CND3	0.125	0.101	0.148	0.242	0.012	2.22E-16	2.22E-16	*
EMO3<->EMO3	0.54	0.504	0.576	0.956	0.019	2.22E-16	2.22E-16	*
EMO3<->HYP3	0.103	0.083	0.123	0.219	0.01	2.22E-16	2.22E-16	*
HYP1<->CND1	0.108	0.076	0.141	0.228	0.017	2.22E-16	2.22E-16	*
HYP1<->HYP1	0.442	0.403	0.482	1	0.02	2.22E-16	2.22E-16	*
HYP2<->CND2	0.168	0.131	0.206	0.346	0.019	2.22E-16	2.22E-16	*
HYP2<->HYP2	0.444	0.402	0.485	0.925	0.021	2.22E-16	2.22E-16	*
HYP3<->CND3	0.14	0.119	0.162	0.313	0.011	2.22E-16	2.22E-16	*
HYP3<->HYP3	0.408	0.381	0.434	0.886	0.013	2.22E-16	2.22E-16	*
PER1<->CND1	0.073	0.04	0.107	0.134	0.017	2.22E-16	2.22E-16	*
PER1<->EMO1	0.114	0.08	0.148	0.2	0.017	6.41E-11	8.74E-11	*
PER1<->HYP1	0.001	-0.029	0.032	0.002	0.016	9.35E-01	9.35E-01	
PER1<->PER1	0.582	0.525	0.639	1	0.029	2.22E-16	2.22E-16	*
PER2<->CND2	0.126	0.083	0.169	0.222	0.022	9.28E-09	2.22E-16	*
PER2<->EMO2	0.218	0.173	0.263	0.381	0.023	2.22E-16	2.22E-16	*
PER2<->HYP2	0.085	0.047	0.123	0.165	0.019	1.00E-05	2.22E-16	*
PER2<->PER2	0.602	0.525	0.679	0.965	0.039	2.22E-16	2.22E-16	*
PER3<->CND3	0.102	0.076	0.128	0.195	0.013	2.22E-16	2.22E-16	*
PER3<->EMO3	0.172	0.144	0.2	0.315	0.014	2.22E-16	2.22E-16	*
PER3<->HYP3	0.077	0.055	0.098	0.162	0.011	2.22E-16	2.69E-12	*
PER3<->PER3	0.553	0.507	0.599	0.936	0.023	2.22E-16	2.22E-16	*

Note: t1 = Age 7; t2 = Age 9; t3 = Age 12

Table S4b. Simultaneous residual correlations and residual variances - NTR

correlations	beta	lower CI	upper CI	stdBeta	robust SE	p.value	p_value adjusted	index
CND1<->CND1	0.429	0.377	0.482	1	0.027	2.22E-16	2.22E-16	*
CND2<->CND2	0.386	0.348	0.424	0.894	0.019	2.22E-16	2.22E-16	*
CND3<->CND3	0.381	0.35	0.413	0.848	0.016	2.22E-16	2.22E-16	*
EMO1<->CND1	0.219	0.174	0.263	0.441	0.023	2.22E-16	2.22E-16	*
EMO1<->EMO1	0.571	0.504	0.638	1	0.034	2.22E-16	2.22E-16	*
EMO1<->HYP1	0.179	0.137	0.221	0.349	0.021	2.22E-16	2.22E-16	*
EMO2<->CND2	0.211	0.182	0.24	0.474	0.015	2.22E-16	2.22E-16	*
EMO2<->EMO2	0.513	0.464	0.562	0.883	0.025	2.22E-16	2.22E-16	*
EMO2<->HYP2	0.168	0.141	0.194	0.36	0.013	2.22E-16	2.22E-16	*
EMO3<->CND3	0.2	0.175	0.224	0.467	0.012	2.22E-16	2.22E-16	*
EMO3<->EMO3	0.48	0.439	0.522	0.837	0.021	2.22E-16	2.22E-16	*
EMO3<->HYP3	0.156	0.136	0.175	0.36	0.01	2.22E-16	2.22E-16	*
HYP1<->CND1	0.202	0.162	0.242	0.455	0.02	2.22E-16	2.22E-16	*
HYP1<->HYP1	0.46	0.41	0.51	1	0.026	2.22E-16	2.22E-16	*
HYP2<->CND2	0.181	0.156	0.206	0.449	0.013	2.22E-16	2.22E-16	*
HYP2<->HYP2	0.424	0.391	0.456	0.878	0.016	2.22E-16	2.22E-16	*
HYP3<->CND3	0.176	0.157	0.196	0.458	0.01	2.22E-16	2.22E-16	*
HYP3<->HYP3	0.389	0.365	0.413	0.819	0.012	2.22E-16	2.22E-16	*
PER1<->CND1	0.277	0.229	0.324	0.56	0.024	2.22E-16	2.22E-16	*
PER1<->EMO1	0.305	0.252	0.358	0.536	0.027	2.22E-16	2.22E-16	*
PER1<->HYP1	0.232	0.186	0.278	0.454	0.023	2.22E-16	2.22E-16	*
PER1<->PER1	0.568	0.506	0.629	1	0.031	2.22E-16	2.22E-16	*
PER2<->CND2	0.241	0.211	0.27	0.538	0.015	2.22E-16	2.22E-16	*
PER2<->EMO2	0.302	0.267	0.336	0.584	0.018	2.22E-16	2.22E-16	*
PER2<->HYP2	0.197	0.168	0.225	0.419	0.014	2.22E-16	2.22E-16	*
PER2<->PER2	0.519	0.478	0.561	0.88	0.021	2.22E-16	2.22E-16	*
PER3<->CND3	0.21	0.185	0.234	0.491	0.013	2.22E-16	2.22E-16	*
PER3<->EMO3	0.277	0.247	0.307	0.579	0.015	2.22E-16	2.22E-16	*
PER3<->HYP3	0.183	0.161	0.204	0.424	0.011	2.22E-16	2.22E-16	*
PER3<->PER3	0.477	0.441	0.513	0.818	0.018	2.22E-16	2.22E-16	*

Note: t1 = Age 7; t2 = Age 9; t3 = Age 12

Table S5. Within-person regressions - TEDS

direction	beta	lower CI	upper CI	stdBeta	robust SE	p-value	p_value adjusted	index
CND1->CND2	0.166	0.071	0.261	0.159	0.048	6.15E-04	2.19E-03	*
CND1->EMO2	0.15	0.078	0.222	0.14	0.037	4.49E-05	3.89E-04	*
CND1->HYP2	0.096	0.03	0.163	0.099	0.034	4.65E-03	9.91E-03	*
CND1->PER2	0.08	-0.001	0.161	0.072	0.041	5.20E-02	9.24E-02	
CND2->CND3	0.144	0.054	0.234	0.147	0.046	1.74E-03	5.07E-03	*
CND2->EMO3	0.032	-0.043	0.107	0.032	0.038	4.01E-01	5.71E-01	
CND2->HYP3	0.118	0.055	0.182	0.13	0.032	2.68E-04	1.71E-03	*
CND2->PER3	-0.008	-0.083	0.067	-0.008	0.038	8.40E-01	9.27E-01	
EMO1->CND2	0.029	-0.04	0.098	0.029	0.035	4.11E-01	5.71E-01	
EMO1->EMO2	0.144	0.064	0.224	0.14	0.041	4.29E-04	2.19E-03	*
EMO1->HYP2	-0.003	-0.066	0.059	-0.004	0.032	9.17E-01	9.47E-01	
EMO1->PER2	0.013	-0.061	0.087	0.012	0.038	7.30E-01	8.98E-01	
EMO2->CND3	0.109	0.038	0.18	0.115	0.036	2.50E-03	6.68E-03	*
EMO2->EMO3	0.173	0.09	0.256	0.177	0.043	4.86E-05	3.89E-04	*
EMO2->HYP3	0.077	0.013	0.141	0.087	0.033	1.84E-02	3.47E-02	*
EMO2->PER3	0.115	0.037	0.192	0.115	0.04	3.70E-03	8.46E-03	*
HYP1->CND2	0.109	0.026	0.192	0.097	0.042	1.00E-02	2.00E-02	*
HYP1->EMO2	0.036	-0.045	0.117	0.031	0.042	3.84E-01	5.71E-01	
HYP1->HYP2	0.244	0.156	0.331	0.234	0.045	4.60E-08	9.96E-07	*
HYP1->PER2	0.015	-0.081	0.11	0.012	0.049	7.65E-01	9.06E-01	
HYP2->CND3	0.117	0.04	0.195	0.111	0.04	3.06E-03	7.53E-03	*
HYP2->EMO3	0.038	-0.039	0.116	0.035	0.039	3.31E-01	5.30E-01	
HYP2->HYP3	0.224	0.143	0.305	0.229	0.041	6.23E-08	9.96E-07	*
HYP2->PER3	0.074	-0.005	0.154	0.067	0.041	6.80E-02	1.14E-01	
PER1->CND2	0.001	-0.067	0.069	0.001	0.035	9.73E-01	9.73E-01	
PER1->EMO2	0.117	0.05	0.184	0.117	0.034	6.00E-04	2.19E-03	*
PER1->HYP2	-0.022	-0.084	0.041	-0.024	0.032	4.93E-01	6.57E-01	
PER1->PER2	0.165	0.063	0.266	0.159	0.052	1.48E-03	4.75E-03	*
PER2->CND3	0.004	-0.063	0.071	0.004	0.034	9.08E-01	9.47E-01	
PER2->EMO3	0.018	-0.056	0.092	0.019	0.038	6.28E-01	8.03E-01	
PER2->HYP3	-0.008	-0.072	0.056	-0.009	0.033	8.14E-01	9.27E-01	
PER2->PER3	0.156	0.068	0.244	0.16	0.045	5.05E-04	2.19E-03	*

Note: 1 = Age 7; 2 = Age 9; 3 = Age 12; p_value adjusted = Benjamini-Hocberg FDR correction; * = estimate survives FDR

Table S5b. Within-person regressions - NTR

direction	beta	lower CI	upper CI	stdBeta	robust SE	p-value	p_value adjusted	index
CND1->CND2	0.296	0.2	0.393	0.296	0.049	1.86E-09	9.92E-09	*
CND1->EMO2	0.116	0.034	0.198	0.1	0.042	5.39E-03	1.33E-02	*
CND1->HYP2	0.094	0.021	0.167	0.089	0.037	1.14E-02	2.59E-02	*
CND1->PER2	0.126	0.042	0.209	0.107	0.043	3.09E-03	9.22E-03	*
CND2->CND3	0.31	0.227	0.392	0.303	0.042	2.25E-13	2.40E-12	*
CND2->EMO3	0.057	-0.016	0.131	0.05	0.038	1.28E-01	2.16E-01	
CND2->HYP3	0.068	-0.001	0.136	0.065	0.035	5.21E-02	9.81E-02	
CND2->PER3	0.023	-0.051	0.097	0.02	0.038	5.41E-01	6.19E-01	
EMO1->CND2	-0.01	-0.067	0.047	-0.011	0.029	7.34E-01	7.79E-01	
EMO1->EMO2	0.24	0.156	0.324	0.238	0.043	2.30E-08	9.21E-08	*
EMO1->HYP2	0.033	-0.019	0.085	0.036	0.027	2.17E-01	3.02E-01	
EMO1->PER2	0.094	0.029	0.158	0.092	0.033	4.64E-03	1.24E-02	*
EMO2->CND3	0.029	-0.032	0.089	0.033	0.031	3.54E-01	4.72E-01	
EMO2->EMO3	0.259	0.175	0.343	0.261	0.043	1.83E-09	9.92E-09	*
EMO2->HYP3	0.02	-0.037	0.076	0.022	0.029	4.90E-01	6.00E-01	
EMO2->PER3	0.097	0.033	0.162	0.097	0.033	3.17E-03	9.22E-03	*
HYP1->CND2	0.01	-0.051	0.07	0.01	0.031	7.54E-01	7.79E-01	
HYP1->EMO2	-0.023	-0.089	0.044	-0.02	0.034	5.06E-01	6.00E-01	
HYP1->HYP2	0.258	0.181	0.334	0.252	0.039	5.05E-11	4.04E-10	*
HYP1->PER2	0.006	-0.065	0.077	0.005	0.036	8.68E-01	8.68E-01	
HYP2->CND3	0.016	-0.04	0.071	0.016	0.028	5.79E-01	6.39E-01	
HYP2->EMO3	0.05	-0.009	0.11	0.046	0.03	9.61E-02	1.71E-01	
HYP2->HYP3	0.363	0.299	0.428	0.366	0.033	0.00E+00	0.00E+00	*
HYP2->PER3	0.046	-0.017	0.108	0.041	0.032	1.51E-01	2.20E-01	
PER1->CND2	0.044	-0.014	0.101	0.05	0.029	1.36E-01	2.17E-01	
PER1->EMO2	0.08	0.017	0.144	0.08	0.032	1.26E-02	2.68E-02	*
PER1->HYP2	0.044	-0.015	0.103	0.048	0.03	1.44E-01	2.20E-01	
PER1->PER2	0.214	0.14	0.288	0.21	0.038	1.29E-08	5.89E-08	*
PER2->CND3	0.078	0.016	0.14	0.089	0.032	1.40E-02	2.80E-02	*
PER2->EMO3	0.119	0.054	0.185	0.121	0.033	3.34E-04	1.19E-03	*
PER2->HYP3	0.022	-0.035	0.078	0.024	0.029	4.51E-01	5.77E-01	
PER2->PER3	0.322	0.25	0.394	0.324	0.037	0.00E+00	0.00E+00	*

Note: 1 = Age 7; 2 = Age 10; 3 = Age 12; p_value adjusted = Benjamini-Hocberg FDR correction; * = estimate survives FDR

Table 6a. Model fit Indices - TEDS sex differences

	chisq	df	chisqPval	rmsea	srmr	cfi
Equal regressions between groups	62.0019	44	0.0379	0.0098	0.0139	0.9991
Unconstrained	18.41	12	0.1038	0.0112	0.0059	0.9997

Table 6b. Nested model comparison - TEDS sex differences

	Df	AIC	BIC	Chisq	Chisq diff	Df diff	Pr(>Chisq)
Unconstrained	12	168835.278	170020.278	18.41	NA	NA	NA
Equal regressions between groups	44	168814.87	169774.155	62.0019	34.7552	32	0.338

Table 6c. Nested model comparison - NTR sex differences

	chisq	df	chisqPval	rmsea	srmr	cfi
Equal regressions between groups	137.8591	44	0	0.0158	0.0136	0.9989
Unconstrained	55.6022	12	0	0.0206	0.0063	0.9995

Table 6d. Model fit Indices - NTR sex differences

	Df	AIC	BIC	Chisq	Chisq diff	Df diff	Pr(>Chisq)
Unconstrained	12	287238.234	288539.82	55.6022	NA	NA	NA
Equal regressions between groups	44	287256.491	288310.156	137.8591	50.8299	32	0.0185

Table S7a. Within-person regressions - NTR males

direction	beta	lower CI	upper CI	stdBeta	robust SE	p-value	p_value adjusted	index
CND1->CND2	0.376	0.241	0.51	0.369	0.069	4.57E-08	2.44E-07	*
CND1->EMO2	0.138	0.031	0.245	0.127	0.055	1.17E-02	2.87E-02	*
CND1->HYP2	0.084	-0.019	0.186	0.086	0.052	1.09E-01	1.93E-01	
CND1->PER2	0.13	0.016	0.243	0.118	0.058	2.49E-02	5.69E-02	nominal
CND2->CND3	0.408	0.305	0.511	0.396	0.052	7.55E-15	1.42E-13	*
CND2->EMO3	0.112	0.026	0.199	0.104	0.044	1.07E-02	2.85E-02	*
CND2->HYP3	0.115	0.027	0.203	0.114	0.045	1.01E-02	2.85E-02	*
CND2->PER3	0.065	-0.028	0.158	0.059	0.047	1.73E-01	2.63E-01	
EMO1->CND2	0.01	-0.08	0.099	0.01	0.045	8.33E-01	8.33E-01	
EMO1->EMO2	0.335	0.219	0.45	0.335	0.059	1.42E-08	9.11E-08	*
EMO1->HYP2	0.009	-0.074	0.092	0.01	0.043	8.31E-01	8.33E-01	
EMO1->PER2	0.104	0.008	0.201	0.102	0.049	3.47E-02	7.39E-02	nominal
EMO2->CND3	0.037	-0.056	0.131	0.039	0.048	4.32E-01	6.29E-01	
EMO2->EMO3	0.378	0.259	0.497	0.374	0.061	4.78E-10	3.82E-09	*
EMO2->HYP3	0.014	-0.067	0.096	0.015	0.042	7.30E-01	8.06E-01	
EMO2->PER3	0.157	0.062	0.253	0.154	0.049	1.27E-03	4.52E-03	*
HYP1->CND2	0.02	-0.079	0.118	0.018	0.05	6.94E-01	8.06E-01	
HYP1->EMO2	-0.02	-0.119	0.082	-0.016	0.051	7.15E-01	8.06E-01	
HYP1->HYP2	0.262	0.142	0.381	0.254	0.061	1.76E-05	8.06E-05	*
HYP1->PER2	0.042	-0.068	0.152	0.036	0.056	4.54E-01	6.31E-01	
HYP2->CND3	0.011	-0.067	0.089	0.011	0.04	7.77E-01	8.29E-01	
HYP2->EMO3	-0.02	-0.095	0.065	-0.013	0.041	7.14E-01	8.06E-01	
HYP2->HYP3	0.358	0.267	0.448	0.343	0.046	8.88E-15	1.42E-13	*
HYP2->PER3	0.022	-0.063	0.106	0.019	0.043	6.14E-01	7.86E-01	
PER1->CND2	0.031	-0.054	0.115	0.033	0.043	4.78E-01	6.37E-01	
PER1->EMO2	0.075	-0.013	0.162	0.076	0.045	9.38E-02	1.77E-01	
PER1->HYP2	0.065	-0.022	0.151	0.073	0.044	1.41E-01	2.26E-01	
PER1->PER2	0.202	0.097	0.307	0.203	0.054	1.61E-04	6.42E-04	*
PER2->CND3	0.077	-0.009	0.164	0.082	0.044	8.01E-02	1.60E-01	
PER2->EMO3	0.124	0.042	0.207	0.125	0.042	3.06E-03	9.78E-03	*
PER2->HYP3	0.06	-0.017	0.138	0.065	0.04	1.27E-01	2.14E-01	
PER2->PER3	0.316	0.218	0.413	0.315	0.05	2.42E-10	2.58E-09	*

Note: 1 = Age 7; 2 = Age 10; 3 = Age 12; p_value adjusted = Benjamini-Hocberg FDR correction; * = estimate survives FDR; nominal = $\alpha < 0.05$

Table S7b. Within-person regressions - NTR females

direction	beta	lower CI	upper CI	stdBeta	robust SE	p-value	p_value adjusted	index
CND1->CND2	0.196	0.075	0.317	0.198	0.062	1.55E-03	9.90E-03	*
CND1->EMO2	0.093	-0.03	0.215	0.07	0.063	1.38E-01	2.46E-01	
CND1->HYP2	0.109	0.013	0.205	0.096	0.049	2.57E-02	6.85E-02	nominal
CND1->PER2	0.136	0.018	0.254	0.106	0.06	2.35E-02	6.84E-02	nominal
CND2->CND3	0.165	0.044	0.286	0.164	0.062	7.49E-03	3.00E-02	*
CND2->EMO3	-0.04	-0.164	0.08	-0.032	0.062	4.99E-01	6.39E-01	
CND2->HYP3	0.002	-0.096	0.1	0.002	0.05	9.73E-01	9.89E-01	
CND2->PER3	-0.04	-0.152	0.063	-0.035	0.055	4.18E-01	5.81E-01	
EMO1->CND2	-0.02	-0.083	0.049	-0.022	0.034	6.16E-01	7.04E-01	
EMO1->EMO2	0.177	0.065	0.289	0.174	0.057	1.94E-03	1.04E-02	*
EMO1->HYP2	0.041	-0.022	0.104	0.047	0.032	1.98E-01	3.34E-01	
EMO1->PER2	0.079	-0.002	0.161	0.08	0.042	5.66E-02	1.29E-01	
EMO2->CND3	0.044	-0.031	0.119	0.058	0.038	2.53E-01	4.05E-01	
EMO2->EMO3	0.171	0.058	0.284	0.176	0.058	3.06E-03	1.40E-02	*
EMO2->HYP3	0.014	-0.059	0.087	0.018	0.037	7.09E-01	7.71E-01	
EMO2->PER3	0.037	-0.045	0.12	0.039	0.042	3.76E-01	5.73E-01	
HYP1->CND2	0	-0.066	0.067	0.001	0.034	9.89E-01	9.89E-01	
HYP1->EMO2	-0.04	-0.124	0.054	-0.03	0.045	4.39E-01	5.85E-01	
HYP1->HYP2	0.25	0.157	0.343	0.249	0.048	1.51E-07	1.61E-06	*
HYP1->PER2	-0.04	-0.124	0.051	-0.032	0.045	4.08E-01	5.81E-01	
HYP2->CND3	0.014	-0.063	0.09	0.016	0.039	7.23E-01	7.71E-01	
HYP2->EMO3	0.105	0.015	0.195	0.092	0.046	2.26E-02	6.84E-02	nominal
HYP2->HYP3	0.356	0.264	0.448	0.383	0.047	2.73E-14	8.74E-13	*
HYP2->PER3	0.08	-0.011	0.171	0.071	0.047	8.62E-02	1.62E-01	
PER1->CND2	0.062	-0.008	0.133	0.078	0.036	8.29E-02	1.62E-01	
PER1->EMO2	0.086	-0.004	0.177	0.08	0.046	6.20E-02	1.32E-01	
PER1->HYP2	0.021	-0.051	0.094	0.023	0.037	5.63E-01	6.93E-01	
PER1->PER2	0.236	0.137	0.334	0.225	0.05	2.66E-06	2.13E-05	*
PER2->CND3	0.088	0.007	0.169	0.114	0.041	3.23E-02	7.95E-02	nominal
PER2->EMO3	0.126	0.025	0.227	0.126	0.052	1.45E-02	5.16E-02	nominal
PER2->HYP3	-0.02	-0.097	0.057	-0.024	0.039	6.11E-01	7.04E-01	
PER2->PER3	0.34	0.241	0.44	0.346	0.051	2.39E-11	3.83E-10	*

Note: 1 = Age 7; 2 = Age 10; 3 = Age 12; p_value adjusted = Benjamini-Hocberg FDR correction; * = estimate survives FDR; nominal = $\alpha < 0.05$

Table 8a. Model fit Indices - wfRICLPM TEDS

	chisq	df	chisqPval	rmsea	srmr	cfi
Full model	848.8839	428	0	0.0153	0.0335	0.9926
Pruned model	879.3469	458	0	0.0148	0.0337	0.9926
Null model	903.2612	460	0	0.0151	0.0341	0.9922

Table 8b. Nested models comparison - wfRICLPM TEDS

	Df	AIC	BIC	Chisq	Chisq diff	Df diff	Pr(>Chisq)
Full model	428	325127.231	326676.298	848.8839	NA	NA	NA
Pruned model	458	325097.694	326435.525	879.3469	22.5334	30	0.8339
Null model	460	325117.609	326441.356	903.2612	18.1226	2	1.00E-04

Table 8c. Model fit Indices - wfRICLPM NTR

	chisq	df	chisqPval	rmsea	srmr	cfi
Full model	955.6536	428	0	0.0122	0.0314	0.9972
Pruned model	1011.7082	459	0	0.012	0.0314	0.9971
Null model	1023.991	460	0	0.0121	0.0314	0.9971

Table 8b. Nested models comparison - wfRICLPM NTR

	Df	AIC	BIC	Chisq	Chisq diff	Df diff	Pr(>Chisq)
Full model	428	537429.595	539128.572	955.6536	NA	NA	NA
Pruned model	459	537423.649	538883.225	1011.7082	35.4228	31	0.2674
Null model	460	537433.932	538885.786	1023.991	7.3618	1	0.0067

Table 9a. MZ Cross-twin cross-trait correlations of variables under study - TEDS

	CON_t1b	HYP_t1b	PEp_t1b	EMp_t1b	CON_t2b	HYP_t2b	PEp_t2b	EMp_t2b	CON_t3b	HYP_t3b	PEp_t3b	EMp_t3b
CON_t1a	0.747	0	0	0	0	0	0	0	0	0	0	0
HYP_t1a	0.352	0.582	0	0	0	0	0	0	0	0	0	0
PEp_t1a	0.265	0.212	0.57	0	0	0	0	0	0	0	0	0
EMp_t1a	0.258	0.223	0.214	0.611	0	0	0	0	0	0	0	0
CON_t2a	0.428	0.312	0.161	0.166	0.784	0	0	0	0	0	0	0
HYP_t2a	0.334	0.491	0.184	0.194	0.456	0.726	0	0	0	0	0	0
PEp_t2a	0.239	0.232	0.344	0.185	0.306	0.276	0.68	0	0	0	0	0
EMp_t2a	0.232	0.239	0.232	0.351	0.333	0.306	0.309	0.649	0	0	0	0
CON_t3a	0.472	0.325	0.182	0.189	0.485	0.349	0.213	0.253	0.764	0	0	0
HYP_t3a	0.313	0.432	0.189	0.194	0.389	0.514	0.227	0.279	0.417	0.729	0	0
PEp_t3a	0.24	0.236	0.301	0.17	0.189	0.187	0.36	0.244	0.278	0.273	0.709	0
EMp_t3a	0.198	0.217	0.164	0.337	0.213	0.189	0.196	0.394	0.3	0.284	0.3	0.598

Note: a = Sib 1; b = Sib 2; t1 = Age 7; t2 = Age 9; t3 = Age 12

Table 9b. DZ Cross-twin cross-trait correlations of variables under study - TEDS

	CON_t1b	HYP_t1b	PEp_t1b	EMp_t1b	CON_t2b	HYP_t2b	PEp_t2b	EMp_t2b	CON_t3b	HYP_t3b	PEp_t3b	EMp_t3b
CON_t1a	0.443	0	0	0	0	0	0	0	0	0	0	0
HYP_t1a	0.166	0.003	0	0	0	0	0	0	0	0	0	0
PEp_t1a	0.214	0.153	0.29	0	0	0	0	0	0	0	0	0
EMp_t1a	0.208	0.173	0.182	0.382	0	0	0	0	0	0	0	0
CON_t2a	0.265	0.137	0.14	0.18	0.485	0	0	0	0	0	0	0
HYP_t2a	0.141	-0.023	0.134	0.152	0.196	0.123	0	0	0	0	0	0
PEp_t2a	0.157	0.137	0.149	0.122	0.206	0.171	0.392	0	0	0	0	0
EMp_t2a	0.203	0.145	0.152	0.233	0.266	0.193	0.253	0.403	0	0	0	0
CON_t3a	0.273	0.123	0.116	0.168	0.281	0.137	0.192	0.205	0.484	0	0	0
HYP_t3a	0.19	0.028	0.117	0.178	0.2	0.096	0.183	0.21	0.236	0.224	0	0
PEp_t3a	0.188	0.134	0.166	0.153	0.194	0.152	0.241	0.208	0.229	0.21	0.351	0
EMp_t3a	0.162	0.084	0.141	0.204	0.182	0.121	0.169	0.247	0.206	0.19	0.209	0.322

Note: a = Sib 1; b = Sib 2; t1 = Age 7; t2 = Age 9; t3 = Age 12

Table 9c. MZ Cross-twin cross-trait correlations of variables under study - NTR

	CON_t1b	HYP_t1b	PEp_t1b	EMp_t1b	CON_t2b	HYP_t2b	PEp_t2b	EMp_t2b	CON_t3b	HYP_t3b	PEp_t3b	EMp_t3b
CON_t1a	0.824	0	0	0	0	0	0	0	0	0	0	0
HYP_t1a	0.521	0.708	0	0	0	0	0	0	0	0	0	0
PEp_t1a	0.558	0.499	0.691	0	0	0	0	0	0	0	0	0
EMp_t1a	0.459	0.4	0.5	0.686	0	0	0	0	0	0	0	0
CON_t2a	0.621	0.405	0.405	0.322	0.824	0	0	0	0	0	0	0
HYP_t2a	0.414	0.521	0.386	0.282	0.487	0.709	0	0	0	0	0	0
PEp_t2a	0.438	0.399	0.485	0.377	0.548	0.514	0.689	0	0	0	0	0
EMp_t2a	0.36	0.319	0.402	0.476	0.448	0.388	0.484	0.656	0	0	0	0
CON_t3a	0.543	0.328	0.34	0.284	0.605	0.378	0.384	0.333	0.811	0	0	0
HYP_t3a	0.404	0.436	0.347	0.27	0.385	0.502	0.343	0.269	0.505	0.697	0	0
PEp_t3a	0.365	0.364	0.423	0.341	0.369	0.371	0.451	0.374	0.505	0.505	0.691	0
EMp_t3a	0.319	0.295	0.384	0.437	0.36	0.307	0.368	0.477	0.472	0.428	0.517	0.686

Note: a = Sib 1; b = Sib 2; t1 = Age 7; t2 = Age 10; t3 = Age 12

Table 9d. DZ Cross-twin cross-trait correlations of variables under study - NTR

	CON_t1b	HYP_t1b	PEp_t1b	EMp_t1b	CON_t2b	HYP_t2b	PEp_t2b	EMp_t2b	CON_t3b	HYP_t3b	PEp_t3b	EMp_t3b
CON_t1a	0.504	0	0	0	0	0	0	0	0	0	0	0
HYP_t1a	0.327	0.24	0	0	0	0	0	0	0	0	0	0
PEp_t1a	0.417	0.307	0.399	0	0	0	0	0	0	0	0	0
EMp_t1a	0.404	0.342	0.395	0.455	0	0	0	0	0	0	0	0
CON_t2a	0.39	0.257	0.297	0.301	0.477	0	0	0	0	0	0	0
HYP_t2a	0.279	0.171	0.216	0.252	0.319	0.232	0	0	0	0	0	0
PEp_t2a	0.314	0.228	0.267	0.284	0.369	0.287	0.369	0	0	0	0	0
EMp_t2a	0.33	0.273	0.316	0.352	0.408	0.349	0.394	0.473	0	0	0	0
CON_t3a	0.349	0.198	0.255	0.235	0.358	0.212	0.262	0.287	0.496	0	0	0
HYP_t3a	0.234	0.11	0.181	0.218	0.253	0.154	0.2	0.25	0.316	0.23	0	0
PEp_t3a	0.293	0.178	0.246	0.242	0.301	0.199	0.259	0.283	0.372	0.273	0.364	0
EMp_t3a	0.329	0.245	0.301	0.315	0.336	0.262	0.316	0.354	0.405	0.334	0.38	0.482

Note: a = Sib 1; b = Sib 2; t1 = Age 7; t2 = Age 10; t3 = Age 12

Table 10a. Within-person and between-sibling regression estimates from the wfRICLPM - TEDS

direction	beta	lower CI	upper CI	stdBeta	robust SE	p-value	p_value adjusted	index
CNDa1->CNDa2	0.138	0.068	0.207	0.13297737	0.035	9.81E-05	1.05E-03	*
CNDa1->CNDb2	-0.002	-0.067	0.062	-0.0021332	0.033	9.46E-01	9.68E-01	
CNDa1->EMOa2	0.063	0.007	0.118	0.0571944	0.028	2.68E-02	8.13E-02	nominal
CNDa1->EMOb2	0.073	0.019	0.127	0.06651538	0.028	8.19E-03	3.25E-02	*
CNDa1->HYPa2	0.089	0.041	0.138	0.09345268	0.025	2.86E-04	1.91E-03	*
CNDa1->HYPb2	0.04	-0.007	0.086	0.04151165	0.024	9.28E-02	2.33E-01	
CNDa1->PERa2	0.074	0.016	0.132	0.06673742	0.029	1.20E-02	4.28E-02	*
CNDa1->PERb2	0.035	-0.022	0.092	0.03170058	0.029	2.27E-01	4.27E-01	
CNDa2->CNDa3	0.169	0.096	0.242	0.17065674	0.037	5.27E-06	1.12E-04	*
CNDa2->CNDb3	-0.045	-0.112	0.022	-0.0457852	0.034	1.84E-01	4.21E-01	
CNDa2->EMOa3	0.008	-0.048	0.064	0.00759083	0.029	7.84E-01	9.38E-01	
CNDa2->EMOb3	-0.009	-0.065	0.047	-0.008414	0.029	7.61E-01	9.38E-01	
CNDa2->HYPa3	0.106	0.047	0.165	0.11487326	0.03	4.15E-04	2.42E-03	*
CNDa2->HYPb3	-0.002	-0.06	0.056	-0.0018627	0.03	9.54E-01	9.68E-01	
CNDa2->PERa3	0.009	-0.053	0.072	0.00899003	0.032	7.66E-01	9.38E-01	
CNDa2->PERb3	-0.012	-0.072	0.047	-0.0118016	0.03	6.84E-01	9.33E-01	
EMOa1->CNDa2	0.007	-0.039	0.054	0.00732728	0.024	7.64E-01	9.38E-01	
EMOa1->CNDb2	0.001	-0.045	0.047	0.00122194	0.023	9.60E-01	9.68E-01	
EMOa1->EMOa2	0.131	0.072	0.19	0.12789604	0.03	1.26E-05	2.02E-04	*
EMOa1->EMOb2	-0.011	-0.065	0.043	-0.0108392	0.027	6.85E-01	9.33E-01	
EMOa1->HYPa2	0.013	-0.028	0.055	0.01488727	0.021	5.27E-01	8.33E-01	
EMOa1->HYPb2	-0.007	-0.047	0.032	-0.0082141	0.02	7.16E-01	9.38E-01	
EMOa1->PERa2	0.028	-0.025	0.082	0.02720481	0.027	3.00E-01	5.42E-01	
EMOa1->PERb2	-0.014	-0.065	0.037	-0.0135339	0.026	5.89E-01	8.89E-01	
EMOa2->CNDa3	0.069	0.017	0.12	0.07350291	0.026	8.64E-03	3.25E-02	*
EMOa2->CNDb3	0.048	-0.003	0.098	0.05096716	0.026	6.28E-02	1.68E-01	
EMOa2->EMOa3	0.183	0.12	0.246	0.18616619	0.032	1.39E-08	4.44E-07	*
EMOa2->EMOb3	-0.004	-0.063	0.055	-0.0039211	0.03	8.99E-01	9.68E-01	
EMOa2->HYPa3	0.073	0.02	0.126	0.08380688	0.027	6.61E-03	2.82E-02	*
EMOa2->HYPb3	0.034	-0.018	0.086	0.03947086	0.027	1.94E-01	4.27E-01	
EMOa2->PERa3	0.105	0.049	0.16	0.1053955	0.028	2.20E-04	1.76E-03	*
EMOa2->PERb3	0.019	-0.036	0.074	0.01890822	0.028	5.02E-01	8.24E-01	
HYPa1->CNDa2	0.088	0.026	0.15	0.07986594	0.032	5.27E-03	2.41E-02	*
HYPa1->CNDb2	0.004	-0.055	0.062	0.00336244	0.03	9.01E-01	9.68E-01	
HYPa1->EMOa2	0.095	0.036	0.155	0.08189772	0.031	1.74E-03	8.58E-03	*
HYPa1->EMOb2	0.037	-0.021	0.094	0.03130904	0.029	2.10E-01	4.27E-01	
HYPa1->HYPa2	0.218	0.154	0.281	0.21314632	0.032	1.89E-11	1.21E-09	*
HYPa1->HYPb2	0.002	-0.056	0.061	0.002193	0.03	9.41E-01	9.68E-01	
HYPa1->PERa2	0.039	-0.023	0.102	0.033184	0.032	2.18E-01	4.27E-01	
HYPa1->PERb2	0.036	-0.022	0.094	0.03018317	0.03	2.27E-01	4.27E-01	
HYPa2->CNDa3	0.065	0.005	0.125	0.06070608	0.031	3.34E-02	9.29E-02	nominal
HYPa2->CNDb3	0.007	-0.05	0.065	0.00669514	0.029	8.07E-01	9.40E-01	
HYPa2->EMOa3	0.071	0.008	0.134	0.06329438	0.032	2.68E-02	8.13E-02	nominal
HYPa2->EMOb3	-0.015	-0.075	0.045	-0.0135316	0.031	6.22E-01	9.05E-01	
HYPa2->HYPa3	0.149	0.068	0.23	0.14950582	0.041	2.99E-04	1.91E-03	*
HYPa2->HYPb3	0.084	0.009	0.159	0.08442537	0.038	2.80E-02	8.13E-02	nominal
HYPa2->PERa3	0.076	0.011	0.141	0.06691503	0.033	2.23E-02	7.51E-02	nominal
HYPa2->PERb3	0.009	-0.056	0.073	0.00765615	0.033	7.92E-01	9.38E-01	
PERa1->CNDa2	0.005	-0.043	0.053	0.00525089	0.025	8.37E-01	9.57E-01	
PERa1->CNDb2	-0.01	-0.058	0.038	-0.0104461	0.025	6.85E-01	9.33E-01	
PERa1->EMOa2	0.102	0.052	0.153	0.10063477	0.026	7.27E-05	9.31E-04	*
PERa1->EMOb2	0.025	-0.024	0.074	0.02490162	0.025	3.13E-01	5.42E-01	
PERa1->HYPa2	-0.011	-0.053	0.031	-0.0127191	0.021	5.97E-01	8.89E-01	
PERa1->HYPb2	0.019	-0.021	0.059	0.02112345	0.021	3.61E-01	6.07E-01	
PERa1->PERa2	0.139	0.065	0.213	0.13508431	0.038	2.16E-04	1.76E-03	*
PERa1->PERb2	0.033	-0.031	0.096	0.03192491	0.032	3.10E-01	5.42E-01	
PERa2->CNDa3	0.034	-0.019	0.087	0.03674194	0.027	2.09E-01	4.27E-01	
PERa2->CNDb3	0.009	-0.041	0.06	0.01003833	0.026	7.19E-01	9.38E-01	
PERa2->EMOa3	0.046	-0.013	0.106	0.04781837	0.03	1.27E-01	3.00E-01	
PERa2->EMOb3	-0.001	-0.056	0.054	-0.0011599	0.028	9.68E-01	9.68E-01	
PERa2->HYPa3	0.018	-0.04	0.076	0.02136383	0.03	5.33E-01	8.33E-01	
PERa2->HYPb3	-0.004	-0.06	0.051	-0.0050893	0.028	8.78E-01	9.68E-01	
PERa2->PERa3	0.134	0.054	0.214	0.13682848	0.041	1.02E-03	5.46E-03	*
PERa2->PERb3	0.057	-0.01	0.125	0.05866529	0.034	9.47E-02	2.33E-01	

Note: a = Sib 1; b = Sib 2; t1 = Age 7; t2 = Age 9; t3 = Age 12; pvalue adjusted = Benjamini-Hocberg FDR correction; *= estimate survives FDR; nominal = $\alpha < 0.05$

Table 10b. Within-person and between-sibling regression estimates from the wfRICLPM - NTR

direction	beta	lower CI	upper CI	stdBeta	robust SE	p-value	p_value adjusted	index
CNDa1->CNDa2	0.265	0.194	0.336	0.26321359	0.036	2.43E-13	2.22E-12	*
CNDa1->CNDb2	-0.017	-0.079	0.045	-0.016654	0.032	5.97E-01	8.13E-01	
CNDa1->EMOa2	0.06	-0.002	0.121	0.05108393	0.031	5.83E-02	1.97E-01	
CNDa1->EMOb2	0.075	0.017	0.133	0.06404189	0.03	1.18E-02	5.38E-02	nominal
CNDa1->HYPa2	0.06	0.006	0.114	0.05822213	0.027	2.81E-02	1.12E-01	nominal
CNDa1->HYPb2	0.022	-0.027	0.072	0.02168025	0.025	3.70E-01	6.77E-01	
CNDa1->PERa2	0.072	0.011	0.133	0.06199938	0.031	2.03E-02	8.67E-02	nominal
CNDa1->PERb2	0.039	-0.017	0.095	0.03374877	0.028	1.68E-01	4.67E-01	
CNDa2->CNDa3	0.284	0.22	0.349	0.27796539	0.033	0.00E+00	0.00E+00	*
CNDa2->CNDb3	-0.004	-0.058	0.05	-0.0040232	0.028	8.82E-01	9.40E-01	
CNDa2->EMOa3	0.004	-0.049	0.056	0.00307434	0.027	8.96E-01	9.40E-01	
CNDa2->EMOb3	0.025	-0.025	0.075	0.02177773	0.025	3.24E-01	6.56E-01	
CNDa2->HYPa3	0.075	0.023	0.126	0.07165066	0.026	4.41E-03	2.42E-02	*
CNDa2->HYPb3	-0.011	-0.057	0.034	-0.0109286	0.023	6.21E-01	8.19E-01	
CNDa2->PERa3	0.013	-0.041	0.067	0.01151021	0.028	6.30E-01	8.19E-01	
CNDa2->PERb3	-0.026	-0.074	0.023	-0.0221178	0.025	2.98E-01	6.56E-01	
EMOa1->CNDa2	-0.004	-0.041	0.034	-0.0040876	0.019	8.53E-01	9.40E-01	
EMOa1->CNDb2	0.015	-0.022	0.051	0.0167053	0.019	4.35E-01	7.33E-01	
EMOa1->EMOa2	0.225	0.165	0.284	0.22168664	0.03	1.47E-13	1.57E-12	*
EMOa1->EMOb2	-0.024	-0.077	0.028	-0.02413	0.027	3.60E-01	6.77E-01	
EMOa1->HYPa2	-0.004	-0.042	0.035	-0.00424	0.02	8.47E-01	9.40E-01	
EMOa1->HYPb2	0.012	-0.026	0.051	0.01385725	0.02	5.26E-01	7.83E-01	
EMOa1->PERa2	0.066	0.02	0.111	0.06475869	0.023	4.53E-03	2.42E-02	*
EMOa1->PERb2	0.014	-0.029	0.057	0.01381646	0.022	5.26E-01	7.83E-01	
EMOa2->CNDa3	0.029	-0.013	0.072	0.03302273	0.022	1.79E-01	4.78E-01	
EMOa2->CNDb3	0.014	-0.026	0.055	0.01626962	0.021	4.90E-01	7.65E-01	
EMOa2->EMOa3	0.265	0.205	0.326	0.26800766	0.031	0.00E+00	0.00E+00	*
EMOa2->EMOb3	-0.014	-0.065	0.037	-0.0144167	0.026	5.85E-01	8.13E-01	
EMOa2->HYPa3	0.011	-0.03	0.053	0.01262622	0.021	5.91E-01	8.13E-01	
EMOa2->HYPb3	0	-0.041	0.04	-0.0003368	0.021	9.88E-01	9.88E-01	
EMOa2->PERa3	0.084	0.037	0.131	0.08439101	0.024	4.46E-04	2.86E-03	*
EMOa2->PERb3	0.016	-0.029	0.061	0.01599754	0.023	4.89E-01	7.65E-01	
HYPa1->CNDa2	0.005	-0.038	0.048	0.00504575	0.022	8.22E-01	9.40E-01	
HYPa1->CNDb2	0.008	-0.032	0.049	0.00853765	0.021	6.85E-01	8.43E-01	
HYPa1->EMOa2	-0.032	-0.082	0.017	-0.0283859	0.025	2.01E-01	4.99E-01	
HYPa1->EMOb2	0.014	-0.035	0.062	0.01193359	0.025	5.84E-01	8.13E-01	
HYPa1->HYPa2	0.212	0.156	0.269	0.21019865	0.029	1.15E-13	1.48E-12	*
HYPa1->HYPb2	-0.009	-0.06	0.042	-0.0087759	0.026	7.33E-01	8.85E-01	
HYPa1->PERa2	0.033	-0.018	0.085	0.02935451	0.026	2.03E-01	4.99E-01	
HYPa1->PERb2	-0.004	-0.052	0.044	-0.0031909	0.025	8.82E-01	9.40E-01	
HYPa2->CNDa3	0.004	-0.037	0.045	0.00395298	0.021	8.50E-01	9.40E-01	
HYPa2->CNDb3	-0.008	-0.045	0.029	-0.0080556	0.019	6.74E-01	8.43E-01	
HYPa2->EMOa3	0.022	-0.021	0.066	0.01997138	0.022	3.17E-01	6.56E-01	
HYPa2->EMOb3	-0.002	-0.042	0.038	-0.0019329	0.02	9.16E-01	9.46E-01	
HYPa2->HYPa3	0.299	0.25	0.349	0.29560394	0.025	0.00E+00	0.00E+00	*
HYPa2->HYPb3	-0.016	-0.06	0.028	-0.0155703	0.022	4.80E-01	7.65E-01	
HYPa2->PERa3	0.032	-0.013	0.077	0.02835199	0.023	1.67E-01	4.67E-01	
HYPa2->PERb3	-0.021	-0.06	0.017	-0.0190994	0.02	2.80E-01	6.43E-01	
PERa1->CNDa2	0.037	-0.004	0.077	0.04167413	0.021	7.95E-02	2.54E-01	
PERa1->CNDb2	-0.016	-0.053	0.02	-0.018621	0.019	3.81E-01	6.77E-01	
PERa1->EMOa2	0.052	0.005	0.098	0.05085748	0.024	3.01E-02	1.13E-01	nominal
PERa1->EMOb2	0.023	-0.019	0.066	0.02296426	0.022	2.81E-01	6.43E-01	
PERa1->HYPa2	0.041	-0.001	0.083	0.04509274	0.021	5.74E-02	1.97E-01	
PERa1->HYPb2	0.019	-0.019	0.056	0.02056135	0.019	3.28E-01	6.56E-01	
PERa1->PERa2	0.19	0.135	0.246	0.18743672	0.028	1.60E-11	1.28E-10	*
PERa1->PERb2	-0.012	-0.06	0.037	-0.0114287	0.025	6.40E-01	8.19E-01	
PERa2->CNDa3	0.058	0.015	0.101	0.06536797	0.022	8.63E-03	4.25E-02	*
PERa2->CNDb3	-0.004	-0.043	0.035	-0.0043327	0.02	8.49E-01	9.40E-01	
PERa2->EMOa3	0.113	0.067	0.159	0.1139425	0.023	1.35E-06	9.57E-06	*
PERa2->EMOb3	0.02	-0.023	0.062	0.02007721	0.022	3.60E-01	6.77E-01	
PERa2->HYPa3	0.035	-0.006	0.076	0.03875089	0.021	9.26E-02	2.82E-01	
PERa2->HYPb3	0.016	-0.022	0.053	0.01733305	0.019	4.14E-01	7.16E-01	
PERa2->PERa3	0.32	0.268	0.372	0.31996245	0.026	0.00E+00	0.00E+00	*
PERa2->PERb3	0.001	-0.046	0.047	0.00081036	0.024	9.73E-01	9.88E-01	

Note: a = Sib 1; b = Sib 2; t1 = Age 7; t2 = Age 10; t3 = Age 12;pvalue adjusted = Benjamini-Hocberg FDR correction; *= estimate survives FDR nominal = $\alpha < 0.05$

Appendix 6 – Supplementary material for Chapter 6

Supplementary figures

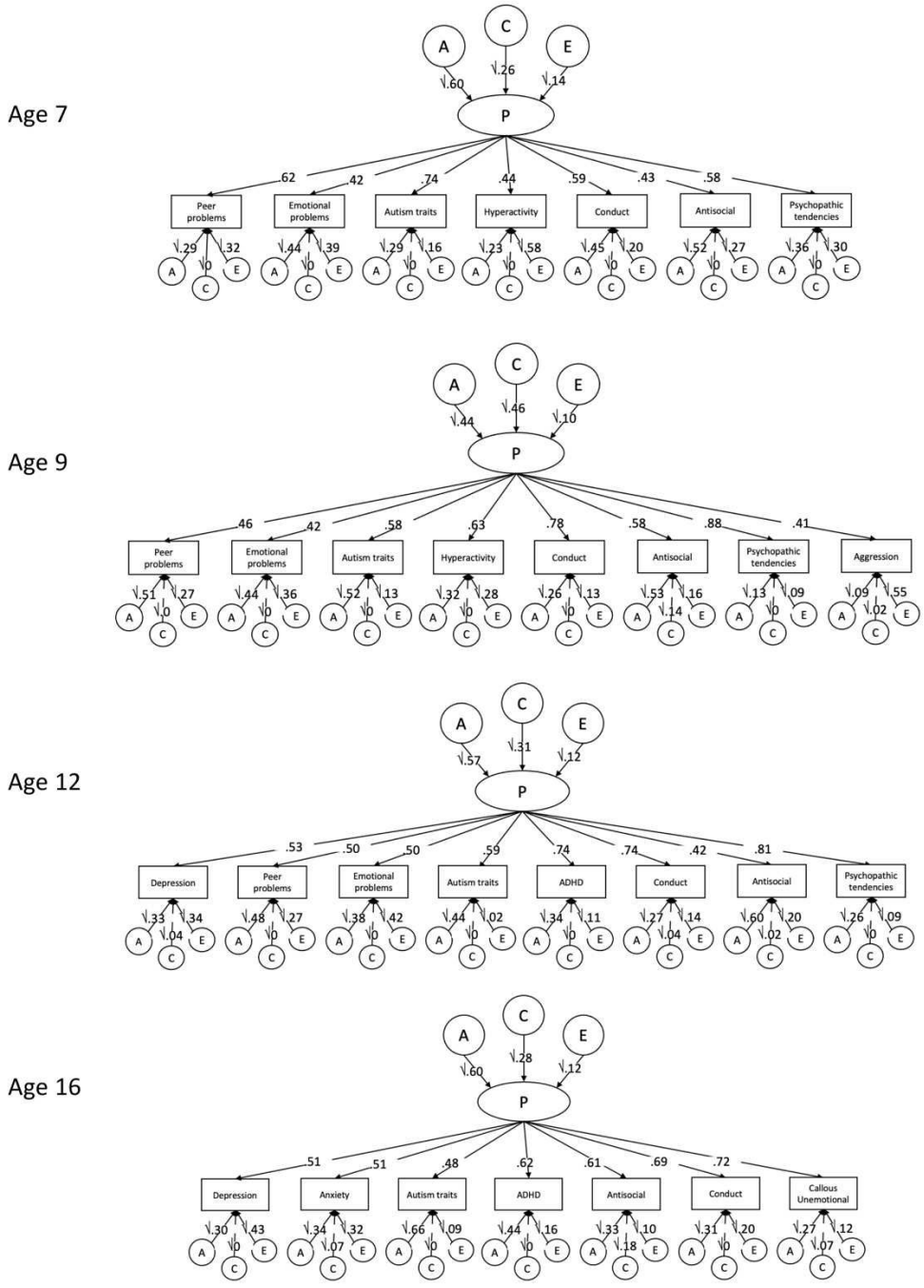


Figure S1. Common pathway model presenting standardized squared path estimates for genetic (A), shared environmental (C) and unique environmental (E) influences on the p-factor obtained from different parent rated psychopathology measures across time (age 7, 9, 12 and 16). In the diagram are also reported measure specific ACE estimates as well as loadings on the p-factor (arrows leaving P).

Supplementary Tables

Table S1. Logistic regression results of droput status vs demographic variables available at recruitment in the full sample

outcome	predictor	b	2.5% CI	97.5% CI	SE	Zscore	Pval	OR	2.5% CI OR	97.5% CI OR	index
parent_7_pc	SEX	0.162	0.096	0.229	0.034	4.78	1.76E-06	1.176	1.1	1.257	*
parent_7_pc	QUAL_father	-0.124	-0.14	-0.108	0.008	-15.164	6.15E-52	0.883	0.869	0.897	*
parent_7_pc	QUAL_mother	-0.084	-0.099	-0.069	0.008	-10.764	5.08E-27	0.919	0.905	0.934	*
parent_7_pc	EL_mother	-0.072	-0.094	-0.05	0.011	-6.319	2.63E-10	0.931	0.91	0.951	*
parent_7_pc	EL_father	-0.054	-0.068	-0.04	0.007	-7.508	6.02E-14	0.948	0.934	0.961	*
parent_7_pc	SES	-0.251	-0.285	-0.217	0.017	-14.365	8.63E-47	0.778	0.752	0.805	*
parent_9_pc	SEX	0.183	0.095	0.272	0.045	4.071	4.68E-05	1.201	1.1	1.312	*
parent_9_pc	QUAL_father	-0.089	-0.11	-0.067	0.011	-8.131	4.24E-16	0.915	0.896	0.935	*
parent_9_pc	QUAL_mother	-0.067	-0.087	-0.047	0.01	-6.488	8.68E-11	0.935	0.917	0.954	*
parent_9_pc	EL_mother	-0.022	-0.051	0.007	0.015	-1.505	1.32E-01	0.978	0.95	1.007	
parent_9_pc	EL_father	-0.036	-0.055	-0.018	0.009	-3.832	1.27E-04	0.965	0.947	0.982	*
parent_9_pc	SES	-0.165	-0.21	-0.12	0.023	-7.216	5.35E-13	0.848	0.811	0.887	*
parent_12_pc	SEX	0.277	0.203	0.351	0.038	7.331	2.28E-13	1.319	1.225	1.42	*
parent_12_pc	QUAL_father	-0.152	-0.17	-0.135	0.009	-17.03	4.89E-65	0.859	0.844	0.874	*
parent_12_pc	QUAL_mother	-0.112	-0.129	-0.095	0.009	-13.133	2.13E-39	0.894	0.879	0.909	*
parent_12_pc	EL_mother	-0.094	-0.119	-0.069	0.013	-7.372	1.68E-13	0.91	0.888	0.933	*
parent_12_pc	EL_father	-0.066	-0.081	-0.05	0.008	-8.282	1.21E-16	0.936	0.922	0.951	*
parent_12_pc	SES	-0.318	-0.356	-0.28	0.019	-16.506	3.31E-61	0.728	0.701	0.756	*
parent_16_pc	SEX	0.275	0.201	0.349	0.038	7.319	2.50E-13	1.317	1.223	1.417	*
parent_16_pc	QUAL_father	-0.15	-0.168	-0.133	0.009	-16.874	7.02E-64	0.86	0.846	0.876	*
parent_16_pc	QUAL_mother	-0.111	-0.128	-0.095	0.009	-13.082	4.16E-39	0.895	0.88	0.91	*
parent_16_pc	EL_mother	-0.09	-0.115	-0.065	0.013	-7.074	1.51E-12	0.914	0.892	0.937	*
parent_16_pc	EL_father	-0.065	-0.081	-0.05	0.008	-8.247	1.62E-16	0.937	0.922	0.952	*
parent_16_pc	SES	-0.314	-0.352	-0.276	0.019	-16.363	3.51E-60	0.731	0.704	0.759	*
teacher_7_pc	SEX	0.132	0.06	0.203	0.037	3.606	3.12E-04	1.141	1.062	1.226	*
teacher_7_pc	QUAL_father	-0.124	-0.141	-0.106	0.009	-14.062	6.47E-45	0.884	0.869	0.899	*
teacher_7_pc	QUAL_mother	-0.084	-0.101	-0.068	0.008	-10.071	7.43E-24	0.919	0.904	0.934	*
teacher_7_pc	EL_mother	-0.072	-0.097	-0.048	0.012	-5.872	4.31E-09	0.93	0.908	0.953	*
teacher_7_pc	EL_father	-0.056	-0.071	-0.04	0.008	-7.182	6.87E-13	0.946	0.932	0.96	*
teacher_7_pc	SES	-0.254	-0.29	-0.217	0.019	-13.529	1.06E-41	0.776	0.748	0.805	*
teacher_9_pc	SEX	0.237	0.136	0.339	0.052	4.574	4.80E-06	1.268	1.146	1.404	*
teacher_9_pc	QUAL_father	-0.083	-0.108	-0.059	0.013	-6.642	3.09E-11	0.92	0.898	0.943	*
teacher_9_pc	QUAL_mother	-0.068	-0.091	-0.044	0.012	-5.723	1.05E-08	0.935	0.913	0.957	*
teacher_9_pc	EL_mother	-0.029	-0.063	0.004	0.017	-1.703	8.86E-02	0.971	0.939	1.004	
teacher_9_pc	EL_father	-0.047	-0.068	-0.026	0.011	-4.359	1.31E-05	0.954	0.934	0.974	*
teacher_9_pc	SES	-0.175	-0.226	-0.123	0.026	-6.66	2.73E-11	0.84	0.798	0.884	*
teacher_12_pc	SEX	0.296	0.207	0.385	0.045	6.518	7.14E-11	1.344	1.23	1.469	*
teacher_12_pc	QUAL_father	-0.166	-0.187	-0.145	0.011	-15.76	5.83E-56	0.847	0.83	0.865	*
teacher_12_pc	QUAL_mother	-0.118	-0.138	-0.098	0.01	-11.561	6.49E-31	0.889	0.871	0.907	*
teacher_12_pc	EL_mother	-0.097	-0.127	-0.067	0.015	-6.365	1.95E-10	0.908	0.881	0.935	*
teacher_12_pc	EL_father	-0.079	-0.098	-0.061	0.01	-8.266	1.39E-16	0.924	0.906	0.941	*
teacher_12_pc	SES	-0.346	-0.391	-0.301	0.023	-15.041	3.93E-51	0.708	0.676	0.74	*
child_9_pc	SEX	0.206	0.116	0.296	0.046	4.489	7.16E-06	1.229	1.123	1.344	*
child_9_pc	QUAL_father	-0.094	-0.115	-0.072	0.011	-8.467	2.53E-17	0.911	0.891	0.931	*
child_9_pc	QUAL_mother	-0.073	-0.094	-0.053	0.01	-7.008	2.43E-12	0.929	0.91	0.949	*
child_9_pc	EL_mother	-0.029	-0.059	0	0.015	-1.939	5.25E-02	0.971	0.943	1	
child_9_pc	EL_father	-0.037	-0.056	-0.018	0.01	-3.86	1.14E-04	0.964	0.946	0.982	*
child_9_pc	SES	-0.179	-0.224	-0.133	0.023	-7.678	1.61E-14	0.836	0.799	0.875	*
child_12_pc	SEX	0.157	0.093	0.222	0.033	4.813	1.48E-06	1.17	1.098	1.248	*
child_12_pc	QUAL_father	-0.123	-0.138	-0.107	0.008	-15.456	6.85E-54	0.884	0.871	0.898	*
child_12_pc	QUAL_mother	-0.083	-0.098	-0.068	0.008	-11.007	3.54E-28	0.92	0.907	0.934	*
child_12_pc	EL_mother	-0.065	-0.087	-0.044	0.011	-5.961	2.51E-09	0.937	0.917	0.957	*
child_12_pc	EL_father	-0.054	-0.068	-0.041	0.007	-7.831	4.85E-15	0.947	0.934	0.96	*
child_12_pc	SES	-0.244	-0.277	-0.211	0.017	-14.434	3.16E-47	0.784	0.758	0.81	*
child_16_pc	SEX	0.42	0.311	0.531	0.056	7.492	6.79E-14	1.523	1.364	1.7	*
child_16_pc	QUAL_father	-0.142	-0.167	-0.117	0.013	-10.973	5.14E-28	0.867	0.846	0.89	*
child_16_pc	QUAL_mother	-0.114	-0.138	-0.09	0.012	-9.219	3.01E-20	0.892	0.871	0.914	*
child_16_pc	EL_mother	-0.069	-0.106	-0.032	0.019	-3.682	2.31E-04	0.933	0.9	0.968	*
child_16_pc	EL_father	-0.066	-0.089	-0.043	0.012	-5.684	1.32E-08	0.936	0.915	0.958	*
child_16_pc	SES	-0.3	-0.355	-0.246	0.028	-10.76	5.30E-27	0.741	0.701	0.782	*

Note. N full sample = 27,444; Outcome is dropout status (yes = 1, no = 0) with respect to complete cases on measures employed in principal component (pc) analyses in Chapter 4 (see also supplementary Table S5 in Appendix 3); Sex was coded as 0 = females, 1 = males; EL = employment level (ranging from 1 = unemployed to 9 = manager); QUAL = highest qualification obtained (ranging from 1= none to 8 = postgrad); SES = socio economic status composite as descirbed in Chapter 3. index = whether association survived Bonferroni correction for multiple testing.

Table S2. Logistic regression results of droput status vs demographic variables available at recruitment in the genoryped sample

outcome	predictor	b	2.5% CI	97.5% CI	SE	Zscore	Pval	OR	2.5% CI OR	97.5% CI OR	index
parent_7_pc	**SEX**	0.111	0.032	0.19	0.04	2.766	5.68E-03	1.118	1.033	1.21	
parent_7_pc	**QUAL_father**	-0.08	-0.1	-0.06	0.01	-7.975	1.52E-15	0.923	0.905	0.941	*
parent_7_pc	**QUAL_mother**	-0.063	-0.082	-0.045	0.009	-6.701	2.07E-11	0.939	0.922	0.956	*
parent_7_pc	**EL_mother**	-0.038	-0.065	-0.011	0.014	-2.784	5.37E-03	0.963	0.937	0.989	
parent_7_pc	**EL_father**	-0.033	-0.05	-0.016	0.008	-3.87	1.09E-04	0.968	0.952	0.984	*
parent_7_pc	**SES**	-0.158	-0.199	-0.118	0.021	-7.602	2.91E-14	0.854	0.819	0.889	*
parent_9_pc	**SEX**	0.124	0.029	0.219	0.049	2.554	1.06E-02	1.132	1.029	1.245	
parent_9_pc	**QUAL_father**	-0.03	-0.053	-0.006	0.012	-2.463	1.38E-02	0.971	0.948	0.994	
parent_9_pc	**QUAL_mother**	-0.036	-0.058	-0.014	0.011	-3.166	1.54E-03	0.965	0.944	0.987	
parent_9_pc	**EL_mother**	0.023	-0.009	0.055	0.016	1.431	1.52E-01	1.023	0.991	1.056	
parent_9_pc	**EL_father**	-0.009	-0.029	0.011	0.01	-0.918	3.59E-01	0.991	0.971	1.011	
parent_9_pc	**SES**	-0.045	-0.093	0.004	0.025	-1.802	7.16E-02	0.956	0.911	1.004	
parent_12_pc	**SEX**	0.252	0.168	0.336	0.043	5.9	3.63E-09	1.286	1.183	1.399	*
parent_12_pc	**QUAL_father**	-0.114	-0.135	-0.094	0.01	-10.918	9.41E-28	0.892	0.874	0.911	*
parent_12_pc	**QUAL_mother**	-0.097	-0.116	-0.077	0.01	-9.809	1.03E-22	0.908	0.89	0.926	*
parent_12_pc	**EL_mother**	-0.066	-0.094	-0.037	0.015	-4.515	6.33E-06	0.936	0.91	0.963	*
parent_12_pc	**EL_father**	-0.047	-0.064	-0.029	0.009	-5.198	2.02E-07	0.954	0.938	0.971	*
parent_12_pc	**SES**	-0.239	-0.282	-0.196	0.022	-10.862	1.74E-27	0.787	0.754	0.822	*
parent_16_pc	**SEX**	0.25	0.167	0.334	0.043	5.882	4.05E-09	1.284	1.182	1.396	*
parent_16_pc	**QUAL_father**	-0.112	-0.132	-0.091	0.01	-10.714	8.72E-27	0.894	0.876	0.913	*
parent_16_pc	**QUAL_mother**	-0.096	-0.115	-0.077	0.01	-9.741	2.02E-22	0.909	0.891	0.926	*
parent_16_pc	**EL_mother**	-0.06	-0.089	-0.032	0.014	-4.149	3.34E-05	0.942	0.915	0.969	*
parent_16_pc	**EL_father**	-0.046	-0.064	-0.029	0.009	-5.144	2.69E-07	0.955	0.938	0.972	*
parent_16_pc	**SES**	-0.234	-0.277	-0.191	0.022	-10.672	1.37E-26	0.791	0.758	0.826	*
teacher_7_pc	**SEX**	0.067	-0.015	0.149	0.042	1.596	1.11E-01	1.069	0.985	1.16	
teacher_7_pc	**QUAL_father**	-0.076	-0.096	-0.056	0.01	-7.368	1.73E-13	0.927	0.908	0.946	*
teacher_7_pc	**QUAL_mother**	-0.061	-0.08	-0.042	0.01	-6.239	4.39E-10	0.941	0.923	0.959	*
teacher_7_pc	**EL_mother**	-0.038	-0.066	-0.01	0.014	-2.642	8.24E-03	0.963	0.937	0.99	
teacher_7_pc	**EL_father**	-0.034	-0.051	-0.017	0.009	-3.842	1.22E-04	0.967	0.95	0.984	*
teacher_7_pc	**SES**	-0.157	-0.199	-0.115	0.022	-7.272	3.54E-13	0.855	0.819	0.892	*
teacher_9_pc	**SEX**	0.181	0.074	0.288	0.055	3.302	9.61E-04	1.198	1.076	1.334	
teacher_9_pc	**QUAL_father**	-0.024	-0.05	0.003	0.013	-1.768	7.70E-02	0.976	0.951	1.003	
teacher_9_pc	**QUAL_mother**	-0.036	-0.06	-0.011	0.013	-2.84	4.51E-03	0.965	0.941	0.989	
teacher_9_pc	**EL_mother**	0.013	-0.022	0.049	0.018	0.737	4.61E-01	1.013	0.978	1.05	
teacher_9_pc	**EL_father**	-0.022	-0.044	0	0.011	-1.929	5.37E-02	0.978	0.957	1	
teacher_9_pc	**SES**	-0.057	-0.112	-0.003	0.028	-2.052	4.01E-02	0.944	0.894	0.997	
teacher_12_pc	**SEX**	0.254	0.159	0.35	0.049	5.208	1.91E-07	1.29	1.172	1.42	*
teacher_12_pc	**QUAL_father**	-0.124	-0.147	-0.101	0.012	-10.621	2.38E-26	0.883	0.863	0.904	*
teacher_12_pc	**QUAL_mother**	-0.096	-0.118	-0.074	0.011	-8.625	6.43E-18	0.908	0.889	0.929	*
teacher_12_pc	**EL_mother**	-0.066	-0.098	-0.033	0.017	-3.951	7.79E-05	0.937	0.906	0.967	*
teacher_12_pc	**EL_father**	-0.06	-0.08	-0.039	0.01	-5.773	7.80E-09	0.942	0.923	0.961	*
teacher_12_pc	**SES**	-0.258	-0.307	-0.209	0.025	-10.306	6.65E-25	0.772	0.735	0.811	*
child_9_pc	**SEX**	0.15	0.053	0.246	0.049	3.035	2.40E-03	1.161	1.055	1.279	
child_9_pc	**QUAL_father**	-0.036	-0.059	-0.012	0.012	-2.941	3.27E-03	0.965	0.942	0.988	
child_9_pc	**QUAL_mother**	-0.043	-0.065	-0.021	0.011	-3.792	1.50E-04	0.958	0.937	0.979	*
child_9_pc	**EL_mother**	0.015	-0.018	0.047	0.016	0.889	3.74E-01	1.015	0.982	1.048	
child_9_pc	**EL_father**	-0.01	-0.031	0.01	0.01	-1.008	3.13E-01	0.99	0.97	1.01	
child_9_pc	**SES**	-0.061	-0.11	-0.012	0.025	-2.42	1.55E-02	0.941	0.896	0.989	
child_12_pc	**SEX**	0.108	0.03	0.186	0.04	2.71	6.72E-03	1.114	1.03	1.204	
child_12_pc	**QUAL_father**	-0.079	-0.099	-0.06	0.01	-7.971	1.58E-15	0.924	0.906	0.942	*
child_12_pc	**QUAL_mother**	-0.064	-0.082	-0.045	0.009	-6.822	8.95E-12	0.938	0.921	0.956	*
child_12_pc	**EL_mother**	-0.029	-0.055	-0.002	0.014	-2.129	3.32E-02	0.972	0.946	0.998	
child_12_pc	**EL_father**	-0.034	-0.051	-0.018	0.008	-4.078	4.54E-05	0.966	0.951	0.982	*
child_12_pc	**SES**	-0.151	-0.191	-0.111	0.021	-7.334	2.23E-13	0.86	0.826	0.895	*
child_16_pc	**SEX**	0.379	0.264	0.494	0.059	6.459	1.05E-10	1.461	1.303	1.639	*
child_16_pc	**QUAL_father**	-0.092	-0.119	-0.064	0.014	-6.604	4.00E-11	0.912	0.888	0.938	*
child_16_pc	**QUAL_mother**	-0.088	-0.114	-0.062	0.013	-6.69	2.23E-11	0.916	0.893	0.94	*
child_16_pc	**EL_mother**	-0.032	-0.071	0.007	0.02	-1.607	1.08E-01	0.969	0.932	1.007	
child_16_pc	**EL_father**	-0.042	-0.066	-0.019	0.012	-3.497	4.71E-04	0.959	0.936	0.981	*
child_16_pc	**SES**	-0.198	-0.255	-0.14	0.029	-6.7	2.08E-11	0.821	0.775	0.87	*

Note. N genotyped sample = 10,346; Outcome is dropout status (yes = 1, no = 0) with respect to complete cases on measures employed in principal component (pc) analyses in Chapter 4 (see also supplementary Table S5 in Appendix 3); Sex was coded as 0 = females 1, = males; EL = employment level (ranging from 1 = unemployed to 9 = manager); QUAL = highest qualification obtained (ranging from 1= none to 8 = postgrad); SES = socio economic status composite as descirbed in Chapter 3. index = whether association survived Bonferroni correction for multiple testing.