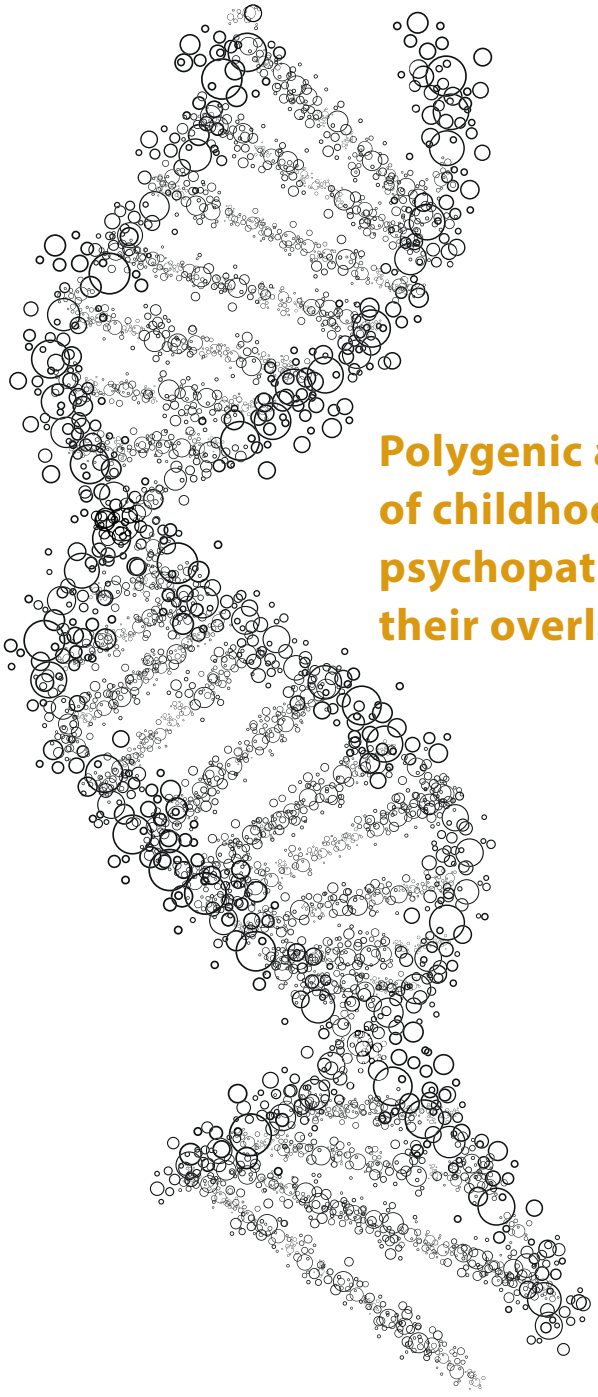


Polygenic analyses of childhood and adult psychopathology, and their overlap

Wonuola A. Akingbuwa

Polygenic analyses of childhood and adult psychopathology, and their overlap

BY

Wonuola A. Akingbuwa

Layout and print: Guus Gijben | Proefschrift-aio.nl

ISBN: 978-94-93270-42-8

VRIJE UNIVERSITEIT

POLYGENIC ANALYSES OF CHILDHOOD AND ADULT PSYCHOPATHOLOGY, AND THEIR OVERLAP

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. J.J.G. Geurts,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Gedrags- en Bewegingswetenschappen
op woensdag 16 februari 2022 om 11.45 uur
in een bijeenkomst van de universiteit,
De Boelelaan 1105

door

Omowonuola Abimbola Akingbuwa

geboren te Lagos, Nigeria

promotoren:	prof.dr. M. Bartels
	prof.dr. C.M. Middeldorp
copromotor:	dr. A.R. Hammerschlag
promotiecommissie:	prof.dr. J.C.N. de Geus
	prof.dr. T. van Amelsvoort
	prof.dr. P. Hoekstra
	dr. A. Ganna
	dr. Z. Ayorech
	dr. J.C. Polderman

paranymphs: Eshim Shahid Jami
Sofieke Kevenaar

Acknowledgements

This work was supported by the European Union's Horizon 2020 research and innovation programme, Marie Skłodowska Curie Actions Innovative Training Networks (MSCA-ITN-2016) under grant agreement No. 721567.

Contents

Chapter 1

General Introduction

Background

Psychiatric disorders are burdensome for a significant proportion of children and adolescents worldwide, affecting not just them, but also their families and society at large. Prevalence rates vary across different disorders like attention deficit hyperactivity disorder (ADHD), anxiety and depressive disorders, autism spectrum disorders (ASD), conduct disorders among others, but overall rates of up to 13.4% have been reported [1, 2]. Globally, psychiatric disorders are the leading cause of disability in children and youth, accounting for a substantial proportion of all years lived with disability (YLDs) and disability-adjusted life years (DALYs), particularly in high-income countries [3, 4].

Course and comorbidity in childhood psychopathology

Research has shown that a substantial proportion of children and adolescents with psychopathology continue to suffer from psychiatric health problems in adulthood, and are at increased risk of psychopathology compared to adults without a history of psychiatric problems [5-9]. This can take the form of homotypic continuity, where the same disorder continues across time, or heterotypic continuity, where childhood disorders precede other psychopathology [10]. Longitudinal cohort-based studies have shown that the onset of adult disorders including major depressive disorder (MDD), schizophrenia (SCZ) and substance use disorders, are often preceded by childhood externalizing problems like attention-deficit/hyperactivity disorder (ADHD) and oppositional defiant disorder (ODD) [11-14] as well as internalizing problems like depression and anxiety [7, 14-16]. Further, the consequences of psychopathology are not restricted to long-term mental health, but appear to extend to socio-economic and physical health outcomes. Prospective studies have shown that children with attention problems, obsessive compulsive disorder, depression, bipolar and anxiety disorders, went on to have reduced educational attainment, increased BMI, and insomnia in later life [5, 6, 8, 17-22].

Comorbidity, the presence of two or more psychiatric disorders in the same individual, is also a common feature of psychopathology in childhood. A common example is observed in children with ADHD, who often have comorbid symptoms of ODD or conduct disorder (CD). Likewise, internalizing problems like anxiety and depression also frequently co-occur with externalizing disorders like ADHD and CD, as well as with each other [23]. Importantly, individuals with comorbid psychiatric

disorders can be at risk of poorer prognosis due to more severe symptoms, as well as longer duration of illness, and more functional disability [24]. Thus, knowledge of the aetiology of childhood psychopathology is crucial in order to understand their development and progression across the lifespan, and ultimately to provide targets for early intervention.

Childhood psychopathology investigated in this thesis encompasses a range of psychiatric disorders and traits that generally manifest across childhood and adolescence. Neurodevelopmental disorders like ADHD and ASD, which by definition have their onset during the developmental period (infancy, childhood, and adolescence), typically start during childhood, while others like depression can start anytime across the lifespan including during childhood [25]. While most disorders can be diagnosed from childhood onwards, it is rare for some, such as substance use disorders, bipolar disorders and schizophrenia, to be diagnosed before the age of 12.

Heritability and genetic architecture of childhood psychopathology

Twin-family and adoption studies have shown that psychiatric disorders and traits aggregate in families and that this aggregation is mainly due to genetic factors [26]. This also applies to childhood psychiatric traits. The proportion of phenotypic variance explained by variation in genetic factors varies by trait, with estimates around 40% for anxiety and symptoms of depression, and up to 70% for ADHD and ASD [26]. The knowledge that psychiatric disorders are heritable has led to new questions, including which specific genetic variants are associated with these traits, and whether such genetic factors can also explain comorbidity/co-occurrence of different traits, as well as stability across development. These are important questions to investigate as variation in genetic variants may eventually give rise to phenotypic differences between individuals.

One such variant type are single nucleotide polymorphisms (SNPs) which are the result of single base changes in the DNA at specific genomic locations. Genome-wide association studies (GWAS) allow us to evaluate associations between a large number of SNPs and complex traits such as a psychiatric disorders, and have been used to show that these disorders are polygenic i.e. they are influenced by many genetic variants [27]. Importantly, the SNPs investigated in GWAS are typically common, i.e. their minor allele is present in more than 1% of the population (minor

allele frequency MAF >1%). GWAS have enabled the identification of many genetic variants associated with psychiatric traits [27]. Moreover, GWAS results also facilitate other genetic methodologies that allow further investigation of the aetiology of complex traits. These include methods for estimating the proportion of phenotypic variance that is explained by measured variants/SNPs (SNP heritability), detecting and quantifying to what extent there is genetic overlap across both psychiatric and related phenotypes (genetic correlation), as well as testing causal relationships between different traits [27-30]. These methods can be used to assess and identify the contributions of genetic variants to the stability of psychopathology across the life span. They have been relatively successful for adult psychiatric traits such as schizophrenia and major depression, as well as anthropometric traits like height, where hundreds of trait-associated loci have been identified, and numerous genetic associations with other traits have been observed [31-33].

One major finding from GWAS is that effect sizes for common variants associated with psychopathology are small, and the amount of variance explained by SNPs, i.e. SNP heritability, is substantially lower than what is estimated from twin studies, even in very large samples [27]. One reason for this is that methods for estimating SNP heritability are limited to effects of genetic variants that are captured or measured on current DNA genotyping arrays used in GWAS. Commonly used SNP arrays contain up to 2,000,000 SNPs, which is substantially less than the 3.2 billion base pairs of the human genome. In general, the more SNPs are measured, the more variance can be explained [34]. This has been shown for traits like schizophrenia where along with increasing sample size, an increase in the number of variants tested has facilitated the identification of more associated variants [31, 35, 36]. Additionally, GWAS only study common variants while other types of genetic variants play a role in the aetiology of psychiatric disorders and traits as well, including rare variants (SNPs with MAF < 0.1%). Rare variants are not measured on genotyping arrays, rather whole genome sequence (WGS) or whole exome sequence (WES) data are needed to assess their contribution to psychiatric traits [37]. Overall, significant advances in statistical and molecular genetics methods have enabled attempts to answer these questions and facilitate further understanding of the aetiology of psychiatric disorders and traits.

This thesis

On the whole, the genetics of childhood psychopathology is understudied compared to adult traits, and many questions remain to be answered. The overarching aim of this thesis is to elucidate the role of genetic factors in the occurrence, course, and comorbidity of psychiatric symptoms across childhood and adolescence. The use of different methodologies makes it possible to pursue different lines of inquiry in order to achieve a well-rounded understanding of the mechanisms underlying psychiatric traits and psychopathology across development. As such, I employed different statistical genetic methods and approaches with the aim of investigating polygenic and environmental contributions to childhood psychopathology, as well as gaining insight into their underlying architecture.

When my PhD began in 2017, the first large GWAS of childhood psychopathology (> 55,000 individuals included) had just been published online [38]. Since then, more studies have been published, with similar aims of identifying genetic variants associated with childhood psychiatric traits and the roles they play in aetiology. Chapter 2 of this thesis focuses on a timely review of these studies, aimed at a clinical readership. Chapters 3 to 5 focus on the role of common genetic factors in the continuity of psychopathology and prediction of long-term outcomes. Chapter 6 compares the effect of common versus rare variants and their underlying biological mechanisms in schizophrenia. Finally, chapter 7 contains a discussion of findings from the thesis. Each chapter is described further in the chapters outlined below.

Chapter 2: Systematic review of molecular genetic studies of child and adolescent psychiatric disorders

In this chapter, we performed a systematic review of studies published from 2008 to 2020 that used statistical genetic methods to evaluate the contribution of common genetic variants to psychiatric disorders and traits. We were interested in studies that investigated childhood onset or childhood measured psychiatric traits with the aim of either 1) identifying common trait-associated genetic variants, 2) estimating SNP heritability, 3) investigating genetic overlap between psychiatric traits or 4) investigating the contribution of genetic factors to the stability of traits across development. Importantly, these studies must have employed one or more of four popular techniques including GWAS, polygenic scores (PGS), genetic relationship matrix restricted maximum likelihood (GREML), or linkage

disequilibrium score regression (LDSC). These methods are described in more detail in the chapter.

***Chapter 3:** Longitudinal analyses of genetic associations between childhood psychopathology and adult traits*

In this chapter, we investigated genetic associations between repeated measures of childhood psychopathology (ADHD symptoms, internalizing problems, and social problems), and polygenic scores (PGS) of adult traits including major depression, bipolar disorder, subjective wellbeing, neuroticism, insomnia, educational attainment, and BMI constructed in almost 43,000 children age 6 to 17 years. We performed further analyses to ascertain whether variables including age, type of childhood psychopathology, measurement instrument, or rater were moderators of the association between the childhood measures and adult trait PGS.

***Chapter 4:** Multivariate analyses of genetic associations between childhood psychopathology and adult traits*

We followed up the analyses from chapter 3 by using structural equation modelling to investigate the extent to which the genetic associations observed between childhood psychopathology and adult traits are explained by correlations between the adult trait PGS, as well as correlations between the childhood measures. While the analyses in chapter 3 may be described as multivariate due to the analyses of multiple traits, the reference to multivariate analyses in this chapter is related to the fact that we simultaneously modelled the associations between all phenotypes and account for the correlations between them.

***Chapter 5:** A genetically informed prediction model for aggression and intentional self-harm*

In this chapter, we combined genetic, environmental and psychosocial (risk) factors to produce a model for the prediction of intentional self-harm, aggression, or a combination of both by age 20. In order to determine which factors most predict self-harm and aggressive behaviours, we combined genetic predictors for aggression, ADHD, ASD, and other psychiatric and anthropometric traits, with environmental and behavioural risk factors such as behavioural problem scores, social difficulty, substance abuse and measures of family dynamics.

***Chapter 6:** Ultra-rare, rare, and common genetic variants implicate negative selection and neuronal processes in the aetiology of schizophrenia*

In this chapter, we investigated the extent to which common variant and rare variant enrichment analyses converge to similar results for schizophrenia. Specifically, we assessed the effects of schizophrenia-associated common and (ultra-)rare protein-truncating variants (PTVs) in multiple gene sets defined for various brain cell types and synaptic functions, as well as a gene set of PTV-intolerant (PI) genes. Finally, we assessed the extent to which gene sets implicated by (ultra-)rare variants were correlated with those implicated by common variants, as this may shed light on whether common and rare variants act on these disorders through the same or different pathways.

***Chapter 7:** Summary and general discussion of findings*

Here I bring together the findings from the previous chapters, focusing on their implications and how they may guide future research in child psychiatric genetics.

Chapter 2

Systematic review of molecular genetic studies of child and adolescent psychiatric disorders

Published as: **Akingbuwa, W. A.**, Hammerschlag, A. R., Bartels, M., & Middeldorp, C. M. (2021). Systematic Review: Molecular Studies of Common Genetic Variation in Child and Adolescent Psychiatric Disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*.

*Supplementary accessible at https://doi.org/10.1016/j.jaac.2021.03.020

ABSTRACT

Objective: We conducted a systematic review of studies using molecular genetics and statistical approaches to investigate the role of common genetic variation in the development, persistence, and comorbidity of childhood psychiatric traits.

Method: A literature review was performed on the Pubmed database, following the Preferred Reporting Items for Meta-Analyses (PRISMA) guidelines. There were 131 studies meeting inclusion criteria, having investigated at least one type of childhood-onset or childhood measured psychiatric disorder or trait with the aim of 1) identifying trait-associated common genetic variants, 2) estimating the contribution of single nucleotide polymorphisms (SNPs) to the amount of variance explained (SNP-heritability), 3) investigating genetic overlap between psychiatric traits, or 4) investigating whether stability in traits or the association with adult traits is explained by genetic factors.

Results: The first robustly associated genetic variants have started to be identified for childhood psychiatric traits. There were substantial contributions of common genetic variants to many traits, with variation in SNP-heritability estimates depending on age and raters. Moreover, genetic variants also appeared to explain comorbidity as well as stability across a range of psychiatric traits in childhood and across the lifespan.

Conclusion: Common genetic variation plays a substantial role in childhood psychiatric traits. Increased sample sizes will lead to increased power to identify genetic variants and to understand genetic architecture, which will ultimately be beneficial to targeted and prevention strategies. This can be achieved by harmonizing phenotype measurements as is already proposed by large international consortia and by including the collection of genetic material in every study.

Key words: child and adolescent psychiatry, genetic variation, child and adolescent genetics, systematic review, molecular genetics

INTRODUCTION

Over the past decade, the field of psychiatric genetics, including childhood psychiatry, has made remarkable progress with many new discoveries. This has been facilitated by rapid progress in molecular genetic methods. That psychiatric disorders are heritable is well established via twin studies [26], with estimates varying from around 70-80% for bipolar disorder and schizophrenia [39, 40], to 40-50% for anxiety and depression [41]. Estimates for childhood onset and childhood-measured phenotypes are equally high. Both attention-deficit/hyperactivity disorder (ADHD) and autism spectrum disorder (ASD) have heritability estimates ranging from 60-90% [42], while estimates for parent-reported childhood anxiety and depressive symptoms average around 40% [43, 44]. Further, twin studies have been used to show that stability and comorbidity among childhood and adolescent psychopathology traits are largely genetically mediated [30, 45].

The most widely applied method to investigate specific genetic variants contributing to heritability is genome-wide association (GWA) analysis, in which millions of common variants are tested for association with a complex trait [46, 47]. Initial GWA analyses showed that large sample sizes are required to identify the typically small, polygenic effects of trait-associated genetic variants [27, 48]. This gave rise to extensive collaborations which collated large amounts of data for genetic analyses in consortia within the field of genetics, like the EArly Genetics and Lifecourse Epidemiology (EAGLE) consortium [49], Psychiatric Genomics Consortium (PGC) [50], and The Social Science Genetics Association Consortium (SSGAC). As a result, genetic variants have been identified for psychiatric disorders including ADHD, ASD, schizophrenia, bipolar disorder, and major depressive disorder (MDD) [32, 38, 51-54]. Furthermore, there has been rapid development of polygenic techniques [55-57] that investigate the joint effect of genetic variants, in order to assess the genetic architecture of traits [27-30]. These studies have provided insight into the contribution of common genetic variation to heritability estimates, as well as the role of genetic factors in the persistence of symptoms over time and in frequently occurring comorbidity.

With increasing sample sizes for childhood phenotypes, it is timely to provide an overview of findings on the contribution of common genetic variants to child and adolescent psychiatric traits. We were specifically interested in studies

that investigate disorders/traits that typically have their onset in childhood e.g. neurodevelopmental disorders such as ASD and ADHD, as well as studies investigating traits that can be diagnosed across the lifespan, but are measured in childhood samples. To this end we systematically reviewed publications using genome-wide approaches to identify common genetic variants underlying vulnerability to these disorders, and studies using polygenic analyses aiming to increase our understanding of factors influencing comorbidity and continuity in psychiatric disorders.

Methods in studies focusing on common genetic variation

Before describing our search strategy and the results of the studies included in the review, we provide brief summaries of popular methods applied in molecular genetic studies focusing on common genetic variation. Extensive descriptions of these methods are provided in recently published reviews, e.g. [27-30].

Identification of common variants

Genome-wide association studies (GWAS)

GWA analyses test the associations between a trait and genetic markers across the genome, usually single base changes in the DNA sequence called single nucleotide polymorphisms (SNPs). If a SNP is significantly more common in cases or controls, this suggests that the SNP in question is associated with the trait and may play a role in its aetiology, conferring risk or protective effects [58, 59]. GWAS are not limited to dichotomous phenotypes, and quantitatively measured traits such as symptom counts can also be investigated using this method. GWAS typically use a significance threshold of 5×10^{-8}, based on an approximation of independent markers that are tested. This stringent threshold means that large samples are required to identify the typically small effects of genetic variants. In order to increase the statistical power to detect associated genetic variants for a given trait, multiple independent GWA analyses performed across distinct cohorts can be meta-analysed.

Estimation of SNP heritability

Linkage disequilibrium score regression (LDSC)

SNP-based heritability indicates what proportion of the variance of a trait is explained by measured SNPs, in contrast to broad heritability estimates based on twin-family studies. LDSC estimates SNP-based heritability using summary data from GWAS [56]. As a result of various evolutionary mechanisms, combinations

of alleles/SNPs can occur. Linkage disequilibrium (LD) occurs when SNPs are non-randomly correlated with other SNPs at different loci i.e. they are more or less frequently associated than would be expected at random [60]. LDSC uses a measure of LD, called LD score, that is estimated for each SNP by taking the sum of correlations between that SNP and all nearby SNPs, and is calculated in an ancestrally similar reference sample when individual genotype data is unavailable for the GWAS sample. The slope from a regression of LD scores on GWAS test statistics is proportional to the SNP-based heritability of the trait examined in the GWAS.

Genetic relationship matrix restricted maximum likelihood (GREML)
GREML, as implemented in Genome-wide complex trait analysis (GCTA) software, estimates the phenotypic variance explained by all measured SNPs simultaneously [57]. First, a genetic relatedness matrix is built for a sample of unrelated individuals, indicating the genetic similarities between all individuals based on their genome-wide genotypes. Next, using a linear mixed model that includes the genetic relatedness matrix, it is calculated to what extent the phenotypic similarity between unrelated individuals is due to their genetic similarity [57, 61].

Because, LD-structure differences between reference and sample data can bias LDSC estimates [28, 62], GREML is generally preferred where individual level is available. However, in the absence of individual level data, and at very large sample sizes, LDSC is more computationally efficient and provides a good alternative [28, 63].

Estimation of shared genetic variance explaining comorbidity or continuity of symptoms over time
GREML and LDSC
Both GREML and LDSC can be extended to bivariate analyses that allow the estimation of the genetic co-heritability, i.e., the amount of variance shared between two traits as a result of genetics, also known as their genetic correlation [64, 65]. These bivariate analyses can be performed both on non-overlapping samples, as well as on traits measured in the same individuals.

Polygenic risk scores (PRS)
To calculate PRS, a GWAS is conducted in a discovery sample to define risk alleles of SNPs and their effect sizes. Next, in an independent target sample, for each

individual, a polygenic risk score is calculated by totalling the number of risk alleles the individual has (based on the discovery GWAS) and weighting the score by the effect sizes of each allele [30]. The proportion of risk alleles included in a score are generally selected based on thresholds depending on the exact methodology used. The PRS represents an individual's genetic liability for a trait.

The method was initially developed to test the theory of polygenic inheritance when the first GWA studies lacked significant effects. PRS based on a GWAS without any or with few significant hits were used to predict the same trait in another sample, in this way showing that there was an effect captured in these common variants that likely would become significant when sample sizes became large enough [55]. Presently, they are generally used to assess genetic associations between different traits or the same trait measured at different ages. Typically, an outcome measure of interest from a target sample (e.g., depression) is regressed on their PRS for another trait (e.g., ADHD) to test the association between them. A significant result suggests that genetic variants common to both traits underlie their association. While LDSC and GREML require tens of thousands of subjects for both sets of traits being investigated, PRS work if the discovery sample is large but the target sample is small (~2,000 subjects at least). This is particularly advantageous when a target trait is rare or expensive to measure.

METHOD

Search strategy

The literature search was conducted on the PubMed database for studies published from 2008 up till the 9th of August 2020, as the most relevant/powerful molecular genetic studies are likely to have been published during the last decade. We included studies that investigated traits that have their onset in childhood, or that investigated traits that can be diagnosed across the lifespan, but were measured in childhood or adolescent samples. We followed the Preferred Reporting Items for Meta-Analyses (PRISMA) guidelines [66] (Figure 1). Search terms included psychiatry and psychopathology outcomes: *autis*, depress*, mood, emotion*, affective disorder*, internali*, anxi*, worry, fear*, obses*, compul*, OCD, panic, phobi*, inhibit*, shy*, withdrawn, behav*, attenti*, inattenti* externalising, externalizing, conduct disorder*, ADHD*, hyperactiv*, impuls*, disruptive*, problem*, aggress*, violen*,*

oppositional, ODD, psychiatr, and psychopatholog**. Additionally, each search included terms that were designed to produce studies using statistical methods to analyse molecular genetic data including: *GWAS, genome-wide*, association stud*, polygenic*, polygenic scores, risk scores, PRS, summary statistics, LD score regress*, LD score, GCTA, GREML, LDSR, LDSC*. Finally, we included terms designed to limit results to childhood and adolescent samples as well as include longitudinal genetic studies, using the following terms*: child*, adolescen*, teen*, youth, develop*, continuity*, stab*, change**.

Study inclusion criteria

The studies included in this review met the following criteria: published in English in a peer-reviewed journal; investigated at least one type of childhood/adolescent-onset or childhood/adolescent measured psychiatric disorder or trait; aimed to 1) identify common trait-/disorder-associated genetic variants, 2) estimate the contribution of SNPs to the amount of variance explained, 3) investigate whether associations between psychiatric traits are explained by genetic factors, or 4) investigate whether the stability in traits or the association with adult traits is explained by genetic factors. Finally, results published were based on analyses using one or more of the following methods: GWAS, LDSC, GREML, and PRS.

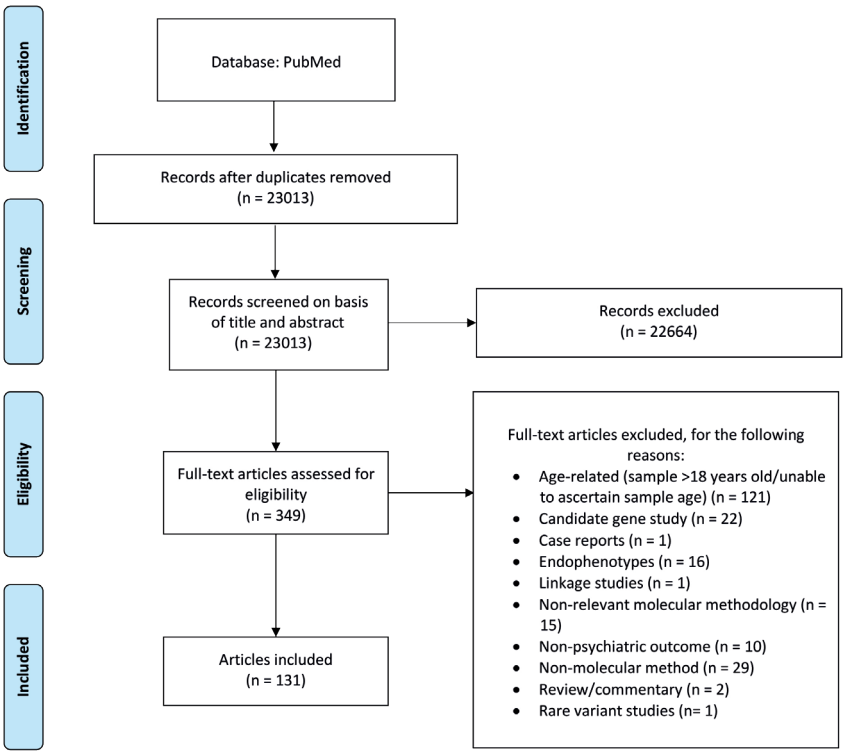


Figure 1. PRISMA Flowchart Showing Selection of Studies For Inclusion In Review

RESULTS

We identified 131 studies that addressed at least one aim of this review. See Table 1 for the proportion of them that used each method, address each trait and assess each relevant research question.

Identification of common variants

Of 50 relevant GWAS studies of childhood psychiatric traits, 15 reported significant genetic variants (Table S1, available online). The most commonly investigated phenotypes were ADHD and ASD, and their related continuous measures. Early studies were family-based, primarily made up of probands/cases and their unaffected parents and/or siblings, with more recent studies additionally including unrelated cases and controls.

Table 1. Characteristics of Studies Included in Review

Research question addressed	Method(s)	Number of studies
Variant identification	GWAS	50
SNP-heritability estimation	GREML, LDSC	34
Genetic contributions to comorbidity	GREML, LDSC	5
	PRS	18
Genetic contributions to stability or associations with adult traits	GREML, LDSC	16
	PRS	63

Note: The sum of studies in this table is greater than the total number of studies included in this review due to multiple studies addressing multiple aims. This table does not account for studies investigating childhood within-trait analyses as they do not strictly fit the aims of the review. GREML = genetic relationship matrix restricted maximum likelihood; GWAS = genome wide association study; LDSC = linkage disequilibrium score regression; PRS = polygenic risk score; SNP = single nucleotide polymorphisms.

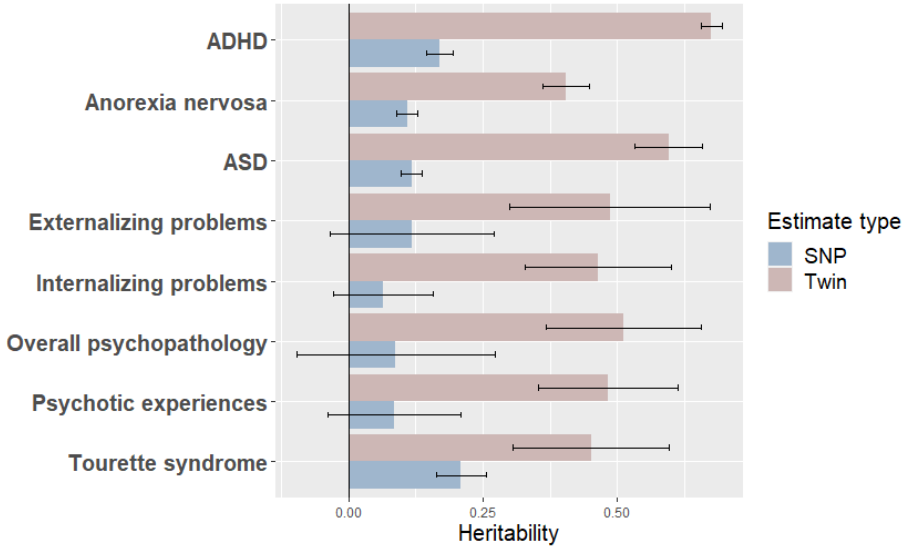
Significant genetic variants were detected for clinical measures of ADHD [38, 67-69], ASD [53, 70-72], anorexia nervosa [73], and Tourette syndrome [74], as well as continuous measures of ASD-related traits/symptoms such as social communication problems [75, 76] and restrictive and repetitive behaviours [77], for depressive symptoms [78], and for the anhedonia domain of self-reported psychotic-like experiences [79]. Only results from the recent ASD [53] and ADHD [38] studies were replicated in independent samples. The ASD GWAS identified three loci, while the ADHD GWAS identified 12 loci. The direction of effect for the top loci from both GWAS were replicated in five cohorts for ASD, and three for ADHD. Further, all 12 loci from the ADHD GWAS were significant in at least one of three replication meta-analyses. It is becoming common practice to perform functional annotation analyses of GWAS results in order to further clarify the biological basis for genetic associations. Such analyses have implicated dopamine regulation and brain development in the aetiology of ADHD and ASD respectively [38, 53]. It is important to note that the most recent findings from the case-control analyses were based on mixed adult and childhood samples, likely in a bid to maximise the power to detect significant variants.

SNP heritability

We identified 34 studies that estimated the SNP heritability of childhood psychopathology traits using GREML or LDSC (Table S2, available online). Analyses of clinically diagnosed traits generally used the same samples in subsequent analyses. We thus report estimates from the most recent studies, with estimates from individual studies described in Table S2, available online. SNP-based heritability estimates for

clinical disorders were based on mixed adult and childhood samples, and were 17%, 11.8%, and 21% for ADHD [69], ASD [53], and Tourette syndrome [74] respectively, while estimates ranged from 11% to 17% for anorexia nervosa depending on the assumed lifetime prevalence of the disorder [73] (Figure 2). Age-stratified analyses of childhood ADHD [69] and ASD [53] yielded heritability estimates of 19% and 4.9% respectively, while sex-stratified ADHD analyses found estimates to be significantly higher in male participants (24.7%) than female participants (12.3%) [68].

Figure 2. Comparison of Single Nucleotide Polymorphisms (SNP)- and Twin-based Heritability Estimates of Childhood Psychiatric Traits



Note: SNP-based estimates are those reported in the current review, while twin/family-based heritability estimates are of similar measures from other studies. Bars represent confidence intervals corresponding to α = 0.05, and are plotted for estimates for which they are provided. Twin-based heritability estimate sources (specific trait names from publications in brackets): ADHD (hyperkinetic disorders) [26], ASD (pervasive developmental disorders) [26], anorexia (eating disorders) [26], externalizing problems (conduct disorder) [26], internalizing problems (depression + anxiety + emotional disorder) [26], overall psychopathology (mental and behavioural disorders) [26], psychotic experiences [80], Tourette syndrome (Tic disorders) [26]. All twin-based estimates were obtained from Polderman et al 2015 using the MaTCH tool (http://match.ctglab.nl/#/home), apart from psychotic experiences, which was obtained from a recent publication that used similar measures and samples to the SNP-based estimates. Abbreviations: ADHD = attention-deficit/hyperactivity disorder; ASD = autism spectrum disorder; SNP = single nucleotide polymorphism.

Estimates for continuous measures were mostly low and non-significant, likely due to a lack of power from individual studies. They also showed more variation across age, rater and methodology, compared to clinically measured traits. We grouped results from relevant studies according to domains and meta-analysed estimates from different traits across these domains, combining estimates across age, rater and methods. We identified domains for ADHD symptoms, ASD symptoms, externalizing problems, internalizing problems, psychotic experiences, and general/overall psychopathology. Meta-analyses results, as well as studies and traits included per domain are described in Figures S1-S6, available online. Heritability estimates from these meta-analyses ranged from 6.48% to 14.5%, and are lower than relevant twin-based heritability estimates (Figure 2).

Aside from directly measured traits, the heritability of latent psychopathology factors were also investigated, with estimates of 38% for a general psychopathology factor, capturing the correlations between parent, teacher, and self-reported measures across multiple domains of internalizing and externalizing problems [81]. This indicates that it is possible to capture the genetic variation that is related to an individual's broad risk of psychopathology. An estimate of 14% was also reported for stable genetic factors affecting emotional problems across childhood and adolescence [82], suggesting that while genetic variants may have varying effects across development, it is also likely that a set of SNPs exist which have effects throughout development. These estimates were meta-analysed in the general psychopathology, and internalizing problems domains respectively.

In summary, estimates for continuous measures showed variation across different population-based samples, while estimates for clinical measures were more stable. Differences between samples in age, rater, and instrument used to measure the continuous outcomes might contribute to the higher variation, although we did not detect overall trends of differences between these variables.

Genetic overlap and stability of traits across time

Genetic factors explaining associations between childhood traits

PRS

Childhood cross-trait analyses were mostly limited to ADHD and ASD (Table S3, available online), with studies showing associations between PRS of ADHD and ASD, and childhood conduct disorder symptoms, irritability, ADHD symptoms, social

communication problems/autistic traits, eating disorder symptoms, anxiety and depression, as well as higher symptom levels in latent externalizing, internalizing, and general psychopathology factors [67, 83-96] (Table 2). Further, female participants with clinical diagnoses of anxiety and depression were found to have higher ADHD PRS than male participants [84], while male, but not female participants, with higher ADHD PRS had higher autistic trait scores [88].

GREML/LDSC

As with the PRS analyses, childhood cross-trait analyses using GREML/LDSC generally focused on ADHD and ASD, with reported genetic correlations of up to 0.37 based on clinical samples [53, 97, 98] (Figure 3, Table S4, available online).

Findings from childhood within-trait analyses do not strictly fit our aims but are described in Table S5, available online. Overall, they show associations between clinical measures of ADHD and ASD in one sample and clinical or continuous measures of the same trait in a different sample, suggesting the same underlying construct. Sex-stratified analyses of ADHD also reported a genetic correlation of almost 1 between male and female participants, suggesting that the same genetic variants underlie ADHD in both sexes [68], while PRS analyses showed higher ADHD PRS in female ADHD cases than male cases in some studies [99, 100], but not others [84].

Genetic factors explaining stability in traits, or associations with adult traits

PRS

PRS analyses investigating the role of genetic factors in the continuity of symptoms across childhood were limited. Longitudinal analyses of aggression found that PRS of aggression were not associated with aggression measured at different ages across childhood [101] (Table S6, available online). However this may have been the result of a stringent threshold at which SNPs were included in the score, resulting in a lower number of SNPs included than usual. We further identified a subset of studies that used PRS to investigate developmental trajectories for childhood psychopathology (Table S7, available online). They showed that higher ADHD PRS was associated with a trajectory of persistent ADHD [102, 103], as well as high-persistent, and increasing trajectories for irritability [104]. In contrast, aggression PRS was not associated with any symptom-defined conduct disorder trajectories, although it moderated the effect of interventions on trajectory class membership [105].

More studies have focused on the association between adult and childhood traits, investigating similar or different symptoms across development, as well as disorder trajectories (Table 2, Table S6 and S7, available online). Associations were reported between PRS of adult traits including schizophrenia, MDD, OCD, anxiety and externalizing disorder, and clinical and continuous measures of the same/similar trait in childhood and adolescence [79, 82, 96, 106-117] or in those at high risk [118, 119], although this was not always the case [92, 107, 108, 120-126]. An exception is bipolar disorder, for which no significant associations with similar childhood traits, such as mania were identified [106]. Still, PRS analyses of bipolar disorder performed in relatives of individuals with bipolar disorder indicated that, as expected, children and siblings had higher PRS for bipolar disorder compared to control participants [127, 128].

Similarly, there were a myriad of cross-trait associations observed between PRS of adult psychiatric traits, and dissimilar childhood traits including ADHD, depression, anxiety, OCD, conduct disorder, ASD, internalizing and externalizing problems, irritability, psychotic-like experiences, binge eating as well as trajectories for increasing, early- and adolescent-onset emotional problems [53, 86, 92, 113, 122, 123, 126, 129-135], though significant associations were not observed in all studies for all pairings [67, 83, 89, 93-95, 102, 108, 115, 123, 132, 136-142] (Table 2, Table S6 and S7, available online). Significant findings were generally more common in analyses with schizophrenia PRS, compared to other adult traits including bipolar disorder, MDD, anxiety and OCD. Given genetic correlations between schizophrenia and bipolar disorder, findings that bipolar disorder PRS were not related to the measured childhood phenotypes while schizophrenia scores were, may be related to higher statistical power for schizophrenia GWAS compared to bipolar disorder. Schizophrenia was the first psychiatric disorder for which samples were large enough to obtain sufficient statistical power. Longitudinal analyses further showed associations between schizophrenia PRS and internalizing and externalizing problems at different ages from age 3 to 16 [92, 122, 134], with one report of an increase in the strength of association with increasing age [134]. Similar longitudinal analyses of depression PRS found that associations with childhood psychopathology were not moderated by age, rater, or the type of childhood psychopathology, suggesting the existence of stable genetic factors that affect multiple traits across the life span [115].

PRS of ADHD and ASD were differentially associated with adult traits including MDD, anxiety, adult ADHD, bipolar disorder schizophrenia, as well as high decreasing trajectory for externalizing behaviours [69, 133, 143, 144]. ADHD PRS was also

found to distinguish bipolar disorder cases with childhood ADHD from controls without bipolar disorder [145].

Table 2. Genetic Associations Between Childhood Psychiatric Traits and Other Psychiatric Traits Using Polygenic risk scores (PRS)

Discovery trait (PRS)	Target trait	Target sample size	Variance explained (%)	Study references
Childhood cross-trait				
ADHD	ASD	1238	0.80	93
	ASD symptoms	1921 – 6854	**0.40 – 3.00**	88, 91
		5653 – 6854	0.00 – 0.10	91, 100
	Eating disorder symptoms	5674 – 5680	**0.10 – 0.13**	87
		5668	0.00	87
	Externalizing problems	394 – 6854	**0.41 – 1.99**	91, 92, 94
		1902 – 7975	0.00 – 0.30	86, 91, 92
	Internalizing problems	6603 – 7975	**0.20 – 0.42**	86, 91
		1843 – 7975	0.00 – 0.27	86, 91, 92
	Irritability	560 – 5584	**0.40 – 1.70**	83
		4023 – 4960	0.01 – 0.20	83
	Neurodevelopmental problems	7975	**0.90**	86
		7975	0.10	86
	Overall psychopathology	6603 – 7975	**0.86 – 1.06**	86, 91
ASD	ADHD	433 – 1688	0.13 – 0.34	67, 93, 94
	ADHD symptoms	1134	**0.77**	89
		394	0.19	94
	Childhood onset schizophrenia	233	**6.48**	96
	Externalizing problems	1902 – 7975	0.00 – 0.13	86, 92
	Internalizing problems	1843 – 7975	0.00 – 0.05	86, 92
	Neurodevelopmental problems	7975	0.00	86
	Overall psychopathology	7975	0.00 – 0.10	86
Childhood-adult within				
Anxiety disorder	Internalizing problems	5703 – 12,220	**0.09 – 0.41**	82, 106
		3755	0.07	106
Bipolar disorder	Bipolar disorder symptoms	3808	0.023	106
Depression	Internalizing problems	7975 – 42,998	**0.17 – 0.30**	86, 115
		932 – 7975	0.10 – 0.60	86, 123
Depressive symptoms	Depressive symptoms	709	**1.50**	116
Externalizing disorder	Externalizing problems	246	**5.00**	114

Table 2: Continued

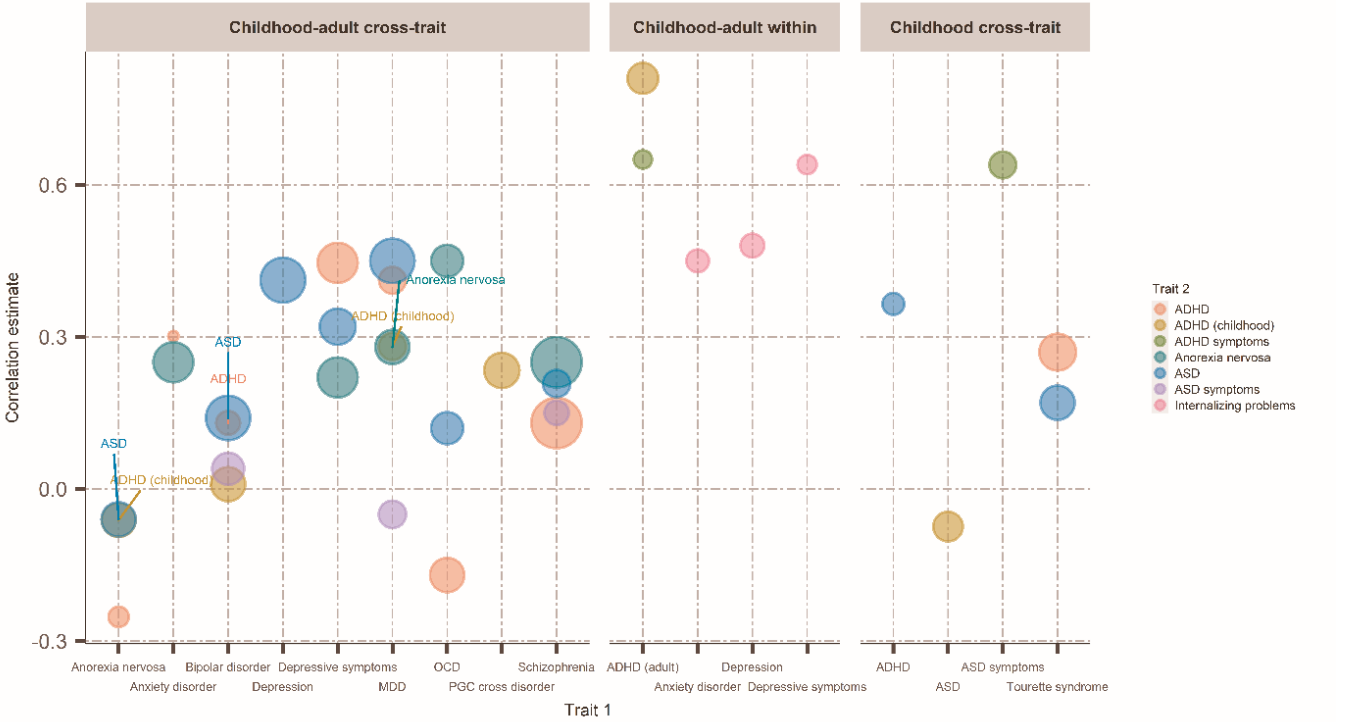
MDD	Depression	466	**5.00**	112
	Depressive symptoms	1450 – 6826	**0.20 – 0.73**	106, 112
	Internalizing problems	5703 – 12,220	**0.44 – 0.48**	82, 106
		1843 – 2202	0.004 – 0.02	92
OCD	OCD symptoms	650 – 3982	**0.23 – 2.28**	106, 117
		650 – 13,400	0.01 – 0.85	106, 117
Schizophrenia	Childhood onset schizophrenia	233	**18.52**	96
	Psychotic symptoms	2096 – 10,098	**0.08 – 0.70**	79, 106, 107, 109
		2133 – 8665	0.00 – 0.30	79, 107, 121
Tics/Tourette syndrome	Tics/Tourette symptoms	1043 – 13,396	**0.12 – 0.46**	106, 139
		4813	0.16	139
Childhood-adult cross-trait				
ADHD	Anxiety disorder	120,362	**0.06**	143
	Bipolar disorder	120,019	0.04	143
	MDD	126,605	**0.11**	143
	Schizophrenia	118,075	0.05	143
ADHD (childhood)	ADHD (adult)	22,406	**0.41**	69
Anorexia nervosa	ADHD symptoms	13,451 – 13,455	0.02 – 0.03	87
Anxiety disorder	ADHD symptoms	5154	0.02	138
ASD	Bipolar disorder	11,810	**0.08**	133
	MDD	16,610	**0.12**	133
	Schizophrenia	17,115	**0.04**	133
Bipolar disorder	ADHD	5422 – 6105	**0.18 – 0.88**	133, 141
		727 – 6102	0.11 – 0.99	67, 129, 141
	ADHD symptoms	1134 – 42,998	0.00 – 0.16	89, 115, 141
	ASD	10,263	**0.08**	133
	ASD symptoms	6128 – 42,998	0.002 – 0.2	115, 141
	Borderline personality disorder traits	5246	0.00	141
	Externalizing problems	1843 – 6133	0.02 – 4.00	92, 141
	Internalizing problems	1843 – 42,998	0.01 – 3.00	92, 115, 141
	Overall psychopathology	6111 – 42,998	0.00 – 0.003	115, 141
	Prosocial behaviour	6138	0.00	141
	Psychotic symptoms	8665	**0.12**	79
		2133 – 10,098	0.00 – 0.1	79, 121
Depression	ADHD symptoms	42,998	**0.25**	115
	ASD symptoms	42,998	**0.16**	115
	Externalizing problems	932 – 7975	0.00 – 0.40	86, 123
	Neurodevelopmental problems	7975	0.00	86
	Overall psychopathology	42,998	**0.17**	115
		7975	0.10 – 0.20	86
Depressive symptoms	Externalizing problems	709	**1.40**	116

Table 2: Continued

Externalizing disorder	ADHD symptoms	246	**7.00**	114
MDD	ADHD	1688	0.25	67
	ADHD symptoms	1134	0.19	89
	Externalizing problems	1843 – 2202	0.01 – 0.08	92
	Irritability	560 – 5584	0.01- 0.10	83
	Psychotic symptoms	6579 – 10,098 6297 – 8665	**0.08 – 0.11** 0.004 – 0.03	79 79
OCD	ADHD symptoms	5154	0.02	138
	Tics/Tourette syndrome	461	-1.20	142
OCD+TS	OCD	580	**1.70**	142
	Tics/Tourette syndrome	461	0.20	142
Polygenic p factor	Overall psychopathology	7026	**0.64 – 0.76**	146
Schizophrenia	ADHD	727 433 – 1688	**0.45** 0.08 – 0.58	129 67, 93, 94
	ADHD symptoms	394 – 2992	0.00 – 0.83	89, 94, 132
	ADHD/ASD	1631	0.30	93
	Anxiety disorder	4107	**0.50**	107
	ASD	10,263 1238	**0.09** 0.2	133 93
	ASD symptoms	3978 – 5137	0.10 – 0.43	147
	Externalizing problems	1154 – 2202 545 – 7975	**0.10 – 1.10** 0.00 – 0.15	92, 123, 132 86, 92, 123, 132
	Internalizing problems	1843 – 7975 932 – 7975	**0.20 – 0.77** 0.00 – 0.40	86, 92, 107 86, 123, 132
	Irritability	1358	**0.10** 0.00	132 132
	MDD	4106	0.005	107
	Neurodevelopmental problems	7975	0.00	86
	OCD	813	**3.17**	130
	Overall psychopathology	7975	**0.20 – 0.40**	86
Tics/Tourette syndrome	ADHD symptoms	6046	0.10	139
	ASD symptoms	6019	0.12	139
	OCD	580	0.04	142
	OCD symptoms	6006	0.11	139

Note: Variance explained from PRS analyses of childhood psychopathology traits, as well as study references for the estimates. Estimates are included for studies that reported them. Boldface type indicates estimates from significant association. Similar target traits were grouped by domain across different studies and ages (see Table S8, available online, for domain classifications). ADHD = attention-deficit/hyperactivity disorder; ASD = autism spectrum disorder; MDD = major depressive disorder; OCD = obsessive-compulsive disorder.

Figure 3. Genetic Correlations Between Childhood Psychiatric Disorders/Traits and Other Psychiatric Traits



Note: Correlation estimates are separated by analysis type. Included in this plot are associations between traits for which genetic correlation estimates are provided in their respective studies. Similar traits were grouped by domain (see Table S8, available online, for domain classifications). Trait pairs for which there were multiple estimates were meta-analysed to a single estimate. Point sizes correspond to the inverse of the estimate standard errors such that larger sizes represent estimates with smaller standard errors. Estimates with overlapping points are labelled for the sake of clarity. Relevant results from Barkhuizen et al [153] are not plotted as they were not reported with standard errors in the original study, while estimates from Anttila et al [98] are excluded due to negative standard errors being reported in the original study. ADHD = attention-deficit/hyperactivity disorder; ASD = autism spectrum disorder; MDD = major depressive disorder; OCD = obsessive-compulsive disorder; PGC = Psychiatric Genomics Consortium

GREML/LDSC

Social communication problems, peer problems and ADHD symptoms showed partly stable genetic effects across ages, with correlations between measures obtained from age 7 to 17 ranging from 0.1 to 1. Comparable estimates were also reported for cross-trait genetic correlations between traits. In all scenarios, correlations were highest at adjacent timepoints [76, 85, 148].

Studies also showed moderate to strong genetic correlations between ADHD, ASD, childhood emotional problems, anorexia nervosa, social communication problems, and symptoms of psychotic experiences including cognitive disorganization and anhedonia, and several adult psychiatric disorders [38, 53, 73, 82, 97, 98, 147, 149-153]. The largest correlations were observed with depression while correlations with bipolar disorder were lower (Figure 3, Table S9, available online).

In summary, although findings regarding genetic overlap were not always consistent, they provide evidence of pleiotropic effects in childhood psychopathology traits, i.e. the existence of a set of genetic variants influencing multiple traits. They also suggest the existence of genetic variants that influence psychopathology across development and across multiple psychiatric traits. Non-significant findings may point to a lack of power in either discovery GWAS, target samples, or both, rather than an absence of pleiotropy. It is also important to highlight that effect sizes for PRS associations were generally low with variance by PRS ranging from 0 to 18%. This is largely a function of the methodology and effect sizes are likely to increase with increasing GWAS sample sizes.

DISCUSSION

In this paper, we review findings from molecular and statistical genetic approaches explaining the contribution of common genetic variants to childhood psychiatric traits. We highlight recent GWA studies which have identified the first robustly associated genetic variants for childhood psychiatric traits. Further, we describe results based on genetic techniques including GREML and LDSC, which have enabled estimations of the heritability based on measured SNPs, and showed substantial contribution of common genetic variants to many childhood psychiatric traits. As well as PRS, these methods have been used to study genetic overlap

across traits and/or across time, which resulted in abundant genetic associations between multiple childhood psychiatric traits, as well as between childhood and adult traits, providing evidence for the presence of shared co-heritability.

On the whole, the identification of trait-associated variants appears to be associated with increasing sample sizes (Table S1, available online). Many GWA studies in this review that did not identify significant variants were likely underpowered. An increase in collaborative efforts and consortia-focused analyses have resulted in increasing sample sizes over the last few years, resulting in the identification of the first robust genetic risk variants for ASD and ADHD. This suggests similar outcomes for other childhood traits in the near future. It is important to highlight that large sample sizes for traits including ADHD and ASD were achieved by combining childhood and adult samples. This increases the power to detect trait-associated variants if these disorders are genetically similar/identical in childhood and later life. Studies have shown moderate to strong correlations between adult depression and anxiety, and childhood emotional problems [82], as well as between adult ADHD and childhood ADHD, suggesting similar underlying architecture [69]. Further, GWAS analyses of ADHD identified significant loci in combined analyses, but not in separate analyses of adult, and childhood ADHD. This was despite the fact that heritability estimates were slightly higher in the separate samples compared to the combined [69], further highlighting the importance of statistical power to detect effects. Nevertheless, there is considerable need for well-powered childhood-sample GWAS and/or age-stratified analyses as other traits may have different architecture across development. Other explanations for the lack of significant findings include heterogeneity and measurement error in phenotype definitions [154]. Heterogeneity may be introduced by the use of different raters and instruments to measure the same psychiatric traits. For example, varying degrees of concordance have been reported for measures of aggression depending on rater, and item content of available measures [155, 156]. Rather than combining different measurements in order to achieve large sample sizes, more stringent phenotyping to obtain more homogeneous phenotypes may contribute to the identification of associated variants and SNP-based heritability [154, 155]. Results from GWAS can also be informative in understanding the underlying genetic architecture and biological mechanisms of childhood psychiatric traits and the identification of genome-wide significant hits is an important first step this, as observed by the implication of dopamine regulation in ADHD [38]. Further, ADHD GWAS results have been utilized

to investigate potential genes and pathways that can be targeted by existing drugs [157]. This study implicated signal transduction and cell adhesion as potential treatment targets, and future studies for other childhood psychiatric disorders may provide potential novel avenues for treatment as well.

For all traits considered in this review, SNP-based heritability estimates from LDSC and GREML are substantially lower than estimates from twin studies. This is in part because both methods are limited to additive effects of causal variants tagged by the common SNPs on current DNA genotyping arrays used in GWAS. Analyses on BMI and height suggest that the difference between family- and SNP-based heritability estimates may be explained by rare variants [158]. It is likely that this also holds for other complex traits like childhood psychiatric traits. Indeed, increased burden of rare and *de novo* variants have been associated with disorders including ASD, ADHD and OCD [159-161], and children carrying specific pathogenic/disorder-associated CNVs have increased frequency of psychiatric disorders including ASD and ADHD, as well as anxiety disorders and oppositional defiant disorders [162, 163]. Sample sizes are still generally low for such analyses in childhood traits, but may increase in future as the cost of sequencing decreases, providing new opportunities to broaden our insight on the genetic architecture of these traits.

A major observation of the current review is a range of within- and cross-trait associations in childhood psychopathology using both PRS and GREML/LDSC. Childhood psychiatric traits were associated with other childhood traits, as well as adult disorders including MDD and schizophrenia. Further, modelling of genome-wide joint architecture of psychiatric disorders identified a factor comprised of childhood-onset disorders including ADHD, ASD and Tourette syndrome, as well as MDD [97]. There was also evidence of a general psychopathology (*p*) factor [146], which has been shown to explain a substantial amount of phenotypic and genetic variance across multiple childhood, and adult psychiatric disorders [164, 165]. Combined with the observed genetic overlap, these findings demonstrate a contribution of pleiotropic genetic effects to the development of psychopathology, and may suggest shared biological pathways. The findings of cross-trait associations may be informative for the validity of current diagnostic practices which define disorders in distinct categories based on the symptoms displayed [166]. Commonly occurring symptom overlap combined with substantial genetic overlap across disorders

suggests a spectrum of psychopathology, and thus the need for a re-evaluation of current diagnostic categories as they may not be accurate [167, 168].

Along with pleiotropic effects, recent studies have also provided evidence of specific genetic effects contributing to psychopathology. A recent study has shown differential genetic and phenotypic associations between ADHD and neurodevelopmental, versus externalising or internalizing disorders, after accounting for the *p* factor in a large sample [169]. Another study showed that the *p* factor explained considerable variance in childhood psychopathology measures, but inclusion of more specific emotional, behavioural and neurodevelopmental factors explained even more variance than just the *p* factor alone. The amount of (phenotypic) variance explained by the different factors differed depending on the childhood measure. For instance, for ADHD, most of the variance was explained by the *p* factor while variance in anxiety/mood problems were explained more by the emotional factor. The strength of associations between PRS of different psychiatric traits and these factors also varied depending on the PRS, and whether the associations were tested in univariate or multivariate model which included PRS of other traits. For example, depression PRS were not independently associated with the *p* factor but were associated with the emotional problems factor, while schizophrenia and ADHD PRS were associated with both the *p* factor and more specific neurodevelopmental and emotional problem factors [86, 91]. Combined with evidence of pleiotropy described above, these results suggest the existence of both general as well as specific genetic factors/variants which are involved in psychiatric aetiology. Future studies combining multivariate methodology with molecular data should focus on investigating and identifying both shared and specific genetic variants across childhood traits.

Some of the observed PRS associations are present at different ages across childhood and adolescence [92, 115, 122, 134]. This suggests the presence of genetic variants affecting psychopathology across the lifetime, not only explaining homotypic continuity of the same disorder across time, but also heterotypic continuity, where one disorder precedes/predicts another at a later time [10]. This may provide opportunities to identify children at risk for chronic course early, and provide targeted treatment. Although they currently explain too little variance to be clinically valid for individual risk prediction, there is potential for PRS to be combined with other risk factors to build a more complete picture of risk profiles and eventually improve disorder risk

prediction. Schizophrenia PRS have shown improved predictive value of psychosis in individuals at high risk [119]. Predictive accuracy will likely increase with increasing sample sizes of genetic studies, and the inclusion of PRS of correlated traits in multi-trait analyses has been shown to improve predictive power for ASD and will likely show similar results for other traits [53].

We conducted a further search on the bioRxiv and medRxiv servers for relevant studies that were not included with the main results, as they are yet to undergo peer-review. We identified 17 additional studies initially published on either server from March 2019 to September 2020. In these studies, there were no genetic variants detected for aggression [155], internalizing symptoms [170], obsessive compulsive traits [171], and total childhood problem scores [172], although 6 loci were identified in a cross-disorder meta-analysis of ADHD, ASD, OCD and Tourette syndrome [173], and 14 in a cross-disorder analyses of MDD and ADHD [174]. Reported SNP-heritabilities were similarly low to moderate, ranging from 2 to 21% for traits including ADHD, obsessive-compulsive traits, aggression, internalizing problems, and psychotic symptoms [155, 170-172, 174-177]. Again, there was evidence of both within- and cross-trait genetic associations [155, 171, 173-185].

Overall we identified numerous positive findings regarding the genetics of childhood psychopathology, particularly relating to cross-trait genetic overlap. However, we would be remiss not to address the important issue of publication bias in our reporting. Publication bias occurs when results from research influence whether or not it is published, such that published studies are skewed in favour of those with positive results. While not formally assessed in this review, we cannot rule out the possibility that our findings are affected by this phenomenon. In addition, while we did not filter on genetic ancestry, there was a clear Eurocentric focus on populations investigated, with a handful of studies also investigating East Asian populations. Future studies and data collection plans should include samples from more diverse ancestry. The accuracy of methods like genetic risk prediction is reduced when discovery and target samples are ancestry divergent [186, 187], which may preclude the use of genomic medicine in individuals of other ancestries.

The results of this review show that understanding of the genetic architecture of childhood psychiatric traits is increasing. Common trait-associated variants are starting to be identified, and studies show abundant genetic overlap between

multiple psychiatric traits. There remain many challenges to further increase our understanding of the genetic architecture of childhood psychiatric traits. Increasing sample sizes in diverse ancestries in order to identify more trait-associated variants is crucial and may be achieved in a variety of ways. For instance, by collecting genetic data in studies that do not have the identification of genetic variants as a primary aim, such as clinical trials and randomised controlled trials; this may also lead to better prediction of treatment outcomes. Moreover, the harmonization of data across studies and study types is crucial, in order to maximise power to detect effect [188, 189]. Continuation of large-scale collaborative consortia efforts to collect longitudinal data beyond what is currently available is also important. Findings from genetic studies have potential to impact disease prediction in children at risk, allowing for the possibility of earlier interventions which may enable them to have a more favourable course.

Chapter 3

Longitudinal analyses of genetic associations between childhood psychopathology and adult traits

Published as: **Akingbuwa, W. A.**, Hammerschlag, A. R., Jami, E. S., Allegrini, A. G., Karhunen, V., Sallis, H., Ask, H., Askeland, R. B., Baselmans, B., Diemer, E., Hagenbeek, F. A. Havdahl, A., Hottenga, J., Mbarek, H., Rivadeneira, F., Tesli, M., van Beijsterveldt, C., Breen, G., Lewis, C. M., Thapar, A., Boomsma, D. I., Kuja-Halkola, R., Reichborn-Kjennerud, T., Magnus, P., Rimfeld, K., Ystrom, E., Jarvelin, M., Lichtenstein, P., Lundstrom, S., Munafò, M. R., Plomin, R., Tiemeier, H., Nivard, M. G., Bartels, M., Middeldorp, C. M., and the Bipolar Disorder and Major Depressive Disorder Working Groups of the Psychiatric Genomics Consortium (2020). Genetic Associations Between Childhood Psychopathology and Adult Depression and Associated Traits in 42 998 Individuals: A Meta-Analysis. *JAMA psychiatry*, *77* (7), 715-728.

*Supplementary materials accessible at https://doi.org/10.1001/jamapsychiatry.2020.0527

ABSTRACT

Importance Adult mood disorders are often preceded by behavioral and emotional problems in childhood. It is yet unclear what explains the associations between childhood psychopathology and adult traits.

Objective To investigate whether genetic risk for adult mood disorders and associated traits is associated with childhood disorders.

Design, Setting, and Participants This meta-analysis examined data from 7 ongoing longitudinal birth and childhood cohorts from the UK, the Netherlands, Sweden, Norway, and Finland. Starting points of data collection ranged from July 1985 to April 2002. Participants were repeatedly assessed for childhood psychopathology from ages 6 to 17 years. Data analysis occurred from September 2017 to May 2019.

Exposures Individual polygenic scores (PGS) were constructed in children based on genome-wide association studies of adult major depression, bipolar disorder, subjective well-being, neuroticism, insomnia, educational attainment, and body mass index (BMI).

Main Outcomes and Measures Regression meta-analyses were used to test associations between PGS and attention-deficit/hyperactivity disorder (ADHD) symptoms and internalizing and social problems measured repeatedly across childhood and adolescence and whether these associations depended on childhood phenotype, age, and rater.

Results The sample included 42 998 participants aged 6 to 17 years. Male participants varied from 43.0% (1040 of 2417 participants) to 53.1% (2434 of 4583 participants) by age and across all cohorts. The PGS of adult major depression, neuroticism, BMI, and insomnia were positively associated with childhood psychopathology (β estimate range, 0.023-0.042 [95% CI, 0.017–0.049]), while associations with PGS of subjective well-being and educational attainment were negative (β, −0.026 to −0.046 [95% CI, −0.020 to −0.057]). There was no moderation of age, type of childhood phenotype, or rater with the associations. The exceptions were stronger associations between educational attainment PGS and ADHD

compared with internalizing problems (Δβ, 0.0561 [Δ95% CI, 0.0318-0.0804]; ΔSE, 0.0124) and social problems (Δβ, 0.0528 [Δ95% CI, 0.0282-0.0775]; ΔSE, 0.0126), and between BMI PGS and ADHD and social problems (Δβ, −0.0001 [Δ95% CI, −0.0102 to 0.0100]; ΔSE, 0.0052), compared with internalizing problems (Δβ, −0.0310 [Δ95% CI, −0.0456 to −0.0164]; ΔSE, 0.0074). Furthermore, the association between educational attainment PGS and ADHD increased with age (Δβ, −0.0032 [Δ 95% CI, −0.0048 to −0.0017]; ΔSE, 0.0008).

Conclusions and Relevance Results from this study suggest the existence of a set of genetic factors influencing a range of traits across the life span with stable associations present throughout childhood. Knowledge of underlying mechanisms may affect treatment and long-term outcomes of individuals with psychopathology.

INTRODUCTION

Longitudinal studies indicate that the onset of mood disorders in adulthood, including depression and bipolar disorder (BD), is often preceded by childhood problems. These include not only internalizing problems, such as depression and anxiety,[190, 191] but also externalizing traits, such as attention-deficit/hyperactivity disorder (ADHD) and aggression.[12, 192, 193] Moreover, both in prospective and retrospective studies, behavioral and emotional problems during childhood and adolescence have been associated with other adult outcomes that are associated with adult mood disorders, including educational attainment (EA),[5, 17, 194, 195] insomnia,[18, 196] subjective well-being (SWB),[197] personality,[198-201] and body mass index (BMI; calculated as weight in kilograms divided by height in meters squared).[202-204]

Both twin/family and molecular genetic studies have reported heritability[26, 98, 205] and stability[44, 82, 206] of psychopathology over time. Studies of BD in high-risk families also show that children of parents with BD are susceptible to psychiatric disorders and symptoms in childhood,[207] adolescence, and early adulthood.[208, 209] These results suggest that genetic factors may underlie the persistence of symptoms or the transition from one disorder to another between childhood and adulthood. Polygenic score (PGS) analyses enable the examination of the genetic association between adult traits and childhood symptoms of psychopathology.

Polygenic scores are aggregate scores of an individual's genetic risk for a trait, calculated by summing risk alleles from a discovery genome-wide association study (GWAS), weighted by their effect sizes.[30] For complex (ie, polygenic) traits influenced by many genetic variants, PGS summarize genetic risk across loci that are not individually significant in a GWAS. A statistically significant association between measured traits and PGS based on another trait suggests a shared genetic etiology. Results of studies using PGS to investigate the association of childhood psychopathology with mood disorders and associated traits vary. Analyses investigating depression and BD PGS have found no evidence of associations with emotional and behavior problems during childhood and adolescence, although there is evidence of association between depression PGS and emotional problems in adulthood.[92, 140, 210] Associations between PGS of EA and ADHD or attention problems have been more consistent, with multiple studies[92, 140, 211, 212] showing strong genetic associations between EA and ADHD or attention problems in childhood and adolescence.

The last 2 years have seen ever-larger GWAS for traits, including major depression (MD),[32, 51] BD,[52] EA,[213] and BMI,[33] consequently increasing accuracy of PGS.[214] Combined with the substantial increase in individuals genotyped in large longitudinal childhood cohorts that assess psychopathology, this provides an opportunity to rigorously investigate whether genetic factors underlie the associations between childhood psychopathology and adult mood disorders and associated nonpsychiatric traits (EA, insomnia, SWB, neuroticism, and BMI) and determine whether this association depends on age. Using 7 childhood population-based cohorts, we studied 42 998 individuals with repeated measures of ADHD symptoms, internalizing, and social problems. We performed meta-analyses to test whether PGS of adult traits are associated with childhood and adolescent psychopathology and whether this association depends on various factors, including age, type of psychopathology, type of scale used to measure psychopathology, and the informant.

METHODS

Participants and Measures

We obtained self-rated or maternal-rated measures of ADHD symptoms, internalizing, and social problems from 7 population-based cohorts (Table 1). Data collection was approved by each cohort's local institutional review or ethics board, waiving the need for informed consent for this study. The starting points of data collection varied, ranging from July 1985 to April 2002. Data analysis was performed from September 2017 to May 2019. Cohort descriptions can be found in the eAppendix 2 in the Supplement.

Genotyping and Polygenic Scores

Genotyping and quality control were performed by each cohort, following common standards (eAppendix 2 in the Supplement). In each cohort, PGS were constructed for the following adult traits: MD,[51] BD,[52] SWB,[215] neuroticism,[215] insomnia,[216] EA,[213] and BMI.[33] Height[33] was included as a control phenotype (eTable 1 in the Supplement contains the GWAS discovery sample size for each trait). To avoid overlap between discovery and target samples, summary statistics omitting the target cohort or cohorts were used. Analyses were limited to individuals of European ancestry.

Polygenic scores were estimated using LDpred, a method that takes into account the level of linkage disequilibrium between measured single-nucleotide variants (SNVs; often called single-nucleotide polymorphisms) to avoid inflation of effect sizes.[217] The method LDpred requires the inclusion of prior probabilities corresponding to the fraction of SNVs thought to be causal, which allows for testing varying proportions of SNVs associated with the outcome of interest. We thus tested a range of priors (0.75, 0.50, 0.30, 0.10, and 0.03) to assess the prior at which assessment was optimal. We restricted analyses to common variants, using SNV inclusion criteria of minor allele frequency greater than 5% and imputation quality of R^2 greater than 0.90.

Table 1. Sample characteristics

Cohort	Approximate age groups	Scale(s)	Phenotype(s) measured	Rater	Sample size
Avon Longitudinal Study of Parents and Children	7, 10, 12, 14, 16	Strength and Difficulties Questionnaire	ADHD symptoms, internalizing problems, social problems	Maternal	6502
Child and Adolescent Twin Study in Sweden	9, 12, 15	Autism-Tics, AD/HD and other comorbidities inventory, Screen for Child Anxiety Related Emotional Disorders, Short Mood and Feelings Questionnaire, Strength and Difficulties Questionnaire	ADHD symptoms, internalizing problems, social problems	Maternal, self	11039
Generation R	6, 10	Achenbach System of Empirically Based Assessment (Child Behaviour Checklist)	ADHD symptoms, internalizing problems, social problems	Maternal	2438
Norwegian Mother and Child Cohort Study	8	Screen for Child Anxiety Related Emotional Disorders, Short Mood and Feelings Questionnaire, Rating Scale for Disruptive Behaviour Disorders	ADHD symptoms, internalizing problems	Maternal	4583
Northern Finland Birth Cohort of 1986	16	Achenbach System of Empirically Based Assessment (Youth Self Report)	ADHD symptoms, internalizing problems, social problems	Self	3409
Netherlands Twin Register	7, 10, 12, 14, 17	Achenbach System of Empirically Based Assessment (Child Behaviour Checklist & Youth Self Report)	ADHD symptoms, internalizing problems, social problems	Maternal, self	5501
Twins Early Development Study	7, 8, 9, 12, 14, 16	SDQ, Conners' Parent Rating Scale	ADHD symptoms, internalizing problems, social problems	Maternal, self	9526

Abbreviation: ADHD, attention-deficit/hyperactivity disorder

Cohort-Specific Association Analyses

In each cohort, associations between childhood psychopathology and adult traits were estimated by regressing each outcome measure (ie, ADHD symptoms, internalizing, and social problems) stratified by age and rater, on the calculated PGS of the 8 adult traits at the 5 priors. A wide variety of surveys were used to further characterize the cohort.[218-224]

Where cohorts included related individuals, regressions were performed using the exchangeable model in generalized estimating equations to correct for relatedness in samples.[225] Scales were coded such that higher scores reflected more childhood problems. Both childhood psychopathology scores and PGS were standardized to a mean of 0 and an SD of 1, allowing for comparable βs across cohorts. Sex, age, batch effects, and genetic principal components (which correct for population stratification) were included as covariates in the regression (eAppendix 2 in the Supplement).

Multivariate Meta-analyses

Meta-analyses were performed using the metafor package in R version 3.6.0 (R Foundation for Statistical Computing).[226] To obtain the prior that provided the strongest estimate of the association with overall childhood psychopathology, we performed a random-effects meta-analysis for each of the 5 priors for each adult-trait PGS. Specifying random effects accounts for heterogeneity in the true associations attributable to factors that contribute to sample variation across cohorts, such as differences in measurements and sample characteristics. Subsequent analyses for each adult trait were conducted based on the selected prior from the previous analysis (ie, the one that provided the highest estimate of the association). As a sensitivity check, we repeated all analyses using a prior of 0.50 and compared these results to those using the prior with the highest estimate. We selected the prior of 0.50, because it represents a reasonable estimation of the proportion of associated SNVs across the different types of complex traits we tested.

To correct for dependency in the outcome variables attributable to repeated measures of the same individuals over time, we specified the variance-covariance matrix between their sampling errors. Because errors were assumed to be independent between cohorts, we combined variance-covariance matrices across cohorts by setting correlations between cohorts to 0 in the matrix, further

accounting for differences between cohorts.[134] To test whether the error covariance matrix alone suitably accounted for differences between cohorts, we applied for each adult trait an analysis of variance (ANOVA) test to compare models with the random effects dropped with those where they were specified along with the error covariance matrix.

Subsequent meta-analyses to test the association between each adult-trait PGS and overall childhood psychopathology (ie, all 3 childhood measures analyzed jointly) were performed on the reduced model (no random effects), if dropping them did not result in a significant loss of fit compared with the full model (random effects plus error covariance matrix). We also tested the association between the PGS and each individual childhood psychopathology measure.

Because both the childhood outcomes, and PGS measures are correlated, we estimated the effective number of tests between both sets of variables under the assumption that they are nonindependent.[227, 228] We corrected the meta-analysis results for multiple testing by applying Bonferroni correction ($P = .05$/number of tests) to the effective number of tests (2015.04 effective tests; $\alpha = 2.48 \times 10^{-5}$) (eTable 2 in the Supplement).

Multimodel Inference Analyses to Identify Moderators

To ascertain whether the variables age, type of childhood psychopathology (ie, ADHD symptoms, internalizing problems, or social problems), measurement instrument (eg, Strength and Difficulties Questionnaire,[218] Achenbach System of Empirically Based Assessment[222]), and rater (ie, maternal or self) moderated association between childhood psychopathology and adult-trait PGS, we performed multimodel inference analyses using the glmulti package in R version 3.6.0.[229] The glmulti package allows the definition of a function that takes into account all potential moderators and generates all possible models for the association of interest, returning the best models based on a specified information criterion; in our study, this was Akaike information criterion.[230] Furthermore, it provides parameter estimates based on all possible models, rather than a single-top model, while considering the relative importance of each potential moderator by weighting them. The averaged model avoids relying too strongly on a single best model.

In summary, for each adult-trait PGS, we selected the prior that provide the strongest estimate of its association with childhood psychopathology by performing random-effects meta-analyses at each prior. This was followed by ANOVA tests to determine whether our error covariance matrix suitably accounted for differences between cohorts. We then performed multivariate meta-analyses testing the associations of PGS of adult traits with childhood psychopathology at all ages. Finally, we performed multimodel inference analyses to ascertain whether moderators affected the association between each adult-trait PGS and childhood psychopathology.

RESULTS

The 7 included cohorts combined participants from the Netherlands, UK, Sweden, Norway, and Finland in a combined sample of 42 998 unique participants aged 6 to 17 years old. The percentage of male participants ranged from 43.0% (1040 of 2417 participants) to 53.1% (2434 of 4583 participants) by age and across all cohorts.

Cohort-Specific Association Analyses

Cohort-specific descriptive statistics and correlation matrices of the 3 psychopathology measures, ADHD symptoms, internalizing problems, and social problems are described in eTables 3, 4, 5, 6, 7, 8, and 9 in the Supplement. Correlation matrices show the observed variability or stability of childhood psychopathology over time. Based on cohorts with multiple or consistent measures of psychopathology across development, we observed moderate correlations across different ages. Estimates were highest for measurements of the same trait at adjacent ages, around 0.50, and lowest between self-rated and maternally rated measures, around 0.20. The results of the univariate analyses in each cohort are displayed in eTables 10, 11, 12, 13, 14, 15, and 16 in the Supplement.

Meta-analyses

Random-effects meta-analyses corresponding to the 5 priors showed that the prior that provided the strongest association estimates were 0.75 for EA and BMI; 0.50 for MD, insomnia, and height; 0.30 for neuroticism; 0.10 for BD; and 0.03 for SWB (eTable 17 in the Supplement). A reduced model (error matrix alone) was used in the multivariate and subsequent analyses for all traits except for the EA and BMI PGS, for which we used the full model (random effect plus the error covariance

matrix). This was because ANOVA tests comparing the full model with the reduced model suggested that the error covariance matrix alone insufficiently accounted for differences between cohorts (ANOVA results, eTable 18 in the Supplement).

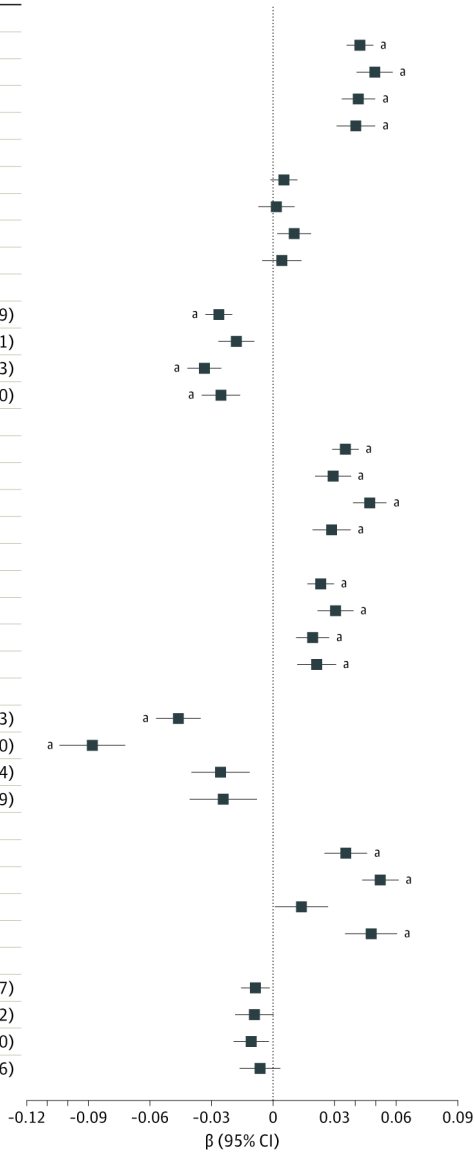
Subsequent meta-analyses of the association between PGS of each adult trait and overall childhood psychopathology (all 3 childhood measures in the same model) showed that the directions of associations were as expected (Figure 1). Significant positive associations were observed for PGS of MD (β, 0.042 [95% CI, 0.036-0.049]; SE, 0.003; $P = 2.48 \times 10^{-37}$; R^2, 0.002), neuroticism (β, 0.035 [95% CI, 0.029-0.042]; SE, 0.003; $P = 1.22 \times 10^{-26}$; R^2, 0.001), insomnia (β, 0.023 [95% CI, 0.017-0.030]; SE, 0.003; $P = 2.36 \times 10^{-12}$; R^2, 0.0005), and BMI (β, 0.035 [95% CI, 0.025-0.046]; SE, 0.005; $P = 2.23 \times 10^{-11}$; R^2, 0.001), while associations for SWB (β, −0.026 [95% CI, −0.020 to −0.033]; SE, 0.003; $P = 1.92 \times 10^{-15}$; R^2, 0.0006) and EA (β, −0.046 [95% CI, −0.035 to −0.057]; SE, 0.006; $P = 6.74 \times 10^{-17}$; R^2, 0.002) were negative. There was no evidence for association with BD PGS (β, 0.005 [95% CI, −0.001 to 0.012]; SE, 0.003; $P = .11$; R^2, 2.50×10^{-5}). No associations were found with the PGS of height.

Moderators

Using model averaging, we considered the effect of 4 moderators (ie, outcome, age, measurement instrument, and rater) across all possible models. Using a *P* value threshold of .0125 (α = .05/number of moderators), we found evidence of moderation for EA and BMI PGS (Table 2). The association between EA PGS and childhood psychopathology varied as a function of outcome, rater, and age. The EA PGS were associated with ADHD symptoms but not internalizing problems (Δβ, 0.0561 [Δ95% CI, 0.0318-0.0804]; ΔSE, 0.0124) or social problems (Δβ, 0.0528 [Δ95% CI, 0.0282-0.0775]; ΔSE, 0.0126); Figure 1). Additionally, the association between ADHD symptoms and EA PGS increased with age (Δβ, −0.0032 [Δ 95% CI, −0.0048 to −0.0017]; ΔSE, 0.0008) in maternal ratings, while self-ratings showed the opposite (Δβ, 0.0463 [Δ95% CI, 0.0315-0.0611]; ΔSE, 0.0075). However, the influence of rater on the associations appears to be driven by a single outlier aged around 17 years in the self-reported data (Figure 2). The association between BMI PGS and childhood psychopathology also varied across outcomes. Associations were strongest with ADHD and social problems (Δβ, −0.0001 [Δ95%CI, −0.0102 to 0.0100]; ΔSE, 0.0052), compared with internalizing problems (Δβ, −0.0310 [Δ95% CI, −0.0456 to −0.0164]; ΔSE, 0.0074). Moderators did not influence associations between the other adult-trait PGS and childhood psychopathology (eTable 19 in the Supplement).

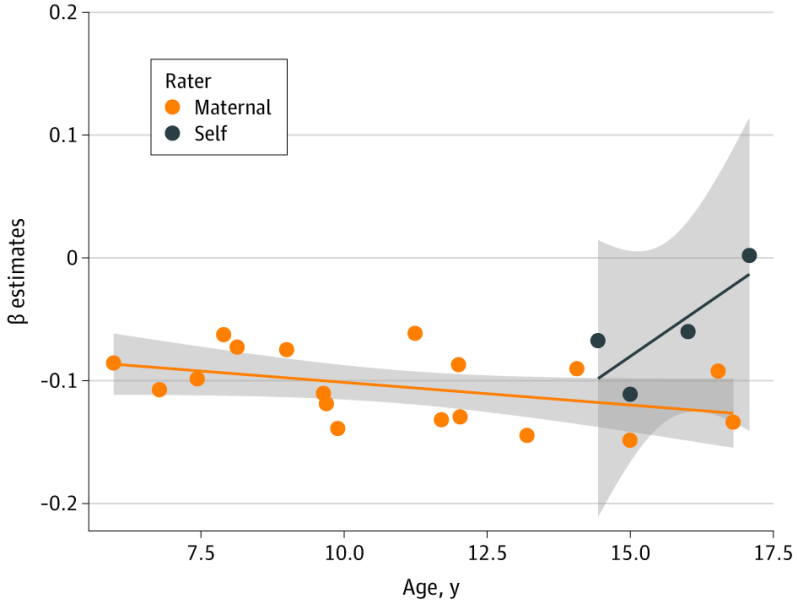
Figure 1. Multivariate Meta-analysis Estimates of the Associations Between Adult Traits and Overall Childhood Psychopathology

Source	β (95% CI)
Major depression	
Combined	0.0423 (0.0358 to 0.0488)
ADHD symptoms	0.0495 (0.0407 to 0.0583)
Internalizing problems	0.0416 (0.0334 to 0.0497)
Social problems	0.0403 (0.0309 to 0.0497)
Bipolar disorder	
Combined	0.0053 (-0.0012 to 0.0119)
ADHD symptoms	0.0016 (-0.0072 to 0.0104)
Internalizing problems	0.0103 (0.0021 to 0.0185)
Social problems	0.0043 (-0.0053 to 0.0138)
Subjective well-being	
Combined	-0.0264 (-0.0329 to -0.0199)
ADHD symptoms	-0.0179 (-0.0266 to -0.0091)
Internalizing problems	-0.0335 (-0.0417 to -0.0253)
Social problems	-0.0254 (-0.0348 to -0.0160)
Neuroticism	
Combined	0.0352 (0.0288 to 0.0417)
ADHD symptoms	0.0292 (0.0205 to 0.0379)
Internalizing problems	0.0471 (0.0390 to 0.0553)
Social problems	0.0285 (0.0192 to 0.0378)
Insomnia	
Combined	0.0232 (0.0167 to 0.0297)
ADHD symptoms	0.0304 (0.0216 to 0.0392)
Internalizing problems	0.0193 (0.0112 to 0.0273)
Social problems	0.0213 (0.0118 to 0.0307)
Educational attainment	
Combined	-0.0461 (-0.0569 to -0.0353)
ADHD symptoms	-0.0880 (-0.1040 to -0.0720)
Internalizing problems	-0.0256 (-0.0398 to -0.0114)
Social problems	-0.0243 (-0.0407 to -0.0079)
Body mass index	
Combined	0.0354 (0.0251 to 0.0458)
ADHD symptoms	0.0523 (0.0434 to 0.0612)
Internalizing problems	0.0138 (0.0009 to 0.0268)
Social problems	0.0478 (0.0352 to 0.0604)
Height	
Combined	-0.0086 (-0.0155 to -0.0017)
ADHD symptoms	-0.0091 (-0.0185 to -0.0002)
Internalizing problems	-0.0106 (-0.0192 to -0.0020)
Social problems	-0.0063 (-0.0163 to -0.0036)



Bars represent confidence intervals corresponding to α = .05. ADHD indicates attention-deficit/hyperactivity disorder. [a]Indicates significance after correction for multiple testing ($\alpha = 2.48 \times 10^{-5}$).

Figure 2. Moderator Effects of Age and Rater on the Association Between Educational Attainment Polygenic Scores and Attention-Deficit/Hyperactivity Disorder



Each point represents β estimates from univariate analyses of the association between educational attainment polygenic scores and attention-deficit/hyperactivity disorder symptoms at different ages. Overall, the negative association becomes stronger with increasing age (Table 2). The gray shadow around the trend line represents the 95% CI of the age effect size.

Sensitivity Analyses

Using a prior of 0.50 sensitivity analyses showed similar results to the main analyses, except for the moderation of outcome on the association with BMI PGS (intercept: β, 0.0439; SE, 0.0087 [95% CI, 0.0269-0.0609]; internalizing problems: Δβ, −0.0257; ΔSE, 0.0130 [Δ 95% CI, −0.0512 to −0.0003]; social problems: Δβ, −0.0018; ΔSE, 0.0055 [Δ 95% CI, −0.0126 to 0.0089]; eFigure in the Supplement). While this was nominally significant ($P = .047$), it did not remain after adjusting for the 4 moderators tested ($\alpha = .0125$; eTable 20 in the Supplement). Results from the main analyses also remained the same when all meta-analyses included random effects.

Table 2. Model-averaged moderator effects for Educational Attainment and Body Mass Index[a]

EDUCATIONAL ATTAINMENT	Estimate	SE	*z* Value	*P* value	ci.lb	ci.ub	Importance
Intercept	-0.0770	0.0092	-8.4072	**4.20x10^{-17}**[b]	-0.0950	-0.0591	1.0000
Rater – self	0.0463	0.0075	6.1370	**8.41x10^{-10}**[b]	0.0315	0.0611	1.0000
Age	-0.0032	0.0008	-4.0563	**4.99x10^{-5}**[b]	-0.0048	-0.0017	0.9896
Outcome – internalizing problems	0.0561	0.0124	4.5239	**6.07x10^{-6}**[b]	0.0318	0.0804	0.9606
Outcome – social problems	0.0528	0.0126	4.2076	**2.58x10^{-5}**[b]	0.0282	0.0775	0.9606
Scale – ATAC	0.0008	0.0016	0.4956	0.6202	-0.0023	0.0039	0.0194
Scale – Conners'	0.0008	0.0016	0.4898	0.6243	-0.0023	0.0039	0.0194
Scale – RS-DBD	0.0007	0.0015	0.4737	0.6357	-0.0022	0.0037	0.0194
Scale – SCARED	0.0001	0.0004	0.1861	0.8524	-0.0007	0.0008	0.0194
Scale – SDQ	-0.0002	0.0004	-0.4316	0.6660	-0.0010	0.0007	0.0194
Scale – SMFQ	-0.0008	0.0016	-0.4923	0.6225	-0.0038	0.0023	0.0194

BMI	Estimate	SE	*z* Value	*p* Value	ci.lb	ci.ub	Importance
Intercept	0.0468	0.0064	7.3531	**1.94x10^{-13}**[b]	0.0343	0.0593	1.0000
Outcome – internalizing problems	-0.0310	0.0074	-4.1744	**2.99x10^{-5}**[b]	-0.0456	-0.0164	0.9374
Outcome – social problems	-0.0001	0.0052	-0.0192	0.9847	-0.0102	0.0100	0.9374
Rater – self	-0.0011	0.0022	-0.5068	0.6123	-0.0055	0.0033	0.0923
Age	7.48x10^{-6}	2.32x10^{-5}	0.3223	0.7473	-3.80x10^{-5}	0.0001	0.0195
Scale – ATAC	-1.42x10^{-9}	3.35x10^{-9}	-0.4241	0.6715	-7.99x10^{-9}	5.14x10^{-9}	8.21x10^{-8}
Scale – Conners'	2.77x10^{-12}	1.62x10^{-9}	0.0017	0.9986	-3.18x10^{-9}	3.19x10^{-9}	8.21x10^{-8}
Scale – RS-DBD	-1.03x10^{-9}	3.12x10^{-9}	-0.3290	0.7422	-7.15x10^{-9}	5.09x10^{-9}	8.21x10^{-8}
Scale – SCARED	-3.32x10^{-9}	6.90x10^{-9}	-0.4809	0.6306	-1.68x10^{-8}	1.02x10^{-8}	8.21x10^{-8}
Scale – SDQ	-1.05x10^{-9}	2.47x10^{-9}	-0.4260	0.6701	-5.90x10^{-9}	3.80x10^{-9}	8.21x10^{-8}
Scale – SMFQ	2.69x10^{-10}	1.67x10^{-9}	0.1612	0.8720	-3.00x10^{-9}	3.54x10^{-9}	8.21x10^{-8}

Abbreviations: A-TAC, Autism-Tics, ADHD, and Other Comorbidities Inventory; BMI, body mass index (calculated as weight in kilograms divided by height in meters squared); RS-DBD, Rating Scale for Disruptive Behavior Disorders; SCARED, Screen for Child Anxiety Related Emotional Disorders; SDQ, Strength and Difficulties Questionnaire; SMFQ, Short Mood and Feelings Questionnaire, ci.lb, confidence interval lower bounds; ci.ub, confidence interval upper bounds

[a] The intercept estimate contains information from the reference variable of each moderator, selected in alphabetical order or with the lowest value, in the case of numerical moderators. Hence the intercept reflects the association estimate between educational attainment or BMI and Achenbach System of Empirically Based Assessment measured, maternally rated attention problems at approximately age 6 years. The other estimates show the change in association estimates depending on the moderator variable. The importance value for each moderator represents their overall support across all models. Moderators present in multiple models with large weights will have higher importance, and the closer this value is to 1, the more important the moderator is for the association being considered.

[b] Values were significant when adjusted for 4 moderators ($\alpha = .05/4 = .0125$).

DISCUSSION

We investigated genetic associations between childhood psychopathology and adult mood disorders and associated traits over time. Using results of well-powered GWAS meta-analyses of adult traits, we calculated PGS in what is, to our knowledge, the largest childhood target sample to date for this type of study (N = 42 998). We revealed strong evidence of associations of PGS for adult MD, SWB, neuroticism, insomnia, EA, and BMI with childhood ADHD symptoms, internalizing problems, and social problems. We found no evidence of associations between BD PGS and childhood psychopathology. In addition, we found no evidence of the moderators age, outcome, measurement instrument, and rater on these associations, except for EA PGS and BMI PGS. While EA PGS was more strongly associated with ADHD symptoms compared with the 2 other outcomes, BMI PGS was more strongly associated with ADHD symptoms and social problems than with internalizing problems. The association between EA PGS and ADHD symptoms increased with age and was stronger for maternal-rated ADHD symptoms compared with self-rated ADHD symptoms.

Our results indicate a consistent pattern of genetic associations between PGS of adult depression and associated traits and childhood psychopathology across age. This has not been observed previously, which is likely partly attributable to the increased power of our larger discovery and target samples compared with previous studies.[92, 210] Moreover, previous studies focused on separate childhood phenotypes[38, 231] as opposed to our approach of simultaneously analyzing multiple childhood problems at different ages. Consistent genetic associations across age suggest a set of genetic variants that influence a range of traits across the life span.

The exceptions to these consistent associations were EA and BMI PGS, which showed moderation on the associations by the different types of childhood outcome. While both were genetically associated with ADHD in accordance with previous research,[38, 140, 211, 212] they were not associated with internalizing problems, or social problems, in the case of EA. The lack of association with internalizing problems was somewhat unexpected, given genetic correlations previously found for BMI and EA with adult MD.[32, 51] These results suggest that genetic associations between EA and BMI and MD may become more apparent after adolescence, while they are already present for childhood ADHD and social problems (for BMI).

We did not identify associations between BD PGS and childhood psychopathology. This is intriguing because moderate genetic correlations with BD have been observed for MD and ADHD, as well as other behavioral-cognitive phenotypes, such as SWB and EA.[98] However, previous analyses of BD PGS also found no associations with continuous measures of psychopathology in childhood[92, 141] or adolescence.[106] These results may be explained by less powerful BD GWAS compared with MD and other traits, which might result in underpowered PGS. Nevertheless, the lack of association with BD PGS may also suggest that genetic risk for BD does not manifest until later in development, but given the higher prevalence rates of childhood psychopathology in offspring of parents with BD, this seems less likely.[209, 232, 233] It will be interesting to see if the observation holds as more powerful GWAS become available for BD.

Limitations

A limitation of our study is that analyses are limited to European ancestry, and therefore results are not generalizable to populations of differing ancestry. Second, associations between PGS and childhood psychopathology measures may be confounded by unaccounted passive gene-environment correlations, an association between a child's genotype and familial environment resulting from parents providing environments that are influenced by their own (parental) genotypes.[137, 234] Consequently, associations observed with adult PGS may be the result of both direct and indirect (environmentally-mediated) genetic effects. Third, dropout may have influenced our results. Previous analyses in longitudinal cohorts have reported negative associations between PGS for schizophrenia, ADHD, and depression and participation in childhood and adolescence.[235, 236] Nonparticipation in adolescence is also associated with higher psychopathology scores at earlier ages.[134] These results suggest that individuals with higher genetic risk for psychiatric disorders and higher childhood psychopathology are more likely to drop out of longitudinal studies. Genetic associations and the magnitude of associations reported may therefore be underestimated. Finally, because we combined data from different cohorts, we introduced heterogeneity in the assessment of childhood psychopathology. However, the meta-regression showed in general, consistent effect sizes across scales and raters. Moreover, combining multiple cohorts resulted in a large sample size, increasing statistical power compared with previous studies, which is a strength of this study.

Conclusions

The general lack of an influence of age and type of childhood psychopathology on our identified associations supports evidence of a common genetic psychopathology factor that remains stable across development.[237] Polygenic scores by themselves are not sufficient to identify individual children at high risk for persistence (they explain <1% of the variance in childhood psychopathology in this study). Nevertheless, these findings are of major importance because the individuals who are affected across the life span with consequences on other outcomes, such as EA and BMI, should be the focus of attention for targeted treatment. Furthermore, PGS could be combined with other risk factors for risk assessment in clinical samples, as was recently done for psychosis risk using schizophrenia PGS.[119] Future studies focusing on samples from high-risk populations are warranted to investigate whether PGS for adult traits, together with other variables, can be used to build risk profiles with reasonable accuracy. These may allow for the stratification of children into high-risk and low-risk groups for persistence, as well as test whether early intervention or more intense treatments for the former group can prevent poor outcomes.[238]

In conclusion, we demonstrate the power of combining genetic longitudinal population data to elucidate developmental patterns in psychopathology. Our study provides novel evidence for the presence of shared genetic factors between childhood psychopathology and depression and associated adult traits, as well as their stability across development. Insight into these associations may aid identification of children at risk for a relatively chronic course of illness, ultimately facilitating targeted treatment to this vulnerable group.

Chapter 4

Multivariate analyses of genetic associations between childhood psychopathology and adult traits

Submitted as: **Akingbuwa, W. A.**, Hammerschlag, A. R., Allegrini, A. G., Sallis, H., Kuja-Halkola, R., Rimfeld, K., Lichtenstein, P., Lundstrom, S., Munafò, M. R., Plomin, R., Nivard, M. G., Bartels, M., & Middeldorp, C. M. (2021). Multivariate analyses of molecular genetic associations between childhood psychopathology and adult mood disorders and related traits.

*Supplementary accessible at https://drive.google.com/drive/folders/1qnSiHRqP-QP4tkiEZS3OsA5zlOKNmMc25?usp=sharing

ABSTRACT

Ubiquitous associations have been detected between different types of childhood psychopathology and polygenic risk scores based on adult psychiatric disorders and related adult outcomes, indicating that genetic factors partly explain the association between childhood psychopathology and adult outcomes. However, these analyses in general do not take into account the correlations between the adult trait polygenic risk scores or between the childhood psychiatric traits. This study aimed to further clarify the influence of genetic factors on continuity of symptoms by accounting for correlations within adult and within childhood traits. Using a multivariate multivariable regression, we analysed associations of childhood attention-deficit/hyperactivity disorder (ADHD), internalizing, and social problems, with polygenic scores (PGS) of adult traits including major depression, bipolar disorder, subjective well-being, neuroticism, insomnia, educational attainment, and body mass index (BMI), derived for 20,539 children aged 8.5 to 10.5 years. After correcting for correlations between traits, major depression PGS were associated with all three childhood traits, i.e., ADHD, internalizing, and social problems. In addition, BMI PGS were associated with ADHD symptoms and social problems, while neuroticism PGS were only associated with internalizing problems and educational attainment PGS were only associated with ADHD symptoms. PGS of bipolar disorder, subjective well-being, and insomnia were not associated with any childhood traits. Our findings suggest that associations between childhood psychopathology and adult traits like insomnia and subjective well-being may be primarily driven by genetic factors that influence adult major depression. Additionally, specific childhood phenotypes are genetically associated with educational attainment, BMI and neuroticism.

INTRODUCTION

Psychiatric disorders cause significant distress and impaired functioning. They are also highly comorbid, with extensive phenotypic and symptom overlap. Comorbidity and symptom overlap has been observed between a range of disorder types including mood disorders like depression and anxiety [239, 240], childhood-onset neurodevelopmental disorders like attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), and Tourette syndrome [241], as well as between ADHD and anxiety disorders and depression [242, 243]. Importantly, a substantial proportion of children and adolescents with psychopathology continue to have psychiatric disorders in adulthood, as well as poorer outcomes related to physical health and functional outcomes, including higher body mass index (BMI), and lower educational attainment among others [5, 6, 8, 244-246]. Thus psychopathology traits are correlated with each other, and are linked to increased risk for negative outcomes, both related to mental health and beyond.

Using both twin- and molecular-based analyses, studies have shown genetic influences on the stability and continuity of psychopathology traits including attention problems, anxiety, and depression over time. Indeed there is evidence of genetic influence both for homotypic continuity (when a disorder is predicted by itself at a later time point) and heterotypic continuity (when one disorder predicts another at a later time point, e.g., childhood anxiety is associated with schizophrenia later in life) [44, 45, 78, 115, 206, 247]. Many studies investigating such genetic associations between childhood psychopathology and adult traits have employed polygenic scores (PGS), which index an individual's genetic risk for a trait based on previously determined effect sizes for alleles associated with the trait [30]. They have been used to show that shared genetic overlap likely underlies associations between childhood psychopathology and adult mood disorders including depression and anxiety, as well as related traits like neuroticism, insomnia, and subjective well-being [115, 248]. Furthermore, PGS have also been used to demonstrate genetic overlap between childhood psychopathology and mood disorder-related functional outcomes, such as educational attainment, and BMI [92, 115, 212].

Crucially, these associations are typically analysed in univariate analyses. However, both the adult, and the childhood traits are phenotypically and genetically correlated [86, 98, 164, 249-252]. This raises the question of whether the ubiquitous genetic

associations observed are genuine or whether they are driven by unaccounted correlations between related traits. For instance, a previous study reported genetic associations between major depression PGS and childhood ADHD, internalizing, and social problems [115]. However, it is possible that the association with all three childhood traits is explained primarily by an association between major depression PGS and internalizing problems, with the associations between major depression PGS and ADHD symptoms and social problems being the result of correlations between the three traits rather than genuine genetic associations between ADHD and social problems, and adult major depression. Knowledge of how underlying correlations influence genetic associations may provide insight into trans-diagnostic continuity of psychopathology across the lifespan and can be of importance for building prediction models for outcomes of childhood psychopathology.

In the current study, we performed a preregistered (https://osf.io/7nkw8) multivariate analysis to investigate genetic associations between childhood ADHD symptoms, internalizing, and social problems, and adult depression and related traits. Specifically, we were interested in exploring how accounting for the correlations between the adult trait PGS and between the childhood measures affects previously observed univariate genetic associations between them. As previous analyses largely showed no age effects in associations between childhood psychopathology and PGS of adult traits, we focused the current analysis at the age at which we had the most combined data which was at age 9-10. We obtained maternal-rated data for 20,539 children across three cohorts.

METHODS

Participants and Measures

Maternal-rated measures of ADHD symptoms, internalizing, and social problems were obtained for children aged 9 – 10 years from four population-based cohorts including the Avon Longitudinal Study of Parents and Children (ALSPAC) [253-255], Child and Adolescent Twin Study in Sweden (CATSS) [256], Netherlands Twin Register (NTR) [257], and Twins Early Development Study (TEDS) [258] (Table 1). CATSS, NTR and TEDS are population based twin cohorts while ALSPAC is a population based birth cohort that recruited all pregnant women in the former county of Avon

with an expected due date between April 1991 and December 1992 Childhood psychopathology was measured in ALSPAC and TEDS using the hyperactivity-inattention, emotional symptoms, and peer relationship problems subscales of the Strength and Difficulties Questionnaire (SDQ) [218], while in the NTR, the attention, internalizing, and social problems subscales of the Child Behaviour Checklist (CBCL) [222] were used. In CATSS, the AD/HD module of the Autism-Tics, AD/HD and other comorbidities inventory [219], was used to measure ADHD symptoms. For internalizing problems, the Screen for Child Anxiety Related Emotional Disorders (SCARED) [220] was selected over the Short Mood and Feelings Questionnaire (SMFQ) [221]. This is because while they both had comparable psychometric properties, the SCARED measures symptoms over the past three months, which is more in line with the longer-term measures of the CBCL (two months) and SDQ (six months) used by other cohorts, compared to the SMFQ which measures symptoms over the past two weeks. The CATSS cohort did not have a measure of social problems at age 9-10.

Genotyping and quality control were performed by each cohort according to common standards and have been previously described [115]. We obtained PGS for traits including major depression [51], bipolar disorder [52], subjective well-being, neuroticism [215], insomnia [216], educational attainment [213] and BMI [33], calculated using LDpred [217]. LDpred allows the inclusion of prior probabilities which correspond to the assumed proportion of genetic variants thought to be causal for a given trait. We used PGS at the most predictive priors per trait, determined from previous univariate analyses [115]. Data collection was approved by each cohort's local institutional review or ethics board, waiving the need for informed consent for this study. Analyses were limited to individuals of European ancestry.

Statistical Analyses

The main model tested is described in Figure 1. The model represents a multivariate regression with three dependent and seven independent variables, as well as additional covariates. The dependent variables are the maternal-rated measures of ADHD symptoms, internalizing, and social problems, while the independent variables are PGS of major depression, bipolar disorder, subjective well-being, neuroticism, insomnia, educational attainment, and BMI. Multivariate multivariable regression analyses were performed in R using path specification in the OpenMx package [259-262]. Full information maximum likelihood (FIML) estimation

[263], optimized in OpenMx was used to account for missingness in the outcome (childhood measures) data. We also accounted for the effects of sex, age, genetic principal components (to correct for population stratification), genotyping chip, and batch effects on the childhood measures, by including them as covariates in the model (Table 1).

Table 1. Sample characteristics

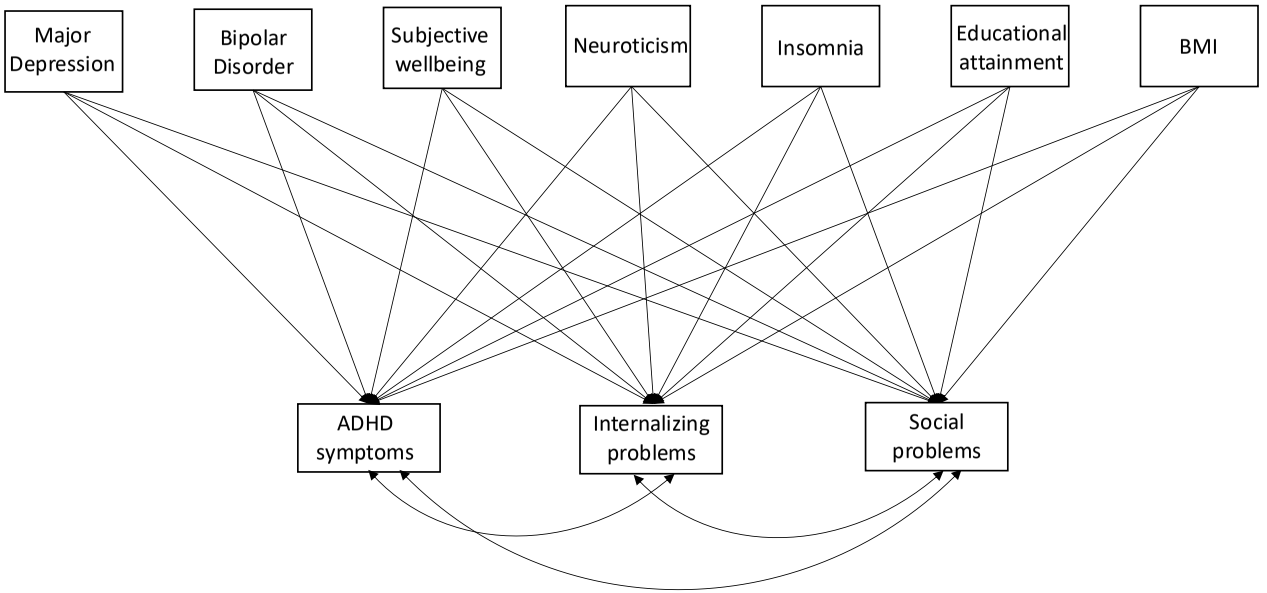
Cohort	Phenotype(s) measured	Scale(s)	Sample size	Covariates included in regression model
ALSPAC	ADHD symptoms, internalizing problems, social problems	SDQ	5025	10 genetic PCs, age, sex
CATSS	ADHD symptoms, internalizing problems	A-TAC, SCARED	7284	10 genetic PCs, sex
NTR	ADHD symptoms, internalizing problems, social problems	ASEBA-CBCL	3652	10 genetic PCs, genotyping chip, age, sex
TEDS	ADHD symptoms, internalizing problems, social problems	SDQ	4578	10 genetic PCs, genotyping chip, genotyping batch, age, sex

Abbreviations: ALSPAC, Avon Longitudinal Study of Parents and Children; CATSS, Child and Adolescent Twin Study in Sweden; NTR, Netherlands Twin Register; TEDS, Twins Early Development Study; ASEBA, Achenbach System of Empirically Based Assessment [222]; A-TAC, Autism-Tics, AD/HD and other comorbidities inventory [219]; CBCL, Child Behaviour Checklist [222]; SDQ, Strength and Difficulties Questionnaire [218]; SCARED, Screen for Child Anxiety Related Emotional Disorders [220]; PCs, principal components.

Both the childhood measures and the PGS were scaled so that they each had a mean of zero and standard deviation of 1. This allowed for data to be jointly analysed across cohorts using a multi-group model, which aggregates fit statistics from separate submodels specified for each cohort. Correlations and regression coefficients were constrained to be equal across cohorts, while estimates for the PCs, genotyping chip and batch effects, as well as their variances which were estimated separately per cohort. We corrected for relatedness in the twin samples (CATSS, NTR, TEDS) by estimating the cross-twin covariance for each outcome measure, as well as cross-twin cross-trait covariances.

We adjusted our significance threshold to account for multiple testing, using Bonferroni adjustment ($\alpha = 0.05$/number of tests), where the number of tests is the number of outcome measures multiplied by the number of predictors ($\alpha = 0.05/(3 \times 7)=0.00238$).

Figure 1. Multivariate model to be tested



***Note**: This figure includes only main predictor and outcome measures but does not include various covariates accounted for in the model. Note: BMI, body mass index; ADHD, attention deficit hyperactivity disorder.*

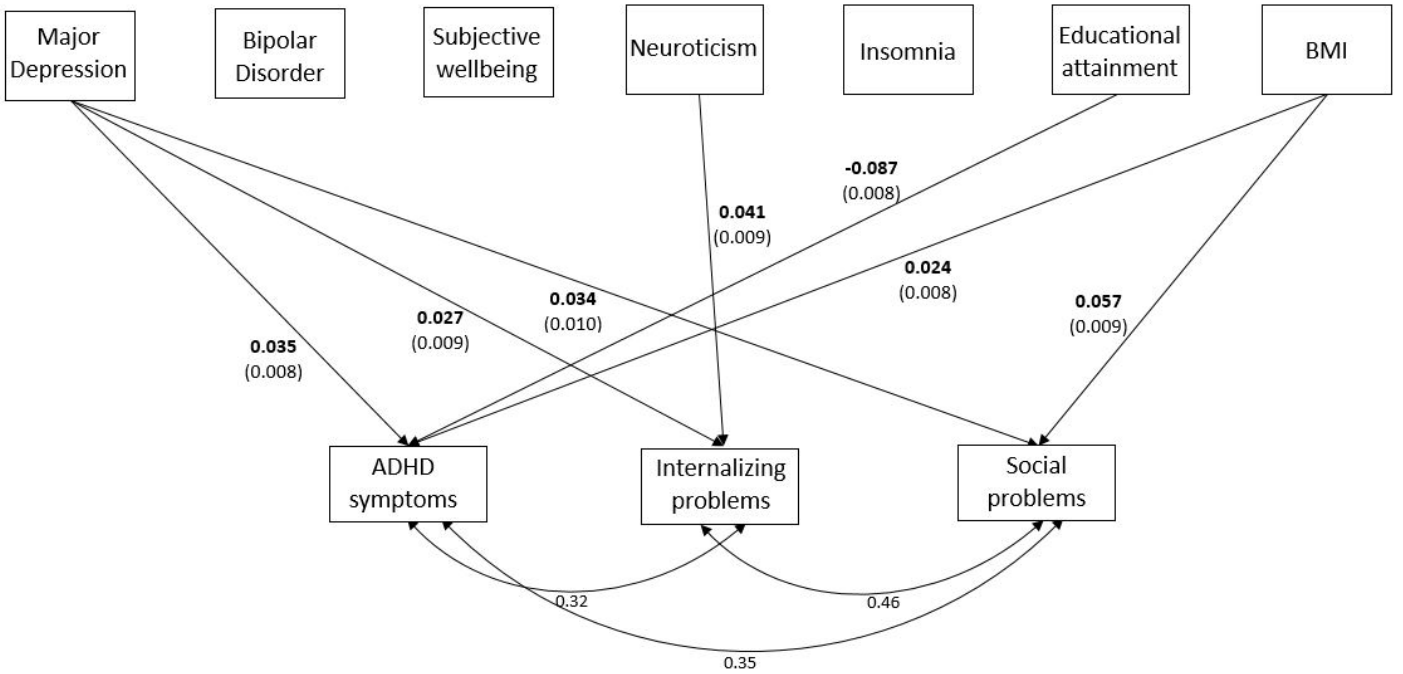
RESULTS

Across all cohorts, 20,539 children were included in the current analyses. Their ages ranged from 8.5 to 10.5 years. Full descriptive statistics per cohort for age and childhood measures, as well as sex-based information are provided in Supplementary Tables 1 and 2.

Associations between adult trait PGS and childhood traits

We fitted a multivariate multivariable regression model investigating associations between the three childhood outcome measures, and PGS at a prior of 0.75 for educational attainment and BMI, 0.5 for major depression, and insomnia, 0.3 for neuroticism, 0.1 for bipolar disorder, and 0.03 for subjective wellbeing. We observed moderate phenotypic correlations between the childhood measures; 0.32 between ADHD symptoms and internalizing problems, 0.35 between ADHD symptoms and social problems and 0.46 between internalizing and social problems (Figure 2). Negative correlations between the PGS ranged from -0.009 to -0.305 while positive correlations ranged from 0.011 to 0.306 (Table 2). The pattern of correlations between the adult trait PGS was similar to those seen in previous analyses, with high correlations between variables on the depression-wellbeing spectrum including neuroticism, and lower associations with other traits like BMI, educational attainment and bipolar disorder [98, 205, 215, 216]. Further, insomnia, subjective wellbeing, and neuroticism were also correlated with each other, although to a slightly lesser extent.

Figure 2. Results of multivariate modelling



Note: *Path coefficients from multivariate model showing significant associations between PGS of adult traits and childhood psychopathology measures. Standard errors of association estimates are in brackets. Note: BMI, body mass index; ADHD, attention deficit hyperactivity disorder.*

Table 2. Polygenic scores correlation matrix

	Major depression	Bipolar disorder	Subjective wellbeing	Neuroticism	Insomnia	Educational attainment	BMI
Major depression	1	0.184	-0.215	0.306	0.191	-0.125	0.05
Bipolar disorder	0.184	1	-0.03	0.068	0.014	0.068	-0.009
Subjective wellbeing	-0.215	-0.03	1	-0.305	-0.118	0.047	0.011
Neuroticism	0.306	0.068	-0.305	1	0.244	-0.152	-0.082
Insomnia	0.191	0.014	-0.118	0.244	1	-0.152	0.04
Educational attainment	-0.125	0.068	0.047	-0.152	-0.152	1	-0.201
BMI	0.05	-0.009	0.011	-0.082	0.04	-0.201	1

Note: matrix represents the average correlation between the scaled PGS of the adult traits across four cohorts

After correction for multiple testing (α = 0.00238), we observed significant positive associations between BMI PGS and ADHD symptoms (beta = 0.024, 95% C.I = 0.008 to 0.039, SE = 0.008, p = 0.002) and social problems (beta = 0.057, 95% C.I. = 0.039 to 0.076, SE = 0.009 , p = 1.37×10^{-09}), between major depression PGS and ADHD symptoms (beta = 0.035, 95% C.I. = 0.019 to 0.051, SE = 0.008, p = 2.23×10^{-05}), internalizing (beta = 0.027, 95% C.I. = 0.010 to 0.044, SE = 0.009, p = 0.002), and social problems (beta = 0.034, 95% C.I. = 0.014 to 0.053, SE = 0.010, p = 0.001), and finally between neuroticism and internalizing problems (beta = 0.041, 95% C.I. = 0.024 to 0.059, SE = 0.009, p = 4.97×10^{-06}). We also observed significant negative associations between educational attainment PGS and ADHD symptoms (beta = -0.087(95% C.I. = -0.071 to -0.102), SE = 0.008, p = 2.45×10^{-28}) (Figure 2). Other associations between childhood measures and PGS were not statistically significant (Table 3).

Table 3. Results from multivariate regression model

	ADHD symptoms				Internalizing problems				Social problems			
	β(SE)	*p* value	ci.lb	ci.ub	β(SE)	*p* value	ci.lb	ci.ub	β(SE)	*p* value	ci.lb	ci.ub
Major depression	**0.035(0.008)**	**$2.23x10^{-05}$**	**0.019**	**0.051**	**0.027(0.009)**	**0.002**	**0.010**	**0.044**	**0.034(0.010)**	**0.001**	**0.014**	**0.053**
Bipolar disorder	-0.002(0.008)	0.743	-0.018	0.013	-0.004(0.008)	0.626	-0.020	0.012	0.009(0.009)	0.330	-0.009	0.028
Subjective well-being	0.004(0.008)	0.639	-0.012	0.019	-0.002(0.009)	0.832	-0.019	0.015	-0.006(0.010)	0.561	-0.025	0.013
Neuroticism	0.004(0.008)	0.614	-0.012	0.021	**0.041(0.009)**	**$4.97x10^{-06}$**	**0.024**	**0.059**	0.029(0.010)	0.004	0.009	0.049
Insomnia	0.008(0.008)	0.334	-0.008	0.023	-0.004(0.008)	0.610	-0.021	0.012	0.002(0.010)	0.826	-0.017	0.021
Educational attainment	**-0.087(0.008)**	**$2.45x10^{-28}$**	**-0.102**	**-0.071**	-0.025(0.009)	0.003	-0.042	-0.009	-0.015(0.010)	0.129	-0.034	0.004
BMI	**0.024(0.008)**	**0.002**	**0.008**	**0.039**	0.011(0.008)	0.190	-0.005	0.027	**0.057(0.009)**	**$1.37x10^{-09}$**	**0.039**	**0.076**
Sex	**0.210(0.009)**	**$7.43x10^{-115}$**	**0.192**	**0.228**	**-0.081(0.010)**	**$1.99x10^{-16}$**	**-0.100**	**-0.061**	**0.074(0.011)**	**$3.40x10^{-11}$**	**0.052**	**0.096**
Age	-0.010(0.009)	0.294	-0.028	0.009	0.005(0.010)	0.584	-0.014	0.024	-0.003(0.010)	0.776	-0.022	0.017

Estimates for all model constrained variables. Note: β(SE), estimate of regression association and accompanying standard error from multivariate model; ci.lb, lower bound of 95% CI; ci.ub, upper bound of 95% CI. Bold estimates represent significant associations at Bonferroni-corrected threshold. Assessment of overall model fit suggested an acceptable to good fit based on RMSEA (0.047) but not CFI (-0.075) and TLI (-0.013).

DISCUSSION

So far, studies have primarily used univariate analyses to investigate genetic associations between childhood psychopathology and PGS of adult mood disorders and related traits like neuroticism, insomnia and subjective-well-being, as well as functional outcomes like educational attainment and BMI [115, 248]. In the current study, we performed a multivariate multivariable regression analysis with the aim of exploring how underlying correlations between these variables influences the strength/presence of previously observed associations. Using a multivariate model, we accounted for correlations between the PGS of adult traits as well as correlations between childhood ADHD symptoms, internalizing, and social problems. We found that major depression PGS were significantly associated with all three measures of childhood psychopathology. In addition, BMI PGS were positively associated with ADHD symptoms and social problems, and neuroticism PGS were positively associated with internalizing problems, while educational attainment PGS were negatively associated with ADHD symptoms. Previously reported associations of childhood psychopathology with PGS of insomnia, neuroticism, and subjective well-being were largely no longer present.

We observed differential genetic associations between childhood psychopathology and adult traits, with all childhood problems investigated associated with genetic risk for major depression. On the other hand, genetic risk for traits like neuroticism, educational attainment and BMI appeared to be related to specific childhood psychopathology measures. The non-specific association of childhood psychopathology with depression PGS suggests that there are genetic variants associated with depression and shared across the three childhood traits, which might be indicative of a dimensional structure of psychopathology where any type of childhood psychopathology is linked to genetic risk for depression. To some extent, we observed the same for PGS of BMI, which showed associations with social problems and ADHD symptoms i.e. there are genetic variants associated with BMI which are shared with both traits. However we did not observe this with PGS of educational attainment, and neuroticism, which were associated with only ADHD symptoms and internalizing problems respectively. This indicates that there are also specific genetic factors that are associated with educational attainment and ADHD symptoms, and with neuroticism and internalizing problems, which are

not shared with the other childhood traits. This is despite the fact that we observed modest correlations between the childhood traits (Figure 2).

These findings highlight the importance of both general and unique genetic factors to the understanding of psychiatric aetiology. Moreover, these results also suggest that many of the previously detected genetic associations between childhood traits and PGS of adult depression-related traits may be the result of their genetic correlations with depression. An exception was neuroticism PGS, which were still associated with internalizing symptoms. Additionally, we observed no associations between bipolar disorder and childhood psychopathology, despite the fact that bipolar disorder also shows moderate genetic correlations with major depression [98]. This may be due to a lack of power in the bipolar disorder discovery GWAS.

We showed that the use of multivariate methodology is important in furthering our understanding of genetic mechanisms underlying psychopathology across childhood and adulthood, but also associations between childhood psychopathology traits and functional outcomes in adulthood. Importantly, genetic risk for depression appeared to be linked to a myriad of childhood psychopathology traits, suggesting shared heritability across development. While this is perhaps expected for associations with internalizing problems, observed cross-disorder associations between major depression PGS and ADHD and social problems have implications for trans-diagnostic continuity across development. It contests the view of psychiatric traits or disorders as enduring discreet conditions, and raises clinically important questions as to the validity of distinct diagnostic boundaries. The observed substantial phenotypic correlations between the childhood traits may hint at symptom overlap, while non-specific associations with depression suggest shared genetic risk for them. Neither of these are strongly supportive of categorical classifications of psychopathology.

The observed associations may also be indicative of a causal association between childhood measures and depression in adulthood, which warrants future analyses of causality. The independent effect of neuroticism PGS on internalizing problems, on top of the effect of PGS for major depression is also interesting in this regard. It could be speculated that the measurement of internalizing problems in childhood is more reflective of a trait of emotional instability just like neuroticism, than of a depressive state like major depression. Furthermore, in conceptualizing causal factors underlying comorbidity between childhood psychopathology, negative emotionality (also

known as neuroticism) has been proposed to be a common feature underlying all childhood psychopathology [23, 264]. Interestingly we only observe associations between neuroticism PGS and internalizing problems. However the nature of PGS is such that the variance that they explain is very small. This means that it is likely/certain that associations observed do not reflect the total genetic overlap between neuroticism and childhood psychopathology. Replication of this result with PGS from larger GWAS are necessary.

Our findings regarding educational attainment and BMI replicate well established findings for genetic overlap between reduced educational attainment and ADHD symptoms in childhood [92, 211, 212], as well as for BMI and childhood psychopathology, particularly ADHD [98, 143]. Genetic analyses of causal mechanisms between ADHD and BMI have so far been inconclusive, with evidence of causality in both directions [265-267]. Analyses of causality between ADHD and educational attainment are fewer still, with one study showing evidence of bidirectional causal associations [268]. We add to the growing body of literature supporting associations between genetic risk for psychopathology, and health and sociodemographic outcomes in later life. The effect sizes reported were generally quite small which perhaps suggest that interpretations of our findings should be made cautiously. Nevertheless, more studies with a focus on causality are crucial, as knowledge of causal mechanisms may eventually inform clinical interventions, as well as risk for adverse effects of functional outcomes in the long-term.

Our study had some limitations. PGS analyses have been shown to include the effects of passive gene-environment correlation – an association between a child's genotype and familial environment as result of parents providing environments that are influenced by their own genotypes [137], which are unaccounted for in the present study and may have affected our findings. Secondly, while PGS involve aggregating the effects of many trait-associated variants, they are not informative about which specific genetic variants drive the observed associations and further fine-mapping and variant prioritization analyses are required to shed more light on this. Further, the small proportion of variance explained by the PGS means that they are currently unable to be used clinically. However, the aim of the current study was primarily to investigate the underlying genetic architecture. Finally, the case samples from the major depression GWAS used to construct the PGS in the current study were ascertained using minimal phenotyping. Minimal phenotyping involves leveraging information from sources

including hospital registers, self-reported symptoms, help seeking, or medication, in order to maximise statistical power to detect genetic variants. Major depression defined through minimal phenotyping has been shown to have different genetic architecture from strictly/clinically defined major depressive disorder (MDD), with genetic loci that are not specific to MDD [269]. Therefore, our findings regarding major depression may be a function of the non-specific nature of genetic factors associated with minimally phenotyped depression. However, major depression defined in this manner shows strong correlation with MDD, as well as good PGS-based prediction of MDD in independent samples [51, 269]. Nevertheless, similar analyses using clinical measures of MDD are important to further confirm our findings.

Results from this study show differential genetic associations between childhood psychopathology and adult depression and related traits, which may be suggestive of both shared and unique genetic factors underlying these associations. Future studies combining multivariate methodology with molecular data should focus on further unravelling these effects not just for psychopathology traits, but also associated functional and non-psychiatric outcomes such as educational attainment, and BMI.

Chapter 5

A genetically informed prediction model for aggression and intentional self-harm

In preparation as: Tate, A.*, **Akingbuwa, W. A.***, Hammerschlag, A. R., van Beijsterveldt, C., Pool, R., Lichtenstein, P., … & Bartels, M. A Genetically Informed Prediction Model for Self-Harm & Aggression in Teens.

*shared first author

**Supplementary materials accessible at https://drive.google.com/drive/folders/1qnSiHRqPQP4tkiEZS3OsA5zlOKNmMc25?usp=sharing

ABSTRACT

Self-harm and aggressive behaviours cause significant personal and societal burden. They are linked to a myriad of adverse outcomes in later life, making prediction of these behaviours an important endeavour. The current study aimed to create a model to predict self-harm and aggressive behaviours in late adolescence. Our data featured a training sample of 5,990 twins from the Child and Adolescent Twin Study in Sweden (CATSS) and an external validation sample of 1,975 individuals from the Netherland Twin Register (NTR). Using a combination of genetic, environmental, and psychosocial predictors derived from parental and self-report data we created a stacked ensemble model that contained a gradient boosted machine, random forest, elastic net, and a neural network, Model performance was assessed using area under the receiver operating characteristic curve (AUC). The neural network model was ultimately not included in the ensemble model. Model performance was uniform between the datasets (train set 0.981 [0.995 – 0.965]; tune set 0.683 [0.624 – 0.731]; test set 0.663 [0.616 – 0.721]; NTR data 0.729 [0.703 – 0.761]), suggesting generalizability in population-based samples across Northern Europe. Additionally, we evaluated variable importance of the predictors in the gradient boosted machines and random forest models which showed that aggression in mid-adolescence, as well as genetic risk for psychiatric traits indexed by polygenic scores were most important to the model.

Ultimately, our model would not be suitable for clinical use. However we improved on the performance of current prediction models that predict self-harm and aggression as well as show that genetic variables may have a role to play in predictive models of adolescent psychopathology.

INTRODUCTION

Aggressive behaviour and self-harm behaviours, with or without suicidal intent, cause significant disruptions on a personal and societal level [270-272]. Although both behaviors are often seen as separate constructs, there is evidence for an intrinsic link [273-275]. Many overlapping risk factors for these behaviours have been implicated across psychological symptoms, and home environmental factors [276]. Self-harm and aggression have been shown to be associated with internalizing symptoms such as depression, substance abuse, family dysfunction, neglect, abuse and maltreatment amidst a myriad of risk factors [276-280]. However, sex is a major distinguishing risk factor as females are more likely to report self-harm while men have higher instances of aggressive behaviour and criminal acts [281, 282]. Given their severity and enmeshment, it is of interest to create a model that can determine who is most likely to self-harm or exhibit aggressive behaviour.

Overlap between self-harm and aggression

Impulsivity is a common thread between the co-occurrence of aggressive behaviour and self-harm [283, 284]. For example, individuals with emotional dysregulation, i.e. an inability to regulate emotional intensity combined with impulsive and maladaptive behaviour to escape unwanted feelings, are at the greatest risk for intentional self-harm, and spontaneous aggression [285-289]. This is observed in psychiatric disorders such as attention-deficit hyperactivity disorder (ADHD) and borderline personality disorder [290-293]. Thus, certain traits may suggest a tendency towards impulsive-aggressive behaviour in a subset of individuals. In other words, the inability to control their behaviour in response to extreme, irritable emotions leaves both themselves and others at risk for victimization and harm. Additionally, there is evidence for genetic influence on self-harm and aggressive behaviours [294-296], though little is known regarding genetic overlap between them.

Predicting self-harm and aggressive behaviour

Many studies have looked into predicting suicide or aggressive behaviour. While many models exist for prediction of suicide behaviours in the clinical population, there are relatively fewer studies which examine self-harm in population-based samples [297, 298]. Additionally, while research in forensic psychology has worked to predict recidivism and violent criminal behaviour in general [299-301], specific prediction of aggression is less common. Moreover, only a handful of studies using a clinical population have

examined these interlinked behaviours together as an outcome [302, 303]. Therefore, it is of interest to create a prediction model that combines aggressive behaviours, and self-harm based on a large scale, epidemiological sample. A combination of genetic, environmental, and psychosocial factors obtained from epidemiological cohorts would theoretically allow for a highly generalizable, comprehensive model that could further inform future models for clinical prediction and decision-making.

Polygenic scores (PGS), which represent an aggregate score of an individual's genetic risk for a trait based on effect sizes from genome-wide association studies (GWAS), can be used to measure genetic overlap between traits [304]. As genetic associations have been reported between self-harm behaviours and psychiatric traits like anxiety, depression, schizophrenia and subjective well-being [294], as well as between aggression and psychiatric traits including ADHD autism spectrum disorder (ASD), well-being, and neuroticism [296], incorporating genetic risk factors like PGS may add information that improves prediction.

Thus, our goal is to create a binary model that can predict who will have intentionally self-harmed, and/or show high levels of aggression when individuals are 18 years old. While aggressive behaviour tends to be childhood- and adolescent-limited, there is a subset of individuals for who aggressive behaviour persists into adulthood. This trajectory is associated with poorer outcomes in adulthood [305]. Thus identifying those who remain aggressive at late adolescence or older may identify those who will remain aggressive throughout their lifetime [306]. Additionally, given the significant overlap between risk factors and instances of co-occurrence, combining aggression and self-harm as an outcome for a tool to identify high-risk individuals may have a knock-on effect for reducing the likelihood of either behaviour.

METHODS

Participants

A total of 9,433 participants from population-based twin cohorts, who completed self-report questionnaires about self-harm, suicidal ideation and aggression between ages 17 and 21 were included in this study. Our sample comprised of 6,669 participants from the Child and Adolescent Twin Study in Sweden (CATSS)

[307] and 2,764 participants from the Netherlands Twin Register (NTR) [308]. Further cohort descriptions are provided in the Supplementary text.

Measures

CATSS: Aggression and self-harm was measured using the Life History of Aggression Checklist [309]. A score of 15 out of a possible 40 was used as a cut-off for determining aggression cases [309]. Self-harm was determined using two binary questions, "Deliberately attempted to injure yourself physically when you were angry or despondent" and "Deliberately attempted to kill yourself when you were angry or despondent". Participants who endorsed either question were classified as self-harming.

We included 19 predictors collected at age 9 or 12 (first wave) and 15 (second wave) which included psychiatric symptoms, parent and child relationship characteristics, as well as substance use were derived from the Autism Tics ADHD and other Comorbidities (ATAC) [219], Statin Child Monitoring (SCM) [310], Strengths and Difficulties Questionnaire (SDQ) [311], Parent Child Relationship Inventory [312], the Reactive-Proactive Aggression Questionnaire respectively [313] and self-reported drug and alcohol use. These variables were continuous or naturally binary, e.g. sex or lifetime history of trying marijuana, thus no cut-offs were used for any of the predictors. A full list of variables used can be found in Supplementary Table 1.

NTR: Aggression and self-harm were measured using the Young Adult Self Report (YASR) and Adult Self Report (ASR) of the Achenbach System of Empirically Based Assessment (ASEBA) [314]. The aggression cut-off was derived using a T-score cut off of 64 from their respective aggressive behaviour subscales [314], which was equivalent to the 91.24% percentile of the NTR sample. Self-harm/suicidal ideation was measured using two related items from the internalizing problems subscales ("I deliberately try to hurt or kill myself") and ("I think about killing myself"), where items were rated 0 (Not at all/Never/Not true), 1 (Somewhat true/Sometimes true) or 2 (Very true/Often true) by participants. Participants were classed as having the self-harm outcome if they rated either item anything other than 0.

The predictors, collected at age 12 (first wave) and 16 (second wave), were derived from demographic information, as well as the Child Behaviour Checklist (CBCL) and

Youth Self Report (YSR) [222]. Similar to CATSS, variables were kept continuous when possible and no cut-offs were used to create the predictors.

Polygenic scores

Polygenic scores (PGS) were constructed using summary data from recent GWAS. A complete list of traits, and associated GWAS on which PGS are based can be found in Supplementary Table 2. Leave-out summary statistics excluding CATSS and/or NTR data samples were generated for any traits for which they were included in the discovery GWAS. Analyses were limited to individuals of European ancestry. Genotyping and quality control were performed in both samples and are described in the Supplementary text.

PGS were derived using LDpred, which accounts for the linkage disequilibrium between single nucleotide polymorphisms (SNPs) to avoid inflation of effect sizes [217]. LDpred requires the specification of prior probabilities which correspond to the fraction of SNPs from the discovery samples considered causal with the trait, and we created scores at a range of priors (0.01, 0.1, 0.3, 0.5, 1). Typically, the prior with the strongest association with an outcome is selected for subsequent prediction. However this can vary depending on the trait the PGS is based on and the outcome being investigated, and may also lead to type 1 errors if multiple testing is unaccounted for. In order to reduce the complexity inherent in having multiple PGS predictors and outcome variables, we performed principal component analysis (PCA) on all priors for each trait PGS, and included the first principal component (PCA-PGS) for each trait in our model according to Coombes, et al. [315]. PCA analysis is an unsupervised machine learning technique which reduces the dimensionality of datasets while maintaining as much variability as possible; the resulting principal components (PCs) represent a certain amount of variation within the dataset. The first PC can be interpreted to represent the most variation within the data [316]. This method has been shown to prevent overfitting each PGS to each outcome and removes the need to select a single prior across all PGS [315]. We subsequently created a genetic general psychopathology score [317] by performing PCA analysis on the PCA-PGS scores related to mental health.

Data Pre-processing

All analysis were performed in R. First, all predictor variables, except for our binary variable sex, were scaled in each cohort order to account for variations in

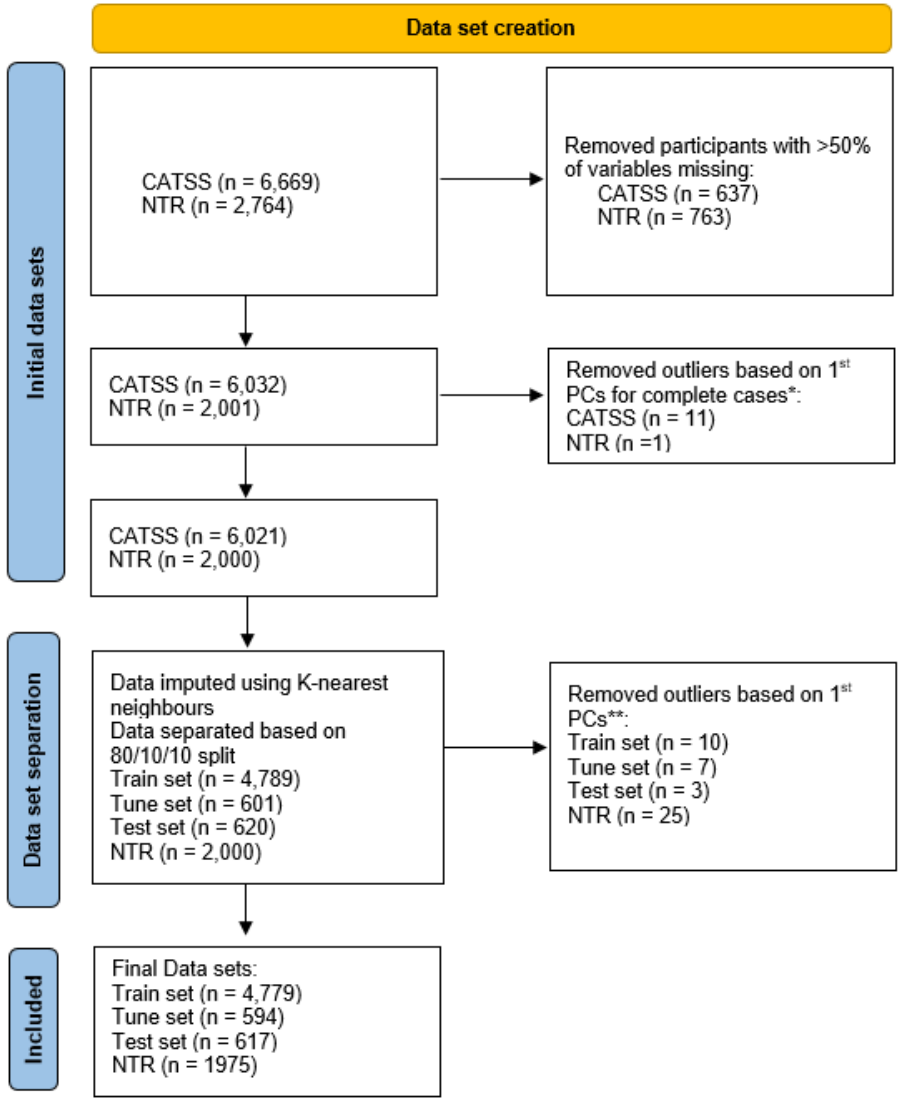
measurement tools. Participants with more than half of variables missing were removed from the analysis (Figure 1). This was done for two reasons: first, although a standard strategy is to maximize the number of data points this may not be the best approach when considering data quality, and a model can only be as good as the data included. Secondly, we used K-nearest neighbours for imputation, which requires a certain number of complete cases, as the full dataset did not adequately provide the correct ratio between participants with missing data and complete cases, removing those with more than half the data missing provided us with a suitable ratio.

Next, PCA analysis was completed on the entire CATSS dataset to determine the distances between those who were classified as aggressive and/or self-harming as well as to identify outliers [318]. Outliers were determined through the first principal component, participants with values that fell outside of the four quartiles were removed. This process was repeated separately in the NTR dataset.

Next, CATSS was broken down into a training set, a tuning set, and a test set based on an 80/10/10 split. As twins are more similar to each other compared to other participants, we stratified based on family number in order to prevent overfitting that could occur from twin pairs being separated between the data subsets. Additionally, we stratified on the outcome to ensure that the outcome proportion was balanced across the data subsets. Descriptive statistics were checked to determine the consistency of the split. The NTR dataset was used as an external validation set to determine the generalizability of the model across Northern Europe, and was thus not split in this manner.

Missing predictor data was imputed using K-nearest neighbours, a k of 6 was chosen by finding the square root of the number of columns in our dataset [319]. In order to avoid bias during this step, the test set as well as the external validation data were imputed separately. The outcome variable was not included as an informative imputation variable. As PCA analysis requires complete cases, we wanted to ensure that there were no outliers in those with missing data that did not have complete cases. Therefore, we completed another check for outliers in the imputed datasets to capture outliers. This was repeated separately in the NTR data and test set. This led to a total sample size of 7,965 participants (N training set = 4,779; N tuning set = 594; N test set = 617; N Dutch data = 1975) (Figure 1).

Figure 1. Flow chart of data set creation



* Principal component analysis (PCA) was completed separately for CATSS and NTR. Participants with a 1st principal component (PC) score outside the first four quartiles were used to determine outlier status.

**3 separate PCA analyses were completed: combined train and tune set, test set, and NTR.

Abbreviations: CATSS = Child and Adolescent Twin Study of Sweden; NTR = Netherlands Twin Register

Statistical Analysis

Main Analysis

R package H2O [320] was used for all supervised machine learning analyses. Model performance was determined by area under the receiver operating characteristic curve (AUC), with the threshold determined by the optimal F1 threshold. This is the threshold where sensitivity and specificity are balanced at their highest point. AUC is a popular measure of predictive accuracy with values ranging from 0.5 (random guess), to 1.0 (perfect prediction). A generally accepted heuristic for prediction using AUC is that an AUC > 0.9 suggests excellent model prediction, 0.8-0.9 is good, 0.7-0.8 is fair, and <0.7 is poor [321]. However, this rule of thumb is context specific and an AUC above 0.95 is desirable for medical use. During the model creation process, each model was trained using the training set and the tune set was used to evaluate performance at each iteration. Neither the test set nor the external validation set (the NTR dataset) were used during this process.

First, as our dataset had an imbalance between those with and without an outcome, we added weights to the training set to improve the positive predictive performance of the model [322]. A weight of 3 for those with an outcome and a weight of 1 for those without an outcome was determined by taking the number of the majority class over the minority class. This means that during the learning process each model resampled individuals in the training set with an outcome three times, while those without an outcome were not resampled.

We created a stacked ensemble model, i.e. a model which combines input predictions from separate models, that contained a gradient boosted machine (GBM), random forest (RF), elastic net, and a neural network (NN) [323]. These models were selected based on their availability in H2O. Each model was trained separately using cross validation (CV) with 3 folds in the training set until the AUC did not improve by 0.0001 for five rounds based on performance within the CV folds. The model was then put to the tuning test to determine if the model training process should continue. A combination of grid search and random search was used to tune the hyper-parameters of each model (Supplementary Tables 3 – 6). Scaled variable importance scores were obtained from our tree-based models GBM and RF, with the overall variable importance rankings determined using the average of scores across both models. The variable importance in models built with H2O can be interpreted as the improvement in the squared error when the feature is split

on a node [320]. As these scores were calculated during the learning process these values are solely based on the training set. These values were then scaled for ease of interpretation.

Once each of the 6 models reached satisfactory performance in the tune set, the models were combined into a stacked ensemble model, and tested on the test and the NTR data. The model was not modified after this step.

In order to account for population stratification, we also included the first five genetic PCs as predictors in our model.

Sensitivity analysis

To determine the stability of the variable importance for the model we trained a GBM and RF using the NTR data. For these analyses we used the same analysis process and range of parameters as the main analysis. However, no validation set was created so the data was split based on an 80/20 split stratified by family ID as well as the outcomes. The test set was reserved until a satisfactory performance was found for each model based on the performance of each cross-validation fold.

RESULTS

Descriptive Statistics

The data was well balanced between the training, tune, and test set (Table 1). The age of the sample at the measurement of the outcomes ranged from 17 to 21. The percentage of those with aggression or self-harm varied between the CATSS and NTR data sets. While the CATSS data had a much higher proportion of individuals who reported self-harm (CATSS = 12.94%; NTR = 2.38%), the NTR data had an increased proportion of individuals who were classified as aggressive (CATSS = 7.46%; NTR = 14.53%). The final proportion of individuals with aggression in the NTR does not correspond to the previously described T-score percentile because the T-score percentile was based on the total NTR sample while the final proportion is based on the sample after data cleaning. Overall, 25.19% of individuals in the CATSS data set were classed to have endorsed the outcome, compared to 18.78% in the NTR data set. PCA analysis did not show a clear distinction between individuals with and without the outcome (Supplementary Figure 1).

We further investigated the discrepancy in the self-harm outcome by performing a logistic regression where self-harm was regressed on cohort and measurement year. We observed a significant positive association with measurement year with the NTR showing increased endorsement of self-harm items with time (β, 0.063; SE, 0.015; $P = 2.41 \times 10^{-05}$), and a significant negative association with cohort (β, -1.561; SE, 0.165; $P < 2.00 \times 10^{-16}$). This may be a combination of differences in the start of data collection for both cohorts (1987 in the NTR and 2004 in CATSS) as well as differences in the wording of the questions across both cohorts, and may indicate that the higher rate observed in CATSS could be partly due to the later year of measurement.

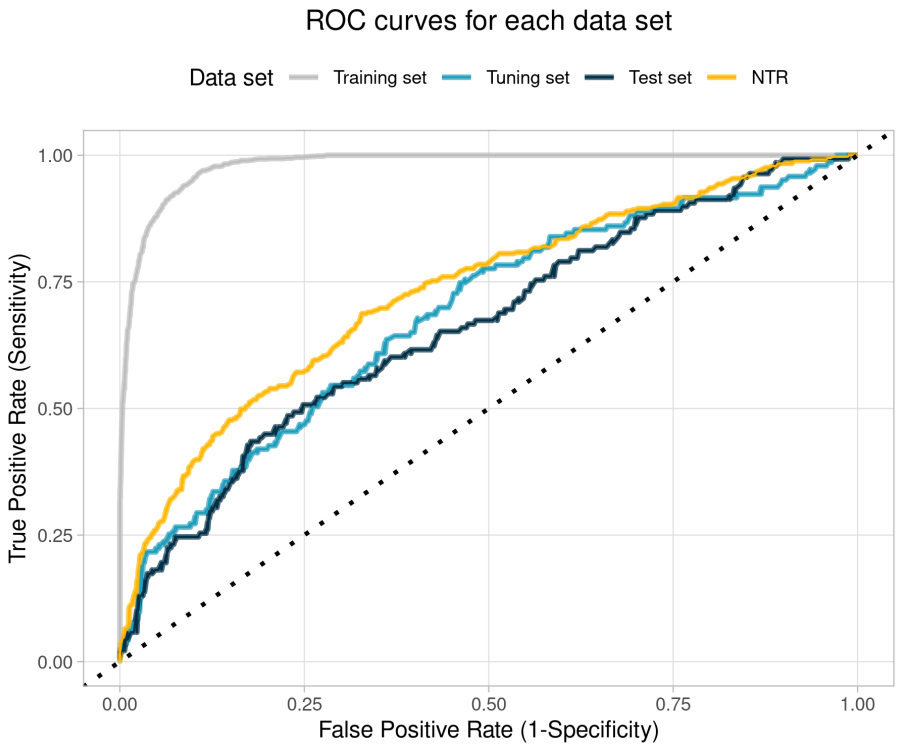
Table 1. Descriptive statistics for all datasets

Description	N (% Female)	N Outcome (%)	Self-Harm	Aggression	Both
Total data	7,965 (59.38%)	1,880 (23.60%)	822 (10.32%)	734 (9.22%)	324 (4.07%)
NTR	1,975 (69.06 %)	371 (18.78%)	47 (2.38%)	287 (14.53%)	37 (1.87%)
CATSS	5,990 (56.19%)	1,509 (25.19%)	775 (12.94%)	447 (7.46%)	287 (4.79%)
Training set	4,779 (56.02%)	3,551 (25.7%)	617 (12.91%)	375 (7.85%)	236 (4.94%)
Tune set	594 (56.06%)	143 (24.07%)	83 (13.97%)	30 (5.05%)	30 (5.05%)
Test set	617 (57.70%)	479 (23.79%)	75 (11.65%)	42 (7.49%)	21 (4.66%)

Model Performance

Ultimately, three models were included in the ensemble model: GBM (AUC [1000 bootstrap, 95% CIs]; train set 0.868 [0.837– 0.870]; tune set 0.661 [0.601 – 0.861]), RF (train set 1.000 [1.000 – 1.000]; tune set 0.676 [0.622 – 0.724]), and elastic net (train set 0.683 [0.630 – 0.728]; tune set 0.682 [0.664 – 0.700]). The neural network model (train set 1.000 [1.000 – 1.000]; tune set 0.606 [0.564 – 0.669]) did not have a suitable performance. Overall the model performance for the ensemble model was uniform between the datasets (train set 0.981 [0.995 – 0.965]; tune set 0.683 [0.624 – 0.731]; test set 0.663 [0.616 – 0.721]; NTR data 0.729 [0.703 – 0.761]) (Figure 2). For each dataset the positive predictive value (PPV) was lower than the negative predictive value (NPV) (Train set PPV = 56.3%, NPV = 99.9%; Tune set PPV = 33.9%, NPV = 87.3%: Test set PPV = 28.9% NPV = 84.6%; NTR data PPV = 26.6%, NPV = 91.4%) (See Supplementary Table 7 for breakdown based on NTR data). Based on a threshold of 0.219 determined by the maximum F1 statistic, the sensitivity (train set = 0.998; tune set = 0.755; test set = 0.667; NTR data = 0.801) and specificity (train set = 0.732; tune set = 0.532; test set = 0.528; NTR data = 0.489) of the models varied extensively.

Figure 2. ROC curves for each data set

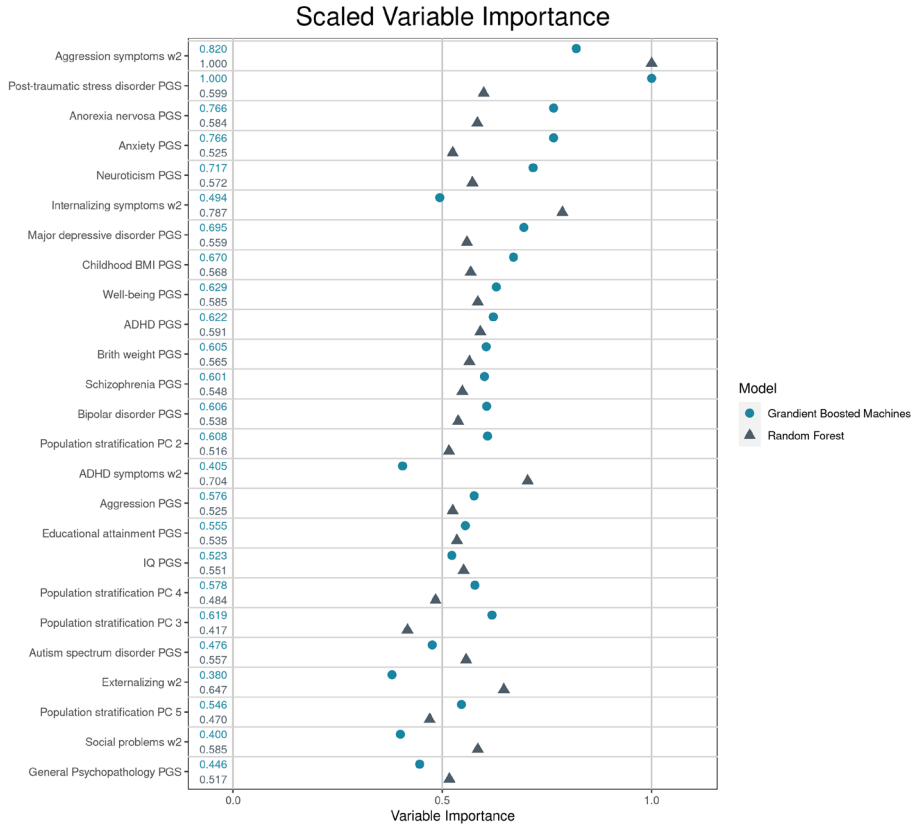


Note: Training set AUC (95% CI) = 0.981 (0.995 – 0.965); Tune set = 0.663 (0.616 – 0.721); Test set = 0.729 (0.703 – 0.761); Netherlands Twin Register (NTR) = 0.729 (0.703 – 0.761)

Variable Importance

Importance rankings across GBM and RF were inconsistent, although aggression at age 15/16 was ranked as most informative overall, followed by PGS of psychiatric traits, and then internalizing problems at age 15/16. Although confidence intervals overlapped across both models, RF was the better performing model and generally rated self-reported psychopathology symptoms at age 15/16 as the most informative, compared to PGS variables, parent reported symptoms at age 9/12, and environmental variables. On the other hand, aggressive symptoms at age 15/16 as well as various PGS variables were rated highest by GBM (Figure 3, Supplementary Table 8).

Figure 3. Scaled variable importance for the top 25 scores



Variable Importance represents the reduction in mean squared error when the variable was split on a node. Abbreviations: w1 = Measured at wave 1; w2 = Measured at wave 2; PGS = Polygenic score; PC = Principal component

Gradient Boosted Machines (GBM) Final AUC [1000 bootstrap, 95% CIs] train set: 0.868 [0.837– 0.870]; GBM Final AUC tune set: 0.661 [0.601 – 0.861]

Random Forest (RF) Final AUC train set: 1.000 [1.000 – 1.000]

RF Final AUC tune set: 0.676 [0.622 – 0.724])

Sensitivity analysis

The sensitivity analysis showed an over reliance on self-reported aggression at age 15/16 for both the GBM and RF models (Supplementary Figure 2; Supplementary Table 9). Additionally, the model performance was poor on the combined CATSS data set for both the GBM (AUC [1000 bootstrap, 95% CIs]; NTR train set 0.937 [0.924 – 0.950]; NTR test set 0.758 [0.694– 0.817]; CATSS data 0.585 [0.568 – 0.602]) and RF model (NTR train set 0.932 [0.995 – 0.965]; NTR test set 0.727 0.653– 0.789]; CATSS data 0.598 [0.580 – 0.614]).

DISCUSSION

In the current study we created a model to identify adolescents at a high risk of self-harm and/or aggressive behaviour, using a wide range of predictors including parental- and self-reported symptoms of psychopathology, behavioural measures, and genetic risk for different traits and psychiatric disorders. By training the model in the CATSS sample and validating it in the NTR sample, we tested the cross-cultural/external prediction of the model. Although the model had a lower final AUC of 0.663 in the CATSS test set, it performed satisfactorily in the NTR sample with a final AUC of 0.729, uniform with the CATSS training and tune set. A clinical cut-off of 80% for sensitivity and 50% for specificity has been previously proposed [324], and although the sensitivity (CATSS 0.667; NTR 0.801) and specificity (CATSS 0.528; NTR 0.489) was partially met in the samples, the performance across both metrics did not suffice in either dataset. However the model performance in the NTR suggests it is generalizable across national twin registers in the Netherlands and Sweden.

While the specificity and sensitivity of the model remain independent of the sample size, the PPV (the probability of correctly predicting the presence of the outcome) and NPV (the probability of correctly predicting the absence of the outcome) are influenced by the absolute number of cases and non-cases. Thus, we expected the PPV scores (CATSS 28.9%; NTR 26.6%) to be lower than the NPV scores (CATSS 84.6%; NTR 91.4%) for each of the datasets as the number of cases were relatively small compared to non-cases, i.e. class imbalance. The PPV scores indicate that the model was correct 29% and 27% of the time when it classified participants as having the outcome. This should be considered in the context of the true prevalence of the outcome which was 23.6% in the entire sample.

Overall, our study shows comparable or marginally improved prediction compared to previous studies investigating both self-harm and aggression outcomes in clinical settings. A previous study which developed a clinical risk assessment implemented by psychologists for self-harm and aggression had an average AUC of 0.63 and 0.66 respectively [302]. Another clinical study predicting self-harm and aggression in patients reported a PPV of 24% [303]. Our model shows improved prediction and arguably provides a less time-intensive approach geared towards the general population, as we did not use a clinical sample nor data from clinical

interviews. We also observed improved prediction when comparing our model to models examining only one of the outcomes, i.e. aggression OR self-harm but not both. For example, our model had improved performance when compared to a study examining aggression in a prison sample [325]. Moreover, based on systematic reviews of studies conducted primarily in psychiatric populations and military veterans, our study improves upon the weighted AUC of 0.61 for self-injurious behaviour and a pooled PPV of 26.3% [297, 298, 326]. Thus, our study shows improvement compared to past models, and its usefulness in a clinical setting is worth testing to see if prediction improves.

As we investigated a range of predictor types across many years, we were also interested in examining the predictiveness of key risk factor domains: home environment, genetic, behavioural, and psychiatric. Unfortunately the variable importance rankings for the GBM and RF models were somewhat inconsistent and did not paint a clear picture of which predictors could be considered most informative, though aggressive and internalizing symptoms at age 15/16, as well as PGS of psychiatric traits ranked highest on average. The high ranking of PGS of various traits may be indicative of genetic associations between the PGS and the outcome variables. In general, the variance explained from PGS of psychiatric traits are not high enough to be clinically relevant. Particularly, PGS of childhood psychopathology traits like ADHD and ASD are based on GWAS with relatively low sample sizes, and the variance explained will likely increase as GWAS become more powerful. These results support that an increase in power should be an important focus in the field of psychiatric genetics, as our results suggest that they are already more important than some other variables. Thus, while on their own PGS are currently not powerful enough to be clinically useful, our results suggest that they may provide additional information which improves prediction when combined with other variables/risk factors [327]. Clinically, such prediction may be useful in selecting individuals at highest risk for a chronic course or non-response to treatment. Finally, the sensitivity analyses where the model was based on the NTR sample showed a similar trend to the RF model in the main analyses. Self-reported symptoms of psychopathology measured at the second wave were the most important predictors of the outcome. However this should be interpreted with care as this model heavily relied on aggression measured at the second wave (Supplementary Figure 2). This might be because the prevalence of aggression in the NTR data was much higher than self-harm.

The strengths of this study include the diverse, longitudinal data across and the use of two international twin cohorts. Our study featured self and parental reported measures at multiple age points as well as genetic data in the form of PGS for many psychiatric disorders and other physical traits. Moreover, we were able to validate our model through an externally collected data source, which was able to provide evidence that the performance of the model was not driven by overfitting, i.e. the model closely fitting the training data as to not be generalizable to new data. However, our study comes with caveats. First, it is likely that our model would be improved by a larger sample size. Similarly, it is likely that additional variables related to emotional dysregulation and additional psychiatric symptoms would also increase our models performance. Thus, future studies which seek to create a model to predict self-harm and aggression in early adulthood should include more predictors, especially during middle teenage years. As we only used self-reported predictors for this time point, it would also be of interest to also include parent or teacher reported predictors. Next, due to power concerns we do not distinguish self-harm with suicidal intent from self-harm without suicidal intent. This distinction may also be an interesting avenue for future research. Finally, our measure of self-harm were somewhat inconsistent across both cohorts, as the measure of self-harm in the NTR also included suicidal ideation. While self-harm and suicidal ideation are linked and highly correlated [294, 328], they are not exactly the same.

Conclusions

Our model improved upon previous prediction models examining aggression and self-harm. The results suggest that aggressive behaviour in mid-adolescence is a key indicator for later aggression and self-harm behaviours. This upholds previous literature that aggressive behaviour in childhood is an indicator for a higher risk of negative outcomes later in life [329-331]. Additionally, internalizing problems in mid-adolescence as well as PGS of psychiatric traits were highly informative to the model. Future studies should investigate multiple raters, e.g. teachers, self, or parents, for questionnaires across each time point to examine whether the predictive importance of psychopathology in mid-adolescence is a function of timing or rater. For example, previous studies have shown that patterns of co-occurence between childhood aggression and internalizing/externalizing problems were largely rater independent [332]. However, SNP-based heritability estimates of self-rated aggression have been shown to be higher than maternal-

rated aggression [296]. Finally, additional work should be done distinguishing between the different classes i.e. multilevel models with aggression and self-harm as separate outcomes.

As of now, by and large, machine learning models are not ready for clinical use in psychiatric clinics and our model is no different [273]. However, we improve upon the performance of current prediction models that look at self-harm and aggression individually and provide a population-based model which combines the two.

Chapter 6

Ultra-rare, rare and common genetic variants implicate negative selection and neuronal processes in the aetiology of schizophrenia

Under revision as: **Akingbuwa, W. A.**, Hammerschlag, A. R., Bartels, M., Nivard, M. G.*, & Middeldorp, C. M.* (2021). Ultra-rare, rare, and common genetic variant analysis converge to implicate negative selection and neuronal processes in the aetiology of schizophrenia.

*shared last author

**Supplementary materials accessible at https://doi.org/10.1101/2021.05.26.21257794

ABSTRACT

Both common and rare genetic variants (minor allele frequency > 1% and < 0.1% respectively) have been implicated in the aetiology of schizophrenia. In this study, we integrate single-cell gene expression data with publicly available Genome-Wide Association Study (GWAS) and exome sequenced data in order to investigate in parallel, the enrichment of common and (ultra-)rare variants related to schizophrenia in several functionally relevant gene sets. Four types of gene sets were constructed 1) protein-truncating variant (PTV)-intolerant (PI) genes 2) genes expressed in brain cell types and neurons ascertained from mouse and human brain tissue 3) genes defined by synaptic function and location and 4) intersection genes, i.e., PI genes that are expressed in the human and mouse brain cell gene sets. We show that common as well as (ultra-)rare schizophrenia-associated variants are overrepresented in PI genes, in excitatory neurons from the prefrontal cortex and hippocampus, medium spiny neurons, and genes enriched for synaptic processes. We also observed stronger enrichment in the intersection genes. Our findings suggest that across the allele frequency spectrum, genes and genetic variants likely to be under stringent selection, and those expressed in particular brain cell types, are involved in the same biological pathways influencing the risk for schizophrenia.

INTRODUCTION

Schizophrenia is a severe and highly heritable psychiatric disorder with onset in late adolescence or early adulthood. It is associated with early mortality and greatly reduced fertility [333, 334], which put selective pressure on genetic variants related to schizophrenia. Despite apparent negative selection, schizophrenia remains highly heritable, with a relatively high prevalence [335]. Its severe clinical presentation and persistence despite negative selection make understanding the nature of the genetic effects on schizophrenia essential.

There is abundant evidence that both common (minor allele frequency > 1%) and rare genetic variants are related to schizophrenia [336-338], with the effects of the variants inversely correlated to their frequency. Rare novel variants can have larger effect, while common variants can only persist in the presence of negative selection if their effects are small [339, 340]. However, common variants, despite smaller individual effects, collectively explain a substantial proportion of the total genetic variance in schizophrenia [338, 341]. The presence of negative selection results in extreme polygenicity where substantial portions of the genome carry variants with tiny individual effects on schizophrenia, yet only critical (core) genes when perturbed by an influential mutation would strongly impact the disorder [342].

The genetic architecture of schizophrenia has forced two separate lines of enquiry into its genetic aetiology. Genome wide associations studies (GWAS) have successfully targeted common variant with small individual effects [54, 338, 343]. Importantly, they have resulted in valuable leads for functional follow-up studies of individual loci. For example, in-depth study of the lead genome wide hit for SCZ in the Major Histocompatibility Complex (MHC) locus has implicated complement component 4 (C4) gene expression and possibly synaptic pruning in puberty in the aetiology of schizophrenia [344]. While further analysis of the schizophrenia locus in the *SLC39A8* gene implicated manganese (Mn) related brain phenotypes in the aetiology of schizophrenia [345].

At the same time, whole exome sequencing (WES) has been used to identify rare mutations with larger effects. Since the variants of interest are rare, these studies require equally large samples. To reduce the multiple testing burden, early research has focused on specific classes of variants for which one can assume a deleterious

(risk increasing) effect a-priori. Specifically, researchers have leveraged modest sample sizes by focusing on *singleton* (i.e. only observed once) variants in genes intolerant to mutations, that are predicted to be protein-truncating variants (PTVs) (i.e. disruptive and likely lead to loss of gene function). These variants and genes are the ones most likely to increase the risk of schizophrenia when perturbed by a mutation of consequence [336, 346-348]. This has been borne out by results from these studies, with a recent study implicating 10 genes in which ultra-rare variants are significantly associated with schizophrenia [336].

These parallel lines of genetic inquiry based on different analytical strategies share a common goal: increased understanding of the neuro-biology of schizophrenia. Subsequently, functional genomic analyses are crucial for understanding the pathways and mechanisms via which these disorder-associated genetic variants may act. Functional analyses of common variants from GWAS have implicated the brain and brain-expressed genes in the aetiology of schizophrenia [54, 338]. WES studies, which make it possible to test for a burden of rare variants across shared functional units such as genes or gene sets, have similarly implicated the brain in schizophrenia aetiology [346, 347]. More recently, high throughput single-cell RNA sequencing techniques, which are able to provide expression profiles of individual brain cells at greater resolution, have been developed. They allow prioritization of specific brain cell types associated with disorders or traits. Single-cell expression data of mouse and human brain cells reveal that disorder-associated common and rare variants are enriched in genes expressed in (excitatory) neurons more than in other (non-neuronal) brain cells [336, 349, 350].

In this preregistered study (https://osf.io/uyv2s), we integrated single-cell gene expression data with results from GWAS and exome sequenced data in order to investigate, in parallel, the enrichment of common and (ultra-)rare variants related to schizophrenia in specific brain cell types. We investigated whether trait-associated common and (ultra-)rare variants were enriched in classes of genes that are functionally relevant for schizophrenia. These included sets of genes expressed in different brain cell types and neurons, as well as PTV-intolerant (PI) genes, i.e. genes under stringent selection. As synaptic functions have previously been implicated in the aetiology of schizophrenia, we included gene sets based on synaptic processes and composition. Finally, we investigated gene sets made up of the intersection of PI genes and the brain expressed genes (i.e. PI genes

that are expressed in the brain) as these are potentially smaller gene sets rich in genes related to the biology of schizophrenia and therefore of considerable value in follow-up analysis, if proven relevant. By synchronizing the functional analyses across common and rare variants, the current study attempts to answer two questions: 1) Do common and (ultra-)rare variant gene set and cell type enrichment analyses converge to similar results for schizophrenia and if so, 2) What gene sets are implicated across both (ultra-)rare and common variant analyses? The current analyses may shed light on whether common and (ultra-)rare variants reveal unique aspects of the aetiology of schizophrenia, or implicate the same pathways.

METHODS

Data and sources

GWAS data

To identify gene sets enriched for common variants, we obtained summary data from large publicly available GWAS results of schizophrenia in individuals of European ancestry [54], as well as those of East Asian ancestry [341].

Exome sequencing data

To identify gene sets enriched for (ultra-)rare variants, we obtained genotype and phenotype data from the Swedish Schizophrenia Exome Sequencing Project [346], a case-control sample of 12380 unrelated Swedish individuals. Cases primarily had diagnoses of schizophrenia, although a small proportion of individuals were diagnosed with bipolar disorder. See data availability for more information on this data set.

Gene sets

Protein-truncating variant (PTV)-intolerant (PI) genes were obtained from the Genome Aggregation Database (gnomAD), and ascertained using the probability of loss-of-function intolerance (pLI) metric. We selected genes with pLI > 0.9, producing a list of 3063 genes [351].

Human brain cell gene sets were based on single-nucleus RNA-sequence (snRNA-seq) data generated on the Genotype-Tissue Expression (GTEx) project brain tissues [352]. We included a total of 14 cell types as ascertained in the study referenced.

Sources and processing of expression data are described there. Excluding sporadic genes and genes with low expression, for the 14 cell types we selected the top 1600 (roughly 15%) differentially expressed genes in each cell type, which likely cover all, or most, genes that have a vital function in a specific cell type.

Mouse brain cell gene sets were based on data obtained from a previous study [350]. Extensive description of sources and processing of expression data are described there. In that study, cells were assigned to level 1 classification, with subtypes of level 1 assigned as level 2 on the basis of single-cell RNA-sequence (scRNA-seq) data and clustering analyses. We focused our analyses on the 24 level 1 gene sets. As the scRNA-seq data were from mouse brains, we mapped the gene homologs using the human-mouse homolog reference from Mouse Genome Informatics. Similar to the human brain gene sets, we selected the top 1600 differentially expressed genes in each cell type.

Synaptic gene sets were selected based on synaptic gene ontology from the SynGo database [353], including gene sets defined by cellular component, i.e. the location in which the genes are active, or by biological process, i.e. the synaptic processes/ functions they influence. In order to ensure that gene sets were powered enough to detect significant effects, we selected only gene sets containing 50 or more genes, resulting in a total of 35 gene sets.

PI x brain cell gene sets contained the intersection genes that are PTV-intolerant and are present in each human and mouse brain cell gene set. Although the PI x brain cell type are smaller than either the PI-gene set or the brain cell type specific gene set, they are potentially more strongly enriched given the genes are (1) relevant to brain function and (2) under strong negative selection.

In total 112 different gene sets were included in the analyses. All genes included in each gene set are available in Supplementary Table 1.

Common variant analyses

Common variant enrichment was evaluated using competitive analyses in MAGMA (v1.08b) [354]. MAGMA is a commonly used program for gene and gene set analysis that applies the principles of linear regression. It works by first computing gene-level associations in which *p*-values for individuals SNPs around a gene are

averaged, while taking linkage disequilibrium (LD) structure into account. LD is estimated using ancestry-appropriate reference panels for each population investigated. We used the European panel of 1000 Genomes Project phase 3 for LD estimation in the GWAS based on individuals of European ancestry, and the East Asian panel in the GWAS based on individuals of East Asian ancestry. Subsequently, gene-based *p*-values were converted to z scores to test associations between each gene set and schizophrenia diagnosis. For each GWAS summary dataset, we excluded SNPs with INFO <0.8, as well as duplicate SNPs. Gene location information with start and stop sites were obtained from the MAGMA website, with no windows specified around the genes.

Rare variant analyses

Analyses of exome sequenced data including QC was mainly performed using Hail 0.2 (Hail Team. Hail 0.2.62-84fa81b9ea3d. https://github.com/hail-is/hail/commit/84fa81b9ea3d). QC generally followed those described in a previous study [347], as well as those detailed here: https://astheeggeggs.github.io/BipEx/index.html. Full description, including variant annotation and (ultra-)rare variant definitions are provided in the Supplementary text.

Gene set (association) analyses

We assessed (ultra-)rare variant enrichment in each gene set using logistic regression, testing the association between the burden of (ultra-)rare variants, and schizophrenia diagnosis. Sex and 10 genetic PCs were included as covariates in each set of analyses. We excluded individuals who were more than 4 median absolute deviations from the study specific median number of synonymous (ultra-) rare variants. The significance threshold was set to 5% false discovery rate in analyses across all gene sets per variant allele frequency.

We deviated from the pre-registration outlined by omitting parallel analyses for autism spectrum disorder. As analyses of autism exome sequence data unexpectedly required considerably more processing than the schizophrenia dataset, publication of the current valuable results for schizophrenia would be unreasonably delayed.

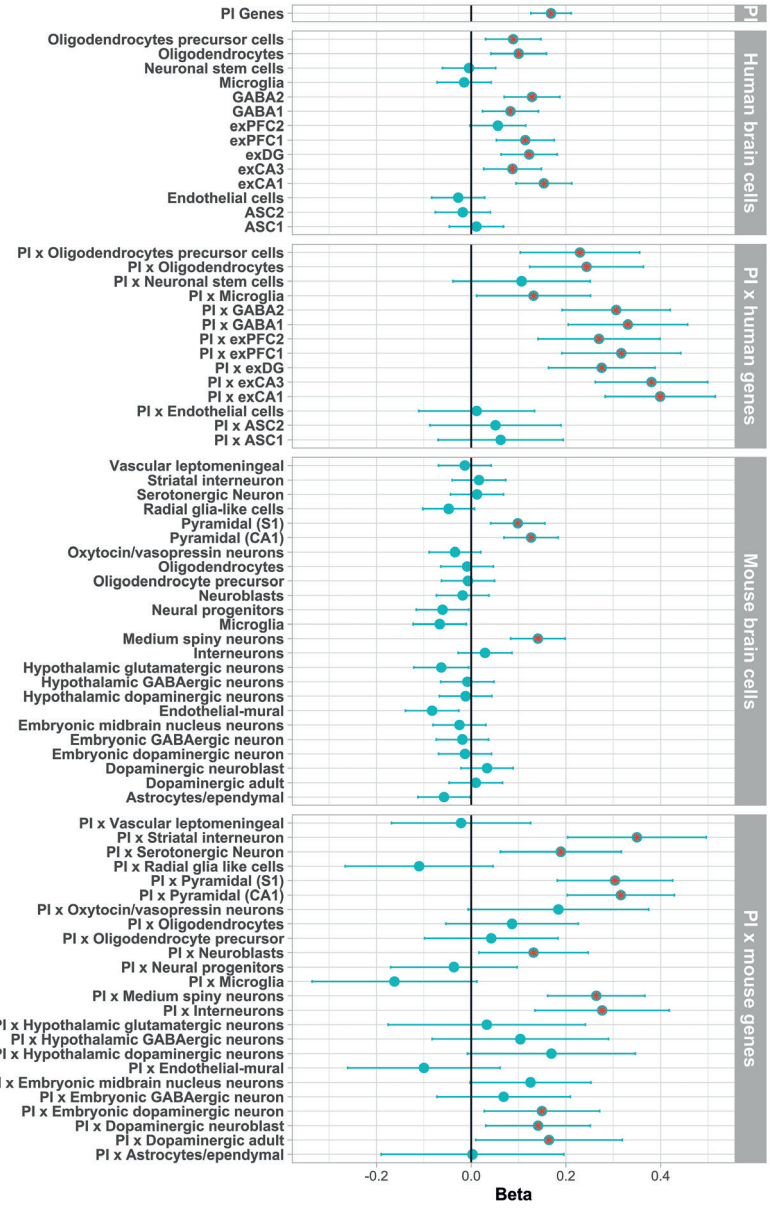
RESULTS

Common variants

We used MAGMA to evaluate the enrichment of common schizophrenia-associated variants in PI genes, as well as genes expressed in brain cells and synapses. In the European sample, schizophrenia-associated common variants were strongly enriched in PI genes, as well as genes highly expressed in human brain cells. Enrichment in genes expressed in mouse brain cells was also present across the different cell types, but not to the same extent as the human brain cells. Across brain cell types, we observed higher association betas at the intersection of PI and brain cell expressed genes suggesting stronger enrichment, although confidence intervals were largely overlapping (Figure 1). Wider standard errors in these analyses likely reflect the loss of power from selecting only intersecting genes. Across the human brain cell types analysed, enrichment was strongest for genes expressed in excitatory neurons from the hippocampus (pyramidal and granule) and prefrontal cortex, as well as GABAergic interneurons and oligodendrocyte cells. In mouse brain cells enrichment was strongest in pyramidal cortical and hippocampal neurons, as well as medium spiny neurons (a type of GABAergic neuron). Again we observed more significantly enriched gene sets in the PI and mouse brain interaction gene sets, additionally implicating dopaminergic, and serotonergic neurons as well as interneurons. Finally, we observed significant enrichment in genes associated with synaptic processes. Common variants were most enriched in postsynaptic cellular components and biological processes. The standard errors for these associations were wide, likely due to the lower number of genes per gene set (Supplementary Figure 1).

In analyses of individuals of East-Asian ancestry we observed a similar pattern of results as the in the analyses of European individuals, in that there was stronger enrichment at the intersection of PI and brain cell expressed genes. However, we only observed significant enrichment in four gene sets, including the PI gene set (Supplementary Figures 2 and 3). This may have been related to power as this GWAS had a smaller sample size.

Figure 1. Common variant enrichment in PI and brain cell gene sets.



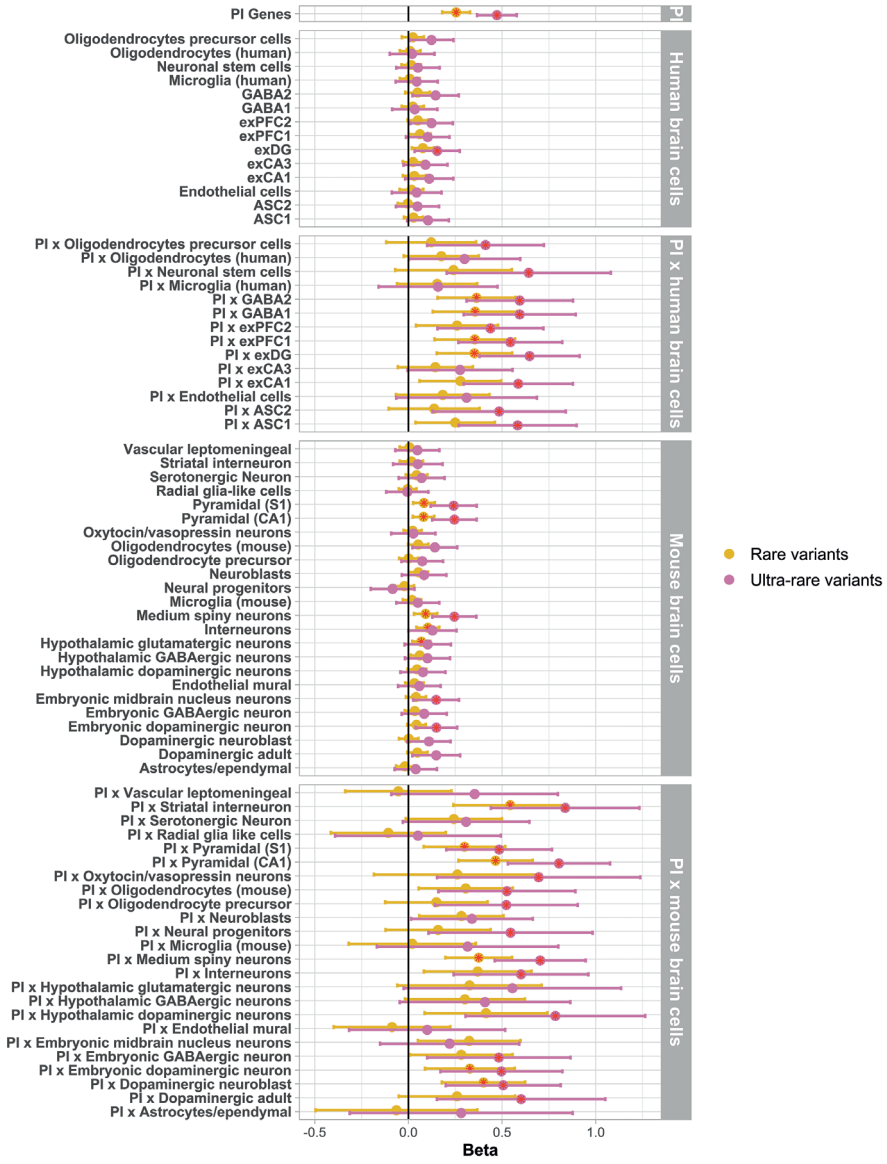
Red stars denote significant gene sets after multiple testing correction. ASC=astrocytes, exCA1/exCA3=pyramidal neurons from the Hippocampal Cornu Ammonis regions, exDG=granule neurons from the Hippocampal dentate gyrus region, exPFC1/exPFC2=pyramidal neurons from the prefrontal cortex, GABA1/GABA2=GABAergic interneurons. Although not included in the figure, the synaptic gene sets were included in multiple testing correction.

(Ultra-)rare variants

After QC, 10592 individuals (4623 cases and 5969 controls) were included in the ultra-rare variants analyses, while 10553 individuals (4603 cases and 5950 controls) were included in the rare variant analyses. The sample sizes were different across both sets of analyses due to the exclusion of individuals who were more than 4 median absolute deviations from the study specific median number of synonymous variants. We tested the association between the burden of rare (AF < 0.1%) or ultra-rare PTVs and schizophrenia diagnosis. Ultra-rare variants were those observed in 1 out of 188,023 individuals (our sample + gnomAD + DiscovEHR). Schizophrenia cases had a significantly higher burden of both ultra-rare ($\beta = 0.082$, SE = 0.017, $P = 2.79 \times 10^{-6}$) and rare ($\beta = 0.026$, SE = 0.007, $P = 0.0004$) PTVs compared to controls. As a negative control, burden scores for synonymous variants were computed, which were not significantly different between cases and controls for ultra-rare ($\beta = 0.013$, SE = 0.009, $P = 0.151$) or rare ($\beta = 0.0008$, SE = 0.002, $P = 0.718$) variants.

Next, we assessed (ultra-)rare variant enrichment in each gene set using logistic regression, testing the association between the burden of (ultra-)rare variants and schizophrenia diagnosis. Both rare and ultra-rare variants were significantly enriched in PI genes. Overall, there was greater enrichment of ultra-rare PTVs compared to rare PTVs across gene sets analysed. Similar to common variants, enrichment of rare and ultra-rare PTVs was greater at the intersection of PI genes and genes expressed in mouse and human brain cells (Figure 2), though in this case there was more enrichment of genes expressed in mouse brain cells compared to human brain cells. Across the brain cell types, enrichment was again strongest for genes expressed in excitatory neurons from the hippocampus and prefrontal cortex, as well as GABAergic interneurons, oligodendrocyte cells, and medium spiny neurons. We also observed enrichment in synaptic genes (Supplementary Figure 4). Complete regression results for PTVs and synonymous variants are described in Supplementary Tables 2 – 5.

Figure 2. (Ultra-)rare variant enrichment in PI and brain cell gene sets.

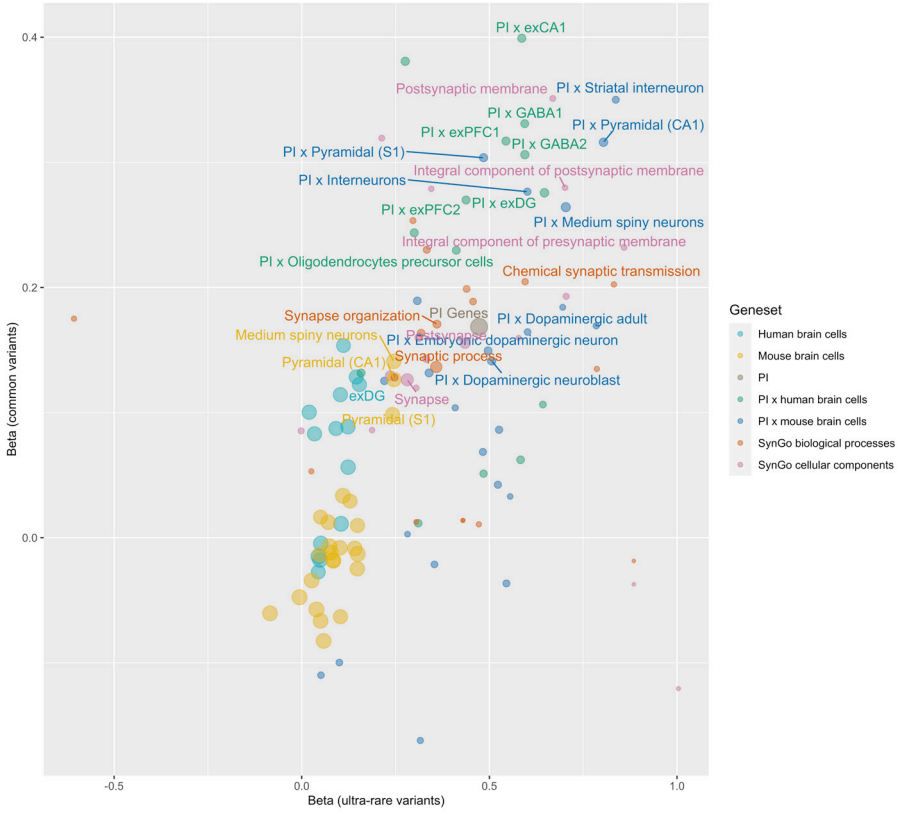


Red stars denote significant gene sets after multiple testing correction. ASC=astrocytes, exCA1/exCA3=pyramidal neurons from the Hippocampal Cornu Ammonis regions, exDG=granule neurons from the Hippocampal dentate gyrus region, exPFC1/exPFC2=pyramidal neurons from the prefrontal cortex, GABA1/GABA2=GABAergic interneurons. Although not included in the figure, the synaptic gene sets were included in multiple testing correction.

Convergence of enrichment across allele frequency spectrum

We investigated whether common and (ultra-)rare variants converged to similar results by rank correlating the association betas across the three allele frequencies for each gene set. We also evaluated overlapping gene set enrichment across the allele frequency spectrum.

Figure 3. Correlation between gene set enrichment in common vs ultra-rare variants.

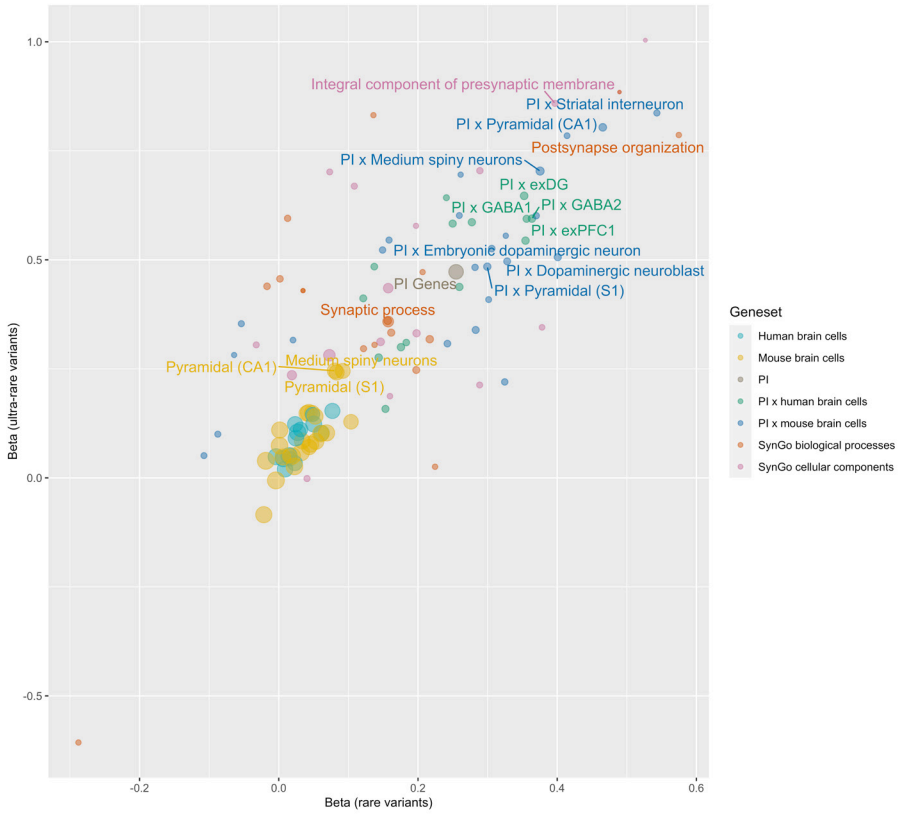


Point sizes represent the weight assigned to each correlation estimate, obtained by calculating the inverse of the product of both standard errors. Correlation estimate is 0.537, while weighted correlation is 0.708. Labelled gene sets are significantly enriched across both common and rare variants

We observed moderate to high correlations of effect sizes across the three pairs of comparisons. Correlation estimates ranged from 0.515 to 0.740, and from 0.561 to 0.802 when weighted by the inverse of standard errors from each

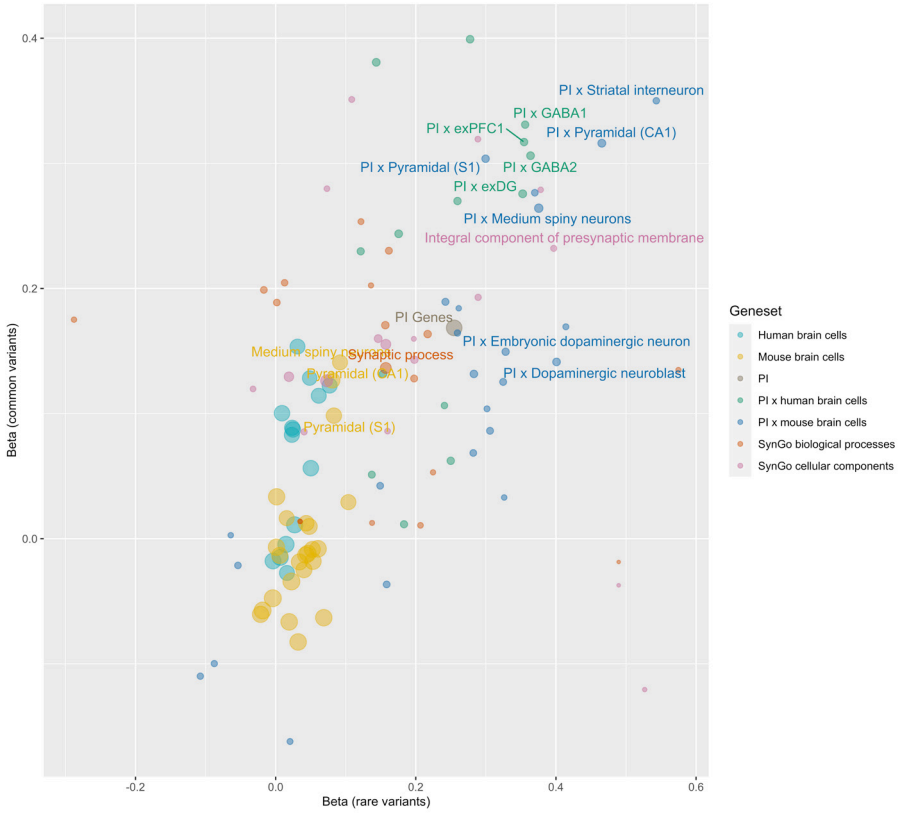
association estimate (weighting precise estimates more heavily than imprecise estimates when comparing across common and rare enrichment results) (Figures 3 – 5). Additionally, 16 gene sets were significantly enriched across all three allele frequencies, implicating PI genes, medium spiny neurons, and pyramidal neurons from the mouse brain, as well as multiple intersection gene sets, across the allele frequency spectrum. 28 gene sets were significantly enriched across common and ultra-rare variants, 17 across rare and ultra-rare variants, and 16 across common and rare variants.

Figure 4. Correlation between gene set enrichment in rare vs ultra-rare variants.



Point sizes represent the weight assigned to each correlation estimate, obtained by calculating the inverse of the product of both standard errors. Correlation estimate is 0.740, while weighted correlation is 0.802. Labelled gene sets are significantly enriched across both common and rare variants

Figure 5. Correlation between gene set enrichment in common vs rare variants.



Point sizes represent the weight assigned to each correlation estimate, obtained by calculating the inverse of the product of both standard errors. Correlation estimate is 0.515, while weighted correlation is 0.561. Labelled gene sets are significantly enriched across both common and rare variants

DISCUSSION

We used GWAS and whole exome analyses of schizophrenia to investigate whether common and (ultra-)rare PTV enrichment converge to similar results in terms of what gene sets are implicated across analyses. We observed partial convergence across the gene sets significantly enriched for common and (ultra-)rare variants, in that multiple gene sets were significantly enriched across all three variant classes. Enrichment analyses implicated mainly excitatory neurons from the prefrontal cortex and hippocampus, medium spiny neurons, and GABAergic neurons, as well

as PI genes, synaptic components, and processes. Moreover, across all three allele frequencies, enrichment was stronger in the gene sets containing the intersection of brain cell types and PI genes compared to the brain cell gene sets.

Brain cell enrichment findings are consistent with findings from previous analyses of both common and (ultra-)rare variants associated with schizophrenia [336, 338, 349, 350]. Overlapping significant enrichment between ultra-rare and common variants provides additional evidence of some convergence in genes and biological mechanisms implicated by genetic variants across the allelic spectrum. Recent analyses showed significant enrichment of ultra-rare variants in genes implicated by schizophrenia GWAS, and that two genes implicated in rare variant analyses also showed associations in the schizophrenia GWAS [336]. Additionally, we showed that genes likely to be under stringent selection (PI genes) are implicated in both common and (ultra-)rare variants, while stronger enrichment in the intersection of brain cell types and PI genes suggests that PI genes are generally important, but even more so in these cell types. These gene sets, particularly the intersection gene sets, potentially provide a manageable set of genes and biological processes to target for follow-up analyses.

An important factor to consider in light of our and similar findings, is how much progress the results represent with regards to disease biology. Analyses of common variant cell-type enrichment support current distinctions between neurological disorders like Parkinson's disease and Alzheimer's disease, versus psychiatric disorders, as they have shown different association patterns. Parkinson's disease has implicated cholinergic and monoaminergic neurons, Alzheimer's disease has implicated microglial cells, while psychiatric disorders like schizophrenia have implicated excitatory neurons [349, 355, 356]. Findings across psychiatric disorders are less clear. Analyses of individual psychiatric disorders suggest similar patterns of cell-type associations for disorders including schizophrenia, bipolar disorder, MDD, and anorexia nervosa, although enrichment was generally strongest for schizophrenia [73, 349]. However, analyses of psychiatric disorder factors have shown that common variant enrichment in excitatory and GABAergic genes from human brain cells is limited to a psychotic disorder factor (comprised of schizophrenia and bipolar disorder), with no enrichment observed in compulsive (anorexia nervosa, obsessive compulsive disorder, Tourette syndrome), neurodevelopmental (attention deficit/hyperactivity disorder, autism spectrum disorder, post-traumatic

stress disorder, and MDD), or internalizing factors (post-traumatic stress disorder, MDD and anxiety disorders) [357]. Importantly, similar neuronal enrichment has been observed for non-psychiatric cognitive traits including intelligence, educational attainment, neuroticism, and body mass index (BMI), which show modest but robust genetic correlations with psychiatric disorders including schizophrenia [349, 358-360].

These findings suggest that at the current resolution of analyses (expression differences between all cell types), common variant enrichment in genes predominantly expressed in neurons is non-specific and pervasive across various behavioural traits, although the lack of findings for most psychiatric disorders may also point to differences in statistical power across the GWAS. Future analyses, for example based on more comprehensive single-cell sequencing of all neuron subtypes, could identify genes that are specific to certain developmental stages in order to seek out cell types, cell functions or developmental phases that are specific to schizophrenia. Such analyses using whole-brain developmental expression profiles have shown enrichment of schizophrenia-associated common variants in the prefrontal cortex during early midfetal development [361]. Limits on data availability currently make it difficult to investigate whether these similarities across various traits in common variant cell type enrichment translate to (ultra-) rare variant enrichment, and is a significant avenue for future research, although brain-expressed genes have also been found to be enriched for ultra-rare variants associated with educational attainment [362]. This is vital as the importance/ contribution of variant classes along the allelic spectrum may vary depending on phenotype.

Our study, as well as other results [336, 338], further suggest that polygenicity, where very many genetic loci are implicated in a disorder like schizophrenia, complicates the search for individual risk loci in both common and (ultra-)rare variant analyses. As gene set analyses are strongly correlated between GWAS and (ultra-)rare variant analysis, it is not unlikely that analyses at the level of the gene show similar correlations between common and (ultra)rare variants. Strong correlations at the level of the gene may call for meta-analysis across GWAS and rare variant studies, or more subtle information integration that accounts for both the weight of evidence, as well as the regions/features/functions of the gene that are influenced by rare or common variation.

Our findings should be considered in light of some limitations. Firstly, our findings are limited by the current definitions of ultra-rare variants, and results might be subject to change if definitions change as more data becomes available. Secondly, our analyses are likely affected by the size of gene sets. Larger gene sets likely have more power to detect effects compared to the smaller gene sets. This was particularly evident in the interaction and synaptic gene sets which had wide confidence intervals. Finally, our analyses were limited to individuals of European ancestry, making it not generalizable to individuals of other ancestry. Overall similar analyses typically contain limited non-European samples, with non-European samples making up 20% of the most recent schizophrenia GWAS [338] and 26% of the most recent WES-based schizophrenia analyses [336].

In conclusion, we show that there is at least partial overlap in the genes disrupted by both common and (ultra-)rare variants associated with schizophrenia suggesting involvement of the same biological mechanisms. Genes influencing neuronal processes as well as genes likely to be under stringent selection are implicated in schizophrenia aetiology across common and (ultra-)rare variants. Future studies integrating information across the allele frequency spectrum might prove useful in furthering our understanding of schizophrenia aetiology.

Chapter 7

Conclusion and general discussion of findings

Childhood and adolescence are important developmental stages that can set the tone for long-term outcomes across the lifespan. In particular, children with psychopathology have an increased risk of continued psychopathology in adulthood compared to the general population [5-9]. Moreover, psychopathology in childhood has been repeatedly identified as a risk factor for other adverse health and socio-economic outcomes. Therefore an understanding of the genetic underpinnings of such traits could be crucial in identifying children most at risk of adverse outcomes, as well as in developing intervention or prevention strategies that may prove useful to them. Over the past decade, there has been major progress in the field of molecular and statistical genetics that facilitated research on the genetic aetiology of psychopathology and psychiatric traits. However, for various reasons, including availability of larger datasets and less developmental variance, the bulk of this research has been on adult traits. Therefore, we performed studies with the aim of investigating genetic mechanisms underlying the persistence of psychiatric traits in children and adolescents, as well as understanding how they develop into adulthood. By performing analyses across the allele frequency spectrum, we further investigated both common and rare genetic variants and their contribution to psychopathology. In this chapter I summarise the major findings from these studies as well as their implications and I discuss them in the context of other findings in the field. Finally, I discuss potential future research directions in childhood psychopathology and psychiatric genomics, informed by findings from the studies in the thesis.

As with initial GWAS of adult traits, the first GWAS of childhood psychopathology traits did not identify any convincing trait-associated genetic variants as a result of small sample sizes. However, results from analyses of SNP-based heritability were low but promising, thus hinting at the possibility of eventually identifying trait-associated variants. Similarly, polygenic scores (PGS) which were initially developed to test the theory of polygenic inheritance, were able to predict the same trait in another sample even when based on GWAS with few or no significant hits. This further hinted at effects captured in common variants which could become significant with appropriately large sample sizes.

By the start of my PhD in 2017, studies using various types of polygenic analyses had been performed for over a decade, with incremental improvements and findings with increased sample sizes in each new study. The systematic review

performed in **chapter 2** was born out of an interest in assessing how far the field had come with regards to findings on molecular genetic analyses of childhood psychopathology. Specifically, we were interested in the extent to which molecular and statistical genetic approaches could be used to explain development, stability, and comorbidity of childhood psychiatric traits. We showed that larger sample sizes facilitate the discovery of trait-associated genetic variants, as well as provide evidence of heritability based on measured genetic variants, known as SNP-based heritability. The review also showed that these SNP-heritability estimates are lower than estimates from twin studies, a known phenomenon across all traits GWAS called the "missing heritability" problem, which was already well known for adult psychopathology [363]. Finally, we observed abundant cross-trait associations across psychopathology traits, suggesting a role for pleiotropic genetic effects (where multiple traits are influenced by the same genetic variant) in the aetiology of psychopathology across development. On the whole this review showed that while substantial progress has been made regarding the genetics of childhood psychopathology, gaps in knowledge remain, and various avenues for research need to be exploited in order to further our understanding. These include analyses of causality, more inclusion of non-European populations in analyses, analyses of larger, phenotypically homogeneous samples, as well as integrating information from rare genetic variants in analyses.

Concurrent with the review described in **chapter 2**, we were interested in the genetic underpinnings of observed phenotypic associations between psychopathology in childhood and adult mood disorders, and related functional outcomes. Building further on the findings described in the review, in **chapter 3** we used polygenic scores (PGS) to investigate whether genetic risk for adult mood disorders and related traits are associated with childhood psychopathology. Combining childhood data from multiple longitudinal birth and population cohorts across Europe, we assembled the largest sample of a study of this kind (N=42,998), testing associations between PGS of adult major depression, bipolar disorder, neuroticism, insomnia, subjective wellbeing, educational attainment, and body mass index (BMI), and phenotypic measures of childhood attention-deficit/hyperactivity disorder (ADHD) symptoms, internalizing, and social problems. We showed a consistent pattern of mostly stable genetic associations between adult trait PGS and childhood psychopathology across age, indicating the existence of a set of genetic factors that influence psychopathology and related traits across the lifespan. Additionally, we showed

differential associations between educational attainment and BMI PGS, and types of childhood psychopathology. Specifically, educational attainment PGS was only associated with ADHD symptoms, while BMI PGS was associated with ADHD symptoms and social problems but not internalizing problems.

As both the adult and childhood traits investigated are genetically correlated, it raised the question of whether the ubiquitous genetic associations observed in **chapter 3** are influenced by correlations between related traits. Thus, in **chapter 4**, we extended our previous analyses by performing multivariate multivariable regression analyses of all the traits from our previous study using OpenMx, which allowed us to account for correlations between the childhood traits and between the adult trait PGS. Again, we observed differential associations between educational attainment and BMI PGS, and childhood psychopathology, i.e. educational attainment PGS was associated with ADHD symptoms, while BMI PGS was associated with ADHD symptoms and social problems. Crucially, previously observed associations between neuroticism, insomnia, and wellbeing PGS and childhood psychopathology measures were no longer present. Only major depression PGS remained associated with all three childhood psychopathology measures, suggesting that while shared genetic factors have a role in our understanding of psychopathology, unique genetic factors are likely also important. Additionally, neuroticism only remained associated with internalizing problems which might be indicative of shared item- or symptom-level measures of both traits.

One of the major aims of studying psychopathology is prediction of who will eventually develop psychopathology. While genetics clearly play a role in psychiatric aetiology, non-genetic/environmental variables also play a role and potentially represent modifiable risk factors for targeted intervention or prevention strategies. Unfortunately such factors are difficult to account for in typical molecular genetic studies. Prediction models represent one method of assessing the relative contribution of different variable types in the development of an outcome. We used this strategy in **chapter 5**, where we combined genetic and early life environmental and psychosocial variables in order to predict aggression and/or self-harm in late adolescence/early adulthood. Although as expected the model performance indicated it would not be suitable for clinical use, our results suggested that aggressive symptoms in mid-adolescence as well as PGS of psychiatric disorders were important predictors of aggression and/or

self-harm. We also showed that a combination of different types of predictors, including both genetic, environmental and psychosocial factors, provided the best model for predicting our outcome.

The analyses described so far largely involved common genetic variants (MAF > 1%) and their role in psychopathology. However, studies have shown that the contribution of rare variants is non-trivial. In **chapter 6** we compared schizophrenia-associated common and (ultra-)rare variant enrichment in gene sets that are functionally relevant for schizophrenia. Our analyses implicated genes likely to be under stringent selection, as well as those expressed in excitatory and medium spiny neurons. Moreover, results showed partial convergence across common and (ultra-)rare variants, indicating that the same biological mechanisms are implicated across the allele frequency spectrum. While studying schizophrenia may represent a deviation from the childhood traits studied in previous chapters, it is one of few traits with appropriate data and statistical power for this kind of analysis and provides proof of concept for how analyses like these might be useful in childhood traits as well.

General discussion and implications of findings

Pervasive genetic overlap across psychopathology

One of the main findings from **chapters 2, 3** and **4** is of pervasive genetic overlap amongst a variety of traits across age, indicating pleiotropy. Pleiotropy occurs when a single genetic variant or gene influences multiple phenotypes. These findings are consistent with numerous studies showing the existence of widespread pleiotropy across psychiatric traits in general, with a recent GWAS analysing over 500 traits including psychiatric traits finding that more than 60% of trait-associated genes were pleiotropic [364]. In **chapters 3** and **4**, we show genetic associations between childhood psychopathology and multiple adult traits not restricted to psychiatric disorders/traits. These results have implications for the continuity of psychopathology across the lifespan, as well as showing that beyond psychopathology, genetics also underlie associations between childhood psychopathology and adult health and socio-economic status (SES) outcomes like BMI and educational attainment.

Pleiotropy can be the result of multiple genetic mechanisms including horizontal pleiotropy – where a genetic variant has direct biological influence on multiple

traits, and vertical pleiotropy – where a variant influences one trait and that trait has a causal effect on another [166, 365]. While the polygenic analyses described in these chapters can provide insight into the genetic architecture of traits, they are not informative about the specific mechanism(s) of pleiotropy that might be at play. Yet this knowledge may be important to be able to use genetic information for future clinical applications. Additionally they are not able to indicate which specific variants or genome regions are responsible for the observed effects. One way in which such pleiotropic variants may be identified is via cross-disorder GWAS meta-analyses. This method has been recently applied in joint analyses of disorders including anorexia nervosa, ADHD, ASD, bipolar disorder, major depression, obsessive compulsive disorder (OCD), schizophrenia, and Tourette's syndrome (TS), performed by the Psychiatric Genomics Consortium's Cross-Disorder Group (PGC-CDG) [97]. Of the 146 genome-wide significant loci identified in this study, 109 (about 75%) were found to be pleiotropic, affecting two or more disorders, while 23 loci (about 16%) affected four or more disorders. The early-onset neurodevelopmental disorders including ASD, ADHD, and TS were implicated in 36%, 16%, and 14% of the pleiotropic loci respectively. The identification of these variants link the occurrence of pleiotropy to specific genetic variants, and represent an important step in understanding this phenomenon and its contribution/role in explaining psychopathology.

Importantly, pleiotropy underlying cross-trait associations may have clinical implications. One of the most pertinent is the implication for current clinical boundaries on which diagnoses are based. Disorders are generally categorised as distinct syndromes based on the presence of specific symptoms as well as other indicators. However, the observation of widespread genetic overlap across different traits as shown by numerous studies raises questions as to the validity of current diagnostic classifications. While useful and practical, current research results indicate the need for a different, more nuanced approach to current classification systems. Namely, one that is also informed by genetic aetiology, particularly with regards to disorders like depression where symptom profiles are highly heterogeneous across patients. For example, genetic information may be leveraged to better define more homogeneous disorder subtypes that better reflect underlying aetiology. Additionally, pleiotropy may have implications for drug development in that development of drugs targeting pleiotropic genes or loci might mean that these drug therapies are useful for treating a range of disorders.

Shared and unique genetic factors likely underlie childhood psychopathology

Observations of correlation, comorbidity, and pleiotropy, have fuelled a line of inquiry into the general factor of psychopathology, so-called *p* factor; the idea that psychopathology exists along a shared continuum and that a broad latent dimension captures variance across all psychopathology traits [164, 165, 366]. Both phenotypic and genetic analyses of psychopathology have provided evidence for this phenomenon, with the *p* factor found to be associated with a range of psychopathology in childhood [164, 251]. These findings further highlight the role of shared genetic factors to the architecture of childhood psychopathology. However, in **chapter 4** we observed that while major depression was genetically associated with all childhood measures of psychopathology investigated, educational attainment, BMI, and neuroticism were genetically associated with specific childhood phenotypes. Further, the finding that major depression PGS remained associated with all three childhood psychopathology measures was interesting and suggests that previously observed associations of childhood psychopathology with adult trait PGS are influenced by their genetic correlations with depression, and that genetic factors influencing adult major depression are associated with childhood psychopathology over and above genetic factors influencing other outcomes. The results suggest that genetic variants associated with depression are important for all childhood psychopathology measures investigated. However, it would be interesting to know if other childhood psychopathology measures, for example externalizing traits, show this association with depression as well. Modest genetic correlations have been observed with childhood aggression for example [296]. Furthermore, our findings highlight the importance of multivariate methods in building a more complete picture of genetic associations.

Together, our findings suggest that a combination of shared and specific genetic factors may underlie childhood psychopathology and explain their genetic overlap with adult traits. This finding is in line with studies that show that while different psychopathology traits typically load unto a common factor, there is evidence of unique contributions as well [251]. For example, in **chapter 2** we describe a study that showed differential genetic and phenotypic associations between ADHD and neurodevelopmental disorders, versus externalising or internalizing disorders, after accounting for the *p* factor [169], as well as another study that showed that while the *p* factor explained considerable variance in childhood psychopathology measures, inclusion of more specific emotional, behavioural and neurodevelopmental factors

explained even more variance than just the *p* factor alone [367]. Additionally, factor analyses of genetic covariance across psychiatric disorders suggest that three [97] or four [357] correlated genetic factor models, including one characterised primarily by childhood-onset neurodevelopmental disorders including ASD and ADHD, best explain genetic overlap between disorders.

Prediction of psychopathology; PGS as clinically relevant risk factors

A major aim of studying psychopathology is psychiatric risk prediction. Analyses of heritability has shown that while complex traits are influenced by genes, they also have substantial environmental components. Thus accurate prediction of psychopathology is unlikely to be achieved using only genetic information. In **chapter 5** we showed that a combination of genetic, behavioural, environmental, and psychiatric variables predict aggression and self-harm behaviours in late adolescence/early adulthood. Importantly, we show that genetic predictors indexed via PGS provide important information for improving prediction, suggesting a potential role for them in the future as predictive risk factors.

Genetic prediction of complex traits like psychiatric disorders is a potential avenue for clinical use of PGS, potentially allowing for early identification of risk, better characterization of psychopathology, and effective treatment. The use of PGS as clinical measures is attractive for many reasons. They are relatively easy to obtain, non-invasive, immutable, and the cost of generating them is getting cheaper as genotyping techniques also become cheaper. However, the predictive power of PGS does not appear clinically relevant as of yet. As observed in **chapter 5**, in our prediction model for self-harm and aggressive behaviours where PGS of psychiatric disorders ranked highly, the overall model performance was only average. In **chapter 3**, we also observed that the best predictive PGS of adult traits explained <1% of the variance in childhood psychopathology. This is the result of many different factors, including that this particular analysis represented cross-trait predictions of traits at a different age, although as described in **chapter 2**, similarly low variance explained has been reported in studies performing within-trait prediction.

Though the scale of GWAS has greatly improved over the last decade, adequate sample sizes are still lacking for most traits for GWAS to produce clinically useful PGS, with childhood traits particularly lagging behind until fairly recently [170, 296]. Even then, large sample sizes have only been achieved for population-based samples,

with polygenic scores in these studies explaining between 0.036% to 0.44% of phenotypic variance. It is possible that increased prediction might be achieved using a target sample with clinical diagnosis. Recent multivariable analyses showed that PGS of educational attainment and smoking initiation jointly explained up to 5.9% of variance when predicting case-control status in a clinical psychiatric child and adolescent [368], higher than observed in other similar studies using population-based samples. The highest variance explained estimate from univariate analyses (3.99% for educational attainment) in the same study was also higher than in comparable analyses using larger population-based samples. However, large enough sample sizes are generally difficult to obtain for psychiatric disorders, and more so childhood psychiatric samples, given the relatively low prevalence of these disorders.

Some of the most predictive PGS of complex traits include those for height and educational attainment, with the proportion of variance explained by each up to 24% and 13% respectively [369]. While substantial, these estimates are still well off the estimates from twin and family based studies [26]. Moreover, there is evidence that these values might be optimistic, as recent studies have shown that indirect parental effects might influence PGS prediction. Non-inherited parental genes which influence the environments they select for their children (genetic nurture) has been shown to impact prediction in cognitive traits like intelligence and educational attainment [137, 370, 371], which are modestly correlated with many psychiatric traits. Although yet to be proven, this may also be the case for (childhood) psychiatric traits.

So far, PGS of childhood psychiatric phenotypes, including ADHD and ASD, have been able to achieve group-level discrimination, i.e. distinguish cases from controls [38, 53], and schizophrenia PGS has been shown to be significantly higher in adolescents at high risk of psychosis who went on to develop psychosis [119]. However, individual-level prediction is currently not feasible, and PGS are unlikely to ever be stand-alone predictors of disorders. Nevertheless, they remain a useful research tool to investigate genetic overlap in traits for which samples are not large enough to employ other methods. They have potential as biomarkers in psychiatry and could eventually be used in combination with other risk factors to improve disorder risk prediction [327, 372], as we do in **chapter 5**. Improving phenotyping in order to decrease the heterogeneity of samples included in GWAS, as well as increasing GWAS sample sizes are some of the most viable avenues of improving their predictive ability. Unfortunately, there is a lack of large clinical samples of childhood psychiatric traits

compared to population-based or adult samples; which remains a stumbling block for improving PGS prediction and further assessing their clinical utility. Prediction based on GWAS of population-based samples is unlikely to ever be precise enough for stand-alone clinical use as prevalence rates of disorders are too low in such samples. Additionally, clinical samples are more likely to carry a broader range of genetic variation associated with a given disorder. A focus on high-risk groups, i.e. clinical samples, is warranted to address this. The results from Jansen et al., (2021) indicate the importance of large scale genetic data collection in clinical samples, in order to facilitate well powered studies in high-risk samples.

The role of rare variants

One of the reasons why PGS prediction remains limited is because the GWAS on which they are based are limited to common genetic variants with allele frequencies of at least 1% in the population. However, other types of genetic variants also play a role in the genetic architecture of psychiatric traits including rare, *de novo*, and structural variants. Next generation sequencing of protein-coding regions of the genome has allowed for the identification of trait-associated genes or sets of genes which are enriched for rare deleterious/disruptive variants [348]. Recent work has suggested that rare variants may offer an insight into genes central to the biology of complex disorders but which cannot be studied from common variant analysis. Due to strong negative selection, common variants that influence key genes are not allowed to persists in the population [342]. On the other hand, newer variants with more deleterious impact arise due to high mutation rates and are necessarily rare as they have not had time to be removed from the population [339, 340]. This complementary line of enquiry is important to pursue as genetic variants at different allele frequencies likely show differential contribution to the architecture of psychiatric disorders. Moreover, association studies across the allele frequency spectrum have had differing levels of success for different disorders. For example, while rare variant analyses have been more successful and implicated more genes for ASD and developmental disorders compared to common variant studies, studies on schizophrenia have generally had success across both rare and common variant analyses [373]. GWAS of ASD have identified 5 associated loci compared to 270 for schizophrenia, while over 100 genes enriched for rare coding variants have been identified for ASD (especially ASD comorbid with intellectual disability) compared to 10 for schizophrenia [31, 336, 373]. Overall, large copy number variants (CNVs) have provided the strongest evidence for specific trait-associated rare variants. While there are important methodological

considerations regarding these findings, statistical power not being the least of them, it is not impossible that different classes of variants are differentially important for different disorders.

As gene discovery studies require large sample sizes to gain enough power to identify risk genes for disorders, gene-set analyses provide a statistical advantage. By aggregating genomic effects from multiple variants, we increase the power to detect associations with a trait/phenotype. Additionally, as gene sets can be constructed to correspond to biological functions, mechanisms, pathways, and/or networks, associations with a phenotype are informative for potential mechanism discovery, as we demonstrate in **chapter 6**. We used gene sets to show that schizophrenia-associated common variants as well as (ultra-)rare protein-truncating variants (PTVs) are significantly enriched in PTV-intolerant (PI) genes which are likely to be under stringent selection, as well as in excitatory neurons from the prefrontal cortex and hippocampus, medium spiny neurons, and genes enriched for synaptic processes. Our results suggested some convergence in the pathways/mechanisms underlying schizophrenia aetiology across the allele frequency spectrum. Additionally, we evaluated findings from other studies regarding common variant enrichment of similar gene sets to those we investigated in other psychiatric and neurological disorders, to assess the specificity of ours and similar findings. We showed that the same genes and brain cell types are implicated in other psychiatric disorders as well as cognitive traits like intelligence and educational attainment, further indicating, perhaps unsurprisingly, that pleiotropy remains prevalent at this level of analysis. Additionally, cross-disorder analyses have also used gene sets to show that pleiotropic risk loci were significantly enriched in pathways related to neurodevelopment [97]. Pleiotropic effects have also been observed for rare variants, with studies showing that genes enriched for rare disruptive variants overlap across schizophrenia, ASD, and intellectual disability [374].

Future directions and considerations

Disentangling shared and specific genetic factors

Overall, the current body of evidence regarding genetic overlap suggests that while pleiotropy is clearly an important factor underlying psychopathology, specific genetic factors likely also play an important role. Future studies focused on further disentangling these effects will be important in better addressing or explaining genetic overlap across psychopathology. For example, recent analyses

leveraging the genetic factor structure underlying anxiety and depression were able to disentangle their shared genetic architecture, identifying shared genetic regions associated with both traits, regions associated with each, and regions associated with both but via separate variants [375]. More studies like this will be useful in further unravelling shared and unique effects amongst traits. This is especially crucial for traits like schizophrenia and bipolar disorder, which represent distinct phenotypes/diagnoses, but are highly genetically correlated. Additionally, analyses that are stratified across age, such as in **chapter 3**, are needed to shed light on developmental mechanisms underlying childhood psychopathology.

Causal inference in psychopathology across development

An important consideration regarding pleiotropy is that it might be induced by causation (vertical pleiotropy). Standard polygenic analyses are not informative on the direction of causation in observed associations, or if the genetic associations are the result of causal mechanisms in the first place. Genetic associations could also be the result of cross-trait assortative mating, where genes for two traits are passed down and inherited together due to mating between people who score high on one trait and those who score high on the second [376]. Nevertheless, knowledge regarding potential causal mechanisms is crucial given the long-term implications of psychopathology in childhood. Additionally, evidence exists for disorders such as bipolar disorder, where onset in childhood was associated with poorer outcomes, compared to adult-onset bipolar disorder [377]. Crucially, poorer outcomes extend beyond psychopathology, and we showed in **chapters 3 and 4** that childhood psychopathology is not only correlated with adult major depression, but with health and socio-economic status (SES) outcomes including BMI and educational attainment. An understanding of causal mechanisms is therefore crucial for identifying children most at risk for unfavourable outcomes in adulthood, and to ensure proper early mitigations. Methods like Mendelian randomization [378] use genetic markers (either a single SNP or a combination of multiple SNPs such as PGS) that are associated with a modifiable exposure as instrumental variables in order to investigate whether the association between the exposure and an outcome is causal. In **chapter 4** we highlight recent studies that have tried to disentangle causal relationships between ADHD and BMI [265-267], and between ADHD and educational attainment [268], with limited success. More investigations in this vein are necessary as identification of modifiable causal variables may provide useful avenues for targeted intervention or prevention.

How PGS may be useful for clinical purposes

Improved PGS prediction represents a very important avenue for identifying children at the greatest risk for continuity not just related to psychopathology, but also adverse health and SES outcomes. It is therefore important to begin to consider how PGS could be used should they become viable clinical tools. They are unlikely to have any preventative utility; most treatments for psychiatric disorders cannot be administered prophylactically and moreover, in most cases, genetic risk alone is not sufficient to lead to the development of a disorder. As we show in **chapter 5**, a combination of genetic, behavioural, and environmental factors produced the best performing predictive model of aggression and/or self-harm. However, they could for example be used in the early stages of disorders where other risk factors are present, but a specific diagnosis is not possible [372]. This could be particularly important in disorders for which early intervention may improve long-term outcomes, for example stratifying children into groups according to their risk for persistence, as well as resulting treatment requirements. Should PGS eventually become useful in developing preventative strategies, for example in combination with family history and genetic nurturing effects, ethical concerns regarding screening otherwise healthy children for risk of psychopathology, make it unlikely to be an immediately viable approach. Another important consideration, or even a huge limitation, regarding PGS and any potential future use is that currently, most GWAS are performed in individuals of European ancestry (up to 80% Europeans in studies that include samples of multiple ancestries). Unsurprisingly, studies that have subsequently analysed the predictive value of PGS across different populations have found that prediction of individual risk is better in Europeans than non-Europeans [187]. If PGS became clinically viable biomarkers today, they would likely only be useful for those of European ancestry. We (scientists) have ethical and moral obligations to prioritise diversifying samples in genomic studies to ensure that this does not become another avenue in which social/societal inequalities are further exacerbated.

A holistic approach to understanding aetiology

It is clear that analyses of common variants alone will be limited in sufficiently furthering our understanding of the genetic aetiology of psychopathology. Even in non-psychiatric traits like height where analyses are based on large sample sizes, only a fraction of its 80% twin-based heritability estimate has been captured by measured common variants [33]. While it is possible that this is due to various

methodological factors related to common variant analyses, it is also clear that a proportion of heritability is likely explained by rare variants. In fact, a recent study has used whole genome sequence data to show that for height and BMI, integration of common and rare variant information moved heritability estimates closer to those estimated in twin and family studies [379]. It is possible that this is also the case for psychiatric traits, which in turn has implications for any future genetic-based prediction of psychopathology. This is also in line with findings that show that for disorders such as schizophrenia and ASD, as well as severe neurodevelopmental disorders, both common and rare variants contributed to the risk for these disorders [380-382]. An important caveat to this is that height is a much more precise phenotype to measure than psychiatric disorders, and phenotypic heterogeneity may reflect genotypic heterogeneity which blurs the estimate. All in all, a combination of common and rare variants appears to underlie psychopathology, and an incorporation of both might prove to be the most predictive genetic variable. For example, combining burden scores based on rare variants with PGS from common variants to see if this increases prediction/ variance explained.

Finally, there are epistatic genetic effects as well as environmental interactions at play which underlie disorder aetiology. They are largely unaccounted for in the methods described so far and represent complementary lines of research that should be pursued to ensure more complete picture of childhood psychopathology.

Conclusions

It is clear that much progress has been and is being made with regards to understanding the genetics of childhood psychopathology. An increase in the availability of ethnically diverse research samples at different developmental stages represents a key step to further this, and will aid the identification of trait-associated variants in the first instance. Beyond this, a move from genetic variation to disease mechanisms and biological pathways is crucial, something that gene sets can hint at, but resolution beyond that is also important. As we still observe widespread pleiotropy at that level, it is important to go beyond this as the mechanisms underlying these traits might be different even though the same gene sets and thus pathways are implicated. The hope is that in the coming years, increasing sample sizes along with the integration of genomic and functional approaches will converge onto specific systems, cell types, mechanisms, and

developmental stages that better explain disorder aetiology. Although widespread polygenicity and pleiotropy in psychiatric disorders make translating genetic findings to biological understanding difficult, leveraging both in order to further our understanding of psychopathology, and eventually enhance clinical utility of research findings, will be important in the years to come.

References

1. Polanczyk GV, Salum GA, Sugaya LS, Caye A, Rohde LA. Annual Research Review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of Child Psychology and Psychiatry*. 2015;56(3):345-365.
2. Silva SA, Silva SU, Ronca DB, Gonçalves VSS, Dutra ES, Carvalho KMB. Common mental disorders prevalence in adolescents: A systematic review and meta-analyses. *PloS one*. 2020;15(4):e0232007. doi:10.1371/journal.pone.0232007
3. Erskine H, Moffitt TE, Copeland W, et al. A heavy burden on young minds: the global burden of mental and substance use disorders in children and youth. *Psychological medicine*. 2015;45(7):1551-1563.
4. Baranne ML, Falissard B. Global burden of mental disorders among children aged 5–14 years. *Child and Adolescent Psychiatry and Mental Health*. 2018/04/12 2018;12(1):19. doi:10.1186/s13034-018-0225-4
5. Costello EJ, Maughan B. Annual research review: optimal outcomes of child and adolescent mental illness. *Journal of Child Psychology and Psychiatry*. 2015;56(3):324-341. doi:10.1111/jcpp.12371
6. McLeod GF, Horwood LJ, Fergusson DM. Adolescent depression, adult mental health and psychosocial outcomes at 30 and 35 years. *Psychological medicine*. 2016;46(7):1401.
7. Castagnini AC, Foldager L, Caffo E, Thomsen PH. Early-adult outcome of child and adolescent mental disorders as evidenced by a national-based case register survey. *European Psychiatry*. 2016/10/01/ 2016;38:45-50. doi:https://doi.org/10.1016/j.eurpsy.2016.04.005
8. Asselmann E, Wittchen HU, Lieb R, Beesdo-Baum K. Sociodemographic, clinical, and functional long-term outcomes in adolescents and young adults with mental disorders. *Acta Psychiatrica Scandinavica*. 2018/01/01 2018;137(1):6-17. doi:10.1111/acps.12792
9. Copeland WE, Wolke D, Shanahan L, Costello EJ. Adult functional outcomes of common childhood psychiatric problems: a prospective, longitudinal study. *JAMA psychiatry*. 2015;72(9):892-899.
10. Shevlin M, McElroy E, Murphy J. Homotypic and heterotypic psychopathological continuity: a child cohort study. *Social Psychiatry and Psychiatric Epidemiology*. 05/26 12/17/received 05/17/accepted 2017;52(9):1135-1145. doi:10.1007/s00127-017-1396-7
11. Chronis-Tuscano A, Molina BSG, Pelham WE, et al. Very Early Predictors of Adolescent Depression and Suicide Attempts in Children With Attention-Deficit/Hyperactivity Disorder. *Archives of General Psychiatry*. 2010;67(10):1044-1051. doi:10.1001/archgenpsychiatry.2010.127
12. Meinzer MC, Lewinsohn PM, Pettit JW, et al. Attention Deficit-Hyperactivity Disorder in Adolescence Predicts Onset of Major Depressive Disorder through Early Adulthood. *Depression and anxiety*. 02/19 2013;30(6):546-553. doi:10.1002/da.22082
13. Klein RG, Mannuzza S, Olazagasti MAR, et al. Clinical and Functional Outcome of Childhood Attention-Deficit/Hyperactivity Disorder 33 Years Later. *Archives of General Psychiatry*. 2012;69(12):1295-1303. doi:10.1001/archgenpsychiatry.2012.271
14. Gundel LK, Pedersen CB, Munk-Olsen T, Dalsgaard S. Longitudinal association between mental disorders in childhood and subsequent depression – A nationwide prospective cohort study. *Journal of Affective Disorders*. 2018/02/01/ 2018;227:56-64. doi:https://doi.org/10.1016/j.jad.2017.10.023

15. Essau CA, Lewinsohn PM, Olaya B, Seeley JR. Anxiety disorders in adolescents and psychosocial outcomes at age 30. *Journal of Affective Disorders*. 2014/07/01/ 2014;163:125-132. doi:https://doi.org/10.1016/j.jad.2013.12.033
16. Ranøyen I, Lydersen S, Larose TL, et al. Developmental course of anxiety and depression from adolescence to young adulthood in a prospective Norwegian clinical cohort. *European Child & Adolescent Psychiatry*. 2018/11/01 2018;27(11):1413-1423. doi:10.1007/s00787-018-1139-7
17. Polderman TJC, Boomsma DI, Bartels M, Verhulst FC, Huizink AC. A systematic review of prospective studies on attention problems and academic achievement. *Acta Psychiatrica Scandinavica*. 2010;122(4):271-284. doi:doi:10.1111/j.1600-0447.2010.01568.x
18. Goldman-Mellor S, Gregory AM, Caspi A, et al. Mental health antecedents of early midlife insomnia: evidence from a four-decade longitudinal study. *Sleep*. 2014;37(11):1767-1775.
19. Gordon CT, Fabiano GA. The Transition of Youth with ADHD into the Workforce: Review and Future Directions. *Clin Child Fam Psychol Rev*. 2019/09/01 2019;22(3):316-347. doi:10.1007/s10567-019-00274-4
20. Pérez-Vigil A, Fernández de la Cruz L, Brander G, et al. Association of Obsessive-Compulsive Disorder With Objective Indicators of Educational Attainment: A Nationwide Register-Based Sibling Control Study. *JAMA Psychiatry*. 2018;75(1):47-55. doi:10.1001/jamapsychiatry.2017.3523
21. Melaku YA, Appleton S, Reynolds AC, et al. Association Between Childhood Behavioral Problems and Insomnia Symptoms in Adulthood. *JAMA Netw Open*. Sep 4 2019;2(9):e1910861. doi:10.1001/jamanetworkopen.2019.10861
22. Chen L, Huang Y, Kasen S, Skodol A, Cohen P, Chen H. Impact of Adolescent Personality Disorders on Obesity 17 Years Later. *Psychosomatic Medicine*. 2015;77(8):921-926. doi:10.1097/psy.0000000000000228
23. Rhee SH, Lahey BB, Waldman ID. Comorbidity Among Dimensions of Childhood Psychopathology: Converging Evidence From Behavior Genetics. *Child Development Perspectives*. 2015;9(1):26-31. doi:https://doi.org/10.1111/cdep.12102
24. Connor DF, Steeber J, McBurnett K. A review of attention-deficit/hyperactivity disorder complicated by symptoms of oppositional defiant disorder or conduct disorder. *J Dev Behav Pediatr*. Jun 2010;31(5):427-40. doi:10.1097/DBP.0b013e3181e121bd
25. Morris-Rosendahl DJ, Crocq M-A. Neurodevelopmental disorders-the history and future of a diagnostic concept *Dialogues in clinical neuroscience*. 2020;22(1):65-72. doi:10.31887/DCNS.2020.22.1/macrocq
26. Polderman TJC, Benyamin B, de Leeuw CA, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*. 05/18/online 2015;47:702. doi:10.1038/ng.3285 https://www.nature.com/articles/ng.3285#supplementary-information
27. Visscher PM, Wray NR, Zhang Q, et al. 10 years of GWAS discovery: biology, function, and translation. *The American Journal of Human Genetics*. 2017;101(1):5-22.
28. Maier R, Visscher P, Robinson M, Wray N. Embracing polygenicity: a review of methods and tools for psychiatric genetics research. *Psychological medicine*. 2018;48(7):1055-1067.
29. Martin AR, Daly MJ, Robinson EB, Hyman SE, Neale BM. Predicting polygenic risk of psychiatric disorders. *Biological psychiatry*. 2018;

30. 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):1068-1087. doi:10.1111/jcpp.12295
31. Ripke S, Walters JTR, O'Donovan MC. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *medRxiv*. 2020:2020.09.12.20192922. doi:10.1101/2020.09.12.20192922
32. Howard DM, Adams MJ, Clarke T-K, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature neuroscience*. 2019/03/01 2019;22(3):343-352. doi:10.1038/s41593-018-0326-7
33. Yengo L, Sidorenko J, Kemper KE, et al. Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet*. Oct 15 2018;27(20):3641-3649. doi:10.1093/hmg/ddy271
34. Loos RJF. 15 years of genome-wide association studies and no signs of slowing down. *Nature Communications*. 2020/11/19 2020;11(1):5900. doi:10.1038/s41467-020-19653-5
35. Schizophrenia Working Group of the Psychiatric Genomics C, Ripke S, Neale BM, et al. Biological Insights From 108 Schizophrenia-Associated Genetic Loci. *Nature*. 07/22 2014;511(7510):421-427. doi:10.1038/nature13595
36. Schizophrenia Psychiatric Genome-Wide Association Study C. Genome-wide association study identifies five new schizophrenia loci. *Nature genetics*. 2011;43(10):969-976. doi:10.1038/ng.940
37. Timpson NJ, Greenwood CMT, Soranzo N, Lawson DJ, Richards JB. Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nature Reviews Genetics*. 2018/02/01 2018;19(2):110-124. doi:10.1038/nrg.2017.101
38. Demontis D, Walters RK, Martin J, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature genetics*. 2019;51(1):63.
39. Lichtenstein P, Yip BH, Björk C, et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *The Lancet*. 2009/01/17/ 2009;373(9659):234-239. doi:https://doi.org/10.1016/S0140-6736(09)60072-6
40. Hilker R, Helenius D, Fagerlund B, et al. Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide Danish Twin Register. *Biological Psychiatry*. 2018/03/15/ 2018;83(6):492-498. doi:https://doi.org/10.1016/j.biopsych.2017.08.017
41. Smoller JW, Block SR, Young MM. Genetics of anxiety disorders: the complex road from DSM to DNA. *Depression and anxiety*. 2009;26(11):965-975.
42. Freitag CM, Asherson P, Hebebrand J. Behavioural genetics of childhood disorders. *Current topics in behavioral neurosciences*. 2012;12:395-428. doi:10.1007/7854_2011_178
43. Rice F, Harold G, Thapar A. The genetic aetiology of childhood depression: a review. *Journal of child Psychology and Psychiatry*. 2002;43(1):65-79.
44. Nivard M, Dolan C, Kendler K, et al. Stability in symptoms of anxiety and depression as a function of genotype and environment: a longitudinal twin study from ages 3 to 63 years. *Psychological medicine*. 2015;45(5):1039-1049. doi:10.1017/S003329171400213X
45. Hannigan L, Walaker N, Waszczuk M, McAdams T, Eley T. Aetiological influences on stability and change in emotional and behavioural problems across development: a systematic review. *Psychopathology review*. 2017;4(1):52-108. doi:10.5127/pr.038315

46. Pearson TA, Manolio TA. How to interpret a genome-wide association study. *Jama*. 2008;299(11):1335-1344.
47. Manolio TA. Genomewide association studies and assessment of the risk of disease. *New England journal of medicine*. 2010;363(2):166-176.
48. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *The American Journal of Human Genetics*. 2012;90(1):7-24.
49. Middeldorp CM, Felix JF, Mahajan A, et al. The Early Growth Genetics (EGG) and EArly Genetics and Lifecourse Epidemiology (EAGLE) consortia: design, results and future prospects. *European Journal of Epidemiology*. 2019/03/01 2019;34(3):279-300. doi:10.1007/s10654-019-00502-9
50. Sullivan PF, Agrawal A, Bulik CM, et al. Psychiatric Genomics: An Update and an Agenda. *Am J Psychiatry*. Jan 1 2018;175(1):15-27. doi:10.1176/appi.ajp.2017.17030283
51. Wray NR, Ripke S, Mattheisen M, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet*. May 2018;50(5):668-681. doi:10.1038/s41588-018-0090-3
52. Stahl EA, Breen G, Forstner AJ, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature Genetics*. 2019/05/01 2019;51(5):793-803. doi:10.1038/s41588-019-0397-8
53. Grove J, Ripke S, Als TD, et al. Identification of common genetic risk variants for autism spectrum disorder. *Nature genetics*. 2019;51(3):431.
54. Pardiñas AF, Holmans P, Pocklington AJ, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nature genetics*. 2018;50(3):381-389. doi:10.1038/s41588-018-0059-2
55. The International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. 10.1038/nature08185. *Nature*. 08/06/print 2009;460(7256):748-752. doi:http://www.nature.com/nature/journal/v460/n7256/suppinfo/nature08185_S1.html
56. Bulik-Sullivan BK, Loh P-R, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*. 02/02/online 2015;47:291. doi:10.1038/ng.3211 https://www.nature.com/articles/ng.3211#supplementary-information
57. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *American journal of human genetics*. 2011;88(1):76-82. doi:10.1016/j.ajhg.2010.11.011
58. Sullivan PF. The Psychiatric GWAS Consortium: Big Science Comes to Psychiatry. *Neuron*. 2010;68(2):182-186. doi:10.1016/j.neuron.2010.10.003
59. Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP, Zondervan KT. Basic statistical analysis in genetic case-control studies. *Nature protocols*. 02/03 2011;6(2):121-133. doi:10.1038/nprot.2010.182
60. Slatkin M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nature reviews Genetics*. 2008;9(6):477-485. doi:10.1038/nrg2361
61. Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*. 06/20/online 2010;42:565. doi:10.1038/ng.608 https://www.nature.com/articles/ng.608#supplementary-information

62. Ni G, Moser G, Schizophrenia Working Group of the Psychiatric Genomics C, Wray NR, Lee SH. Estimation of Genetic Correlation via Linkage Disequilibrium Score Regression and Genomic Restricted Maximum Likelihood. *American journal of human genetics*. 2018;102(6):1185-1194. doi:10.1016/j.ajhg.2018.03.021
63. Evans LM, Tahmasbi R, Vrieze SI, et al. Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nature genetics*. 2018;50(5):737-745. doi:10.1038/s41588-018-0108-x
64. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An Atlas of Genetic Correlations across Human Diseases and Traits. *Nature genetics*. 09/28 2015;47(11):1236-1241. doi:10.1038/ng.3406
65. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics (Oxford, England)*. 2012;28(19):2540-2542. doi:10.1093/bioinformatics/bts474
66. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009;339:b2700. doi:10.1136/bmj.b2700
67. Hawi Z, Yates H, Pinar A, et al. A case–control genome-wide association study of ADHD discovers a novel association with the tenascin R (TNR) gene. *Translational psychiatry*. 2018;8(1):284.
68. Martin J, Walters RK, Demontis D, et al. A genetic investigation of sex bias in the prevalence of attention-deficit/hyperactivity disorder. *Biological psychiatry*. 2018;83(12):1044-1053.
69. Rovira P, Demontis D, Sánchez-Mora C, et al. Shared genetic background between children and adults with attention deficit/hyperactivity disorder. *Neuropsychopharmacology*. 2020/09/01 2020;45(10):1617-1626. doi:10.1038/s41386-020-0664-5
70. Anney R, Klei L, Pinto D, et al. A genome-wide scan for common alleles affecting risk for autism. *Human molecular genetics*. 2010;19(20):4072-4082.
71. Lu ATH, Cantor RM. Allowing for sex differences increases power in a GWAS of multiplex Autism families. *Molecular Psychiatry*. 2012/02/01 2012;17(2):215-222. doi:10.1038/mp.2010.127
72. Liu X, Shimada T, Otowa T, et al. Genome-wide Association Study of Autism Spectrum Disorder in the East Asian Populations. *Autism Research*. 2016;9(3):340-349. doi:10.1002/aur.1536
73. Watson HJ, Yilmaz Z, Thornton LM, et al. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nature Genetics*. 2019/08/01 2019;51(8):1207-1214. doi:10.1038/s41588-019-0439-2
74. Yu D, Sul JH, Tsetsos F, et al. Interrogating the Genetic Determinants of Tourette's Syndrome and Other Tic Disorders Through Genome-Wide Association Studies. *Am J Psychiatry*. Mar 1 2019;176(3):217-227. doi:10.1176/appi.ajp.2018.18070857
75. St Pourcain B, Whitehouse AO, Ang WQ, et al. Common variation contributes to the genetic architecture of social communication traits. *Molecular autism*. 2013;4(1):34.
76. St Pourcain B, Skuse DH, Mandy WP, et al. Variability in the common genetic architecture of social-communication spectrum phenotypes during childhood and adolescence. *Molecular autism*. 2014;5(1):18.

77. Cantor RM, Navarro L, Won H, Walker RL, Lowe JK, Geschwind DH. ASD restricted and repetitive behaviors associated at 17q21. 33: genes prioritized by expression in fetal brains. *Molecular psychiatry*. 2018;23(4):993.
78. Sallis H, Evans J, Wootton R, et al. Genetics of depressive symptoms in adolescence. *BMC psychiatry*. 2017;17(1):321. doi:10.1186/s12888-017-1484-y
79. Pain O, Dudbridge F, Cardno AG, et al. Genome-wide analysis of adolescent psychotic-like experiences shows genetic overlap with psychiatric disorders. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. 2018;177(4):416-425.
80. Zavos HMS, Freeman D, Haworth CMA, et al. Consistent Etiology of Severe, Frequent Psychotic Experiences and Milder, Less Frequent Manifestations: A Twin Study of Specific Psychotic Experiences in Adolescence. *JAMA Psychiatry*. 2014;71(9):1049-1057. doi:10.1001/jamapsychiatry.2014.994
81. Neumann A, Pappa I, Lahey BB, et al. Single nucleotide polymorphism heritability of a general psychopathology factor in children. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2016;55(12):1038-1045. e4.
82. Cheesman R, Purves KL, Pingault J-B, Breen G, Plomin R, Eley TC. Extracting stability increases the SNP heritability of emotional problems in young people. *Translational psychiatry*. 2018;8(1):223.
83. Riglin L, Eyre O, Cooper M, et al. Investigating the genetic underpinnings of early-life irritability. *Translational psychiatry*. 2017;7(9):e1241.
84. Martin J, Taylor MJ, Rydell M, et al. Sex-specific manifestation of genetic risk for attention deficit hyperactivity disorder in the general population. *Journal of Child Psychology and Psychiatry*. 2018;59(8):908-916.
85. Stergiakouli E, Smith GD, Martin J, et al. Shared genetic influences between dimensional ASD and ADHD symptoms during child and adolescent development. *Molecular autism*. 2017;8(1):18. doi:10.1186/s13229-017-0131-2
86. Riglin L, Thapar AK, Leppert B, et al. Using Genetics to Examine a General Liability to Childhood Psychopathology. *Behavior Genetics*. 2019/12/11 2019;doi:10.1007/s10519-019-09985-4
87. Yao S, Kuja-Halkola R, Martin J, et al. Associations Between Attention-Deficit/Hyperactivity Disorder and Various Eating Disorders: A Swedish Nationwide Population Study Using Multiple Genetically Informative Approaches. *Biological Psychiatry*. 2019;86(8):577-586. doi:10.1016/j.biopsych.2019.04.036
88. Serdarevic F, Tiemeier H, Jansen PR, et al. Polygenic Risk Scores for Developmental Disorders, Neuromotor Functioning During Infancy, and Autistic Traits in Childhood. *Biological Psychiatry*. 2020/01/15/ 2020;87(2):132-138. doi:https://doi.org/10.1016/j.biopsych.2019.06.006
89. Chang S, Yang L, Wang Y, Faraone SV. Shared polygenic risk for ADHD, executive dysfunction and other psychiatric disorders. *Translational Psychiatry*. 2020/06/09 2020;10(1):182. doi:10.1038/s41398-020-00872-9
90. Rai D, Culpin I, Heuvelman H, et al. Association of Autistic Traits With Depression From Childhood to Age 18 Years. *JAMA Psychiatry*. 2018;75(8):835-843. doi:10.1001/jamapsychiatry.2018.1323
91. Brikell I, Larsson H, Lu Y, et al. The contribution of common genetic risk variants for ADHD to a general factor of childhood psychopathology. *Molecular Psychiatry*. 2020/08/01 2020;25(8):1809-1821. doi:10.1038/s41380-018-0109-2

92. Jansen PR, Polderman TJC, Bolhuis K, et al. 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*. 2018;59(1):39-47. doi:10.1111/jcpp.12759
93. Jansen AG, Dieleman GC, Jansen PR, Verhulst FC, Posthuma D, Polderman TJC. Psychiatric Polygenic Risk Scores as Predictor for Attention Deficit/Hyperactivity Disorder and Autism Spectrum Disorder in a Clinical Child and Adolescent Sample. *Behavior Genetics*. 2019/07/25 2019;doi:10.1007/s10519-019-09965-8
94. Vuijk PJ, Martin J, Braaten EB, et al. Translating Discoveries in Attention-Deficit/Hyperactivity Disorder Genomics to an Outpatient Child and Adolescent Psychiatric Cohort. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2019/08/14/ 2019;doi:https://doi.org/10.1016/j.jaac.2019.08.004
95. Nigg JT, Karalunas SL, Gustafsson HC, et al. Evaluating chronic emotional dysregulation and irritability in relation to ADHD and depression genetic risk in children with ADHD. *Journal of Child Psychology and Psychiatry*. 2020/02/01 2020;61(2):205-214. doi:10.1111/jcpp.13132
96. Ahn K, An SS, Shugart YY, Rapoport JL. Common polygenic variation and risk for childhood-onset schizophrenia. *Molecular Psychiatry*. 2016/01/01 2016;21(1):94-96. doi:10.1038/mp.2014.158
97. Lee PH, Anttila V, Won H, et al. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell*. 2019/12/12/ 2019;179(7):1469-1482. e11. doi:https://doi.org/10.1016/j.cell.2019.11.020
98. Anttila V, Bulik-Sullivan B, Finucane HK, et al. Analysis of shared heritability in common disorders of the brain. *Science*. 2018;360(6395):eaap8757. doi:10.1126/science.aap8757
99. Hamshere ML, Langley K, Martin J, et al. High loading of polygenic risk for ADHD in children with comorbid aggression. *American journal of psychiatry*. 2013;170(8):909-916.
100. Martin J, Hamshere ML, Stergiakouli E, O'Donovan MC, Thapar A. Genetic risk for attention-deficit/hyperactivity disorder contributes to neurodevelopmental traits in the general population. *Biological psychiatry*. 2014;76(8):664-671.
101. Elam KK, Clifford S, Shaw DS, Wilson MN, Lemery-Chalfant K. Gene set enrichment analysis to create polygenic scores: a developmental examination of aggression. *Translational Psychiatry*. 2019/09/02 2019;9(1):212. doi:10.1038/s41398-019-0513-7
102. Riglin L, Collishaw S, Thapar AK, et al. Association of Genetic Risk Variants With Attention-Deficit/Hyperactivity Disorder Trajectories in the General Population. *JAMA Psychiatry*. 2016;73(12):1285-1292. doi:10.1001/jamapsychiatry.2016.2817
103. Manfro AG, Santoro M, Polanczyk GV, et al. Heterotypic trajectories of dimensional psychopathology across the lifespan: the case of youth-onset attention deficit/hyperactivity disorder. *Journal of Child Psychology and Psychiatry*. 2019;60(5):533-544.
104. Lucy Riglin, Ph.D. ,, Olga Eyre, M.R.C.Psych .,, Ajay K. Thapar, M.R.C.G.P. , Ph.D. ,, et al. Identifying Novel Types of Irritability Using a Developmental Genetic Approach. *American Journal of Psychiatry*. 2019;176(8):635-642. doi:10.1176/appi.ajp.2019.18101134
105. Shaw DS, Galán CA, Lemery-Chalfant K, et al. Trajectories and Predictors of Children's Early-Starting Conduct Problems: Child, Family, Genetic, and Intervention Effects. *Development and Psychopathology*. 2019;31(5):1911-1921. doi:10.1017/S0954579419000828

106. Taylor MJ, Martin J, Lu Y, et al. Association of genetic risk factors for psychiatric disorders and traits of these disorders in a Swedish population twin sample. *JAMA psychiatry*. 2018;
107. Jones HJ, Stergiakouli E, Tansey KE, et al. Phenotypic Manifestation of Genetic Risk for Schizophrenia During Adolescence in the General Population. *JAMA psychiatry*. 2016;73(3):221-228. doi:10.1001/jamapsychiatry.2015.3058
108. Jones HJ, Heron J, Hammerton G, et al. Investigating the genetic architecture of general and specific psychopathology in adolescence. *Translational Psychiatry*. 2018/08/08 2018;8(1):145. doi:10.1038/s41398-018-0204-9
109. Velthorst E, Froudist-Walsh S, Stahl E, et al. Genetic risk for schizophrenia and autism, social impairment and developmental pathways to psychosis. *Translational psychiatry*. 2018;8(1):204-204. doi:10.1038/s41398-018-0229-0
110. Pries LK, Klingenberg B, Menne-Lothmann C, et al. Polygenic liability for schizophrenia and childhood adversity influences daily-life emotion dysregulation and psychosis proneness. *Acta psychiatrica Scandinavica*. 2020;141(5):465-475. doi:10.1111/acps.13158
111. Musci RJ, Masyn KE, Benke K, Maher B, Uhl G, Ialongo NS. The effects of the interplay of genetics and early environmental risk on the course of internalizing symptoms from late childhood through adolescence. *Development and psychopathology*. 2016;28(1):225-237.
112. Halldorsdottir T, Piechaczek C, Soares de Matos AP, et al. Polygenic Risk: Predicting Depression Outcomes in Clinical and Epidemiological Cohorts of Youths. *American Journal of Psychiatry*. 2019:appi. ajp. 2019.18091014.
113. Musliner KL, Mortensen PB, McGrath JJ, et al. Association of Polygenic Liabilities for Major Depression, Bipolar Disorder, and Schizophrenia With Risk for Depression in the Danish Population. *JAMA psychiatry*. 2019;76(5):516-525.
114. Salvatore JE, Aliev F, Bucholz K, et al. Polygenic risk for externalizing disorders: Gene-by-development and gene-by-environment effects in adolescents and young adults. *Clinical Psychological Science*. 2015;3(2):189-201.
115. Akingbuwa WA, Hammerschlag AR, Jami ES, et al. Genetic Associations Between Childhood Psychopathology and Adult Depression and Associated Traits in 42998 Individuals: A Meta-Analysis. *JAMA Psychiatry*. Apr 15 2020;doi:10.1001/jamapsychiatry.2020.0527
116. Su J, Kuo SIC, Bucholz KK, et al. Understanding Mechanisms of Genetic Risk for Adolescent Internalizing and Externalizing Problems: The Mediating Role of Parenting and Personality. *Twin Research and Human Genetics*. 2018;21(4):310-321. doi:10.1017/thg.2018.36
117. Bralten J, Widomska J, Witte WD, et al. Shared genetic etiology between obsessive-compulsive disorder, obsessive-compulsive symptoms in the population, and insulin signaling. *Translational Psychiatry*. 2020/04/27 2020;10(1):121. doi:10.1038/s41398-020-0793-y
118. Taylor JH, Asabere N, Calkins ME, et al. Characteristics of youth with reported family history of psychosis spectrum symptoms in the Philadelphia Neurodevelopmental Cohort. *Schizophrenia Research*. 2019/12/26/ 2019;doi:https://doi.org/10.1016/j.schres.2019.12.021
119. Perkins DO, Olde Loohuis L, Barbee J, et al. Polygenic Risk Score Contribution to Psychosis Prediction in a Target Population of Persons at Clinical High Risk. *Am J Psychiatry*. Nov 12 2020;177(2):155-163. doi:10.1176/appi.ajp.2019.18060721

120. Zammit S, Hamshere M, Dwyer S, et al. A population-based study of genetic variation and psychotic experiences in adolescents. *Schizophrenia bulletin*. 2013;40(6):1254-1262.
121. Sieradzka D, Power RA, Freeman D, et al. Are genetic risk factors for psychosis also associated with dimension-specific psychotic experiences in adolescence? *PloS one*. 2014;9(4):e94398-e94398. doi:10.1371/journal.pone.0094398
122. Riglin L, Collishaw S, Richards A, et al. The impact of schizophrenia and mood disorder risk alleles on emotional problems: investigating change from childhood to middle age. *Psychological medicine*. 2018;48(13):2153-2158.
123. Ensink JBM, de Moor MHM, Zafarmand MH, et al. Maternal environmental risk factors and the development of internalizing and externalizing problems in childhood: The complex role of genetic factors. *American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*. 2020;183(1):17-25. doi:10.1002/ajmg.b.32755
124. Hannigan LJ, Pingault JB, Krapohl E, McAdams TA, Rijsdijk FV, Eley TC. Genetics of co-developing conduct and emotional problems during childhood and adolescence. *Nat Hum Behav*. Jul 2018;2(7):514-521. doi:10.1038/s41562-018-0373-9
125. Kwong ASF, López-López JA, Hammerton G, et al. Genetic and Environmental Risk Factors Associated With Trajectories of Depression Symptoms From Adolescence to Young Adulthood. *JAMA Netw Open*. 2019;2(6):e196587-e196587. doi:10.1001/jamanetworkopen.2019.6587
126. Rice F, Riglin L, Thapar AK, et al. Characterizing developmental trajectories and the role of neuropsychiatric genetic risk variants in early-onset depression. *JAMA psychiatry*. 2018;
127. Fullerton JM, Koller DL, Edenberg HJ, et al. Assessment of first and second degree relatives of individuals with bipolar disorder shows increased genetic risk scores in both affected relatives and young At-Risk Individuals. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. 2015;168(7):617-629.
128. Wilcox HC, Fullerton JM, Glowinski AL, et al. Traumatic stress interacts with bipolar disorder genetic risk to increase risk for suicide attempts. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2017;56(12):1073-1080.
129. Hamshere ML, Stergiakouli E, Langley K, et al. Shared polygenic contribution between childhood attention-deficit hyperactivity disorder and adult schizophrenia. *The British Journal of Psychiatry*. 2013;203(2):107-111. doi:10.1192/bjp.bp.112.117432
130. Costas J, Carrera N, Alonso P, et al. Exon-focused genome-wide association study of obsessive-compulsive disorder and shared polygenic risk with schizophrenia. *Translational psychiatry*. 2016;6(3):e768.
131. Solmi F, Mascarell MC, Zammit S, Kirkbride JB, Lewis G. Polygenic risk for schizophrenia, disordered eating behaviours and body mass index in adolescents. *The British Journal of Psychiatry*. 2019:1-6.
132. Riglin L, Collishaw S, Richards A, et al. Schizophrenia risk alleles and neurodevelopmental outcomes in childhood: a population-based cohort study. *The Lancet Psychiatry*. 2017;4(1):57-62.
133. Cross-Disorder Group of the Psychiatric Genomics C. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet*. 02/28 2013;381(9875):1371-1379. doi:10.1016/S0140-6736(12)62129-1

134. Nivard MG, Gage SH, Hottenga JJ, et al. Genetic Overlap Between Schizophrenia and Developmental Psychopathology: Longitudinal and Multivariate Polygenic Risk Prediction of Common Psychiatric Traits During Development. *Schizophrenia Bulletin*. 2017:sbx031.
135. Lucy Riglin, Ph.D. ,, Gemma Hammerton, Ph.D. ,, Jon Heron, Ph.D. ,, et al. Developmental Contributions of Schizophrenia Risk Alleles and Childhood Peer Victimization to Early-Onset Mental Health Trajectories. *American Journal of Psychiatry*. 2019;176(1):36-43. doi:10.1176/appi.ajp.2018.18010075
136. Sudre G, Frederick J, Sharp W, et al. Mapping associations between polygenic risks for childhood neuropsychiatric disorders, symptoms of attention deficit hyperactivity disorder, cognition, and the brain. *Molecular psychiatry*. 2019:1.
137. 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*. 2019/08/01/ 2019;105(2):351-363. doi:https://doi.org/10.1016/j.ajhg.2019.06.006
138. Burton CL, Wright L, Shan J, et al. SWAN scale for ADHD trait-based genetic research: a validity and polygenic risk study. *Journal of Child Psychology and Psychiatry*. 2019;
139. Abdulkadir M, Mathews CA, Scharf JM, et al. Polygenic risk scores derived from a Tourette syndrome genome-wide association study predict presence of tics in the Avon Longitudinal Study of Parents and Children cohort. *Biological psychiatry*. 2019;85(4):298-304.
140. Krapohl E, Euesden J, Zabaneh D, et al. Phenome-wide analysis of genome-wide polygenic scores. *Molecular Psychiatry*. 08/25 05/20/received 07/06/revised 07/14/accepted 2016;21(9):1188-1193. doi:10.1038/mp.2015.126
141. Mistry S, Escott-Price V, Florio AD, Smith DJ, Zammit S. Genetic risk for bipolar disorder and psychopathology from childhood to early adulthood. *Journal of affective disorders*. 2019;246:633-639.
142. Yu D, Mathews CA, Scharf JM, et al. Cross-disorder genome-wide analyses suggest a complex genetic relationship between Tourette's syndrome and OCD. *The American journal of psychiatry*. 2015;172(1):82-93. doi:10.1176/appi.ajp.2014.13101306
143. Du Rietz E, Coleman J, Glanville K, Choi SW, O'Reilly PF, Kuntsi J. Association of polygenic risk for attention-deficit/hyperactivity disorder with co-occurring traits and disorders. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. 2018;3(7):635-643. doi:10.1016/j.bpsc.2017.11.013
144. Li JJ. Assessing phenotypic and polygenic models of ADHD to identify mechanisms of risk for longitudinal trajectories of externalizing behaviors. *Journal of Child Psychology and Psychiatry*. 2019/11/01 2019;60(11):1191-1199. doi:10.1111/jcpp.13071
145. Grigoroiu-Serbanescu M, Giaroli G, Thygesen JH, et al. Predictive power of the ADHD GWAS 2019 polygenic risk scores in independent samples of bipolar patients with childhood ADHD. *Journal of Affective Disorders*. 2020/03/15/ 2020;265:651-659. doi:https://doi.org/10.1016/j.jad.2019.11.109
146. Allegrini AG, Cheesman R, Rimfeld K, et al. The p factor: genetic analyses support a general dimension of psychopathology in childhood and adolescence. *Journal of child psychology and psychiatry, and allied disciplines*. 2020;61(1):30-39. doi:10.1111/jcpp.13113
147. St Pourcain B, Robinson EB, Anttila V, et al. ASD and schizophrenia show distinct developmental profiles in common genetic overlap with population-based social communication difficulties. *Molecular Psychiatry*. 2018/02/01 2018;23(2):263-270. doi:10.1038/mp.2016.198

148. St Pourcain B, Haworth CMA, Davis OSP, et al. Heritability and genome-wide analyses of problematic peer relationships during childhood and adolescence. *Human genetics*. 2015;134(6):539-551. doi:10.1007/s00439-014-1514-5
149. Cross-Disorder Group of the Psychiatric Genomics C. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Analysis. *Nat Genet*. 09//print 2013;45(9):984-994. doi:10.1038/ng.2711 http://www.nature.com/ng/journal/v45/n9/abs/ng.2711.html#supplementary-information
150. Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24. 32 and a significant overlap with schizophrenia. *Molecular autism*. 2017;8:1-17.
151. O'Connell KS, Shadrin A, Bahrami S, et al. Identification of genetic overlap and novel risk loci for attention-deficit/hyperactivity disorder and bipolar disorder. *Molecular Psychiatry*. 2019/12/02 2019;doi:10.1038/s41380-019-0613-z
152. Ohi K, Otowa T, Shimada M, Sasaki T, Tanii H. Shared genetic etiology between anxiety disorders and psychiatric and related intermediate phenotypes. *Psychological Medicine*. 2020;50(4):692-704. doi:10.1017/S003329171900059X
153. Barkhuizen W, Pain O, Dudbridge F, Ronald A. Genetic overlap between psychotic experiences in the community across age and with psychiatric disorders. *Translational Psychiatry*. 2020/03/09 2020;10(1):86. doi:10.1038/s41398-020-0765-2
154. Ormel J, Hartman CA, Snieder H. The genetics of depression: successful genome-wide association studies introduce new challenges. *Translational Psychiatry*. 2019/03/15 2019;9(1):114. doi:10.1038/s41398-019-0450-5
155. Ip HF, van der Laan CM, Brikell I, et al. Genetic Association Study of Childhood Aggression across raters, instruments and age. *bioRxiv*. 2019:854927. doi:10.1101/854927
156. Hendriks AM, Ip HF, Nivard MG, et al. Content, diagnostic, correlational, and genetic similarities between common measures of childhood aggressive behaviors and related psychiatric traits. *Journal of Child Psychology and Psychiatry*. 2020/02/20 2020;n/a(n/a) doi:10.1111/jcpp.13218
157. Hegvik T-A, Waløen K, Pandey SK, Faraone SV, Haavik J, Zayats T. Druggable genome in attention deficit/hyperactivity disorder and its co-morbid conditions. New avenues for treatment. *Molecular Psychiatry*. 2019/10/18 2019;doi:10.1038/s41380-019-0540-z
158. Wainschtein P, Jain DP, Yengo L, et al. Recovery of trait heritability from whole genome sequence data. *bioRxiv*. 2019:588020.
159. Sanders SJ, He X, Willsey AJ, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron*. 2015;87(6):1215-1233. doi:10.1016/j.neuron.2015.09.016
160. Cappi C, Brentani H, Lima L, et al. Whole-exome sequencing in obsessive-compulsive disorder identifies rare mutations in immunological and neurodevelopmental pathways. Original Article. *Translational Psychiatry*. 03/29/online 2016;6:e764. doi:10.1038/tp.2016.30 https://www.nature.com/articles/tp201630#supplementary-information
161. Williams NM, Franke B, Mick E, et al. Genome-wide analysis of copy number variants in attention deficit hyperactivity disorder: the role of rare variants and duplications at 15q13.3. *The American journal of psychiatry*. 2012;169(2):195-204. doi:10.1176/appi.ajp.2011.11060822

162. Niarchou M, Chawner SJRA, Doherty JL, et al. Psychiatric disorders in children with 16p11.2 deletion and duplication. *Translational Psychiatry*. 2019/01/16 2019;9(1):8. doi:10.1038/s41398-018-0339-8
163. Chawner SJ, Owen MJ, Holmans P, et al. Genotype–phenotype associations in children with copy number variants associated with high neuropsychiatric risk in the UK (IMAGINE-ID): a case-control cohort study. *The Lancet Psychiatry*. 2019;6(6):493-505.
164. Caspi A, Houts RM, Belsky DW, et al. The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clin Psychol Sci*. 2014;2(2):119-137. doi:10.1177/2167702613497473
165. Selzam S, Coleman JRI, Caspi A, Moffitt TE, Plomin R. A polygenic p factor for major psychiatric disorders. *Translational Psychiatry*. 2018/10/02 2018;8(1):205. doi:10.1038/s41398-018-0217-4
166. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. *Nature reviews Genetics*. 06/11 2013;14(7):483-495. doi:10.1038/nrg3461
167. Waszczuk MA, Eaton NR, Krueger RF, et al. Redefining phenotypes to advance psychiatric genetics: Implications from hierarchical taxonomy of psychopathology. *Journal of Abnormal Psychology*. 2019:No Pagination Specified-No Pagination Specified. doi:10.1037/abn0000486
168. Sullivan PF, Geschwind DH. Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. *Cell*. 2019/03/21/ 2019;177(1):162-183. doi:https://doi.org/10.1016/j.cell.2019.01.015
169. Du Rietz E, Pettersson E, Brikell I, et al. Overlap between attention-deficit hyperactivity disorder and neurodevelopmental, externalising and internalising disorders: separating unique from general psychopathology effects. *The British Journal of Psychiatry*. 2020:1-8. doi:10.1192/bjp.2020.152
170. Jami ES, Hammerschlag AR, Ip HF, et al. Genome-wide association meta-analysis of childhood and adolescent internalising symptoms. *medRxiv*. 2020:2020.09.11.20175026. doi:10.1101/2020.09.11.20175026
171. Burton CL, Lemire M, Xiao B, et al. Genome-wide Association Study of Pediatric Obsessive-Compulsive Traits: Shared Genetic Risk between Traits and Disorder. *bioRxiv*. 2020:858241. doi:10.1101/858241
172. Neumann A, Nolte IM, Pappa I, et al. A genome-wide association study of total child psychiatric problems scores. *medRxiv*. 2020:2020.06.04.20121061. doi:10.1101/2020.06.04.20121061
173. Yang Z, Wu H, Lee PH, et al. Cross-disorder GWAS meta-analysis for Attention Deficit/Hyperactivity Disorder, Autism Spectrum Disorder, Obsessive Compulsive Disorder, and Tourette Syndrome. *bioRxiv*. 2019:770222. doi:10.1101/770222
174. Powell V, Martin J, Thapar A, Rice F, Anney RJL. Investigating regions of shared genetic variation in attention deficit/hyperactivity disorder and major depressive disorder: A GWAS meta-analysis. *bioRxiv*. 2020:2020.04.22.054908. doi:10.1101/2020.04.22.054908
175. Loohuis LO, Mennigen E, Ori A, et al. Genetic and clinical analyses of psychosis spectrum symptoms in a large multi-ethnic youth cohort reveal significant link with ADHD. *bioRxiv*. 2019:814087. doi:10.1101/814087

176. Demontis D, Walters R, Rajagopal VM, et al. Identification of risk variants and characterization of the polygenic architecture of disruptive behavior disorders in the context of ADHD. *bioRxiv*. 2019:791160. doi:10.1101/791160
177. López MLR, Franke B, Klein M. Localizing regions in the genome contributing to ADHD, aggressive and antisocial behavior. *bioRxiv*. 2019:750091. doi:10.1101/750091
178. Hübel C, Abdulkadir M, Herle M, et al. Binge-eating disorder, anorexia nervosa, and constitutional thinness differ in their associations with anthropometric and psychiatric polygenic scores. *medRxiv*. 2020:2020.03.24.20042648. doi:10.1101/2020.03.24.20042648
179. Palmer CE, Loughnan RJ, Makowski C, et al. Delineating genetic and familial risk for psychopathology in the ABCD study. *medRxiv*. 2020:2020.09.08.20186908. doi:10.1101/2020.09.08.20186908
180. Rabinowitz JA, Campos AI, Benjet C, et al. Depression polygenic scores are associated with major depressive disorder diagnosis and depressive episode in Mexican adolescents. *medRxiv*. 2020:2020.05.31.20118711. doi:10.1101/2020.05.31.20118711
181. Martin J, Khramtsova EA, Goleva SB, et al. Examining sex-differentiated genetic effects across neuropsychiatric and behavioral traits. *bioRxiv*. 2020:2020.05.04.076042. doi:10.1101/2020.05.04.076042
182. Lussier AA, Hawrilenko M, Wang M-J, et al. Genetic susceptibility for major depressive disorder associates with trajectories of depressive symptoms across childhood and adolescence. *medRxiv*. 2020:2020.03.02.19007088. doi:10.1101/2020.03.02.19007088
183. Yilmaz Z, Schaumberg K, Halvorsen M, et al. Predicting Eating Disorder and Anxiety Symptoms Using Anorexia Nervosa and Obsessive-Compulsive Disorder Polygenic Scores. *medRxiv*. 2020:2020.07.02.20142844. doi:10.1101/2020.07.02.20142844
184. Karcher NR, Paul SE, Johnson EC, et al. Psychotic-like experiences and polygenic liability in the ABCD Study. *medRxiv*. 2020:2020.07.14.20153551. doi:10.1101/2020.07.14.20153551
185. Hannigan LJ, Askeland RB, Ask H, et al. Genetic liability for schizophrenia and childhood psychopathology in the general population. *medRxiv*. 2020:2020.05.01.20086215. doi:10.1101/2020.05.01.20086215
186. Duncan L, Shen H, Gelaye B, et al. Analysis of polygenic risk score usage and performance in diverse human populations. *Nature communications*. 2019;10(1):1-9.
187. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*. 2019/04/01 2019;51(4):584-591. doi:10.1038/s41588-019-0379-x
188. Fusar-Poli P, Bauer M, Borgwardt S, et al. European college of neuropsychopharmacology network on the prevention of mental disorders and mental health promotion (ECNP PMD-MHP). *European Neuropsychopharmacology*. 2019/12/01/ 2019;29(12):1301-1311. doi:https://doi.org/10.1016/j.euroneuro.2019.09.006
189. Lavoie S, Allott K, Amminger P, et al. Harmonised collection of data in youth mental health: Towards large datasets. *Aust N Z J Psychiatry*. Jan 2020;54(1):46-56. doi:10.1177/0004867419844322
190. Kim-Cohen J, Caspi A, Moffitt TE, Harrington H, Milne BJ, Poulton R. Prior juvenile diagnoses in adults with mental disorder: developmental follow-back of a prospective-longitudinal cohort. *Archives of general psychiatry*. 2003;60(7):709-717.
191. Rao U, Chen L-A. Characteristics, correlates, and outcomes of childhood and adolescent depressive disorders. *Dialogues in Clinical Neuroscience*. 2009;11(1):45-62.

192. Biederman J, Ball SW, Monuteaux MC, et al. New Insights Into the Comorbidity Between ADHD and Major Depression in Adolescent and Young Adult Females. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2008/04/01/ 2008;47(4):426-434. doi:http://dx.doi.org/10.1097/CHI.0b013e31816429d3
193. Loth AK, Drabick DA, Leibenluft E, Hulvershorn LA. Do childhood externalizing disorders predict adult depression? A meta-analysis. *J Abnorm Child Psychol*. 2014;42(7):1103-1113.
194. Erickson J, El-Gabalawy R, Palitsky D, et al. Educational attainment as a protective factor for psychiatric disorders: findings from a nationally representative longitudinal study. *Depression and anxiety*. 2016;33(11):1013-1022.
195. Breslau J, Lane M, Sampson N, Kessler RC. Mental disorders and subsequent educational attainment in a US national sample. *Journal of Psychiatric Research*. 2008/07/01/ 2008;42(9):708-716. doi:http://dx.doi.org/10.1016/j.jpsychires.2008.01.016
196. Riemann D. Insomnia and comorbid psychiatric disorders. *Sleep Medicine*. 2007;8:S15-S20.
197. Bartels M, Cacioppo JT, van Beijsterveldt TC, Boomsma DI. Exploring the association between well-being and psychopathology in adolescents. *Behavior genetics*. 2013;43(3):177-190.
198. Kendler KS, Gatz M, Gardner CO, Pedersen NL. Personality and major depression: a Swedish longitudinal, population-based twin study. *Archives of general psychiatry*. 2006;63(10):1113-1120.
199. Rosenström T, Gjerde LC, Krueger RF, et al. Joint factorial structure of psychopathology and personality. *Psychological medicine*. 2018:1-10.
200. Aldinger M, Stopsack M, Ulrich I, et al. Neuroticism developmental courses-implications for depression, anxiety and everyday emotional experience; a prospective study from adolescence to young adulthood. *BMC psychiatry*. 2014;14(1):210.
201. Newton-Howes G, Horwood J, Mulder R. Personality characteristics in childhood and outcomes in adulthood: findings from a 30 year longitudinal study. *Australian & New Zealand Journal of Psychiatry*. 2015;49(4):377-386.
202. Hasler G, Pine D, Gamma A, et al. The associations between psychopathology and being overweight: a 20-year prospective study. *Psychological medicine*. 2004;34(6):1047-1057.
203. Anderson SE, Cohen P, Naumova EN, Must A. Association of depression and anxiety disorders with weight change in a prospective community-based study of children followed up into adulthood. *Archives of pediatrics & adolescent medicine*. 2006;160(3):285-291.
204. Fuemmeler BF, Østbye T, Yang C, McClernon FJ, Kollins SH. Association between Attention-Deficit/Hyperactivity Disorder (ADHD) symptoms and obesity and hypertension in early adulthood: A population-based study. *International journal of obesity (2005)*. 2011;35(6):852.
205. Jansen PR, Watanabe K, Stringer S, et al. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nature Genetics*. 2019:1. doi:10.1038/s41588-018-0333-3
206. Kan K-J, Dolan CV, Nivard MG, et al. Genetic and environmental stability in attention problems across the lifespan: evidence from the Netherlands twin register. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2013;52(1):12-25. doi:10.1016/j.jaac.2012.10.009

207. Birmaher B, Axelson D, Goldstein B, et al. Psychiatric disorders in preschool offspring of parents with bipolar disorder: the Pittsburgh Bipolar Offspring Study (BIOS). *American Journal of Psychiatry*. 2010;167(3):321-330.
208. Hillegers MH, Reichart CG, Wals M, Verhulst FC, Ormel J, Nolen WA. Five-year prospective outcome of psychopathology in the adolescent offspring of bipolar parents. *Bipolar disorders*. 2005;7(4):344-350.
209. Mesman E, Nolen WA, Reichart CG, Wals M, Hillegers MH. The Dutch bipolar offspring study: 12-year follow-up. *American Journal of Psychiatry*. 2013;170(5):542-549.
210. Riglin L, Collishaw S, Richards A, et al. The impact of schizophrenia and mood disorder risk alleles on emotional problems: investigating change from childhood to middle age. *Psychological medicine*. 2017:1-6.
211. de Zeeuw EL, van Beijsterveldt CE, Glasner TJ, et al. Polygenic scores associated with educational attainment in adults predict educational achievement and ADHD symptoms in children. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. 2014;165(6):510-520. doi:10.1002/ajmg.b.32254
212. Stergiakouli E, Martin J, Hamshere ML, et al. Association between polygenic risk scores for attention-deficit hyperactivity disorder and educational and cognitive outcomes in the general population. *International journal of epidemiology*. 2017;46(2):421-428. doi:10.1093/ije/dyw216
213. Lee JJ, Wedow R, Okbay A, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*. 2018/08/01 2018;50(8):1112-1121. doi:10.1038/s41588-018-0147-3
214. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS genetics*. 2013;9(3):e1003348.
215. Okbay A, Baselmans BML, De Neve J-E, et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. Article. *Nat Genet*. 06//print 2016;48(6):624-633. doi:10.1038/ng.3552 http://www.nature.com/ng/journal/v48/n6/abs/ng.3552.html#supplementary-information
216. Hammerschlag AR, Stringer S, de Leeuw CA, et al. Genome-wide association analysis of insomnia complaints identifies risk genes and genetic overlap with psychiatric and metabolic traits. Article. *Nature Genetics*. 06/12/online 2017;49:1584. doi:10.1038/ng.3888 https://www.nature.com/articles/ng.3888#supplementary-information
217. Vilhjálmsson BJ, Yang J, Finucane HK, et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *The American Journal of Human Genetics*. 2015;97(4):576-592. doi:10.1016/j.ajhg.2015.09.001
218. Goodman R. The Strengths and Difficulties Questionnaire: a research note. *Journal of child psychology and psychiatry*. 1997;38(5):581-586. doi:10.1111/j.1469-7610.1997.tb01545.x
219. Larson T, Anckarsäter H, Gillberg C, et al. The autism-tics, AD/HD and other comorbidities inventory (A-TAC): further validation of a telephone interview for epidemiological research. *BMC psychiatry*. 2010;10(1):1. doi:10.1186/1471-244X-10-1
220. Birmaher B, Khetarpal S, Brent D, et al. The screen for child anxiety related emotional disorders (SCARED): Scale construction and psychometric characteristics. *Journal of the American Academy of Child & Adolescent Psychiatry*. 1997;36(4):545-553. doi:10.1097/00004583-199704000-00018

221. Sharp C, Goodyer IM, Croudace TJ. The Short Mood and Feelings Questionnaire (SMFQ): a unidimensional item response theory and categorical data factor analysis of self-report ratings from a community sample of 7-through 11-year-old children. *J Abnorm Child Psychol*. 2006;34(3):365-377. doi:10.1007/s10802-006-9027-x
222. Achenbach TM. Achenbach system of empirically based assessment (ASEBA). *The encyclopedia of clinical psychology*. 2014:1-8.
223. Silva RR, Alpert M, Pouget E, et al. A rating scale for disruptive behavior disorders, based on the DSM-IV item pool. *Psychiatric Quarterly*. 2005;76(4):327-339.
224. Conners CK, Sitarenios G, Parker JD, Epstein JN. The revised Conners' Parent Rating Scale (CPRS-R): factor structure, reliability, and criterion validity. *J Abnorm Child Psychol*. 1998;26(4):257-268.
225. Minică CC, Dolan CV, Kampert MM, Boomsma DI, Vink JM. Sandwich corrected standard errors in family-based genome-wide association studies. *European Journal of Human Genetics*. 2015;23(3):388-394.
226. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *Journal of statistical software*. 2010;36(3)
227. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *The American Journal of Human Genetics*. 2004;74(4):765-769.
228. Derringer J. A simple correction for non-independent tests. 2018;
229. Calcagno V, de Mazancourt C. glmulti: an R package for easy automated model selection with (generalized) linear models. *Journal of statistical software*. 2010;34(12):1-29.
230. Wagenmakers E-J, Farrell S. AIC model selection using Akaike weights. journal article. *Psychonomic Bulletin & Review*. February 01 2004;11(1):192-196. doi:10.3758/bf03206482
231. Baselmans BML, Willems YE, van Beijsterveldt CEM, et al. Unraveling the Genetic and Environmental Relationship Between Well-Being and Depressive Symptoms Throughout the Lifespan. Original Research. *Frontiers in Psychiatry*. 2018-June-14 2018;9(261) doi:10.3389/fpsyt.2018.00261
232. Singh MK, DelBello MP, Stanford KE, et al. Psychopathology in children of bipolar parents. *Journal of affective disorders*. 2007;102(1-3):131-136.
233. Birmaher B, Axelson D, Monk K, et al. Lifetime psychiatric disorders in school-aged offspring of parents with bipolar disorder: the Pittsburgh Bipolar Offspring study. *Archives of general psychiatry*. 2009;66(3):287-296.
234. Mostafavi H, Harpak A, Conley D, Pritchard JK, Przeworski M. Variable prediction accuracy of polygenic scores within an ancestry group. *BioRxiv*. 2019:629949.
235. Martin J, Tilling K, Hubbard L, et al. Association of genetic risk for schizophrenia with nonparticipation over time in a population-based cohort study. *American journal of epidemiology*. 2016;183(12):1149-1158.
236. Taylor AE, Jones HJ, Sallis H, et al. Exploring the association of genetic factors with participation in the Avon Longitudinal Study of Parents and Children. *International journal of epidemiology*. 2018;47(4):1207-1216.
237. Allegrini AG, Cheesman R, Rimfeld K, et al. The p factor: genetic analyses support a general dimension of psychopathology in childhood and adolescence. *Journal of Child Psychology and Psychiatry*. 0(0)doi:10.1111/jcpp.13113

238. Chatterjee N, Shi J, García-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nature Reviews Genetics*. 2016;17(7):392.
239. Johansson R, Carlbring P, Heedman Å, Paxling B, Andersson G. Depression, anxiety and their comorbidity in the Swedish general population: point prevalence and the effect on health-related quality of life. *PeerJ*. 2013;1:e98-e98. doi:10.7717/peerj.98
240. Tiller JWG. Depression and anxiety. *Medical Journal of Australia*. 2013;199(S6):S28-S31. doi:https://doi.org/10.5694/mja12.10628
241. Huisman-van Dijk HM, Schoot Rvd, Rijkeboer MM, Mathews CA, Cath DC. The relationship between tics, OC, ADHD and autism symptoms: A cross- disorder symptom analysis in Gilles de la Tourette syndrome patients and family-members. *Psychiatry research*. 2016;237:138-146. doi:10.1016/j.psychres.2016.01.051
242. D'Agati E, Curatolo P, Mazzone L. Comorbidity between ADHD and anxiety disorders across the lifespan. *International Journal of Psychiatry in Clinical Practice*. 2019/10/02 2019;23(4):238-244. doi:10.1080/13651501.2019.1628277
243. Gnanavel S, Sharma P, Kaushal P, Hussain S. Attention deficit hyperactivity disorder and comorbidity: A review of literature. *World J Clin Cases*. 2019;7(17):2420-2426. doi:10.12998/wjcc.v7.i17.2420
244. Oerlemans AM, Wardenaar KJ, Raven D, Hartman CA, Ormel J. The association of developmental trajectories of adolescent mental health with early-adult functioning. *PloS one*. 2020;15(6):e0233648. doi:10.1371/journal.pone.0233648
245. Ormel J, Oerlemans AM, Raven D, et al. Functional outcomes of child and adolescent mental disorders. Current disorder most important but psychiatric history matters as well. *Psychological Medicine*. 2017;47(7):1271-1282. doi:10.1017/S0033291716003445
246. Copeland WE, Alaie I, Jonsson U, Shanahan L. Associations of Childhood and Adolescent Depression With Adult Psychiatric and Functional Outcomes. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2021;60(5):604-611. doi:10.1016/j.jaac.2020.07.895
247. Stergiakouli E, Smith GD, Martin J, et al. Shared genetic influences between dimensional ASD and ADHD symptoms during child and adolescent development. *Molecular autism*. 2017;8(1):18. doi:10.1186/s13229-017-0131-2
248. Kwong ASF, Morris TT, Pearson RM, et al. Polygenic risk for depression, anxiety and neuroticism are associated with the severity and rate of change in depressive symptoms across adolescence. *Journal of Child Psychology and Psychiatry*. 2021;doi:https://doi.org/10.1111/jcpp.13422
249. Baselmans BML, van de Weijer MP, Abdellaoui A, et al. A Genetic Investigation of the Well-Being Spectrum. *Behavior Genetics*. 2019/05/01 2019;49(3):286-297. doi:10.1007/s10519-019-09951-0
250. Lee PH, Anttila V, Won H, et al. Genome wide meta-analysis identifies genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *bioRxiv*. 2019:528117. doi:10.1101/528117
251. Allegrini AG, Cheesman R, Rimfeld K, et al. The p factor: genetic analyses support a general dimension of psychopathology in childhood and adolescence. *Journal of Child Psychology and Psychiatry*. 2020;61(1):30-39. doi:10.1111/jcpp.13113

252. Bartels M, Hendriks A, Mauri M, et al. 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*. 2018;27(9):1105-1121.
253. Boyd A, Golding J, Macleod J, et al. Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *International journal of epidemiology*. 2013;42(1):111-127. doi:10.1093/ije/dys064
254. Fraser A, Macdonald-Wallis C, Tilling K, et al. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol*. Feb 2013;42(1):97-110. doi:10.1093/ije/dys066
255. Northstone K, Lewcock M, Groom A, et al. The Avon Longitudinal Study of Parents and Children (ALSPAC): an update on the enrolled sample of index children in 2019. *Wellcome Open Research*. 2019;4doi:10.12688/wellcomeopenres.15132.1
256. Anckarsäter H, Lundström S, Kollberg L, et al. The Child and Adolescent Twin Study in Sweden (CATSS). *Twin Research and Human Genetics*. 2012;14(6):495-508. doi:10.1375/twin.14.6.495
257. Ligthart L, van Beijsterveldt CEM, Kevenaar ST, et al. The Netherlands Twin Register: Longitudinal Research Based on Twin and Twin-Family Designs. *Twin Research and Human Genetics*. 2019;22(6):623-636. doi:10.1017/thg.2019.93
258. Rimfeld K, Malanchini M, Spargo T, et al. Twins Early Development Study: A Genetically Sensitive Investigation into Behavioral and Cognitive Development from Infancy to Emerging Adulthood. *Twin Research and Human Genetics*. 2019;22(6):508-513. doi:10.1017/thg.2019.56
259. Pritikin JN, Hunter MD, Boker S. Modular Open-Source Software for Item Factor Analysis. *Educ Psychol Meas*. Jun 1 2015;75(3):458-474. doi:10.1177/0013164414554615
260. Neale MC, Hunter MD, Pritikin JN, et al. OpenMx 2.0: Extended Structural Equation and Statistical Modeling. *Psychometrika*. Jun 2016;81(2):535-49. doi:10.1007/s11336-014-9435-8
261. Hunter MD. State Space Modeling in an Open Source, Modular, Structural Equation Modeling Environment. *Structural Equation Modeling: A Multidisciplinary Journal*. 2018/03/04 2018;25(2):307-324. doi:10.1080/10705511.2017.1369354
262. *OpenMx 2.17.4 User Guide*. 2020.
263. Enders CK, Bandalos DL. The Relative Performance of Full Information Maximum Likelihood Estimation for Missing Data in Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*. 2001/07/01 2001;8(3):430-457. doi:10.1207/S15328007SEM0803_5
264. Lahey BB, Krueger RF, Rathouz PJ, Waldman ID, Zald DH. A hierarchical causal taxonomy of psychopathology across the life span. *Psychological bulletin*. 2017;143(2):142-186. doi:10.1037/bul0000069
265. Leppert B, Riglin L, Dardani C, et al. ADHD genetic liability and physical health outcomes-A two-sample Mendelian randomization study. *bioRxiv*. 2019:630467.
266. Martins-Silva T, Vaz JdS, Hutz MH, et al. Assessing causality in the association between attention-deficit/hyperactivity disorder and obesity: a Mendelian randomization study. *International Journal of Obesity*. 2019/04/18 2019;doi:10.1038/s41366-019-0346-8

267. Liu C, Schoeler T, Davies NM, et al. Are there causal relationships between ADHD and BMI? Evidence from multiple genetically informed designs. *medRxiv*. 2020;
268. Dardani C, Leppert B, Riglin L, et al. Is genetic liability to ADHD and ASD causally linked to educational attainment? *BioRxiv*. 2020;
269. Cai N, Revez JA, Adams MJ, et al. Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nature Genetics*. 2020/04/01 2020;52(4):437-447. doi:10.1038/s41588-020-0594-5
270. Brownstein HH. *The social reality of violence and violent crime*. 2000.
271. Kinchin I, Doran CM, Hall WD, Meurk C. Understanding the true economic impact of self-harming behaviour. *The Lancet Psychiatry*. 2017;4(12):900-901.
272. Ferrey AE, Hughes ND, Simkin S, et al. The impact of self-harm by young people on parents and families: a qualitative study. *BMJ open*. 2016;6(1):e009631.
273. Fazel S, Wolf A, Palm C, Lichtenstein P. Violent crime, suicide, and premature mortality in patients with schizophrenia and related disorders: a 38-year total population study in Sweden. *The lancet psychiatry*. 2014;1(1):44-54.
274. Laporte N, Ozolins A, Westling S, et al. Deliberate self-harm behavior among young violent offenders. *PLoS one*. 2017;12(8):e0182258.
275. Sahlin H, Kuja-Halkola R, Bjureberg J, et al. Association between deliberate self-harm and violent criminality. *JAMA psychiatry*. 2017;74(6):615-621.
276. Shafti M, Taylor PJ, Forrester A, Pratt D. The Co-occurrence of Self-Harm and Aggression: A Cognitive-Emotional Model of Dual-Harm. Hypothesis and Theory. *Frontiers in Psychology*. 2021-February-25 2021;12(415)doi:10.3389/fpsyg.2021.586135
277. Allely CS. The association of ADHD symptoms to self-harm behaviours: a systematic PRISMA review. *BMC psychiatry*. 2014;14(1):133.
278. Hong JS, Espelage DL. A review of research on bullying and peer victimization in school: An ecological system analysis. *Aggression and violent behavior*. 2012;17(4):311-322.
279. Loeber R, Pardini D, Homish DL, et al. The prediction of violence and homicide in young men. *Journal of consulting and clinical psychology*. 2005;73(6):1074.
280. Paul E, Ortin A. Psychopathological mechanisms of early neglect and abuse on suicidal ideation and self-harm in middle childhood. *European Child & Adolescent Psychiatry*. 2019/10/01 2019;28(10):1311-1319. doi:10.1007/s00787-019-01287-8
281. Moffitt TE, Caspi A. Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Development and psychopathology*. 2001;13(2):355-375.
282. Moran P, Coffey C, Romaniuk H, et al. The natural history of self-harm from adolescence to young adulthood: a population-based cohort study. *The Lancet*. 2012;379(9812):236-243.
283. Lockwood J, Daley D, Townsend E, Sayal K. Impulsivity and self-harm in adolescence: a systematic review. *European child & adolescent psychiatry*. 2017;26(4):387-402.
284. O'Connor RC, Kirtley OJ. The integrated motivational–volitional model of suicidal behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2018;373(1754):20170268.
285. Weitzman ML, McHugh RK, Otto MW. The association between affect amplification and urgency. *Depression and anxiety*. 2011;28(12):1105-1110.
286. Jordan JT, Samuelson KW, Tiet QQ. Impulsivity, painful and provocative events, and suicide intent: testing the interpersonal theory of suicide. *Suicide and Life-Threatening Behavior*. 2018;

287. Anestis MD, Joiner TE. Examining the role of emotion in suicidality: Negative urgency as an amplifier of the relationship between components of the interpersonal–psychological theory of suicidal behavior and lifetime number of suicide attempts. *Journal of affective disorders*. 2011;129(1-3):261-269.
288. Harmer B, Lee S, Duong TVH, Saadabadi A. Suicidal Ideation. *StatPearls*. StatPearls Publishing Copyright © 2021, StatPearls Publishing LLC.; 2021.
289. Bresin K. Impulsivity and aggression: A meta-analysis using the UPPS model of impulsivity. *Aggression and Violent Behavior*. 2019/09/01/ 2019;48:124-140. doi:https://doi.org/10.1016/j.avb.2019.08.003
290. Retz W, Rösler M. The relation of ADHD and violent aggression: What can we learn from epidemiological and genetic studies? *International Journal of Law and Psychiatry*. 2009;32(4):235-243.
291. Brent DA, Johnson BA, Perper J, et al. Personality disorder, personality traits, impulsive violence, and completed suicide in adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry*. 1994;33(8):1080-1086.
292. Zanarini MC, Frankenburg FR, Hennen J, Reich DB, Silk KR. Axis I comorbidity in patients with borderline personality disorder: 6-year follow-up and prediction of time to remission. *American Journal of Psychiatry*. 2004;161(11):2108-2114.
293. Terzi L, Martino F, Berardi D, Bortolotti B, Sasdelli A, Menchetti M. Aggressive behavior and self-harm in Borderline Personality Disorder: The role of impulsivity and emotion dysregulation in a sample of outpatients. *Psychiatry Research*. 2017/03/01/ 2017;249:321-326. doi:https://doi.org/10.1016/j.psychres.2017.01.011
294. Campos AI, Verweij KJH, Statham DJ, et al. Genetic aetiology of self-harm ideation and behaviour. *Scientific Reports*. 2020/06/16 2020;10(1):9713. doi:10.1038/s41598-020-66737-9
295. Pappa I, St Pourcain B, Benke K, et al. A genome-wide approach to children's aggressive behavior: The EAGLE consortium. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. 2016;171(5):562-572.
296. Ip HF, van der Laan CM, Krapohl EML, et al. Genetic Association Study of Childhood Aggression across raters, instruments and age. *bioRxiv*. 2021:854927. doi:10.1101/854927
297. Carter G, Milner A, McGill K, Pirkis J, Kapur N, Spittal MJ. Predicting suicidal behaviours using clinical instruments: systematic review and meta-analysis of positive predictive values for risk scales. *The British Journal of Psychiatry*. 2017;210(6):387-395.
298. Belsher BE, Smolenski DJ, Pruitt LD, et al. Prediction models for suicide attempts and deaths: a systematic review and simulation. *JAMA psychiatry*. 2019;76(6):642-651.
299. Kang H-W, Kang H-B. Prediction of crime occurrence from multi-modal data using deep learning. *PloS one*. 2017;12(4):e0176244.
300. Wolf A, Fanshawe T, Sariaslan A, Cornish R, Larsson H, Fazel S. Prediction of violent crime on discharge from secure psychiatric hospitals: A clinical prediction rule (FoVOx). *European psychiatry*. 2018;47:88-93.
301. Ramesh T, Igoumenou A, Montes MV, Fazel S. Use of risk assessment instruments to predict violence in forensic psychiatric hospitals: a systematic review and meta-analysis. *European psychiatry*. 2018;52:47-53.

302. O'Shea LE, Dickens GL. Predictive validity of the Short-Term Assessment of Risk and Treatability (START) for aggression and self-harm in a secure mental health service: Gender differences. *International Journal of Forensic Mental Health*. 2015;14(2):132-146.
303. Phillips NL, Stargatt R, Brown A. Risk assessment of self-and other-directed aggression in adolescent psychiatric inpatient units. *Australian & New Zealand Journal of Psychiatry*. 2012;46(1):40-46.
304. Fullerton JM, Nurnberger JI. Polygenic risk scores in psychiatry: Will they be useful for clinicians? *F1000Research*. 2019;8
305. Huesmann LR, Dubow EF, Boxer P. Continuity of aggression from childhood to early adulthood as a predictor of life outcomes: implications for the adolescent-limited and life-course-persistent models. *Aggress Behav*. Mar-Apr 2009;35(2):136-149. doi:10.1002/ab.20300
306. Moffitt TE. Life-course persistent and adolescence-limited antisocial behavior. *Causes of conduct disorder and juvenile delinquency*. 2003:49-75.
307. Anckarsäter H, Lundström S, Kollberg L, et al. The child and adolescent twin study in Sweden (CATSS). *Twin Research and Human Genetics*. 2011;14(6):495-508.
308. Ligthart L, van Beijsterveldt CE, Kevenaar ST, et al. The Netherlands Twin Register: longitudinal research based on twin and twin-family designs. *Twin Research and Human Genetics*. 2019;22(6):623-636.
309. Coccaro EF, Berman ME, Kavoussi RJ. Assessment of life history of aggression: development and psychometric characteristics. *Psychiatry research*. 1997;73(3):147-157.
310. Stattin H, Kerr M. Parental monitoring: A reinterpretation. *Child development*. 2000;71(4):1072-1085.
311. Malmberg M, Rydell A-m, Smedje H. Validity of the Swedish version of the Strengths and Difficulties Questionnaire (SDQ-Swe). *Nordic journal of psychiatry*. 2003;57(5):357-363.
312. Gerard AB. *Parent-child relationship inventory (PCRI)*. Western psychological services; 1994.
313. Raine A, Dodge K, Loeber R, et al. The reactive–proactive aggression questionnaire: Differential correlates of reactive and proactive aggression in adolescent boys. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*. 2006;32(2):159-171.
314. Achenbach TM, Rescorla L. Manual for the ASEBA adult forms & profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth …; 2003.
315. Coombes BJ, Ploner A, Bergen SE, Biernacka JM. A principal component approach to improve association testing with polygenic risk scores. *Genetic epidemiology*. 2020;44(7):676-686.
316. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 2016;374(2065):20150202.
317. Caspi A, Houts RM, Belsky DW, et al. The p factor: one general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*. 2014;2(2):119-137.
318. Wold S, Esbensen K, Geladi P. Principal component analysis. *Chemometrics and intelligent laboratory systems*. 1987;2(1-3):37-52.
319. Pan R, Yang T, Cao J, Lu K, Zhang Z. Missing data imputation by K nearest neighbours based on grey relational structure and mutual information. *Applied Intelligence*. 2015;43(3):614-632.

320. Cook D. *Practical machine learning with H2O: powerful, scalable techniques for deep learning and AI*." O'Reilly Media, Inc."; 2016.
321. Mandrekar JN. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology*. 2010/09/01/ 2010;5(9):1315-1316. doi:https://doi.org/10.1097/JTO.0b013e3181ec173d
322. Guo X, Yin Y, Dong C, Yang G, Zhou G. On the class imbalance problem. IEEE; 2008:192-201.
323. Huang JZ. An Introduction to Statistical Learning: With Applications in R By Gareth James, Trevor Hastie, Robert Tibshirani, Daniela Witten. Springer; 2014.
324. Runeson B, Odeberg J, Pettersson A, Edbom T, Jildevik Adamsson I, Waern M. Instruments for the assessment of suicide risk: a systematic review evaluating the certainty of the evidence. *PLoS one*. 2017;12(7):e0180292.
325. Menger V, Spruit M, Van Est R, Nap E, Scheepers F. Machine learning approach to inpatient violence risk assessment using routinely collected clinical notes in electronic health records. *JAMA network open*. 2019;2(7):e196709-e196709.
326. Whiting D, Fazel S. How accurate are suicide risk prediction models? Asking the right questions for clinical practice. *Evidence-based mental health*. 2019;22(3):125-128.
327. Murray GK, Lin T, Austin J, McGrath JJ, Hickie IB, Wray NR. Could Polygenic Risk Scores Be Useful in Psychiatry?: A Review. *JAMA Psychiatry*. 2021;78(2):210-219. doi:10.1001/jamapsychiatry.2020.3042
328. Kirchner T, Ferrer L, Forns M, Zanini D. Self-harm behavior and suicidal ideation among high school students. Gender differences and relationship with coping strategies. *Actas Esp Psiquiatr*. Jul-Aug 2011;39(4):226-35.
329. Heron J, Barker ED, Joinson C, et al. Childhood conduct disorder trajectories, prior risk factors and cannabis use at age 16: birth cohort study. https://doi.org/10.1111/add.12268. *Addiction*. 2013/12/01 2013;108(12):2129-2138. doi:https://doi.org/10.1111/add.12268
330. Odgers CL, Moffitt TE, Broadbent JM, et al. Female and male antisocial trajectories: from childhood origins to adult outcomes. *Dev Psychopathol*. Spring 2008;20(2):673-716. doi:10.1017/s0954579408000333
331. Rivenbark JG, Odgers CL, Caspi A, et al. The high societal costs of childhood conduct problems: evidence from administrative records up to age 38 in a longitudinal birth cohort. *J Child Psychol Psychiatry*. Jun 2018;59(6):703-710. doi:10.1111/jcpp.12850
332. Bartels M, Hendriks A, Mauri M, et al. 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. *Eur Child Adolesc Psychiatry*. Sep 2018;27(9):1105-1121. doi:10.1007/s00787-018-1169-1
333. Laursen TM, Nordentoft M, Mortensen PB. Excess Early Mortality in Schizophrenia. *Annual Review of Clinical Psychology*. 2014/03/28 2014;10(1):425-448. doi:10.1146/annurev-clinpsy-032813-153657
334. Power RA, Kyaga S, Uher R, et al. Fecundity of Patients With Schizophrenia, Autism, Bipolar Disorder, Depression, Anorexia Nervosa, or Substance Abuse vs Their Unaffected Siblings. *JAMA Psychiatry*. 2013;70(1):22-30. doi:10.1001/jamapsychiatry.2013.268
335. Avramopoulos D. Recent Advances in the Genetics of Schizophrenia. *Complex Psychiatry*. 2018;4(1):35-51. doi:10.1159/000488679

336. Singh T, Poterba T, Curtis D, et al. Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia. *medRxiv*. 2020;
337. McGrath J, Saha S, Chant D, Welham J. Schizophrenia: A Concise Overview of Incidence, Prevalence, and Mortality. *Epidemiologic Reviews*. 2008;30(1):67-76. doi:10.1093/epirev/mxn001
338. Ripke S, Walters JT, O'Donovan MC. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *medRxiv*. 2020:2020.09.12.20192922. doi:10.1101/2020.09.12.20192922
339. Gejman PV, Sanders AR, Duan J. The role of genetics in the etiology of schizophrenia. *Psychiatr Clin North Am*. 2010;33(1):35-66. doi:10.1016/j.psc.2009.12.003
340. Uher R. The role of genetic variation in the causation of mental illness: an evolution-informed framework. *Molecular Psychiatry*. 2009/12/01 2009;14(12):1072-1082. doi:10.1038/mp.2009.85
341. Lam M, Chen C-Y, Li Z, et al. Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nature Genetics*. 2019/12/01 2019;51(12):1670-1678. doi:10.1038/s41588-019-0512-x
342. O'Connor LJ, Schoech AP, Hormozdiari F, Gazal S, Patterson N, Price AL. Extreme Polygenicity of Complex Traits Is Explained by Negative Selection. *The American Journal of Human Genetics*. 2019/09/05/ 2019;105(3):456-476. doi:https://doi.org/10.1016/j.ajhg.2019.07.003
343. Liu C, Everall I, Pantelis C, Bousman C. Interrogating the Evolutionary Paradox of Schizophrenia: A Novel Framework and Evidence Supporting Recent Negative Selection of Schizophrenia Risk Alleles. Original Research. *Frontiers in Genetics*. 2019-April-30 2019;10(389)doi:10.3389/fgene.2019.00389
344. Sekar A, Bialas AR, de Rivera H, et al. Schizophrenia risk from complex variation of complement component 4. *Nature*. 2016;530(7589):177-183. doi:10.1038/nature16549
345. Mealer RG, Jenkins BG, Chen C-Y, et al. The schizophrenia risk locus in SLC39A8 alters brain metal transport and plasma glycosylation. *Scientific Reports*. 2020/08/04 2020;10(1):13162. doi:10.1038/s41598-020-70108-9
346. Genovese G, Fromer M, Stahl EA, et al. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. Article. *Nature neuroscience*. 10/03/online 2016;19:1433. doi:10.1038/nn.4402 https://www.nature.com/articles/nn.4402#supplementary-information
347. Ganna A, Satterstrom FK, Zekavat SM, et al. Quantifying the Impact of Rare and Ultra-rare Coding Variation across the Phenotypic Spectrum. *American journal of human genetics*. 2018;102(6):1204-1211. doi:10.1016/j.ajhg.2018.05.002
348. Sanders SJ, Neale BM, Huang H, et al. Whole genome sequencing in psychiatric disorders: the WGSPD consortium. *Nature neuroscience*. 2017/12/01 2017;20(12):1661-1668. doi:10.1038/s41593-017-0017-9
349. Bryois J, Skene NG, Hansen TF, et al. Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease. *Nature Genetics*. 2020/05/01 2020;52(5):482-493. doi:10.1038/s41588-020-0610-9
350. Skene NG, Bryois J, Bakken TE, et al. Genetic identification of brain cell types underlying schizophrenia. *Nature Genetics*. 2018/06/01 2018;50(6):825-833. doi:10.1038/s41588-018-0129-5

351. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020/05/01 2020;581(7809):434-443. doi:10.1038/s41586-020-2308-7
352. Habib N, Avraham-Davidi I, Basu A, et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nature Methods*. 2017/10/01 2017;14(10):955-958. doi:10.1038/nmeth.4407
353. Koopmans F, van Nierop P, Andres-Alonso M, et al. SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron*. 2019/07/17/ 2019;103(2):217-234.e4. doi:https://doi.org/10.1016/j.neuron.2019.05.002
354. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol*. 2015;11(4):e1004219.
355. Efthymiou AG, Goate AM. Late onset Alzheimer's disease genetics implicates microglial pathways in disease risk. *Mol Neurodegener*. 2017;12(1):43-43. doi:10.1186/s13024-017-0184-x
356. Wightman DP, Jansen IE, Savage JE, et al. Largest GWAS (N=1,126,563) of Alzheimer's Disease Implicates Microglia and Immune Cells. *medRxiv*. 2020:2020.11.20.20235275. doi:10.1101/2020.11.20.20235275
357. Grotzinger AD, Mallard TT, Akingbuwa WA, et al. Genetic Architecture of 11 Major Psychiatric Disorders at Biobehavioral, Functional Genomic, and Molecular Genetic Levels of Analysis. *medRxiv*. 2020:2020.09.22.20196089. doi:10.1101/2020.09.22.20196089
358. Coleman JRI, Bryois J, Gaspar HA, et al. Biological annotation of genetic loci associated with intelligence in a meta-analysis of 87,740 individuals. *Molecular Psychiatry*. 2019/02/01 2019;24(2):182-197. doi:10.1038/s41380-018-0040-6
359. Demange PA, Malanchini M, Mallard TT, et al. Investigating the genetic architecture of noncognitive skills using GWAS-by-subtraction. *Nature Genetics*. 2021/01/01 2021;53(1):35-44. doi:10.1038/s41588-020-00754-2
360. Baselmans BM, Jansen R, Ip HF, et al. Multivariate genome-wide analyses of the well-being spectrum. *Nature genetics*. 2019;51(3):445.
361. Clifton NE, Hannon E, Harwood JC, et al. Dynamic expression of genes associated with schizophrenia and bipolar disorder across development. *Transl Psychiatry*. 2019;9(1):74-74. doi:10.1038/s41398-019-0405-x
362. Ganna A, Genovese G, Howrigan DP, et al. Ultra-rare disruptive and damaging mutations influence educational attainment in the general population. *Nat Neurosci*. 2016;19(12):1563-1565. doi:10.1038/nn.4404
363. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461(7265):747.
364. Watanabe K, Stringer S, Frei O, et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nature Genetics*. 2019/09/01 2019;51(9):1339-1348. doi:10.1038/s41588-019-0481-0
365. Lee PH, Feng Y-CA, Smoller JW. Pleiotropy and Cross-Disorder Genetics Among Psychiatric Disorders. *Biological Psychiatry*. 2021/01/01/ 2021;89(1):20-31. doi:https://doi.org/10.1016/j.biopsych.2020.09.026
366. Smith GT, Atkinson EA, Davis HA, Riley EN, Oltmanns JR. The General Factor of Psychopathology. *Annual Review of Clinical Psychology*. 2020/05/07 2020;16(1):75-98. doi:10.1146/annurev-clinpsy-071119-115848

367. Riglin L, Thapar AK, Leppert B, et al. Using Genetics to Examine a General Liability to Childhood Psychopathology. *Behavior Genetics*. 2020/07/01 2020;50(4):213-220. doi:10.1007/s10519-019-09985-4
368. Jansen AG, Jansen PR, Savage JE, et al. The predictive capacity of psychiatric and psychological polygenic risk scores for distinguishing cases in a child and adolescent psychiatric sample from controls. *Journal of child psychology and psychiatry*. 2021;
369. Mostafavi H, Harpak A, Agarwal I, Conley D, Pritchard JK, Przeworski M. Variable prediction accuracy of polygenic scores within an ancestry group. *Elife*. 2020;9:e48376. doi:10.7554/eLife.48376
370. Demange PA, Hottenga JJ, Abdellaoui A, et al. Parental influences on offspring education: indirect genetic effects of non-cognitive skills. *biorxiv*. 2020;
371. Cheesman R, Hunjan A, Coleman JR, et al. Comparison of adopted and nonadopted individuals reveals gene–environment interplay for education in the UK Biobank. *Psychological science*. 2020;31(5):582-591.
372. Wray NR, Lin T, Austin J, et al. From Basic Science to Clinical Application of Polygenic Risk Scores: A Primer. *JAMA Psychiatry*. 2021;78(1):101-109. doi:10.1001/jamapsychiatry.2020.3049
373. Rees E, Owen MJ. Translating insights from neuropsychiatric genetics and genomics for precision psychiatry. *Genome Medicine*. 2020/04/29 2020;12(1):43. doi:10.1186/s13073-020-00734-5
374. Rees E, Creeth HDJ, Hwu H-G, et al. Schizophrenia, autism spectrum disorders and developmental disorders share specific disruptive coding mutations. *bioRxiv*. 2020:2020.09.02.279265. doi:10.1101/2020.09.02.279265
375. Thorp JG, Campos AI, Grotzinger AD, et al. Symptom-level modelling unravels the shared genetic architecture of anxiety and depression. *Nat Hum Behav*. 2021/04/15 2021;doi:10.1038/s41562-021-01094-9
376. Abdellaoui A, Verweij KJH. Dissecting polygenic signals from genome-wide association studies on human behaviour. *Nat Hum Behav*. 2021/05/13 2021;doi:10.1038/s41562-021-01110-y
377. Post RM, Leverich GS, Kupka RW, et al. Early-onset bipolar disorder and treatment delay are risk factors for poor outcome in adulthood. *J Clin Psychiatry*. Jul 2010;71(7):864-72. doi:10.4088/JCP.08m04994yel
378. Zheng J, Baird D, Borges M-C, et al. Recent Developments in Mendelian Randomization Studies. *Curr Epidemiol Rep*. 2017;4(4):330-345. doi:10.1007/s40471-017-0128-6
379. Wainschtein P, Jain D, Zheng Z, et al. Recovery of trait heritability from whole genome sequence data. *bioRxiv*. 2021:588020. doi:10.1101/588020
380. Weiner DJ, Wigdor EM, Ripke S, et al. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nature Genetics*. 2017/07/01 2017;49(7):978-985. doi:10.1038/ng.3863
381. Niemi MEK, Martin HC, Rice DL, et al. Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *Nature*. 2018/10/01 2018;562(7726):268-271. doi:10.1038/s41586-018-0566-4
382. Tansey KE, Rees E, Linden DE, et al. Common alleles contribute to schizophrenia in CNV carriers. *Molecular Psychiatry*. 2016/08/01 2016;21(8):1085-1089. doi:10.1038/mp.2015.143

Dankwoord

It has been a long journey to get to this point. If you are reading this for any reason, you have contributed in some way or other to this achievement. Thank you!

To my promotors Christel Middeldorp and Meike Bartels, thank you for your guidance and supervision over the past 4 years. None of this would have been possible without your insight and support. Your knowledge as well as your trust in me allowed me to grow as a scientist. I am grateful for the opportunity to learn from you and I am an infinitely better scientist as a result of your guidance. To my co-promotor Anke Hammerschlag, the journey from AIO to Dr was made all the smoother because of you. Thank you for always being available to answer questions and provide reassurance when I wasn't sure what I was doing.

My thanks also to the members of my doctoral examination committee, Prof. dr. Eco de Geus, Prof. dr. Teresa van Amelsvoort, Prof. dr. Pieter Hoekstra, Dr. Andrea Ganna, Dr. Ziada Ayorech and Dr. Tinca Polderman. Thank you for the considerable time and effort you put into reading and assessing my thesis.

To everyone who was part of CAPICE, but especially the students, Andrea, Ashley, Elis, Eshim, Hema, Kratika, Laura, Lizzie, Marica, Miljan, Sabrina, and Ville; thank you for the fun times at the meetings and workshops, and especially after. I have so many cherished memories with you, and I look forward to running into you all in the future. Ashley, collaborating with you was a special highlight. I am glad we had the opportunity to do that. Thank you for the amazing learning experience.

I especially want to thank everyone at the Department of Biological Psychology, both and present, for welcoming me and making the past 4 years so enjoyable, so much so that I decided to stay. I am grateful to have colleagues as fantastic as you all. Let's hope we can enjoy some Friday afternoon drinks again soon. To my co-authors across the years, thank you for all the great collaborations and valuable feedback on manuscripts. A special shout out to Lannie, Toos, Bart, Hamdi, René Pool and Jouke Jan for all the help with PGS and data. Much of this would not have been possible without your help. Michel, thank you for the inspiration and encouragement to branch out and learn different skills, as well as your support all the way through. And of course for the job. I learn (and laugh) so much from

conversations with you, both scientific and otherwise, and I look forward to continuing to work with you. Bart, photos of your ugly shoes will haunt me forever, thank you for all the laughs. Fiona, Laura W, Lianne, Lissette, Saskia, and Yayouk; thank you for the gezellig clothing exchange nights. Anne Hendriks, your advice and constant encouragement always means the world to me, thank you. Matthijs, my roomie from day 1, thank you for the laughs, poké-breaks, and all the help with coding. I continue to live in hope that our business venture will one day be successful (even though we do zero marketing for it) #secretsignsforever. Martin, Camiel, and Nicole; thank you for the opportunity to share my love of football (Ajax) and Formula 1 with you guys. I have so many amazing memories as a result. To the other AIOs who I have shared this journey with at one point or another: Anne L, Bodine, Denise, Floris, Hekmat, Hill, Klaasjan, Laura H, Nikki, Ruifang, Selim, Susanne, Veronika, and Zoey; thank you for all the fun gatherings and supportive AIO meetings.

To my Biopsy Girlz; Margot, Sofieke, Perline and Zenab, you have been instrumental in helping me keep my sanity over the past year and a bit. Thank you for all the insightful conversations and debates. There is never a dull moment with you girls. Also thanks for re-introducing me to the travesty that is Twilight (I'm sorry I had to). There is nothing I enjoy more than slagging off movies with you. Margot, Sofieke and Perline, thank you for sharing my TikTok obsession and all the funny videos. I always look forward to your curated fyps. Zenab, conversations with you always make me feel seen, I appreciate you so much. Last but most definitely not least, Eshim, my officemate, flatmate, travel buddy (often simultaneously), and all round amazing friend. I could not have asked for a better person to share this experience with, and meeting you was one of the best parts of my PhD. Thank you for always letting me be my full dramatic self and judging me only a little bit. I am so proud of us and how far we've come. You inspire me in so many ways you'll never know. Thank you for your wonderful friendship and the lekker chats. Amsterdam is not the same without you.

To my friends in London and elsewhere around the world, thank you for always being supportive. Chris, I sort of owe this entire experience to you since you found the advert for this position and encouraged me to apply for it. You did good. Thanks to you, Lissie and Judit for always being only a Whatsapp message away whenever I need it. Marcia and Merinda, the M&S and haircare packages are a

particular highlight of my time here. Dozie and Id, you never let it get too long without checking in with me even though I am notoriously bad at this. I love you guys and am so thankful to have you in life. Sophia, even from across the world you remain the best friend a girl could ask for. With you I can talk about any and everything. I will never regret bonding over SO MANY tv-shows.

Finally, I would like to thank my family. Mum and Dad, you sacrificed so much for me to get where I am today and I will never be able to repay you. I love you and I am so blessed to have you as parents. Thank you for the top quality genes and environment, if I do say so myself. Lolu and Rachel, thank you for always having my back and forever letting me third-wheel with you guys. Thanks also for letting me be stuck with you in the middle of a whole-ass pandemic, even as I wreaked havoc across your home (I bet you don't miss that). Folu, Moyo, James, Gina, and my baby Jan; you guys make me laugh like no one else can. I adore you all and I am lucky to have siblings like you. Aunty Lara and Uncle Yemi, you are the best guardians I could have asked for. Thank you for your consistent love and support. Tomini and Timileyin, I love you guys.

Einde!