Polygenic scores for schizophrenia and educational attainment are associated with behavioural problems in early childhood in the general population

Jansen, Philip R; Polderman, Tinca J C; Bolhuis, Koen; van der Ende, Jan; Jaddoe, Vincent W V; Verhulst, Frank C; White, Tonya; Posthuma, Danielle; Tiemeier, Henning

published in
Journal of Child Psychology and Psychiatry
2018

DOI (link to publisher)
10.1111/jcpp.12759

document version
Publisher's PDF, also known as Version of record

The Journal of Child Psychology and Psychiatry

Journal of Child Psychology and Psychiatry 59:1 (2018), pp 39–47 doi:10.1111/jcpp.12759

Polygenic scores for schizophrenia and educational attainment are associated with behavioural problems in early childhood in the general population

Philip R. Jansen,[1,2] Tinca J.C. Polderman,[2] Koen Bolhuis,[1] Jan van der Ende,[1] Vincent W.V. Jaddoe,[3,4] Frank C. Verhulst,[1,5] Tonya White,[1] Danielle Posthuma,[2,6,*] and Henning Tiemeier[1,7,*]

[1]Department of Child and Adolescent Psychiatry, Erasmus University Medical Center, Rotterdam; [2]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University, Amsterdam; [3]Department of Epidemiology, Erasmus University Medical Center, Rotterdam; [4]Department of Paediatrics, Sophia Children's Hospital, Rotterdam, The Netherlands; [5]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; [6]Department of Clinical Genetics, Amsterdam Neuroscience, VU Medical Center, Amsterdam; [7]Department of Psychiatry, Erasmus University Medical Center, Rotterdam, The Netherlands

Background: Genome-wide association studies in adults have identified numerous genetic variants related to psychiatric disorders and related traits, such as schizophrenia and educational attainment. However, the effects of these genetic variants on behaviour in the general population remain to be fully understood, particularly in younger populations. We investigated whether polygenic scores of five psychiatric disorders and educational attainment are related to emotional and behaviour problems during early childhood. **Methods:** From the Generation R Study, we included participants with available genotype data and behavioural problems measured with the Child Behavior Checklist (CBCL) at the age of 3 (n = 1,902), 6 (n = 2,202) and 10 years old (n = 1,843). Polygenic scores were calculated for five psychiatric disorders and educational attainment. These polygenic scores were tested for an association with the broadband internalizing and externalizing problem scales and the specific CBCL syndrome scale scores. **Results:** Analysis of the CBCL broadband scales showed that the schizophrenia polygenic score was associated with significantly higher internalizing scores at 3, 6 and 10 years and higher externalizing scores at age 3 and 6. The educational attainment polygenic score was associated with lower externalizing scores at all time points and lower internalizing scores at age 3. No associations were observed for the polygenic scores of bipolar disorder, major depressive disorder and autism spectrum disorder. Secondary analyses of specific syndrome scores showed that the schizophrenia polygenic score was strongly related to the Thought Problems scores. A negative association was observed between the educational attainment polygenic score and Attention Problems scores across all age groups. **Conclusions:** Polygenic scores for adult psychiatric disorders and educational attainment are associated with variation in emotional and behavioural problems already at a very early age. **Keywords:** Polygenic scores; psychiatric disorders; educational attainment; childhood behaviour.

Introduction

Childhood emotional and behavioural problems are common and show moderate stability throughout childhood (Basten et al., 2016). Although symptoms may fluctuate over time, early childhood problems have predictive value for psychiatric disorders later in life, as well as academic performance (Van der Ende, Verhulst, & Tiemeier, 2016) and risk-taking behaviour (King, Iacono, & McGue, 2004).

Variation in problem behaviour is influenced substantially by genetic factors. Twin study heritability estimates of behavioural problems are fairly constant over the course of childhood, and show that genetic factors explain around 50% of the variance in externalizing (e.g. aggression and oppositional behaviour) and internalizing (e.g. depression and anxiety) behaviours (van der Valk, van den Oord, Verhulst, & Boomsma, 2003). However, no statistically significant genetic polymorphism has yet been identified specifically for common childhood emotional and behavioural problems (Benke et al., 2014; Middeldorp et al., 2016; Pappa et al., 2015). In contrast, GWAS studies in adult behaviour-related phenotypes have identified variants for a wide variety of traits and show that the majority of neuropsychiatric traits are genetically complex and determined by many genetic variants, mostly of small effect (Plomin, DeFries, Knopik, & Neiderhiser, 2016; Sullivan, Daly, & O'Donovan, 2012).

Educational attainment is an important predictor for a variety of important life outcomes and is closely linked to psychopathology (Kessler, Foster, Saunders, & Stang, 1995). Recently, proxy phenotype methods have targeted educational attainment (i.e. years of schooling) to detect genetic variants related to psychiatric and personality-related traits, including schizophrenia and neuroticism (Okbay et al.,

*Both authors contributed equally to the manuscript.
Conflict of interest statement: See Acknowledgements for disclosures.

2016). It was also shown that in school-aged children, a genetic predisposition to higher educational attainment is associated with higher cognitive performance (Ward et al., 2014).

Although many genetic variants have been identified for adult psychiatric disorders, it is unclear whether the same genetic variants are also associated with problem behaviour at an early age. Moreover, it is unclear whether children at high risk for psychopathology already show differences in behaviour during childhood. Population-based cohort studies in adolescents have demonstrated that genetic risks for psychopathology, quantified by risk scoring methods, correlate with a diverse array of behavioural outcomes (Krapohl et al., 2016). More specifically, studies using schizophrenia polygenic risk scores reported that this genetic risk is associated with negative symptoms in the general adolescent population (Derks, Vorstman, Ripke, Kahn, & Ophoff, 2012; Jones et al., 2016). However, a recent study investigating the schizophrenia polygenic score in young children, suggests that manifestations in behaviour and neurodevelopment may be present much earlier in life (Riglin et al., 2017). In addition, most studies in children have tested polygenic risk scores for single traits, without performing comparisons across traits.

The aim of this study was to investigate whether polygenic risk scores of later life outcomes are associated with early childhood behavioural and emotional problems in the general population. We focus on five psychiatric disorders (schizophrenia, bipolar disorders, major depressive disorder, ADHD and autism spectrum disorder) and educational attainment, as these traits have been shown to be associated with early life problem behaviour (Breslau et al., 2009; Caspi, Moffitt, Newman, & Silva, 1996). We hypothesize that genetic variants, associated with psychiatric disorders and educational attainment, are related to variation in symptoms in internalizing and externalizing domains at early childhood, thus earlier in life than described in most previous studies. This study will provide a better understanding of the association between risk variants and behavioural manifestations early in life and insight into the underlying neurobiology of these traits.

Methods

Study sample

The Medical Ethics Committee of the Erasmus Medical Center approved all study procedures, and parents of the participants provided written informed consent.

This study was conducted within the Generation R cohort, a large population-based longitudinal cohort focused on child development (Kooijman et al., 2016). Emotional and behavioural problems were assessed prospectively at the approximate age of 3, 6 and 10 years in respectively 4,612, 6,199 and 4,770 children. Of these children, 2,964, 3,926 and 3,058 were subsequently selected based on the availability of genotype data. Of these, 1,902, 2,202 and 1,843 children passed genotype quality control procedures and were included in the analyses.

Child behaviour measures

Behaviour problems were assessed with the Child Behavior Checklist (CBCL), a comprehensive list of items about various child emotional and behavioural problems, to be completed by the primary caregiver (Achenbach & Rescorla, 2000). Each CBCL item can be scored as: 0 = 'not true', 1 = 'somewhat or sometimes true', 2 = 'very true or often true'. CBCL items can be scored on two broadband scales: 'Internalizing Problems' and 'Externalizing Problems', and on more specific syndrome scales. During the first and second assessment wave, the preschool CBCL version (CBCL/$1^{1/2}$-5) was used, as most children during the assessment were younger than 6 years, and other versions are not appropriate for this age (Tiemeier et al., 2012). The CBCL/$1^{1/2}$-5 survey consists of 100 problem items. At the third assessment wave, the school-age (CBCL/6-18) version was used, consisting of 120 problem items. An overview of the CBCL syndrome scales and the number of items within each domain are shown in Table S1, available online.

Genotyping and imputation

Genotype calling procedures and subsequent processing for the Generation R Study have been described previously (Medina-Gomez et al., 2015). Briefly, genotype data were either collected from cord blood at birth (Illumina 610K Quad Chip) or via vena puncture (Illumina 660K Quad Chip) during a visit to the research centre. Additional quality control steps were performed on the genotype data in PLINK (Purcell et al., 2007). Variants were filtered for minor allele frequency (MAF < 0.01), Hardy-–einberg disequilibrium ($p < .00001$) and missing rate (> 0.05). Individuals from European descent were selected within 4 standard deviations on the first four genetic principal components of the HapMap Phase II Northwestern European (CEU) population. Individuals were additionally filtered on relatedness, sex mismatch and genotype quality (<5% missing).

Genotypes that passed quality control were prephased with the SHAPEIT software package (Delaneau, Marchini & Zagury, 2012). Phased haplotypes were imputed using IMPUTE v2 (Howie, Donnelly, & Marchini, 2009) against the 1000 Genomes (phase I version 3) as the reference panel. After postimputation filtering (INFO score <0.9), a total number of 6,561,671 variants were considered for further analyses.

To correct for population structure, genetic principal components were calculated in EIGENSOFT (Patterson, Price, & Reich, 2006; Price et al., 2006) on a linkage disequilibrium (LD) pruned set of 107,266 genotyped variants that passed quality control.

Polygenic scoring

To calculate polygenic risk scores in the target sample, we used available genome-wide association results (GWAS) for six traits, including schizophrenia (SCZ), attention-deficit hyperactivity disorder (ADHD), bipolar disorder (BP), major depressive disorder (MDD), autism spectrum disorder (ASD) and educational attainment (EA, years of schooling). A more detailed overview of these GWAS studies with references is provided in Table S2. The Generation R Study was not included in any of the discovery GWAS studies that were used for calculation of the polygenic scores.

Polygenic scores were constructed with PRSice (Euesden, Lewis, & O'Reilly, 2015), an R script for calculating polygenic scores in PLINK, by weighing the number of risk alleles by the SNP effect size. Variants were clumped prior to the calculation of the polygenic scores according to LD to obtain the most significant SNP per LD block (kilobase pair window: 250, LD $r^2 < .1$). *p*-Value thresholds (p_T) for inclusion of genetic

variants in the score varied between $p_T < .01$ and $p_T < 1$. Table S3 shows the number of SNPs that were included in the final polygenic score for each *p*-value threshold. Pearson correlations between the polygenic scores of the six traits are shown in Figure S1.

Polygenic scores were standardized to a mean of 0 and standard deviation of 1 to increase interpretation of the score.

Statistical analyses

Statistical analyses were performed in R statistical software (version 3.2.1) (R Core Team, 2014). Polygenic scores for the six traits were tested individually in a linear regression model for association with CBCL Internalizing and Externalizing Problems, corrected for age, sex and four genetic principal components. Next, we tested the most significant *p*-value threshold of each trait for associations with individual syndrome scales.

To account for varying degrees of skewness in CBCL scores, syndrome scores were transformed at each time point using Box–Cox transformation. This method utilizes maximum likelihood estimation (MLE) to find the optimal transformation parameter to approximate a normal distribution (Sakia, 1992).

False-discovery rate (FDR) was applied to correct for multiple comparisons (Benjamini & Hochberg, 1995). Based on the total number of statistical tests across polygenic scores, *p*-value thresholds, broadband scales and specific syndrome scales, a corrected *p*-value significance threshold was set to *p*FDR = .0083, and *p*-values below this corrected threshold were considered statistically significant.

Results

Sample Characteristics

Characteristics of the study sample at the three assessment waves are shown in Table 1. The three groups had a mean age of 3.0 (*SD* = 0.1), 6.0 (*SD* = 0.4) and 9.7 (*SD* = 0.3) years at the time of the assessment, and sex was equally divided among groups (per cent boys: age 3: 52%, age 6: 50%, age 10: 49%). At all ages, boys scored higher than girls on externalizing problems (mean difference: age 3: 0.12, $p < .001$; age 6: 0.07, $p < .001$; age 10: 0.17, $p < .001$); there were no significant sex differences on the internalizing problem scales at age 3, 6 and 10.

Table 1 Sample characteristics

	Assessment		
	Age 3 *N* = 1,902	Age 6 *N* = 2,202	Age 10 *N* = 1,843
Characteristics			
Mean age, years	3.04 ± 0.09	5.99 ± 0.37	9.69 ± 0.27
Gender, % male	52%	51%	49%
Internalizing problems			
Mean score	4.22 ± 3.86	5.21 ± 5.35	4.57 ± 4.79
Range	0–36	0–49	0–41
Median	3	4	3
Externalizing problems			
Mean score	7.88 ± 5.97	6.87 ± 6.41	3.82 ± 4.73
Score range	0–42	0–43	0–39
Median	7	5	2

Internalizing and externalizing problems

The explained variance (R^2) of the broadband internalizing and externalizing problem scales by the six polygenic scores for all *p*-value thresholds is shown in Figure 1. Table S4A–C show the full regression results for these associations.

Here we highlight the *p*-value threshold for each trait that showed the strongest association (i.e. largest increase in R^2) with the outcome across the different age groups.

Schizophrenia

Analyses in 3-year olds showed that the SCZ polygenic score was significantly associated with higher levels of internalizing problems ($p_T < .5$: $\beta = .061$, $p = .008$) and externalizing problems ($p_T < .5$: $\beta = .067$, $p = .004$). At the age of 6 years, an association between the SCZ polygenic score and internalizing scores ($p_T < .5$: $\beta = .088$, $p = <.001$) and externalizing problems ($p_T < .5$: $\beta = .070$, $p < .001$) was also present. At age 10, we observed again the association with internalizing scores ($p_T < .5$: $\beta = .069$, $p = .003$), whereas the association with externalizing scores was no longer significant ($p_T < .5$: $\beta = .039$, $p = .096$).

ADHD

No significant associations were observed with externalizing or internalizing scores at the age of 3. However, we observed a weak positive association with externalizing scores at age 6 ($p_T < .01$: $\beta = .042$, $p = .044$) that was not significant after correction for multiple comparisons. This association became stronger and significant at the age of 10 ($p_T < .01$: $\beta = .076$, $p = .001$).

Bipolar disorder, major depressive disorder and autism spectrum disorder

No association was observed between the BP, MDD and ASD polygenic scores and externalizing and internalizing scales in any age group.

Educational attainment

The EA polygenic score was negatively associated with externalizing ($p_T < .5$: $\beta = -.089$, $p < 001$) and internalizing scores ($p_T < .5$: $\beta = -.067$, $p = .004$) in 3-year olds. Again at the age of 6 years, the EA polygenic score was associated with lower levels of externalizing problems ($p_T < .5$: $\beta = -.067$, $p = .001$), but not with internalizing problems. Similarly, EA polygenic scores were associated with lower externalizing scores at the age of 10 years ($p_T < .5$: $\beta = -.051$, $p = .027$), but the association at this age did not survive multiple comparison correction. Again at this age, no associations with internalizing scores were observed.

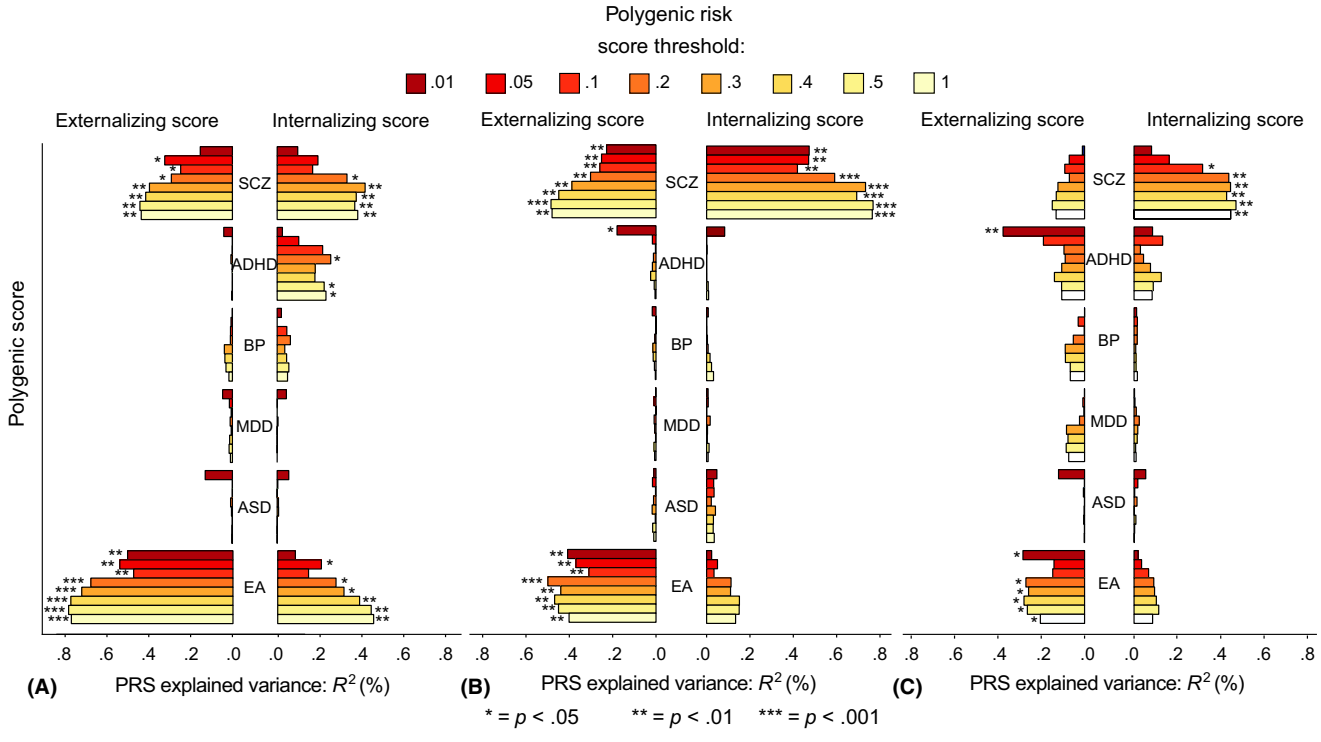

Figure 1 Explained variance (%) in externalizing (A) and internalizing (B) CBCL scores by polygenic scores at the age of 3 years (A) 6 years (B) and 10 years (C). Regression results are corrected for age, gender and four principal components. SCZ, schizophrenia; ADHD, attention-deficit hyperactivity disorder; BP, bipolar disorder; MDD, major depressive disorder; ASD, autism spectrum disorder; EA, educational attainment

Sex interaction

In sensitivity analyses, we tested sex-specific associations of SCZ and EA polygenic scores and the externalizing and internalizing scores; the stratified results are shown in Figure S2A–F. Sex showed a weak interaction with the EA polygenic score on internalizing scores ($p = .026$) and externalizing scores ($p = .024$); however, these results were not significant given the number of tests.

Syndrome scales

To test whether specific syndrome scales of the CBCL were driving the associations between polygenic scores and internalizing and externalizing scores, we performed secondary analyses testing associations between the polygenic scores of SCZ, ADHD and EA and the individual CBCL syndrome scales. Only the *p*-value thresholds that showed the strongest association for SCZ ($p_T < .5$), ADHD ($p_T < .01$) and EA ($p_T < .5$) with the externalizing and internalizing scales in the primary analyses were tested in the secondary analyses. A visual representation of the regression coefficients is shown in Figure 2A–C, and an overview of the full regression results is available in Table S5A–C.

The SCZ polygenic score was mainly associated with higher Emotionally Reactive scores at age 3 ($p_T < .5$, $\beta = .086$, $p < .001$). There was a negative association between the SCZ polygenic score and all internalizing subscales at the age 6, with the strongest association being with Withdrawn scores ($p_T < .5$, $\beta = .072$, $p < .001$). Interestingly, at age 10, there was a strong positive association with Thought Problems scores of the school-aged CBCL version ($p_T < .5$, $\beta = .087$, $p < .001$) (Figure 2C).

As expected, the association between the ADHD polygenic score and externalizing scores at the age 6 was mainly driven by Attention Problems ($p_T < .01$, $\beta = .065$, $p = .002$). At age 10, the ADHD polygenic score was mainly associated with higher levels of aggressive behavior ($p_T < .01$, $\beta = .083$, $p < .001$). The previously observed association with Attention Problems was no longer significant ($p_T < .01$, $\beta = .045$, $p = .051$).

The EA polygenic score showed a strong negative association with Attention Problems scores in 3-year olds ($p_T < .5$, $\beta = -.095$, $p < .001$). This association was observed again at 6 ($p_T < .5$, $\beta = -.082$, $p < .001$) and at age 10 ($p_T < .5$, $\beta = -.082$, $p < .001$) in the school-aged CBCL, in which Attention Problems scores are not part of the externalizing domain.

Discussion

This study presents evidence that in very young children (age 3), a genetic predisposition for psychopathology is associated with more emotional and behavioural problems in the general paediatric population, whereas the polygenic score of EA to lower levels of problem behaviour.

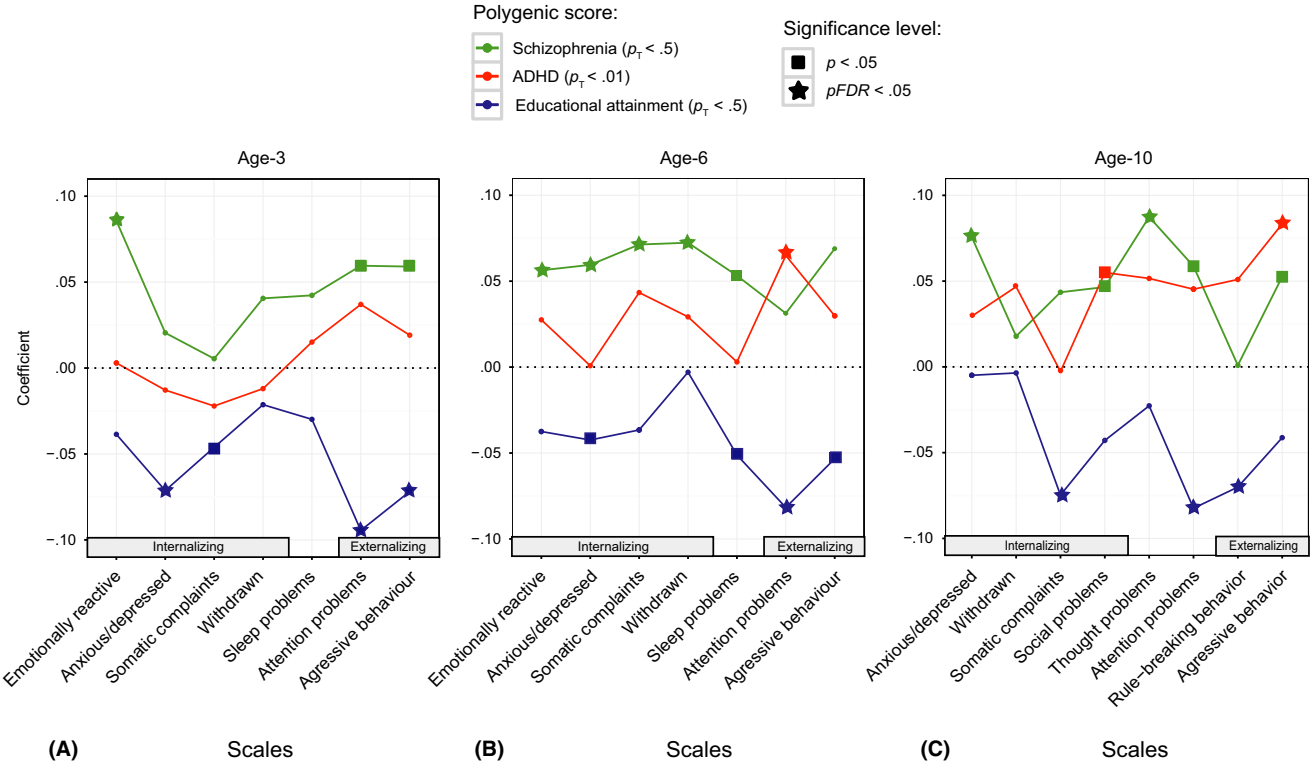

Figure 2 Visual representation of associations between polygenic scores and the specific syndrome scores in the age 3 (A), age 6 (B) and age 10 (C) assessment. Coefficients are standardized betas for the association between polygenic score on the individual syndrome scores, corrected for age, sex and four principal components

The SCZ polygenic score showed an association with internalizing scores from age 3 onwards and was associated with CBCL Thought Problems scores at age 10. Prior research investigating the SCZ polygenic score and problem behaviour has been performed mainly in adolescents (Derks et al., 2012; Jones et al., 2016). Our study shows that genetic predisposition for schizophrenia is associated with variation in behaviour already at an early age, and possibly as early as behaviour can be reliably assessed. This observation is in line with a recent study that reported associations between the SCZ polygenic score and lower cognitive ability, more social impairments and more behavioural problems in 4- to 9-year-old children. (Riglin et al., 2017). Interestingly, this study showed associations with prosocial behaviour and conduct problems at age 4, but no associations with emotional problems were observed. Although we found similar associations with externalizing behaviours, our results add that the genetic predisposition for schizophrenia is associated with higher levels of emotional reactivity in 3-year olds. This discrepancy may be due to more specific items related to the emotional state of the child in the Emotional Reactivity scale of the CBCL.

Given that the incidence of schizophrenia peaks between adolescence and young adulthood, the prominent effect of schizophrenia polygenic scores on internalizing problems at age 6 compared to age 10 was surprising. This suggests that the associations between schizophrenia polygenic risk scores and behaviour in the general population is best captured by the internalizing syndrome scale within the CBCL/$1^{½}$-5 internalizing domain (such as Emotionally Reactive scores) rather than the internalizing scales of the CBCL/6-18, which was used at the assessment at age 10. Differences in the subscale items (e.g. more items related to measures of affect regulation in the CBCL/$1^{½}$-5 Withdrawn scale) may contribute to the different findings between CBCL versions.

The observed association between the SCZ polygenic score and the Thought Problems scale at age 10 is a remarkable new finding and contrasts with two earlier studies in healthy populations that did not find an association between the SCZ polygenic score and positive symptoms (Derks et al., 2012; Jones et al., 2016). The Thought Problems scale, containing items such as 'sees things that aren't there' and 'strange ideas', reflects psychosis-like symptoms and similar behaviour that has previously been found to be a precursor of later life psychosis in prospective studies in similar age groups (Poulton et al., 2000; Welham et al., 2009). Where the two previous studies utilized psychotic symptoms as a binary outcome measure (presence/absence of psychotic experiences), our study aimed to measure psychotic symptoms along a continuum, possibly yielding more power to detect the subtle effects of the polygenic score on psychotic-like experiences in the general population.

We expected to observe similar associations for schizophrenia and bipolar disorder polygenic scores, as the genetic overlap between psychiatric traits is

substantial (Bulik-Sullivan et al., 2015; Purcell et al., 2009). However, these differences are possibly a consequence of differences in sample size of the GWAS for schizophrenia (n = 77,096) compared to bipolar disorder (n = 16,731) and illustrate the importance of a well-powered discovery GWAS for polygenic risk scoring (Dudbridge, 2013). The fact that no associations were observed for ASD and MDD polygenic scores could also be due to the lack of a well-powered GWAS study. However, despite the smallest sample size of the ADHD discovery GWAS (Table S2), we observed evidence for significant associations with Attention Problems and Aggresive Behaviour for one p-value threshold. This finding may be explained by a higher prevalence of ADHD and ADHD-related symptoms in the age range of our study. Moreover, the high specificity of specific CBCL scales for measuring ADHD-related symptoms (such as the Attention Problems scale in the Externalizing Problems scale) may further lead to these observed associations.

Previous studies showed higher ADHD polygenic scores in children with comorbid aggression (Hamshere et al., 2013) and attention problems in the general population (Martin, Hamshere, Stergiakouli, O'Donovan, & Thapar, 2014). Interestingly, we observed that the ADHD polygenic score was related to attention problems in 3- and 6-year olds, but that this association shifted towards more aggression problems at the age of 10 years. While it is possible that these differences are related to the two CBCL versions used, an alternative possibility is that polygenic scores are not necessarily related to a fixed set of symptoms, but related to dynamics and changes during childhood. The association with both the attention and aggression scales could result from a shared genetic aetiology between ADHD symptoms and oppositional defiant disorder (ODD)-related symptoms, which has been described previously in twin research (Tuvblad, Zheng, Raine, & Baker, 2009). Furthermore, distinction between these behavioural scales can be challenging for the parents and may result in classifying attention problems as aggressive behaviour and vice versa.

In our study, genetic variants for educational attainment were negatively associated with externalizing symptoms, which suggests that less externalizing problems early in life may be beneficial for achieving a higher level of education. Indeed, externalizing problems are an important determinant of poor academic performance and have been shown to precede school problems (Van der Ende et al., 2016). Our findings imply a shared genetic aetiology for this association and suggest this is partly explained by higher levels of attention problems. This has also been suggested by prospective studies that showed lower academic achievement in children with symptoms of hyperactivity and inattentiveness (Polderman, Boomsma, Bartels, Verhulst, & Huizink, 2010) and impaired cognitive functioning in individuals with ADHD (Faraone, Ghirardi, Kuja-Halkola, Lichtenstein, & Larsson, 2017). In addition, prior research reported that cognitive ability and behavioural scores are highly intertwined (Blanken et al., 2016) and suggest that the association between educational attainment polygenic scores and lower externalizing problems could result from better cognitive abilities in these children.

Besides a direct association between genetic predisposition and childhood behaviour, the observed associations could partially be explained by the parental polygenic scores: Children with high polygenic scores of psychopathology are more likely to have parents with higher than average polygenic scores. Parental polygenic scores could subsequently lead to differences in environmental factors of the child (e.g. passive gene–environment correlation, including parenting strategies). Studies on this complex interplay suggest that environmental factors such as parenting moderate the associations between polygenic scores and child behaviour (Salvatore et al., 2015). Moreover, studies including genetic data of the mother suggest that the maternal ADHD polygenic score moderates the association between the ADHD polygenic score of the child and educational achievement (Stergiakouli et al., 2017). Applying polygenic scoring in a family-based setting could provide more insight into the dynamic interaction between the genetic profile of the child and the parents, family upbringing such as parenting, and child behaviour.

Our study suggest that the genetic risk for schizophrenia manifests as more internalizing problems, and to lesser degree as more externalizing problems, in children 3 years of age. Recent developmental studies of schizophrenia onset have focused on early puberty and the occurrence of psychotic experiences such as acoustic hallucinations (Zammit et al., 2013). However, based on the current results, we carefully speculate that nonspecific symptoms of emotional reactivity and anxiety may further help to tailor prevention programmes for high-risk children, e.g. as defined by family history.

The strength of the study is that behaviour was assessed at multiple time points, providing information about behaviour during different stages of development. Given that childhood behaviour is dynamic, longitudinal studies are important to study the association between genetic predisposition and behaviour at different ages. Our results illustrate this by showing that associations with behaviour problems were found at specific ages that were not present at an earlier age or disappeared at an older age. Future genetic studies should aim to assess behaviour at multiple time points and study whether changes in behaviour are related to the genetic predisposition of the child.

A limitation of this study is that the analyses were restricted to observations from the primary caregiver. Integration of information from different observers could provide more complete information about the child's behaviour. However, given the young age of

children in our study, we expect scores reported by the primary caregiver to be the most accurate reflection of the children's behaviour. In addition, due to low power of the discovery GWAS study, the observed associations for the ADHD polygenic score were not as robust as those found for schizophrenia and educational attainment. This is illustrated by the observation that associations were only found for the most stringent *p*-value threshold ($p_T < .01$), but lacked broader support from other *p*-value thresholds. In contrast, the associations of schizophrenia and educational attainment showed stronger consistency across multiple *p*-value thresholds.

Conclusion

In conclusion, this study shows that genetic predispositions for psychiatric disorders and educational attainment are associated with early behavioural problems. These associations were present throughout early childhood and at an earlier age than described in most previous studies. Children with a high genetic predisposition for psychiatric traits show specific early manifestations of problem behaviour at a young age, which may further aid the early recognition of precursors of psychopathology in high-risk individuals.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Pearson correlations between PRS for schizophrenia, ADHD, bipolar disorder, major depressive disorder, autism spectrum disorder and educational attainment.

Figure S2. Interaction effect of sex with polygenic scores of educational attainment (EA) and schizophrenia (SCZ) on internalizing and externalizing CBCL scores in the linear regression model at the three time points. Results are corrected for age, and four principal components.

Table S1. Overview of syndrome scores and number of items in the CBCL/1½-5 and CBCL/6-18 questionnaires.

Table S2. Overview of genome-wide association studies used for the calculation of the polygenic risk scores.

Table S3. The final number of SNPs that were included in the polygenic score for each trait after clumping the GWAS results.

Table S4A. Regression results for polygenic risk scores and externalizing and internalizing scores at each *p*-value thresholds during the assessment at the age of 3 years.

Table S4B. Regression results for polygenic risk scores and externalizing and internalizing scores at each *p*-value thresholds during the assessment at the age of 6 years.

Table S5A. Regression results of individual syndrome scales of the CBCL/1½-5 and polygenic risk scores during the assessment at 3 years.

Table S5B. Regression results of individual syndrome scales of the CBCL/1½-5 and polygenic risk scores during the assessment at 6 years.

Table S5C. Regression results of individual syndrome scales of the CBCL/6-18 and polygenic risk scores during the assessment at 10 years.

Acknowledgements

This work was supported by the Sophia Foundation for Scientific Research (SSWO project number: S14-27). The first phase of the Generation R Study is made possible by financial support (Grant number: ZonMW 10.000.1003). The Generation R Study is conducted by the Erasmus Medical Center in close collaboration with the School of Law and Faculty of Social Sciences of the Erasmus University Rotterdam, the Municipal Health Service Rotterdam area, Rotterdam, the Rotterdam Homecare Foundation, Rotterdam and the Stichting Trombosedienst & Artsenlaboratorium Rijnmond (STAR-MDC), Rotterdam. F.C.V. is a contributing author of the Achenbach System of Empirically Based Assessment (ASEBA), for which he has received remuneration. The remaining authors declare that they have no competing or potential conflicts. The authors thank the participants and their parents that contributed to the Generation R Study, and the researchers who were involved in data collection.

Correspondence

Henning Tiemeier, Department of Epidemiology, Erasmus Medical Center, Rotterdam, PO Box 2040 3000 CA, Rotterdam, The Netherlands; Email: h.tiemeier@erasmusmc.nl

Key points

- We tested whether genetic predisposition for psychiatric disorders and educational attainment was related to child emotional and behaviour problem at the ages of 3, 6 and 10 years.
- The polygenic risk score of schizophrenia was associated with more internalizing problems from age 3 onwards, more externalizing problems at age 3 and 6, and more Thought Problems at age 10.
- A higher polygenic score of educational attainment was associated with lower externalizing and internalizing problems and less attention problems.
- The ADHD polygenic score was associated with more attention problems in younger children, and showed an association with more aggression problems scores in older children.

References

Achenbach, T.M., & Rescorla, L.A. (2000). *Manual for ASEBA preschool forms & profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth & Families.

Basten, M., Tiemeier, H., Althoff, R.R., van de Schoot, R., Jaddoe, V.W., Hofman, A., … & van der Ende, J. (2016). The stability of problem behavior across the preschool years: An empirical approach in the general population. *Journal of Abnormal Child Psychology*, *44*, 393–404.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)*, *57*, 289–300.

Benke, K.S., Nivard, M.G., Velders, F.P., Walters, R.K., Pappa, I., Scheet, P.A., … & Middeldorp, C.M. (2014). A genome-wide association meta-analysis of preschool internalizing problems. *Journal of the American Academy of Child and Adolescent Psychiatry*, *53*, 667–676.

Blanken, L.M., White, T., Mous, S.E., Basten, M., Muetzel, R.L., Jaddoe, V.W., … & Tiemeier, H. (2016). Cognitive functioning in children with internalising, externalising and dysregulation problems: A population-based study. *European Child and Adolescent Psychiatry*, *26*, 445–456.

Breslau, J., Miller, E., Breslau, N., Bohnert, K., Lucia, V., & Schweitzer, J. (2009). The impact of early behavior disturbances on academic achievement in high school. *Pediatrics*, *123*, 1472–1476.

Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., … & Neale, B.M. (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, *47*, 1236–1241.

Caspi, A., Moffitt, T.E., Newman, D., & Silva, P.A. (1996). Behavioral observations at age 3 years predict adult psychiatric disorders. Longitudinal evidence from a birth cohort. *Archives of General Psychiatry*, *53*, 1033–1039.

Derks, E.M., Vorstman, J.A., Ripke, S., Kahn, R.S., & Ophoff, R.A. (2012). Investigation of the genetic association between quantitative measures of psychosis and schizophrenia: A polygenic risk score analysis. *Public Library of Science One*, *7*, e37852.

Delaneau, O., Marchini, J., & Zagury, J.F. (2012). A linear complexity phasing method for thousands of genomes. *Nature Methods*, *9*, 179–181.

Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *Public Library of Science Genetics*, *9*, e1003348.

Euesden, J., Lewis, C.M., & O'Reilly, P.F. (2015). PRSice: Polygenic risk score software. *Bioinformatics*, *31*, 1466–1468.

Faraone, S.V., Ghirardi, L., Kuja-Halkola, R., Lichtenstein, P., & Larsson, H. (2017). The familial co-aggregation of attention-deficit/hyperactivity disorder and intellectual disability: A register-based family study. *Journal of the American Academy of Child and Adolescent Psychiatry*, *56*, 167–174.

Hamshere, M.L., Langley, K., Martin, J., Agha, S.S., Stergiakouli, E., Anney, R.J., … & Thapar, A. (2013). High loading of polygenic risk for ADHD in children with comorbid aggression. *American Journal of Psychiatry*, *170*, 909–916.

Howie, B.N., Donnelly, P., & Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *Public Library of Science Genetics*, *5*, e1000529.

Jones, H.J., Stergiakouli, E., Tansey, K.E., Hubbard, L., Heron, J., Cannon, M., … & Zammit, S. (2016). Phenotypic manifestation of genetic risk for schizophrenia during adolescence in the general population. *Journal of the American Medical Association Psychiatry*, *73*, 221–228.

Kessler, R.C., Foster, C.L., Saunders, W.B., & Stang, P.E. (1995). Social consequences of psychiatric disorders, I: Educational attainment. *American Journal of Psychiatry*, *152*, 1026–1032.

King, S.M., Iacono, W.G., & McGue, M. (2004). Childhood externalizing and internalizing psychopathology in the prediction of early substance use. *Addiction*, *99*, 1548–1559.

Kooijman, M.N., Kruithof, C.J., van Duijn, C.M., Duijts, L., Franco, O.H., van IJzendoorn, M.H., … & Jaddoe, V.W. (2016). The Generation R Study: Design and cohort update 2017. *European Journal of Epidemiology*, *31*, 1243–1264.

Krapohl, E., Euesden, J., Zabaneh, D., Pingault, J.B., Rimfeld, K., von Stumm, S., … & Plomin, R. (2016). Phenome-wide analysis of genome-wide polygenic scores. *Molecular Psychiatry*, *21*, 1188–1193.

Martin, J., Hamshere, M.L., Stergiakouli, E., O'Donovan, M.C., & Thapar, A. (2014). Genetic risk for attention-deficit/hyperactivity disorder contributes to neurodevelopmental traits in the general population. *Biological Psychiatry*, *76*, 664–671.

Medina-Gomez, C., Felix, J.F., Estrada, K., Peters, M.J., Herrera, L., Kruithof, C.J., … & Rivadeneira, F. (2015). Challenges in conducting genome-wide association studies in highly admixed multi-ethnic populations: The Generation R Study. *European Journal of Epidemiology*, *30*, 317–330.

Middeldorp, C.M., Hammerschlag, A.R., Ouwens, K.G., Groen-Blokhuis, M.M., St Pourcain, B., Greven, C.U., … & Boomsma, D.I. (2016). A genome-wide association meta-analysis of attention-deficit/hyperactivity disorder symptoms in population-based pediatric cohorts. *Journal of the American Academy of Child and Adolescent Psychiatry*, *55*, 896–905.

Okbay, A., Beauchamp, J.P., Fontana, M.A., Lee, J.L., Pers, T.H., Rietveld, C.A., … & Benjamin, D.J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*, 539–542.

Pappa, I., St Pourcain, B., Benke, K., Cavadino, A., Hakulinen, C., Nivard, M.G., … & Tiemeier, H. (2015). A genome-wide approach to children's aggressive behavior: The EAGLE consortium. *American Journal of Medical Genetics Part B Neuropsychiatric Genetics*, *17*, 562–572.

Patterson, N., Price, A.L., & Reich, D. (2006). Population structure and eigenanalysis. *Public Library of Science Genetics*, *2*, e190.

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

Polderman, T.J., Boomsma, D.I., Bartels, M., Verhulst, F.C., & Huizink, A.C. (2010). A systematic review of prospective studies on attention problems and academic achievement. *Acta Psychiatrica Scandinavica*, *122*, 271–284.

Poulton, R., Caspi, A., Moffitt, T.E., Cannon, M., Murray, R., & Harrington, H. (2000). Children's self-reported psychotic symptoms and adult schizophreniform disorder: A 15-year longitudinal study. *Archives of General Psychiatry*, *57*, 1053–1058.

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

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., … & Sham, P.C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*, 559–575.

Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., & Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, *460*, 748–752.

R Core Team (2014). *R: A language and environment for statistical computing*. Viena, Australia: R foundation for Statistical Computing.

Riglin, L., Collishaw, S., Richards, A., Thapar, A.K., Maughan, B., O'Donovan, M.C., & Thapar, A. (2017). Schizophrenia risk alleles and neurodevelopmental outcomes in childhood: A

population-based cohort study. *The Lancet Psychiatry*, *4*, 57–62.

Sakia, R.M. (1992). The Box-Cox transformation technique: A review. *The Statistician*, *41*, 169–178.

Salvatore, J.E., Aliev, F., Bucholz, K., Agrawal, A., Hesselbrock, V., Hesselbrock, M., … & Dick, D.M. (2015). Polygenic risk for externalizing disorders: Gene-by-development and gene-by-environment effects in adolescents and young adults. *Clinical Psychological Science*, *3*, 189–201.

Stergiakouli, E., Martin, J., Hamshere, M.L., Heron, J., St Pourcain, B., Timpson, N.J., … & Davey Smith, G. (2017). Association between polygenic risk scores for attention-deficit hyperactivity disorder and educational and cognitive outcomes in the general population. *International Journal of Epidemiology*, *46*, 421–428.

Sullivan, P.F., Daly, M.J., & O'Donovan, M. (2012). Genetic architectures of psychiatric disorders: The emerging picture and its implications. *Nature Reviews Genetics*, *13*, 537–551.

Tiemeier, H., Velders, F.P., Szekely, E., Roza, S.J., Dieleman, G., Jaddoe, V.W., … & Verhulst, F.C. (2012). The Generation R Study: A review of design, findings to date, and a study of the 5-HTTLPR by environmental interaction from fetal life onward. *Journal of the American Academy of Child and Adolescent Psychiatry*, *51*, 1119–1135.

Tuvblad, C., Zheng, M., Raine, A., & Baker, L.A. (2009). A common genetic factor explains the covariation among ADHD ODD and CD symptoms in 9–10 year old boys and girls. *Journal of Abnormal Child Psychology*, *37*, 153–167.

van der Valk, J.C., van den Oord, E.J., Verhulst, F.C., & Boomsma, D.I. (2003). Genetic and environmental contributions to stability and change in children's internalizing and externalizing problems. *Journal of the American Academy of Child and Adolescent Psychiatry*, *42*, 1212–1220.

Van der Ende, J., Verhulst, F.C., & Tiemeier, H. (2016). The bidirectional pathways between internalizing and externalizing problems and academic performance from 6 to 18 years. *Development and psychopathology*, *28*, 855–867.

Ward, M.E., McMahon, G., St Pourcain, B., Evans, D.M., Rietveld, C.A., Benjamin, D.J., … & Timpson, N.J. (2014). Genetic variation associated with differential educational attainment in adults has anticipated associations with school performance in children. *Public Library of Science One*, *9*, e100248.

Welham, J., Scott, J., Williams, G., Najman, J., Bor, W., O'Callaghan, M., & McGrath, J. (2009). Emotional and behavioural antecedents of young adults who screen positive for non-affective psychosis: A 21-year birth cohort study. *Psychological Medicine*, *39*, 625–634.

Zammit, S., Kounali, D., Cannon, M., David, A.S., Gunnell, D., Heron, J., … & Lewis, G. (2013). Psychotic experiences and psychotic disorders at age 18 in relation to psychotic experiences at age 12 in a longitudinal population-based cohort study. *American Journal of Psychiatry*, *170*, 742–750.

Accepted for publication: 6 April 2017
First published online: 19 June 2017