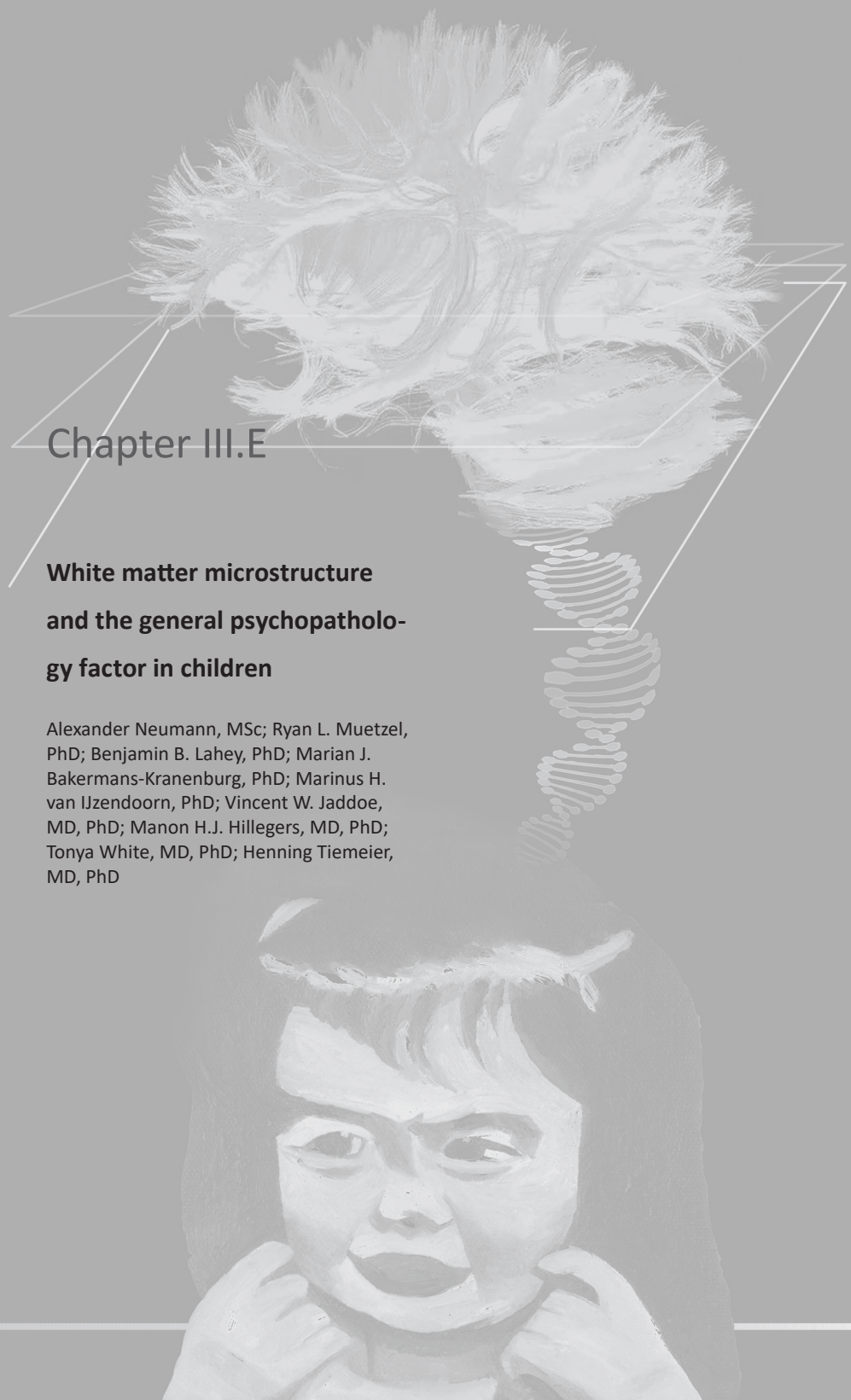


Chapter III.E

White matter microstructure and the general psychopathology factor in children

Alexander Neumann, MSc; Ryan L. Muetzel, PhD; Benjamin B. Lahey, PhD; Marian J. Bakermans-Kranenburg, PhD; Marinus H. van IJzendoorn, PhD; Vincent W. Jaddoe, MD, PhD; Manon H.J. Hillegers, MD, PhD; Tonya White, MD, PhD; Henning Tiemeier, MD, PhD

ABSTRACT

Objective: Co-occurrence of behavioral and emotional problems in childhood is widespread and previous studies have suggested that this reflects vulnerability to experience a range of psychiatric problems, often termed a general psychopathology factor. However, the neurobiological substrate of this general factor is not well understood. We tested the hypothesis that lower overall white matter microstructure is associated with higher levels of the general psychopathology factor in children and less with specific factors.

Method: Global white matter microstructure at age 10 years was related to general and specific psychopathology factors. These factors were estimated using a latent bifactor model with multiple informants and instruments between ages 6-10 years in 3030 children from the population-based birth cohort Generation R. The association of global white matter microstructure and the psychopathology factors was examined with a structural equation model adjusted for sex, age at scan, age at psychopathology assessment, parental education/income and genetic ancestry.

Results: A 1-standard deviation (SD) increase of the global white matter factor was associated with a β=-0.07SD (SE=0.02, $p<0.01$) decrease in general psychopathology. In contrast, a 1-SD increase of white matter microstructure predicted an increase of β=+0.07SD (SE=0.03, $p<0.01$) specific externalizing factor levels. No association was found with the specific internalizing and specific attention factor.

Conclusions: The results suggest that general psychopathology in childhood is related to white matter structure across the brain and not only to specific tracts. Taking into account general psychopathology may also help reveal neurobiological mechanisms behind specific symptoms which are otherwise obscured by comorbidity.

INTRODUCTION

Child psychological problems are commonly grouped into behavioral (externalizing) and emotional (internalizing) problems based on the observation that symptoms within a given domain often co-occur. Nonetheless, it is well known that even across these broadly defined domains, the symptoms correlate substantially.[1,2] Likewise, categorically defined psychiatric disorders co-occur above chance level.[3,4] In recent years, studies in children,[5–7] adolescents,[8,9] and adults[10,11] have suggested that this broadly shared variance can be described by a latent construct which underlies all psychiatric problems: a general psychopathology factor. These studies consistently support the hypothesis that co-occurence of psychiatric problems is explained by both a general propensity to have any problem and by a specific propensity to display characteristics of a certain psychopathology domain.[12]

The question then arises whether this higher order structure of psychopathology is also mirrored in the brain.[12] Zald and Lahey[13] propose a framework in which some brain features underlie the risk to experience any psychiatric problems, while other neural circuits are linked to the occurrence of specific symptoms. One possibility is that global brain characteristics reflect a non-specific psychopathology risk. White matter microstructure, believed to serve as the backbone for fast efficient neural communication, is a possible candidate substrate.

White matter microstrucutre encompasses several neural characteristics important for providing structural connectivity, such as axonal properties and degree of myelination. These characteristics are determined by genetic and environmental factors.[14] White matter differences across several regions were associated with a variety of psychological and psychiatric outcomes, such as IQ and visuospatial abilities,[15] early-onset schizophrenia and bipolar disorder,[16] ADHD[17], anxiety and depression.[18,19] Most studies have only tested the effects of specific tracts. However, given the diversity of tracts identified, the question arises to what extent these associations represent effects of global variation of white matter across the brain. Though the literature is sparse, studies examining whole-brain metrics have demonstrated that global white matter microstructure is associated with cognitive abilities in children,[15] depression in adults,[19] as well as attention and internalizing problems in children born preterm.[20]

These studies of global white matter microsture used traditional definitions of single disorders/domains and did not distinguish between general and specific associations. The use of a latent general psychopathology factor may help better characterize the extent to which white matter microstructure is associated with a general vulnerability to psychopathology. At the same time, specific psychopathology factors can be tested. Internalizing or externalizing factors, which are uncorrelated to the general psychopa-

thology factor, may help to identify links between white matter tracts and specific psychopathology domains.

Against this background, we hypothesized that lower global white matter microstructure across the brain is associated with higher levels of the general psychopathology factor and less with specific psychopathology factors. To test this hypothesis, we measured global white matter microstructure using diffusion tensor imaging (DTI) in 10-year-old children participating in the Generation R Study (GenR), a population-based birth cohort. Global white matter microstructure was quantified as a latent construct, reflecting white matter microstructure of 12 measured tracts. We repeatedly assessed common psychological problems from ages 6-to-10 using mother, father, teacher and child reports, and subsequently estimated general and specific psychopathology factors.

METHOD

Participants and ethical considerations

This study was embedded in GenR,[21,22] a population-based birth cohort. All parents gave informed consent for their children's participation. GenR is conducted in accordance with the Declaration of Helsinki and study protocols have been approved by the Ethics Committee of the Erasmus Medical Center.

Usable DTI scans were available for 3050 children. At least one psychological problem subscale was available for 3030 children. All results are based on this sample of 3030 children, except for the results of the tract-based spatial statistics (TBSS)[23] analysis (n=2996), which required additional quality control (Figure S1). Descriptive statistics can be found in Table 1. A full method description can be found in the Supplementary Methods.

Measures

Child psychological problems

We used the Child Behavior Checklist (CBCL) 1 ½ -5 years[24] to assess child behavioral problems at age 6 years (mean=5.9, SD=0.3) and the CBCL 6-1825 at age 10 years (mean=10, SD=0.3). At the age of 6 years, questionnaires were completed by the primary caregiver (92% mothers). At age 10 years, the questionnaire was filled in by mother and father separately. Teachers assessed children at age 7 years (mean=6.5, SD=1.1) with the Teacher's Rating Form 6-18 years[25]. At age 6 years (mean=6.0, SD=0.4) we conducted the Berkeley Puppet Interview,[26] a semi-structured interactive child interview,

at our research center. At age 10 years (mean=9.8, SD=0.3) years the children rated their problems with the Brief Problem Monitor[27] plus items related to thought problems.

Diffusion Tensor Imaging

Children underwent diffusion tensor imaging at age 10 years (mean=10.1 years, SD=0.6). MRI scans were performed using a 3T General Electric scanner with an 8-channel head coil. Diffusion tensor imaging consisted of a 35-direction echo planar imaging sequence (TR=12,500ms, TE=72ms, FoV=240mm*240mm, acquisition matrix=120*120, slice thickness=2mm, slice number=65, Asset Acceleration Factor=2, b=900s/mm2, 3 b=0 images). We computed fractional anisotropy (FA), mean diffusivity (MD), radial diffusivity (RD), and axial diffusivity (AD).[28] Connectivity distributions for 12, large well-defined and widely reported fiber bundles were derived with probabilistic fiber

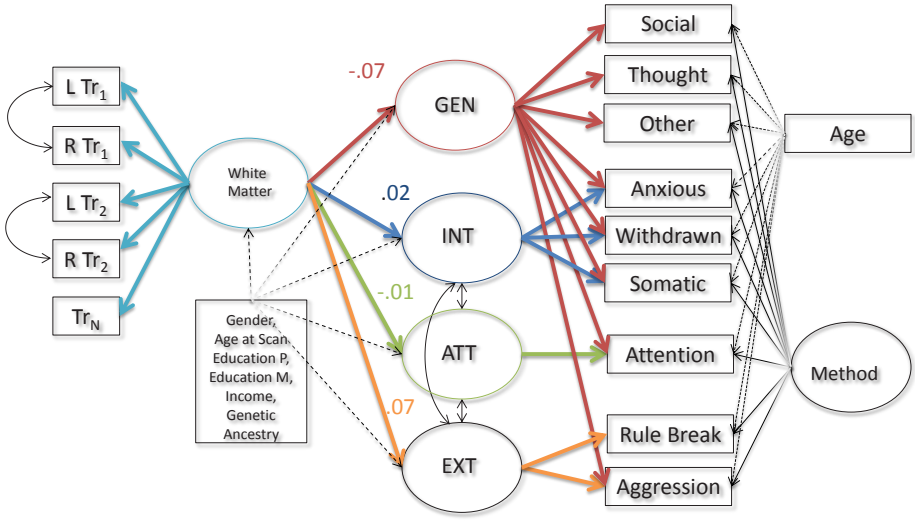


Figure 1: Abbreviated path diagram of the main analysis model. All latent variables (oval shape) are included. Observed variables (square) from the CBCL at age 10 by a single informant are displayed as an example. Observed variables from other instruments and informants, as well as specific tracts were omitted. Numbers displayed are standardized regression coefficients. L Tr = left tract, R Tr = right tract, White matter = Global white matter integrity, Education P = Paternal education, Education M = Maternal education, GEN = General Psychopathology, INT = Internalizing, ATT = Attention, EXT = Externalizing.

tractography.[29,30] For TBSS analyses the DTI images were registered to a study-specific and age-appropriate template.[31]

Measures of IQ, school performance, temperament, happiness and parental psychopathology

We assessed non-verbal IQ with the Snijder-Oomen nonverbal intelligence test at age 6.[32] At the same age we measured temperamental dimensions with the Child Behavior Questionnaire (Very-Short-Form), a parent-rated questionnaire.[33] School performance was assessed by the Cito, a standardized exam at the end of primary school in which language and math skills are tested.[34] Happiness was measured by asking the parents at age 10: "How often was your child happy in the past 4 weeks?". Parental psychopathology was assessed with the Brief Symptom Inventory.35

Statistical Analysis

We used a structural equation model to associate global white matter microstructure with general and specific factors of psychopathology (Figure 1). All models were fitted in R 3.4.136 with the package Lavaan 0.5-23.1109737. We used a maximum likelihood estimator with robust standard errors (MLM) to account for multivariate non-normality.

The general psychopathology factor

The general psychopathology factor was specified to underlie all problem subscales from all instruments and time-points (Table S1). The subscales were also specified to load on one of the specific internalizing, the specific externalizing, or the specific attention scales defined on the basis of the assessment scales. Therefore, the attention subscales from each informant, for example, loaded on the general as well as specific attention factor. Specific psychopathology factors were allowed to correlate among each other, but not with the general psychopathology factor. The specific factors thus represent covariance among subscales that cannot be explained by a general propensity for psychiatric problems. As such the specific factors differ distinctly from the observed broadband scales, e.g. the mother-rated CBCL internalizing and externalizing scores had a correlation of +0.54 (SE=0.01, $p<0.01$), but the specific externalizing and internalizing factors do not correlate positively (r=-0.36, SE=0.04, $p<0.01$) because the shared variance is already captured by the general factor.11 The same holds for the specific attention-specific internalizing correlation (r=-0.47, SE=0.03, $p<0.01$). The specific attention and externalizing factors did not correlate with each other (r=+0.06, SE=0.03, p=0.06). The higher order structure of the model tested can be seen as an extension of the instruments' established factor structure which informed the computation of

the subscales. Importantly, we included additional method factors, which capture the shared variance unique to an informant at a certain age of the child.

The latent factor structure was based on a previous GenR study on 6-to-8 year old children,[6] as well as on models validated in other cohorts[8,11,12]. We performed several additional analyses for further validation and characterization. First, we tested models without a general psychopathology factor (Table S2) and observed a substantial decrease of fit. However, we present association results from a three-factor model (internalizing, externalizing and attention) without the general psychopathology factor as comparison. Second, we explored four models with IQ at age 6, temperament (negative affectivity, surgency, and effortful control measured at age 6), school performance at the end of primary school and happiness as predictors of the psychopathology factors. We previously had associated the IQ and temperament variables with general and specific psychopathology factors in younger children6. As IQ and temperament showed discriminate associations, they can therefore help interpret the latent psychopathology factors. Third, we performed sensitivity analyses adjusted for total intracranial volume, as larger volume is associated with higher FA (e.g. through partial volume effects) and thus could confound global white matter associations.[38] As there is no consensus in the field as to the utility of adjusting DTI scalar metrics for intracranial volume, we present these results as sensitivity analyses.

The global white matter microstructure factor

The global white matter microstructure factor was estimated using the mean FA values of 12 white matter tracts as indicators (Table S3). FA describes how elongated the ellipsoid shape of a diffusion pattern is, with higher values suggesting higher white matter integrity (referred to as higher white matter microstructure in this paper). This model was based on previous studies using GenR data.[15,30] We included the corticospinal tract, inferior longitudinal fasciculus, superior longitudinal fasciculus and uncinate fasciculus of each hemisphere separately in the model. While FA it is good summary measure of white matter microstructure, it can be also helpful to examine diffusivity only perpendicular to the main axis of diffusion (RD) or only alongside it (AD). Higher RD and lower AD are associated with less white matter microstructure. Thus, to better understand results from the FA model, we also estimated global white matter variables based on RD, AD, as well as MD (the average diffusivity in any direction)

Structural Paths

The general and specific psychopathology factors were each simultaneously regressed on the global white matter factor based on FA values to test the associations between white matter microstructure and psychopathology. Figure 1 illustrates the main model, the only model used to test the main hypothesis. All other statistical models were used for exploratory purposes to better interpret the results of the main model. We adjusted all models for several potential confounders in the model, namely sex,

age at scan, assessment age, maternal and paternal education at age 6, income at age 6, and genetic ancestry. All subsequent coefficients are reported as standardized estimates. Since IQ is related to global white matter microstructure,[15] we explored whether associations were specific to psychopathology by including child IQ as a covariate. Additionally, we controlled for maternal/paternal psychopathology (interpersonal sensitivity, depression, anxiety and hostility) to explore to what extent the association is independent of parental characteristics and also to control for potentially remaining rater bias as parents completed several psychopathology measures. We tested for non-linear associations, by fitting a standard regression model with estimated factor scores analogous to the structural paths of the main model, but with the addition of a squared term for the global white matter factor score. We reran the main model with global white matter factor based on MD, RD and AD. P-values were adjusted for multiple testing of four outcomes using false discovery rate (FDR).

Follow-up Analyses

We ran follow-up analyses to investigate the individual contribution of each individual white matter tract by replacing the latent variable global white matter microstructure with the observed FA of a single tract. These follow up analyses had two goals: 1. to test which tracts underlie any observed association with global white matter microstructure and 2. to increase comparability with studies reporting single tract associations. In these exploratory follow-up analyses we computed FDR adjusted p-values for 12 tracts per 4 outcomes (48 tests).

In follow-up analyses, we tested individual voxels (nvoxels=9272) in a TBSS analysis for outcomes associated with global FA values. In these follow-up analyses to the global white matter models, we performed TBSS analyses for outcomes associated with global FA values. We present results for other DTI scalar metrics on a voxel-wise level, if the scalar was significant on a global level. TBSS was performed in FSL39 using a 2mm3 resolution. Adjustment for multiple testing was achieved with permutation testing (nperm=5000) and clusters were formed using the built-in threshold-free cluster enhancement.[40] As FSL does not support latent variables, we estimated psychopathology factor scores based on the main model and adjusted for the same covariates as in the main model.

Measurement invariance

The main analyses are based on the assumption that the latent constructs and associations between them are identical across sex, ancestry (European vs non-European) and socioeconomic status. We tested this assumption by performing measurement invariance analyses of the main model. To this aim we sequentially constrained an increas-

Table 1: Demographics of analysis sample (n=3,030)

	n_{obs}	%
Sex	3030	
Girls	1528	50.4
Household income	2541	
<2800€	835	32.9
2800-4800€	1079	42.5
>4800€	627	24.7
Maternal education	2659	
No or Primary	86	3.2
Secondary	908	34.1
Higher	1665	62.6
Paternal education	2456	
No or Primary	109	4.4
Secondary	833	33.9
Higher	1514	61.6
Genetic ancestry	1889	
Northwestern European	1136	60.1
Child IQ	n_{obs}	mean (SD)
Score	2640	103.3 (14.8)

n_{obs} observed sample size, **SD** standard deviation

Table 2: Psychopathology factors regressed on IQ and temperament

Predictor	n_{obs}	Factor											
		General			Specific Ext			Specific Int			Specific Att		
		β	SE	p	β	SE	p	β	SE	p	β	SE	p
Child IQ	2640	-.12	.02	<.01	-.05	.03	.06	.03	.03	.21	-.15	.02	<.01
School performance	1311	-.12	.02	<.01	-.06	.03	.02	.03	.03	.33	-.31	.02	<.01
Negative Affectivity	2329	.40	.02	<.01	.12	.03	<.01	.15	.03	<.01	-.06	.02	<.01
Surgency	2326	.11	.02	<.01	.20	.02	<.01	-.50	.02	<.01	.18	.02	<.01
Effortful Control	2321	-.13	.02	<.01	-.10	.02	<.01	.06	.02	<.01	-.17	.02	<.01
Happiness	2117	-.23	.02	<.01	-.12	.02	<.01	-.15	.03	<.01	.05	.02	.03
3 Factor Model			-		Externalizing			Internalizing			Attention		
Child IQ	2640	-	-	-	-.10	.02	<.01	-.01	.02	.77	-.18	.02	<.01
School performance	1311	-	-	-	-.12	.03	<.01	.00	.03	.88	-.33	.02	<.01
Negative Affectivity	2329	-	-	-	.26	.02	<.01	.26	.02	<.01	.09	.02	<.01
Surgency	2326	-	-	-	.22	.02	<.01	-.45	.02	<.01	.20	.02	<.01
Effortful Control	2321	-	-	-	-.15	.02	<.01	.01	.02	.61	-.21	.02	<.01
Happiness	2117	-	-	-	-.21	.02	<.01	-.21	.03	<.01	-.05	.02	.02

IQ Intelligence Quotient, **Ext** Externalizing, Int Internalizing, **Att** Attention, **β** Standardized regression coefficient, **SE** standard error
$\mathbf{n}_{obs}$ observed sample size, analysis n = 3,030

Table 3: Psychopathology factors regressed on white matter microstructure

		Factor															
		General				Specific Ext				Specific Int				Specific Att			
Model Name	**H**	**β**	**SE**	**p**	**q**	**β**	**SE**	**p**	**q**	**β**	**SE**	**p**	**q**	**β**	**SE**	**p**	**q**
Global FA (main model)	-	-.07	.02	<.01	<.01	.07	.03	.01	.02	.02	.03	.36	.48	-.01	.02	.77	.77
- *adjusted for: Child IQ and parental psycho-pathology*	-	-.06	.02	<.01	.01	.07	.03	<.01	.01	.02	.03	.47	.60	.01	.02	.60	.60
- *adjusted for: Total Intracranial Volume*	-	-.06	.02	<.01	<.01	.07	.03	<.01	.02	.03	.03	.31	.41	.01	.02	.66	.66
Global MD	-	.01	.02	.78	.78	-.06	.03	.02	.08	.01	.03	.65	.78	-.04	.02	.08	.16
Global RD	-	.03	.02	.12	.24	-.07	.03	.01	.04	-.01	.03	.80	.80	-.02	.02	.37	.49
Global AD	-	-.04	.03	.10	.20	-.02	.03	.39	.39	.03	.03	.39	.39	-.06	.03	.03	.12
Indiviual tracts (FA)																	
Cingulum bundle	L	.02	.02	.38	.65	.03	.02	.25	.56	-.03	.03	.31	.62	.02	.02	.37	.65
	R	.00	.02	.97	.99	.01	.02	.55	.71	-.02	.02	.43	.67	.01	.02	.50	.61
Corticospinal tract	L	-.05	.02	.02	.14	.07	.02	<.01	.03	.01	.03	.69	.82	.00	.02	.86	.90
	R	-.05	.02	.01	.06	.07	.02	<.01	.03	.00	.02	.86	.90	.01	.02	.55	.71
Forceps major	-	-.05	.02	.01	.06	.05	.02	.04	.17	.01	.02	.55	.71	.00	.02	.85	.90
Forceps minor	-	-.07	.02	<.01	.03	.04	.02	.07	.24	-.01	.02	.82	.90	-.01	.02	.77	.87
Inferior longitudi-nal fasciculus	L	-.04	.02	.08	.24	.02	.02	.31	.62	.04	.02	.09	.26	-.02	.02	.38	.65
	R	-.04	.02	.04	.17	.01	.02	.70	.82	.03	.02	.26	.56	-.01	.02	.58	.71
Superior longitu-dinal fasciculus	L	-.03	.02	.12	.33	.05	.02	.02	.12	.02	.02	.36	.65	.01	.02	.53	.71
	R	-.06	.02	<.01	.03	.03	.02	.20	.49	.03	.02	.20	.49	-.02	.02	.41	.67

Table 3: (Continued)

		Factor															
		General				Specific Ext				Specific Int				Specific Att			
Model Name	**H**	**β**	**SE**	**p**	**q**	**β**	**SE**	**p**	**q**	**β**	**SE**	**p**	**q**	**β**	**SE**	**p**	**q**
Uncinate fascic-	L	-.03	.02	.08	.24	.04	.02	.05	.19	-.01	.02	.56	.71	.00	.02	.99	.99
ulus	R	-.04	.02	.04	.17	.05	.02	.04	.17	.02	.02	.52	.71	-.02	.02	.42	.67
3 Factor Model			-			Externalizing				Internalizing				Attention			
Global FA	-	-	-	-	-	.03	.02	.28	-	.01	.03	.71	-	-.04	.02	.12	-

White matter microstructure based on fractional anisotropy, unless otherwise noted. All models are adjusted for sex, age at scan, age at psychopathology assessment, maternal and paternal education, household income and genetic ancestry (n = 3,030).

Ext Externalizing, **Int** Internalizing, **Att** Attention, **h** Hemisphere, **β** Standardized regression coefficient, **SE** standard error, **q** false discovery rate adjusted p-values,
FA Fractional Anisotropy, **MD** mean diffusivity, **RD** radial diffusivity, **AD** axial diffusivity

ing number of parameters. Constraints were judged to significantly worsen fit when the robust CFI dropped by more than 0.01.[41] See Table S8-S10 for models tested.[42,43]

RESULTS

Latent variable loadings

The FA score of all white matter tracts loaded on global white matter microstructure. Loadings ranged from 0.41 (cingulum bundle) to 0.74 (Superior longitudinal fasciculus) (Table S3). Differences in loadings between left and right hemispheres were small, thus both hemispheres contributed about equally to the global white matter construct.

All problem subscales had statistically significant loadings on the general psychopathology factor. Most loadings were moderate to high, in the range of .30 to .70, but some teacher and child self-report loadings at age 6-7 were below .20, see Table S1. The general psychopathology factor model fitted better than the models without the general factor (Table S2). The loadings of the observed problem subscales on the specific factors tended to be lower than on the general factor.

IQ, school performance, temperament and happiness

Children with a higher IQ had lower general psychopathology levels (β=-0.12, SE=0.02, p<0.01) and less specific attention problems (β=-0.15, SE=0.02, p<0.01), but not more or less specific externalizing and internalizing problems (Table 2). Those who performed well at school had lower general psychopathology levels (β=-0.12, SE=0.02, p<0.01), less specific externalizing problems (β=-0.06, SE=0.03, p=0.02) and less specific attention problems (β=-0.31, SE=0.02, p<0.01). Children who scored high on negative affectivity had particularly high levels of the general psychopathology factor (β=+0.40, SE=0.02, p<0.01). Associations of the negative affectivity score with the specific psychopathology factors were much weaker. A different pattern of associations was observed for surgency. Children with higher levels of surgency had lower specific internalizing levels (β=-0.50, SE=0.02, p<0.01) and higher specific externalizing levels (β=+0.20, SE=0.02, p<0.01). Associations of effortful control with all factors were weak. Happier children had lower levels of general psychopathology (β=-0.23, SE=0.02, p<0.01), lower levels of specific externalizing levels (β=-0.12, SE=0.02, p<0.01), and lower levels of specific internalizing levels (β=-0.15, SE=0.03, p<0.01).

Psychopathology factors associations with white matter microstructure

Table 3 summarizes the results of the global white matter microstructure analyses. In the three-factor model, which did not include the general factor, no associations between white matter microstructure and the traditionally defined psychopathology domains (externalizing, internalizing and attention) were found. Next we included the general psychopathology factor in the model. A 1-SD increase of the global white matter

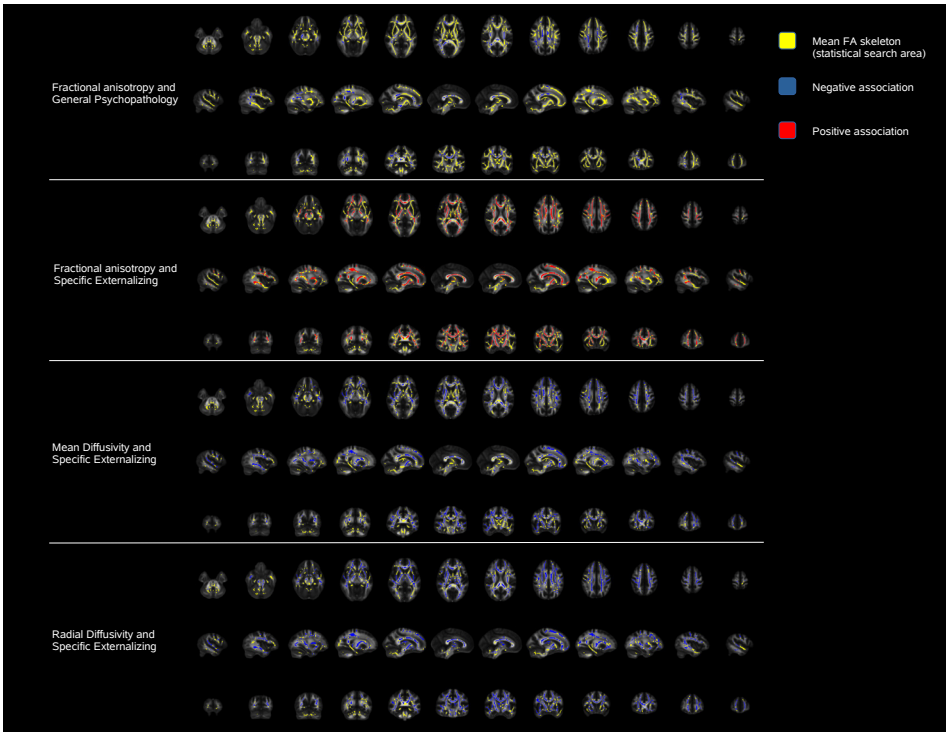


Figure 2: Results of the TBSS analysis (n=2,996). Voxels in the mean FA skeleton (in yellow) were associated with the general psychopathology and specific externalizing factor, using the scalars fractional anisotropy, mean diffusivity and radial diffusivity; these analyses were adjusted for sex, age at scan, maternal and paternal education, household income and genetic ancestry. Voxels with significant p-values after multiple testing correction were coded as blue, if the direction was negative, and red, if the direction was positive.

factor was associated with a β=-0.07SD (SE=0.02, p<0.01, q<0.01) decrease in general psychopathology. In contrast, a 1-SD increase of white matter microstructure predicted an increase of β=+0.07SD (SE=0.03, p=0.01, q=0.02) specific externalizing factor levels. Follow-up analyses showed that this association appears to be more driven by radial diffusivity (β=-0.07SD, SE=0.03, p=0.01, q=0.04), as opposed to axial diffusivity (β=-0.02SD, SE=0.03, p=0.39, q=0.39). Thus, while children with more general psychopathology had lower global white matter microstructure, children with a higher specific externalizing factor had more white matter microstructure. See Figure S2, for scatter plots based on estimated factor scores. Quadratic terms of the global white matter factor scores were not significant for any of the psychopathology factors (Table S4). These associations were largely independent of child IQ, parental psychopathology and total intracranial

volume for both the general factor (β=-0.06, SE=0.02, p<0.01) and the externalizing factor (β=+0.07, SE=0.03, p<0.01), see Table S5 for estimates of all covariates.

The individual white matter tracts were negatively associated with the general psychopathology factor, with the exception of the cingulum bundle, and positively with the specific externalizing factor. The magnitude of associations were mostly of lower magnitude than those of the global white matter factor. The forceps minor and right superior longitudinal fasciculus were associated with general psychopathology after adjustment for false disovery rate. Only the relation of the corticospinal tract with the specific externalizing factor survived correction for multiple testing.

The latent variable models suggest that a global white matter factor based on the FA scalar is negatively associated with the general psychopathology factor and positively with the specific externalizing factor. We therefore explored these two findings further by testing at the voxel level in a TBSS analysis across 9272 voxels of a study-specific white matter skeleton. The results from the voxel-wise analyses were consistent with the global white matter models. For general psychopathology, 1548 (17%) voxels showed a negative association and 0 (0%) a positive when accounting for multiple testing. We found that 85.9% of these voxels formed a single continuous clusters, which was spread across the whole brain (Supplementary Table 6 and Figure 2). It is therefore not possible to define this cluster by specific brain regions, though we observed that voxels were especially represented in the left Inferior longitudinal fasciculus and left corticospinal tract (Table S7). For the specific externalizing factor, 4842 (52%) voxels showed a positive association and 0 (0%) a negative. Because the global white matter factor was also associated with the specific externalizing factor when the structural equation models were based on MD and RD, we tested these scalars in TBSS analyses as well. MD values were significant for 5149 (56%) voxels and RD for 6282 (68%) with all associations being in the negative direction and 0 (0%) positive. Among the defined regions, the forceps minor contained the most associated voxels (Table S7). Depending on the scalar, 97.0% (FA), 99.9% (MD) or 99.7% (RD) of significant voxels formed a single continuous cluster. As with the general psychopathology factor, the cluster was also spread across the whole brain and the global nature was very pronounced (Table S6 and Figure 2).

Measurement invariance

The multi-group analyses showed that the global white matter, general and specific psychopathology constructs did not differ by ancestry (Table S8), sex (Table 9) (strong measurement invariance) or socioeconomic status (Table 10) (strict measurement invariance). We also found no evidence that the associations between white matter and psychopathology factors depended on ancestry, sex or socioeconomic status (relational invariance), i.e. we did not detect interactions in any of the regression parameters. The association between global white matter and general psychopathology was more

than twice as strong in boys (β=-0.11, SE=0.03, p<0.01) than in girls (β=-0.04, SE=0.03, p=0.17), but the difference was not significant (z test: p=0.14).

DISCUSSION

Despite the large sample size of GenR, we did not find associations of white matter microstructure with traditional definitions of externalizing, internalizing and attention latent constructs. However, this changed when taking into account the general psychopathology factor: children with a lower global white matter microstructure had higher levels of general psychopathology. In contrast, more global white matter microstructure was associated with higher levels of the specific externalizing factor. Our finding were not driven by a single white matter tract, but by white matter differences across the brain. Further adjustment for child non-verbal IQ lead to a relatively small reduction of effect size, suggesting that most of the association with psychopathology cannot be explained by IQ.

At age 10 years, the development of many white matter tracts, such as projections of the prefrontal cortex, is still ongoing.[44,45] An altered maturation of white matter microstructure, both delayed or accelerated, at this age might thus be responsible for various psychological problems, ranging from cognitive to behavioral and emotional problems. As we previously reported, lower global white matter microstructure was associated with lower cognition in childhood,[15] and global white matter values were negatively associated with depression.[19] White matter microstructure is highly heritable, especially in younger ages, with heritability estimates in adolescence exceeding those of adulthood.[46] Genetic variants underlying psychopathology potentially influence psychiatric problems by altering white matter microstructure. However, differences in white matter are not only genetically driven. For instance, children in foster care and children who remained institutionalized show differences in microstructure, suggesting environmental effects.[47] White matter microstructure most likely is also a marker for developmental and environmental adversities that underlie psychological problems, or white matter may even mediate these environmental risk effects. The findings implicate that children with a psychiatric problem in one domain, not only are more likely to have psychological problems in another, but are also more likely to have lower white matter microstructure. This suggests that whether a child presents aggressive behavior, attention problems or anxiety, it is not only important to consider psychiatric problems in all domains, but also address other white matter microstructure associated traits, such as cognitive ability. Conversely, prevention of mental health problems and promotion of healthy brain development are expected to have very broad impacts on functioning in many areas.

The negative association of global white matter microstructure with general psychopathology supports the notion that lower white matter microstructure is a marker for

poorer mental health and IQ. However, this is not necessarily the case for all disorders and characteristics. Higher dorsal white matter microstructure ("where pathway") is associated with more visuospatial deficits in Williams syndrome,[48] and developmental increases of FA were associated with lower IQ levels depending on sex and brain region.[49] Furthermore, ADHD is inconsistently associated with higher white matter microstructure in some regions.17At first glance the contrasting positive association with the specific externalizing factor suggests that higher white matter microstructure is unexpectedly a specific risk factor for aggressive and rule-breaking behavior in childhood. However, it should be emphasized that the interpretation of the specific factor is different from traditional internalizing/externalizing factors or broadband scales, which were not associated with white matter microstructure. The specific factors represent the variance which is not shared with any other problem domains. Compared to the traditional externalizing factor, the specific externalizing factor was much less associated with problematic characteristics, such as lower IQ, worse school performance, higher neuroticism or less happiness. In contrast, both the traditional and specific externalizing factor were associated with surgency, i.e. positive affect reactivity, to the same degree. These changes in associations may suggest, that when accounting for general psychopathology, which can be regarded as the extent of problematic behavior, the remaining externalizing factor represents behavior, which is not as problematic, such as assertive behavior. In other words, for a child who only displays aggressive or rule-breaking behavior, but otherwise low levels of depression and anxiety, and an absence of attention problems, the externalizing behavior may be more a reflection of personality rather than psychopathology.

We interpret the associations between white matter microstructure and the psychopathology factors as not-regionally specific based on following observations. First, the associations of the global white matter variable were either stronger or at least as strong as any individual tract. This would not be the case if the results were driven by few specific tracts. Second, the direction of the association of the individual tracts with overall or specific psychopathology were consistently that of the global white matter indicator, except for the cingulum bundle. Third, likewise, in the voxel-wise TBSS analyses, all individual voxels were associated with the general psychopathology or the specific externalizing factor in the direction predicted by the global model. While some regions contained more voxels associated with the general psychopathology factor than others, e.g. left inferior longitudinal fasciculus and left corticospinal tract (general psychopathology) and forceps minor (specific externalizing), nearly all voxels formed a single continuous cluster. This cluster is spread across the whole brain and could not be defined by specific regions.

Few neurobiological studies so far have attempted to distinguish general and specific effects of psychopathological dimensions. Traditional analyses, which rely on symptom counts or diagnoses of internalizing or externalizing problems, typically only estimate the overall association with a single domain. In these studies, it is difficult to

disentangle to what extent the association applies to other psychological domains and to what extent it is specific to the studied domain. The observation that associations are only present when partitioning psychopathology into general and specific effects highlights the usefulness of bifactor models, arguably justifying the increase of functional complexity.[50]

A strength of our study is the stringent adjustment for many potential socioeconomic confounders. However, as with any observational study, residual confounding cannot be ruled out, making a causal interpretation of the associations difficult. Another challenge to the causal interpretation of our findings is that directionality cannot be established with this study. We assumed in our statistical model that white matter microstructure changes underlie the development of psychopathology. To maximize precision and reduce biases of psychopathology factors we incorporated two study waves in the estimation of these factors. However, since imaging was performed during the second wave and we did not investigate changes of psychopathology, we could not test for directionality of effects. It is theoretically plausible that changes in white matter structure are either cause or outcome of psychological problems, or both. Irrespective of the direction, the effect sizes of the global white matter factors were modest, independently explaining less than 1% of the psychopathology factor variances. This may reflect the difficulty in reliably estimating childhood psychopathology. It should be noted, however, that none of the other tested predictors particularly stood out in terms in explanatory power, when carefully controlled for the same variables. This suggest that a multitude of factors are needed, if one wishes to reliably predict levels of psychopathology in the general population.

In summary, global white matter microstructure was associated with lower general psychopathology in school-aged children. At the same time, higher microstructure was also associated with a higher risk for specific externalizing behavior, perhaps better characterized as another trait, e.g. assertiveness. Both associations were independent of socioeconomic status and IQ of the child. This study highlights the importance of distinguishing global measures form specific features for both neurobiological substrates as well as psychiatric symptoms. Pediatric brain imaging studies must carefully control for general psychopathology or psychiatric comorbidity to reliably detect any specific white matter microstructural associations. The global effects identified in childhood emphasize the need for early prevention and promotion of brain and mental health. Further studies are needed to replicate these findings and to investigate the temporal direction of association.

REFERENCES

1. Noordhof A, Oldehinkel AJ, Ormel J. Comorbidity between Internalizing and Externalizing problems in adolescence : fact or artefact? In: In the Absence of a Gold Standard. Groningen; 2010:59-80.
2. Achenbach TM, Ivanova MY, Rescorla LA, Turner L V., Althoff RR. Internalizing/Externalizing Problems: Review and Recommendations for Clinical and Research Applications. J Am Acad Child Adolesc Psychiatry. 2016;55(8):647-656. doi:10.1016/j.jaac.2016.05.012
3. Kessler RC, Chiu WT, Demler O, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry. 2005;62(6):617-627. doi:10.1001/archpsyc.62.6.617
4. Angold A, Costello EJ, Erkanli A, Angold A, Costello EJ. Comorbidity. J child Psychol psychiatry. 1999;40(01):57-87. doi:10.1038/186935d0
5. Olino TM, Dougherty LR, Bufferd SJ, Carlson G a., Klein DN. Testing Models of Psychopathology in Preschool-aged Children Using a Structured Interview-based Assessment. J Abnorm Child Psychol. 2014:1201-1211. doi:10.1007/s10802-014-9865-x
6. Neumann A, Pappa I, Lahey BB, et al. Single Nucleotide Polymorphism Heritability of a General Psychopathology Factor in Children. J Am Acad Child Adolesc Psychiatry. 2016;55(12):1038-1045. doi:10.1016/j.jaac.2016.09.498
7. Lahey BB, Rathouz PJ, Keenan K, Stepp SD, Loeber R, Hipwell AE. Criterion validity of the general factor of psychopathology in a prospective study of girls. J Child Psychol Psychiatry Allied Discip. 2014:1-8. doi:10.1111/jcpp.12300
8. Laceulle OM, Vollebergh W a. M, Ormel J. The Structure of Psychopathology in Adolescence: Replication of a General Psychopathology Factor in the TRAILS Study. Clin Psychol Sci. 2015:1-11. doi:10.1177/2167702614560750
9. Patalay P, Fonagy P, Deighton J, Belsky J, Vostanis P, Wolpert M. A general psychopathology factor in early adolescence. Br J Psychiatry. 2015;207(1):15-22. doi:10.1192/bjp.bp.114.149591
10. Lahey BB, Applegate B, Hakes JK, Zald DH, Hariri AR, Rathouz PJ. Is There a General Factor of Prevalent Psychopathology During Adulthood? J Abnorm Psychol. 2012;121(4):971-977. doi:10.1037/a0028355
11. 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
12. Lahey BB, Krueger RF, Rathouz PJ, Waldman ID, Zald DH. A hierarchical causal taxonomy of psychopathology across the life span. Psychol Bull. 2017;143(2):142-186. doi:10.1037/bul0000069
13. Zald DH, Lahey BB. Implications of the Hierarchical Structure of Psychopathology for Psychiatric Neuroimaging. Biol Psychiatry Cogn Neurosci Neuroimaging. 2017;2(4):310-317. doi:10.1016/j.bpsc.2017.02.003
14. Vuoksimaa E, Panizzon MS, Jr DJH, et al. Heritability of White Matter Microstructure in Late Middle Age: A Twin Study of Tract-Based Fractional Anisotropy and Absolute Diffusivity Indices. Hum Brain Mapp. 2017;38:2026-2036. doi:10.1002/hbm.23502
15. Muetzel RL, Mous SE, van der Ende J, et al. White matter integrity and cognitive performance in school-age children: A population-based neuroimaging study. Neuroimage. 2015;119:119-128. doi:10.1016/j.neuroimage.2015.06.014

16. White T, Langen C, Schmidt M, Hough M, James A. Comparative neuropsychiatry: White matter abnormalities in children and adolescents with schizophrenia, bipolar affective disorder, and obsessive-compulsive disorder. Eur Psychiatry. 2015;30(2):205-213. doi:10.1016/j.eurpsy.2014.10.003
17. Chen L, Hu X, Ouyang L, et al. A systematic review and meta-analysis of tract-based spatial statistics studies regarding attention-deficit/hyperactivity disorder. Neurosci Biobehav Rev. 2016;68(37):838-847. doi:10.1016/j.neubiorev.2016.07.022
18. Albaugh MD, Ducharme S, Karama S, et al. Anxious/depressed symptoms are related to microstructural maturation of white matter in typically developing youths. Dev Psychopathol. 2016;29:1-8. doi:10.1017/S0954579416000444
19. Shen X, Reus LM, Cox SR, et al. Subcortical volume and white matter integrity abnormalities in major depressive disorder: findings from UK Biobank imaging data. Sci Rep. 2017;7(1):5547. doi:10.1038/s41598-017-05507-6
20. Loe IM, Lee ES, Feldman HM. Attention and Internalizing Behaviors in Relation to White Matter in Children Born Preterm. J Dev Behav Pediatr. 2013;34(3):156-164. doi:10.1097/DBP.0b013e3182842122
21. Kooijman MN, Kruithof CJ, van Duijn CM, et al. The Generation R Study: design and cohort update 2017. Eur J Epidemiol. 2016;31(12):1243-1264. doi:10.1007/s10654-016-0224-9
22. Tiemeier H, Velders FP, Szekely E, et al. 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. J Am Acad Child Adolesc Psychiatry. 2012;51(11):1119-1135.e7. doi:10.1016/j.jaac.2012.08.021
23. Smith SM, Jenkinson M, Johansen-Berg H, et al. Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. Neuroimage. 2006;31(4):1487-1505. doi:10.1016/J.NEUROIMAGE.2006.02.024
24. Achenbach TM, Ruffle TM. The Child Behavior Checklist and Related Forms for Assessing Behavioral/Emotional Problems and Competencies. Pediatr Rev. 2000;21(8):265-271. doi:10.1542/pir.21-8-265
25. Achenbach TM, Rescorla LA. Manual for the ASEBA School-Age Forms and Profiles. . (University of Vermont Youth, and Families RC for C, ed.). Burlington, VT; 2001.
26. Arseneault L, Moffitt TE, Caspi A, et al. Strong genetic effects on cross-situational antisocial behaviour among 5-year-old children according to mothers, teachers, examiner-observers, and twins' self-reports. J Child Psychol Psychiatry. 2003;44(6):832-848. doi:Doi 10.1111/1469-7610.00168
27. Achenbach TM, McConaughy SH, Ivanova MY, Rescorla LA. Manual for the ASEBA Brief Problem Monitor (BPM). Burlington, VT ASEBA. 2011.
28. Chang L, Jones DK, Pierpaoli C. RESTORE: Robust Estimation of Tensors by Outlier Rejection. Magn Reson Med. 2005;(53):1088-1095. doi:10.1002/mrm.20426
29. Groot M De, Ikram MA, Akoudad S, et al. Tract-specific white matter degeneration in aging: The Rotterdam Study. Alzheimer's Dement. 2015;11(3):321-330. doi:10.1016/j.jalz.2014.06.011
30. Muetzel RL, Blanken LME, Ende J Van Der, et al. Tracking Brain Development and Dimensional Psychiatric Symptoms in Children: A Longitudinal Population-Based Neuroimaging Study. Am J Psychiatry. 2017. doi:10.1176/appi.ajp.2017.16070813
31. Muetzel RL, Blanken LME, Thijssen S, et al. Resting-state networks in 6-to-10 year old children. Hum Brain Mapp. 2016;37(12):4286-4300. doi:10.1002/hbm.23309
32. Tellegen P, Laros J. The construction and validation of a nonverbal test of intelligence: the

revision of the Snijders-Oomen tests. Eur J Psychol Assess. 1993;9(2):147-157.

33. Putnam SP, Rothbart MK. Development of short and very short forms of the Children's Behavior Questionnaire. J Pers Assess. 2006;87(1):102-112. doi:DOI 10.1207/s15327752jpa8701_09
34. Stichting Cito Instituut voor Toetsontwikkeling Arnhem. Cito Centrale Eindtoets. https://www.cito.nl/onderwijs/primair-onderwijs/centrale-eindtoets/. Published 2018. Accessed September 19, 2018.
35. Derogatis LR, Melisaratos N. The brief symptom inventory: an introductory report. Psychol Med. 1983;13(03):595-605.
36. R Core Team. R: A Language and Environment for Statistical Computing. 2016. https://www.r-project.org/.
37. Rosseel Y. lavaan: An R package for structural equation modeling. J Stat Softw. 2012. http://www.lce.esalq.usp.br/arquivos/aulas/2013/encontro_ppg/Lucia/paper.pdf. Accessed April 13, 2015.
38. Takao H, Hayashi N, Inano S, Ohtomo K. Effect of head size on diffusion tensor imaging. Neuroimage. 2011;57(3):958-967. doi:10.1016/j.neuroimage.2011.05.019
39. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. Fsl. Neuroimage. 2012;62(2):782-790. doi:10.1016/j.neuroimage.2011.09.015
40. Smith SM, Nichols TE. Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage. 2009;44(1):83-98. doi:10.1016/J.NEUROIMAGE.2008.03.061
41. Cheung G, Rensvold R. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equ Model. 2002. http://www.tandfonline.com/doi/abs/10.1207/s15328007sem0902_5. Accessed November 12, 2015.
42. Beaujean A. Latent Variable Modeling Using R: A Step-by-Step Guide. 1st ed. Routledge; 2014.
43. Hirschfeld G, Von Brachel R. Multiple-Group confirmatory factor analysis in R – A tutorial in measurement invariance with continuous and ordinal indicators. Pract Assessment, Res Eval. 2014;19(7):1-11.
44. Simmonds DJ, Hallquist MN, Asato M, Luna B. Developmental stages and sex differences of white matter and behavioral development through adolescence: A longitudinal diffusion tensor imaging (DTI) study. Neuroimage. 2014;92:356-368. doi:10.1016/j.neuroimage.2013.12.044
45. Lebel C, Beaulieu C. Longitudinal Development of Human Brain Wiring Continues from Childhood into Adulthood. J Neurosci. 2011;31(30):10937-10947. doi:10.1523/JNEUROSCI.5302-10.2011
46. Chiang M-C, Mcmahon KL, de Zubicaray GI, et al. Genetics of white matter development: A DTI study of 705 twins and their siblings aged 12 to 29. Neuroimage. 2010;54(3):2308-2317. doi:10.1016/j.neuroimage.2010.10.015
47. Bick J, Zhu T, Stamoulis C, Fox NA, Zeanah C, Nelson CA. Effect of Early Institutionalization and Foster Care on Long-term White Matter Development. JAMA Pediatr. 2015;169(3):211. doi:10.1001/jamapediatrics.2014.3212
48. Hoeft F, Barnea-goraly N, Haas BW, et al. More Is Not Always Better: Increased Fractional Anisotropy of Superior Longitudinal Fasciculus Associated with Poor Visuospatial Abilities in Williams Syndrome. J Neurosci. 2007;27(44):11960-11965. doi:10.1523/JNEUROSCI.3591-07.2007

49. Schmithorst VJ. Developmental sex differences in the relation of neuroanatomical connectivity to intelligence. Intelligence. 2009;37(2):164-173. doi:10.1016/j.intell.2008.07.001
50. Bonifay W, Cai L. On the Complexity of Item Response Theory Models. Multivariate Behav Res. 2017;0(0):1-20. doi:10.1080/00273171.2017.1309262

SUPPLEMENTARY METHODS

Participants

This study was embedded in the Generation R Study.[1,2] Generation R is a population-based birth cohort with the goal of identifying early environmental and genetic determinants of development and child health. All parents gave informed consent for their children's participation. The Generation R Study is conducted in accordance with the World Medical Association Declaration of Helsinki and study protocols have been approved by the Medical Ethics Committee of the Erasmus Medical Center, Rotterdam.

We performed MRI scans in 3996 children. White matter microstructure information from diffusion tensor imaging was available in 3669 children without dental braces.3 405 scans failed automatic or manual quality control. Incidental findings were present for 13 children, we thus did not include them in the present analyses. Early in the data acquisition, 201 children were scanned with an older MRI software version and slightly different sequence parameters which yielded systematically different diffusion values, and were therefore excluded from analyses. At least one psychological problem subscale was available for 3030 children. All results are based on this sample of 3030 children, except for the results of the tract-based spatial statistics (TBSS) analysis (n=2996). Additional quality control excluded 47 children, because parts of the brain were not in the field of view of the scan (usually the lower portion of the cerebellum or the very top of the head) and a further 7 were excluded due to misregistration. See Supplementary Figure 1 for participant flow chart. Demographic and descriptive statistics can be found in Table 1.

MEASURES

Child psychological problems assessed by parents

We used the Child Behavior Checklist (CBCL) 1 ½ -5 years[4] to assess child behavioral problems at age 5.9 years (SD=0.3) and the CBCL 6-185 at age 9.7 years (SD=0.3). At the age of 6 years, questionnaires were completed by the primary caregiver (92% mothers). At age 10 years, the questionnaire was filled in by mother and father each, who indicated for a wide array of statements, whether they were not true (0), somewhat/sometimes true (1) or very/often true (2). The item scores were then summed into several subscales. The CBCL 1 ½-5 includes the subscales: Emotionally Reactive, Anxious/Depressed, Somatic Complaints, Withdrawn, Sleep Problems, Attention Problems, Aggressive Behavior; and a sum score of other items. The CBCL 6-18 includes the subscales: Anxious Depressed, Withdrawn Depressed, Somatic Problems, Social Problems,

Thought Problems, Attention, Rule-breaking, Aggression and again a sum score of other items.

Child psychological problems assessed by teachers

Teachers assessed children at age 6.5 (SD=1.1) with the Teacher's Rating Form (TRF) 6-18 years.[5] They were approached independently of the parents, but with parental consent. The TRF is scored like the CBCL 6-18 and includes the same subscales, but lacks a sum score of other items.

Child psychological problems assessed by children

At age 6.0 (SD=0.4) we conducted the Berkeley Puppet Interview (BPI),[6] a semi-structured interactive child interview, at our research center. The interview is performed with two identical dog hand puppets. The two puppets made opposing statements and the child chose the statement that described him/her best. Scoring was performed using video recordings with high intercoder reliability (average ICC=[0.96-0.98]).[7] Six subscales were calculated: Depression, Separation Anxiety, Overanxious, Oppositional Defiant, Overt Hostility, and Conduct Problems.

At age 9.8 (SD=0.3) years the children rated their problems with the Brief Problem Monitor (BPM).[8] This questionnaire uses items of the CBCL and TRF but is shorter and has only 3 subscales: internalizing, externalizing and attention problems. We also added questions related to thought problems, modeled after the CBCL and TRF items. The items for thought problems scores were: "I cannot put some thoughts out of my head", "I hear sounds or voices that other people do not", "I see things that other people think they are not there", "I save too many things that I do not need", "I have thoughts that other people find strange", "I have thoughts about hurting myself".

Imaging

Children underwent diffusion tensor imaging at age 10.1 (SD=0.6) years. They were first familiarized with the MRI-environment in a mock scanning session. MRI scans were performed using a 3T General Electric scanner (MR750W) with an 8-channel receive-only head coil. Diffusion tensor imaging consisted of a 35-direction echo planar imaging sequence (TR=12,500ms, TE=72ms, FoV=240mm*240mm, acquisition matrix=120*120, slice thickness=2mm, number of slices = 65, Asset Acceleration Factor = 2, b=900s/mm2, 3 b=0 images). In addition, high-resolution T1-weighted sequences, specifically IR-prepared Fast Spoiled Gradient Recalled Sequences with the GE option BRAVO (TR =8.77ms, TE =3.4ms, TI =600ms, Flip Angle=10°, FOV = 220mm x 220mm, Acquisition Matrix = 220 x 220, slice thickness = 1mm, number of slices = 230, voxel size = 1mm x

1mm x 1mm, ARC Acceleration = 2.) were performed and used for estimation of intracranial volume.

We used the functional MRI of the Brain's Software Library (FSL 5.0.9[9]) together with the Camino Diffusion Toolkit[10] for pre-processing of the diffusion tensor images. The following preprocessing steps were applied: adjustment for eddy current-induced artifacts, translation/rotations resulting from minor head motion, and removal of non-brain tissue. The diffusion gradient direction table was rotated with the transformation matrix from the eddy current correction step. We computed fractional anisotropy (FA), mean diffusivity (MD), radial diffusivity (RD), and axial diffusivity (AD) using the RESTORE method[11], as previously described.[12]

Connectivity distributions for 12, large well-defined and widely reported fiber bundles were derived with probabilistic fiber tractography using the FSL plugin AutoPtx[13], as described previously.[12] The first step involved the estimation of diffusion values per voxel, accounting for two fiber orientations.[14] AutoPtx provided a predefined set of seed and target masks, which were aligned to each participant in native space using a nonlinear registration. Based on this information, Probtrackx determined connectivity distributions for 12 large fiber bundles. The values were normalized based on the number of successful seed-to-target attempts. The FA values per voxel were weighted depending on the connectivity distribution, ensuring that voxels most likely to be part of a bundle also contribute the most to the overall connectivity of the bundle.[15] Total intracranial volume was estimated using FreeSurfer 6.0.[16]

Image Quality assessments consisted of automatic and manual checks. DTIPrep (https://www.nitrc.org/projects/dtiprep/) was used to detect artifacts by examining slicewise variation in signal intensity and by visual inspection of the sum-of-squares error (SSE) map from the diffusion tensor. Scans were excluded based on the default settings from the automatic procedure (n=257) and when they contained substantial artifact in the SSE map based on a 0-3 scale from "none" to "severe" (n=140). Furthermore, registration accuracy to standard space and accuracy of the tract reconstructions were visually inspected.[12]

For TBSS analyses the DTI images were registered to a study-specific and age-appropriate template space[17] using FNIRT[18]. Specifically, the FSL FMRIB58 template image was used by warping to our study-specific template, which was based on 130 children without behavioral problems and excellent T1-weighted images. FA data were non-linearly aligned with this FA template using FNIRT with spline interpolation. Afterwards, the warp fields from the FA map were applied to the MD, RD and AD maps. The thresholded (>0.2) mean FA image was used to create a mean FA skeleton. Finally, the voxel level scalars of each participant were projected onto this common skeleton and associat-

ed with the psychopathology outcomes. Proper registration and whole-brain coverage were visually inspected."

Measures of IQ, school performance, temperament, happiness and parental psychopathology

We assessed non-verbal cognitive abilities with the Snijder-Oomen nonverbal intelligence test at age 6.0 (SD=0.4).[19] At the same age we also measured temperamental dimensions (negative affectivity, surgency/extraversion, and effortful control) with the Child Behavior Questionnaire (Very-Short-Form), a parent-rated questionnaire.[20] School performance was assessed by the Cito[21], a standardized exam at the end of primary school in which language and math skills are tested. Happiness was measured by asking the parents at age 10: "How often was your child happy in the past 4 weeks?", who could answer "never", "almost never", "sometimes", "usually" or "always". Maternal and paternal psychopathology were assessed at child age 10 with the Brief Symptom Inventory with four problem subscales: interpersonal sensitivity, depression, anxiety and hostility.[22] Household income and highest achieved education of mother and father were assessed with questionnaires and treated as continuous measures in the analyses at birth and age 6.

Genetic Ancestry

Brain features may differ across ethnicities due to differential exposures. At the same time these exposures may independently affect psychopathology and create a confounding bias. This bias can be adjusted for with categorical information on national origin or genetic information on ancestry. Genetic ancestry was more strongly associated with white matter structure, suggesting that this is a better marker for differential exposures in the ethnic groups. We therefore decided to control for continuous scores of genetic ancestry in our models.

Genetic ancestry was based on single nucleotide polymorphisms (SNP).[23] 518,245 SNPs were measured with Illumina 610K/660W arrays. Quality control included sample (≥97.5%) and SNP call rates (≥95%), minor allele frequency ≥1% and deviations from Hardy-Weinberg equilibrium ($p<10^{-7}$). Four principal components of ancestry (PCA) were derived from multidimensional scaling (n=5731). Participants exceeding 4 SDs difference with the mean European reference level (HapMap CEU) on any of the first four principal components were classified as non-northwestern European (n=760), as opposed to northwestern European (n=1137). No genetic information was available for 1145 (38%) children.

Statistical Analysis

We used a structural equation model to associate global white matter microstructure with general and specific factors of psychopathology. See Figure 1 for an ab-

breviated path diagram. All models were fitted in R 3.4.1[24] with the package Lavaan 0.5-23.11097.[25] We used a maximum likelihood estimator with robust standard errors (MLM) to account for multivariate non-normality. Family structure was adjusted for using a stratified cluster approach in the package lavaan.survey 1.1.3.1[26], specifically with zygosity (monozygotic, dizygotic, non-twin) as stratification variable and family ID as cluster variable. Latent variables were scaled with a marker variable (scale of the first indicator). All subsequent coefficients are reported as standardized estimates. Missing variables were handled with multiple imputations using mice 2.30.[27] All variables featured in the analysis models, plus squared and orthogonalized (from the original variable) white matter microstructure tracts, paternal education/household income at birth, maternal IQ, national origin, and principal components of ancestry 5-20, were considered in the imputation model as predictors. In the analysis sample every participant had valid DTI images and complete information on at least three psychopathology subscales. All participants had complete information on sex, zygosity, MRI scan age and most tracts, except for the corticospinal tract (two missing) and right cingulum bundle (one missing). All other variables used in the analyses and imputation model had various degrees of missingness (see Table 1, 2 and Supplementary Table 1, 3) and missing data for these variables were imputed. We filtered for only robust predictor-target pairs with a minimum spearman correlation of 0.05 using the quickpred() function. We estimated 120 imputations with 30 iterations.

The general psychopathology factor

The general psychopathology factor was specified to underlie all problem subscales from all instruments and time-points. The subscales were also specified to load on one of the specific internalizing, the specific externalizing, or the specific attention scales defined on the basis of the assessment scales. The attention subscales from each informant, for example, loaded on the general as well as specific attention factor. A few subscales did not have paths to any specific factor (See Supplementary Table 1 for the item structure), since the assessment scales did not group them into any higher order domains. Each specific psychopathology factor was allowed to correlate among each other, but not with the general psychopathology factor. The specific factors thus represent covariance among subscales that cannot be explained by a general propensity for psychiatric problems. As such the specific factors differ from the observed broadband scales. The observed domain scores correlate positively, but the specific factors do not correlate positively because the shared variance is already captured by the general factor. The higher order structure of the tested model can be seen as an extension of the instruments' established factor structure which instructed the computation of the subscales.

In addition, we included method factors. All psychopathology subscales of a single informant at a specific assessment age load on one method factor. The subscales of the CBCL 6-18 rated by the mother at age 10 years, for example, load on one method factor.

These method factors capture the shared variance among problem subscales, which is unique to a certain context, (i.e. to an informant at a certain age of the child). These factors were therefore specified to be uncorrelated to all other factors. We chose to model this shared method variance explicitly, because it is specific to a certain rating context.

The latent factor structure was based on the best fitting model in a previous study using Generation R data on 6-8 years old children,[28] as well as on models validated in other cohorts in adolescents and adults.[29–31] However, we performed several additional analyses for further validation and characterization. First, we tested models without a general psychopathology factor. We fitted a model with 2 (internalizing and externalizing) and with 3 factors (internalizing, externalizing and attention). In the 2 factor model attention subscales were specified to load on the externalizing factor, whereas in the 3 factor model attention was specified to load on a separate attention factor. The 3 factor model fit substantially better (see Supplementary Table 2) based on robust CFI, RMSEA and BIC. We thus used 3 specific factors in the main general psychopathology factor model, however, we also present association results from the 3 factor model without the main general psychopathology factor as comparison.

Second, we explored four models with, in order, IQ and temperament (negative affectivity, surgency, and effortful control) measured at age 6, school performance at the end of primary school and happiness as predictors of the psychopathology factors. We previously had associated the IQ and temperament variables with general and specific psychopathology factors in younger children.[28] As IQ and temperament showed discriminate associations, they can therefore help interpret the latent psychopathology factors. Third, we performed sensitivity analyses adjusted for total intracranial volume, as larger volume is associated with higher FA and thus could confound global white matter associations.[32]

The global white matter microstructure factor

The global white matter microstructure factor was estimated using the mean FA values of 12 white matter tracts as indicators (Supplementary Table 3). FA describes how elongated the ellipsoid shape of a diffusion pattern is, with higher values suggesting higher white matter integrity (from here on referred to as higher white matter microstructure). This model was based on a previous studies using Generation R data.[12,15] We included corticospinal, inferior longitudinal fasciculus, superior longitudinal fasciculus and uncinate fasciculus tracts of each hemisphere separately in the model. The error terms of each tract were allowed to correlate between both hemispheres. The hemisphere division is not applicable for the forceps major and minor tracts that cross hemispheres. We tested for non-linear associations in the main model, by estimating a standard regression model analogous to the strucutural paths with estimated factor score, but with the addition of a squared term for the global white matter factor score. While FA it is good summary measure of white matter microstructure, it can be also helpful to examine diffusivity only perpendicular to the main axis of diffusion (RD) or

only alongside it (AD). Higher RD and lower AD are associated with less white matter microstrucutre. Thus, to better understand results from the FA model, we also estimated global white matter variables based on RD, AD, as well as MD (the average diffusivity in any direction).

Structural Paths

The general and specific psychopathology factors were each simultaneously regressed on the global white matter factor based on FA values to test the associations between white matter microstructure and psychopathology. Figure 1 illustrates the main model, the only model used to test the hypothesis. All other statistical models were used for exploratory purposes to better interpret the results of the main model. Since IQ is related to global white matter microstructure,[15] we explored whether associations were specific to psychopathology by including child IQ as a covariate in a separate model. Secondly, we controlled for maternal/paternal psychopathology (interpersonal sensitivity, depression, anxiety and hostility) to explore to what extent the association is independent of parental characteristics and also to control for potentially remaining rater bias as parents completed several psychopathology measures. We reran the main model with global white matter factor based on MD, RD and AD.

We included several potential confounders in all models, namely sex, age at scan, maternal and paternal education at age 6, income at age 6, and genetic ancestry. Paths from these covariates to the latent variable global white matter microstructure, general and specific psychopathology factors were included to adjust for confounding biases. Age at psychopathology assessment was adjusted by including assessment age as covariates of the corresponding psychopathology subscales.

IQ and parental psychopathology

Since IQ was related to global white matter microstructure,[15] we explored whether associations were specific to psychopathology by including child IQ as a covariate. Secondly, we controlled for maternal/paternal psychopathology (interpersonal sensitivity, depression, anxiety and hostility) to explore to what extent the association is independent of parental characteristics and thus control for potentially remaining rater bias caused by half of the psychopathology measures being informed by parents.

Follow-up Analyses

After testing the global white matter factor, we ran follow up analyses to investigate the individual contribution of each individual white matter tract. In each model the latent variable global white matter microstructure was replaced with the observed FA of a single tract. These follow up analyses had two goals: 1. to confirm that most tracks

behave similarly as suggested by the latent global variable and 2. to increase comparability with studies reporting single track associations.

Additionally we tested individual voxels (nvoxels=9272) in a TBSS analysis to further explore the results from the main analysis. In these follow-up analyses to the global white matter models, we performed TBSS analyses for outcomes associated with global FA values. We present results for other DTI scalar metrics on a voxel-wise level, if the scalar was significant on a global level. TBSS was performed in FSL 9 using a 2mm3 resolution. Adjustment for multiple testing was achieved with permutation testing (nperm=5000) and clusters were formed using the built-in threshold-free cluster enhancement [33.] As FSL does not support latent variables, we estimated psychopathology factor scores based on the main model and adjusted for the same covariates as in the main model.

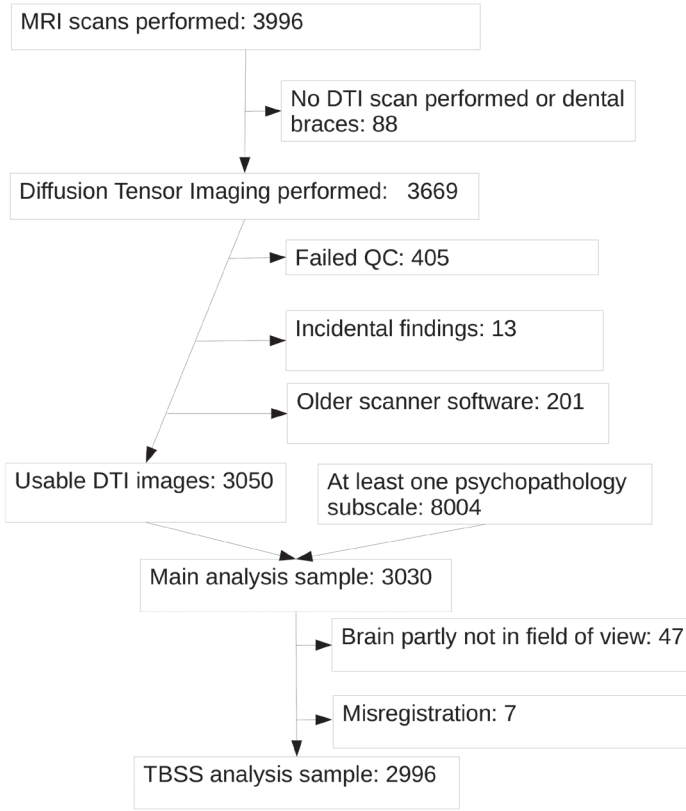
Measurement Invariance

Generation R is a highly ethnically diverse sample. The main analyses are based on the assumption that the latent constructs and associations between them are identical across sex, ancestry (European vs non-European) and socioeconomic status (higher maternal education vs no higher maternal education). We tested this assumption by performing measurement invariance analyses of the main model. To this aim we sequentially constrained an increasing number of parameters. Constraints were judged to significantly worsen fit when the robust CFI dropped by more than 0.01.[34] See Supplementary Table 8-10 for models tested.[35,36] The genetic ancestry invariance models featured 1136 children with northwestern European ancestry and 758 children with other ancestry. The sex invariance analyses featured 1510 boys and 1529 girls. The socioeconomic status invariance analyses included 1665 children of mothers with higher education and 994 children of mothers without. The analyses suggest strong measurement and relational invariance across European and non-European ancestry (Supplementary Table 8), as well as across girls and boys (Supplementary Table 9) and socioeconomic status (Supplementary Table 10), suggesting that loadings, intercepts and regression coefficients are similar across groups. The association between global white matter and general psychopathology was more than twice as strong in boys (β=-0.11, SE=0.03, p<0.01) than in girls (β=-0.04, SE=0.03, p=0.17), but the difference was not significant (z test: p=0.14). Constraining the models further to have equal residual variance between the ancestry groups or sex led to a significant worsening of fit, suggesting that the amount of unexplained variance differs in both groups. In other words, the included psychopathology subscales show a different amount of variation between groups, which could not be explained by the tested variables. Boys have more or less unexplained variance than girls depending on the specific psychopathology subscale and children of European ancestry showed more residual variance than children with

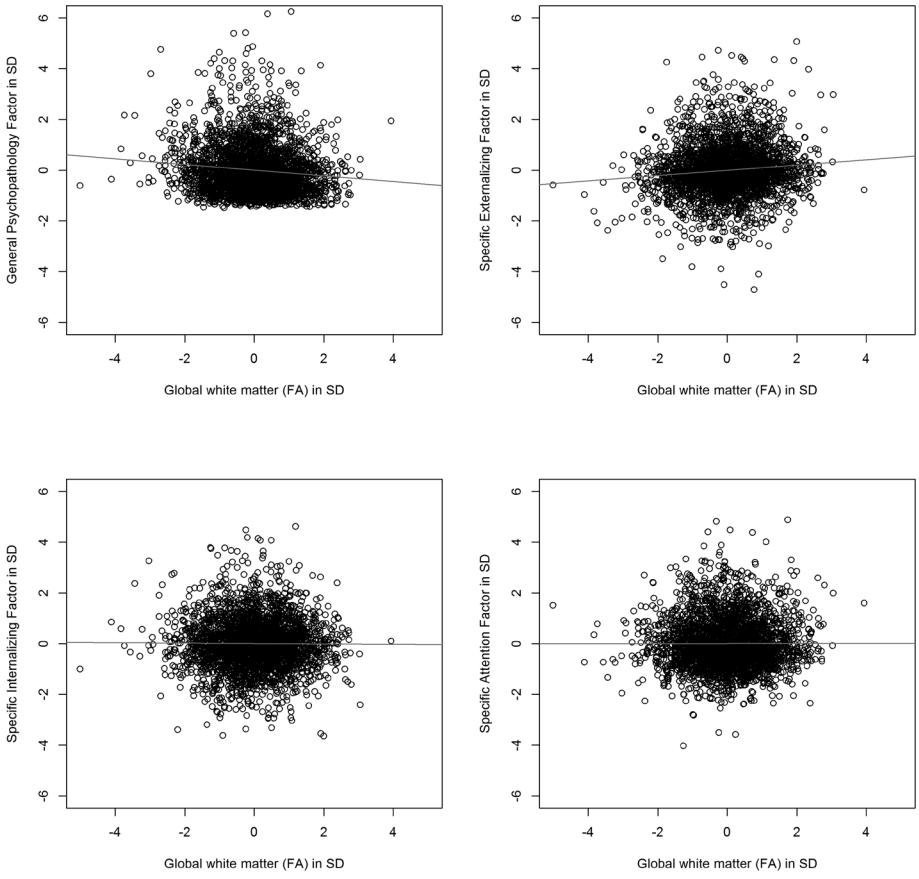
other ancestries across all subscales. We did not find evidence that the residuals differ between socioeconomic status.

Non-response analysis

For 4974 children information on psychopathology was available, but no DTI scans. These children had higher amounts of maternally rated anxiety (2.26 vs 2.10, p=0.03), aggression (3.10 vs 2.70, p<0.01), and attention (3.44 vs 3.08, p<0.01). Furthermore, the percentage of household incomes below 2800€ was higher (43% vs 33%, p<0.01), as well as percentage of mothers with no higher education (48% vs 36%, p<0.01). Whether the relation between white matter microstructure and the psychopathology factors is differentially associated in the this group is unknown given that DTI information is missing and no proxy of sufficient quality are available for imputation. T-tests were computed with the BSDA 1.2.0 package.[37]



Supplementary Figure 1: Participant flow chart



Supplementary Figure 2: Scatter plots between global white matter microstructure and psychopathology factors, based on estimates of factor scores (n=3,030)

Table S 1: Item loadings on general and specific psychopathology factors

Indicators	n_{obs}	Factor											
		General			Specific Int			Specific Ext			Specific Att		
		λ	SE	p	λ	SE	p	λ	SE	p	λ	SE	p
Mother													
Age 6, CBCL													
Emotional Reactivity	2635	.52	.02	<.01	.05	.02	<.01						
Anxious Depressed	2639	.42	.02	<.01	.25	.02	<.01						
Somatic Problems	2637	.40	.03	<.01	.24	.02	<.01						
Withdrawn	2640	.43	.02	<.01	.02	.02	.35						
Attention	2644	.49	.02	<.01							.34	.02	<.01
Aggression	2634	.58	.02	<.01				.27	.02	<.01			
Sleep	2636	.40	.02	<.01									
Other	2648	.59	.02	<.01									
Age 10, CBCL													
Anxious Depressed	2612	.58	.02	<.01	.34	.02	<.01						
Withdrawn Depressed	2604	.49	.02	<.01	.17	.02	<.01						
Somatic Problems	2594	.52	.02	<.01	.35	.02	<.01						
Social Problems	2600	.72	.02	<.01									
Thought Problems	2598	.65	.02	<.01									
Other	2601	.75	.01	<.01									
Attention	2599	.63	.02	<.01							.58	.02	<.01
Rule-breaking	2599	.63	.02	<.01				.32	.03	<.01			
Aggression	2599	.73	.02	<.01				.45	.02	<.01			
Father													
Age 10, CBCL													

Table S 1: (Continued)

Anxious Depressed	2015	.53	.02	<.01	.33	.02	<.01							
Withdrawn Depressed	2010	.40	.03	<.01	.18	.02	<.01							
Somatic Problems	1997	.47	.02	<.01	.32	.02	<.01							
Social Problems	2006	.63	.02	<.01										
Thought Problems	2009	.57	.02	<.01										
Other	2010	.66	.02	<.01										
Attention	2009	.57	.02	<.01								.52	.01	<.01
Rule-breaking	2010	.53	.02	<.01				.28	.02	<.01				
Aggression	2010	.62	.02	<.01				.39	.03	<.01				
Teacher														
Age 7, TRF														
Anxious Depressed	1776	.18	.02	<.01	.26	.02	<.01							
Withdrawn Depressed	1776	.17	.02	<.01	.26	.02	<.01							
Somatic Problems	1770	.14	.02	<.01	.15	.02	<.01							
Social Problems	1776	.26	.02	<.01										
Thought Problems	1776	.21	.03	<.01										
Attention	1776	.32	.02	<.01								.22	.01	<.01
Rule-breaking	1776	.22	.03	<.01				.10	.02	<.01				
Aggression	1776	.26	.03	<.01				.14	.02	<.01				
Child														
Age 6, BPI														
Depression	2796	.17	.02	<.01	.13	.02	<.01							

Table S 1: (Continued)

Separation Anxiety	2791	.10	.02	<.01	.17	.02	<.01							
Overanxious-ness	2798	.15	.02	<.01	.18	.02	<.01							
Oppositionality	2795	.11	.02	<.01				.13	.02	<.01				
Hostility	2800	.16	.02	<.01				.12	.02	<.01				
Conduct Prob-lems	2795	.16	.02	<.01				.13	.02	<.01				
Age 10, BPM														
Attention	2465	.39	.02	<.01							.45	.02	<.01	
Externalizing	2457	.42	.02	<.01				.30	.02	<.01				
Internalizing	2462	.35	.02	<.01	.20	.02	<.01							
Thought Prob-lems	2465	.32	.02	<.01										

λ standardized loading, **SE** standard error, **Int** Internalizing, **Ext** Externalizing, **Att** Attention, $\mathbf{N}_{obs}$ observed sample size of indicator, analysis n = 3,030

Table S 2: Model fit indices of models with and without the general psychopathology factor (n=3,030)

Model	robust CFI	BIC	robust RMSEA
2-factor	T.804	752891	.052 [.051, .053]
3-factor	.829	750612	.049 [.048, .050]
General Factor	.876	746645	.042 [.041, .043]

2-factor The 2-factor model contains internalizing and externalizing factors and no general factor
3-factor Same as 2-factor, but attention items load on a separate attention factor instead of on an externalizing factor
General Factor The general factor model is the same as the 3-factor, but contains a general psychopathology factor. This is the main model used for hypothesis testing (see Methods).
robust CFI robust version of the Comparative Fit Index.
BIC Bayesian Information Criterion
robust RMSEA robust version of the Root Mean Square Error of Approximation. A 90% confidence interval is given.

Table S 3: White matter tract descriptives and loadings on global white matter microstructure

Tract FA values	Hemisphere	n_{obs}	Mean	SD	λ	SE	p
Cingulum bundle	Left	3030	.42	.04	.46	.02	<.01
	Right	3029	.37	.04	.41	.02	<.01
Corticospinal tract	Left	3028	.54	.02	.49	.02	<.01
	Right	3028	.53	.02	.50	.02	<.01
Forceps major	-	3030	.57	.03	.42	.02	<.01
Forceps minor	-	3030	.60	.03	.53	.02	<.01
Inferior longitudinal fasciculus	Left	3030	.43	.02	.57	.02	<.01
	Right	3030	.44	.02	.62	.01	<.01
Superior longitudinal fasciculus	Left	3030	.40	.02	.68	.01	<.01
	Right	3030	.40	.02	.74	.01	<.01
Uncinate fasciculus	Left	3030	.39	.03	.46	.02	<.01
	Right	3030	.40	.03	.57	.02	<.01

SD standard deviation, **λ** standardized loading, **SE** standard error
n_{obs} observed sample size of indicator, analysis n = 3,030

Table S 4: Psychopathology factor scores regressed on white matter microstructure scores (quadratic model)

	Factor Scores											
	General			Specific Ext			Specific Int			Specific Att		
Predictor	β	SE	p	β	SE	p	β	SE	p	β	SE	p
Global FA score	-.09	.02	<.01	.08	.02	<.01	.03	.02	.08	-.01	.02	.57
Global FA score squared	-.00	.01	.99	-.02	.01	.19	-.00	.01	.71	.02	.01	.06

Factor scores were z-score standardized. All models are adjusted for sex, age at scan, maternal and paternal education, household income and genetic ancestry (n = 3,030).
Ext Externalizing, **Int** Internalizing, **Att** Attention, **β** Standardized regression coefficient, **SE** standard error

Table S 5: Standardized path coefficients of covariates in IQ and parental psychopathology adjusted model

Outcome	Predictor	β	SE	p
General Psychopathology	Global FA	-.06	.02	<.01
	Sex (Female)	-.11	.02	<.01
	Age at scan	.03	.02	.18
	Paternal education	-.06	.03	.03
	Maternal education	-.07	.03	<.01
	Income	-.09	.03	<.01
	Ancestry Component 1	.00	.02	.85
	Ancestry Component 2	.01	.02	.59
	Ancestry Component 3	-.11	.02	<.01
	Ancestry Component 4	-.04	.02	.04
	IQ	-.10	.02	<.01
	Maternal interpersonal sensitivity	.19	.04	<.01
	Maternal depression	-.06	.04	.13
	Maternal anxiety	.11	.04	<.01
	Maternal hostility	.13	.04	<.01

Table S 5: (Continued)

Outcome	Predictor	β	SE	p
	Paternal interpersonal sensitivity	.06	.04	.12
	Paternal depression	.05	.04	.22
	Paternal anxiety	.03	.03	.29
	Paternal hostility	.11	.03	<.01
Specific Externalizing	Global FA	.07	.03	<.01
	Sex (Female)	-.09	.02	<.01
	Age at scan	.00	.03	.99
	Paternal education	.03	.03	.31
	Maternal education	.09	.03	<.01
	Income	-.04	.03	.23
	Ancestry Component 1	-.01	.03	.72
	Ancestry Component 2	.03	.02	.27
	Ancestry Component 3	.03	.02	.31
	Ancestry Component 4	-.01	.02	.60
	IQ	-.05	.03	.03

Table S 5: (Continued)

Outcome	Predictor	β	SE	p
	Maternal interpersonal sensitivity	-.09	.04	.03
	Maternal depression	-.12	.05	<.01
	Maternal anxiety	-.04	.04	.39
	Maternal hostility	.19	.04	<.01
	Paternal interpersonal sensitivity	.07	.04	.08
	Paternal depression	-.05	.04	.22
	Paternal anxiety	-.08	.04	.04
	Paternal hostility	.05	.03	.16
Specific Internalizing	Global FA	.02	.03	.47
	Sex (Female)	.27	.02	<.01
	Age at scan	.00	.03	.86
	Paternal education	.09	.03	<.01
	Maternal education	.02	.03	.59
	Income	-.09	.03	<.01
	Ancestry Component 1	.02	.03	.55

Table S 5: (Continued)

Outcome	Predictor	β	SE	p
	Ancestry Component 2	-.07	.03	<.01
	Ancestry Component 3	.10	.03	<.01
	Ancestry Component 4	.01	.03	.83
	IQ	.03	.02	.29
	Maternal interpersonal sensitivity	.03	.04	.49
	Maternal depression	-.04	.05	.41
	Maternal anxiety	.16	.05	<.01
	Maternal hostility	-.05	.04	.24
	Paternal interpersonal sensitivity	.06	.04	.15
	Paternal depression	.08	.04	.08
	Paternal anxiety	.00	.04	.94
	Paternal hostility	.00	.04	.90
Specific Attention	Global FA	.01	.02	.60
	Sex (Female)	-.17	.02	<.01
	Age at scan	-.03	.02	.24

Table S 5: (Continued)

Outcome	Predictor	β	SE	p
	Paternal education	-.05	.03	.10
	Maternal education	-.01	.03	.61
	Income	.00	.03	.93
	Ancestry Component 1	.07	.02	<.01
	Ancestry Component 2	.04	.02	.05
	Ancestry Component 3	-.11	.02	<.01
	Ancestry Component 4	-.03	.02	.11
	IQ	-.15	.02	<.01
	Maternal interpersonal sensitivity	-.09	.03	<.01
	Maternal depression	.07	.04	.08
	Maternal anxiety	-.04	.04	.23
	Maternal hostility	.00	.03	.96
	Paternal interpersonal sensitivity	-.02	.03	.49
	Paternal depression	-.04	.04	.23
	Paternal anxiety	-.01	.03	.71

Table S 5: (Continued)

Outcome	Predictor	β	SE	p
	Paternal hostility	-.03	.03	.33
Global FA	Sex (Female)	-.03	.02	.13
	Age at scan	.15	.02	<.01
	Paternal education	-.01	.03	.60
	Maternal education	.03	.03	.21
	Income	.04	.03	.17
	Ancestry Component 1	.05	.02	.04
	Ancestry Component 2	.08	.02	<.01
	Ancestry Component 3	.00	.02	.93
	Ancestry Component 4	-.01	.02	.71
	IQ	.12	.02	<.01
	Maternal interpersonal sensitivity	.04	.03	.19
	Maternal depression	-.08	.03	.02
	Maternal anxiety	.05	.03	.09
	Maternal hostility	-.01	.03	.84

Table S 5: (Continued)

Outcome	Predictor	β	SE	p
	Paternal interpersonal sensitivity	.02	.03	.47
	Paternal depression	-.02	.03	.62
	Paternal anxiety	-.02	.03	.46
	Paternal hostility	.00	.03	.85

β Standardized regression coefficient, **SE** standard error

Table S 6: Clustering of significantly associated voxels

					Minimum p Voxel Coordinates			Center of Gravity Voxel Coordinates		
Scalar	Factor	n_{voxels}	Total %	Min p	X	Y	Z	X	Y	Z
FA	GPF	1330	14.34	.008	27	42	45	40.2	53.7	48.1
		111	1.20	.044	38	54	31	35.6	56.6	39.7
		94	1.01	.037	63	42	49	63.0	44.2	40.0
		12	0.13	.046	61	61	28	61.5	59.4	27.6
		1	0.01	.050	35	63	45	35.0	63.0	45.0
FA	EXT	4699	50.68	<.001	40	58	31	44.1	57.7	46.2
		76	0.82	.035	52	49	42	48.7	54.6	42.5
		35	0.38	.037	40	59	44	40.4	55.9	44.0
		20	0.22	.045	49	54	53	48.5	55.0	52.7
		7	0.08	.048	41	52	40	41.0	53.7	40.1
		3	0.03	.047	42	55	38	42.7	54.7	37.7
		1	0.01	.050	60	57	37	60.0	57.0	37.0
		1	0.01	.050	60	56	39	60.0	56.0	39.0
MD	EXT	5144	55.48	<.001	37	59	55	44.4	58.5	45.2
		3	0.03	.049	49	51	41	49.0	51.3	41.0
		2	0.02	.049	50	54	44	49.5	54.0	44.0
RD	EXT	6262	67.54	<.001	68	45	28	44.6	58.0	45.0
		11	0.12	.049	42	42	24	42.1	42.7	25.5
		8	0.09	.049	36	38	22	35.3	36.8	21.9
		1	0.01	.050	43	44	22	43.0	44.0	22.0

Voxel-wise analysis was performed with TBSS adjusted for sex, age at scan, maternal and paternal education, household income and genetic ancestry (n = 2,996). Voxel coordinates are in reference to the Generation R atlas (https://www.nitrc.org/projects/genr).
FA fractional anisotropy, **MD** mean diffusivity, **RD** radial diffusivity, **GPF** General Psychopathogy Factor, **Ext** Specific Externalizing Factor, $\mathbf{n_{voxels}}$ Number of significantly associated voxels contained in cluster, **Total %** nvoxels divided by total number of voxels tested (nttotal_voxels = 9,272) **Min p** p-value of voxel with the lowest p value adjusted for multiple testing

Table S 7: Number of significant voxels per region

		Factor			
		(DTI scalar)			
		General	Specific EXT		
Region	**Hemisphere**	**FA**	**FA**	**MD**	**RD**
Anterior thalamic radiation	Left	0	1	1	2
	Right	9	2	3	5
Cingulum cingulate gyrus	Left	0	51	12	2
	Right	3	36	0	36
Corticospinal tract	Left	117	8	9	10
	Right	4	8	8	10
Forceps major	-	7	11	2	7
Forceps minor	-	12	95	66	102
Inferior f ronto-occipital fasciculus	Left	23	1	2	1
	Right	2	6	18	13
Inferior longitudinal f asciculus	Left	124	2	4	2
	Right	1	1	1	1
Superior longitudinal fasciculus	Left	8	7	24	9
	Right	9	9	13	10
Superior longitudinal fasciculus temporal part	Left	0	0	1	0
	Right	4	6	7	6
Uncinate fasciculus	Left	9	0	1	3
	Right	0	34	35	3

Voxel-wise analysis was performed with TBSS adjusted for sex, age at scan, maternal and paternal education, household income and genetic ancestry (n = 2,996). Regions are based on the John Hopkins University Atlas in MNI-152 space.
Ext Externalizing, **FA** Fractional Anisotryopy, **MD** mean diffusivity, **RD** radial diffusivity, **AD** axial diffusivity

Table S 8: Ancestry invariance analysis ($n_{Europen}$ = 1136, $n_{non\text{-}European}$ = 758)

Model	Constraints	robust CFI	ΔCFI
Configural	no	.906	-
Weak	Loadings	.904	.002
Strong	Loadings, Intercepts	.903	.001
Strong and Relational	Loadings, Intercepts, Regressions	.900	.003
Strict and Relational	Loadings, Intercepts, Regressions, Residuals	.889	.011

ΔCFI represents the change in CFI from the more complex to more constrained model

Table S 9: Sex invariance analysis (n_{boys} = 1510, n_{girls} = 1529)

Model	Constraints	robust CFI	ΔCFI
Configural	no	.898	-
Weak	Loadings	.893	.005
Strong	Loadings, Intercepts	.889	.004
Strong and Relational	Loadings, Intercepts, Regressions	.886	.003
Strict and Relational	Loadings, Intercepts, Regressions, Residuals	.871	.015

ΔCFI represents the change in CFI from the more complex to more constrained model

Table S 10: SES invariance analysis ($n_{low\ or\ middle}$ = 994, n_{high} = 1665)

Model	Constraints	robust CFI	ΔCFI
Configural	no	.887	-
Weak	Loadings	.885	.002
Strong	Loadings, Intercepts	.885	.000
Strong and Relational	Loadings, Intercepts, Regressions	.883	.002
Strict and Relational	Loadings, Intercepts, Regressions, Residuals	.880	.003

ΔCFI represents the change in CFI from the more complex to more constrained model

REFERENCES

1. Jaddoe VW V, van Duijn CM, Franco OH, van der Heijden AJ, van Iizendoorn MH, de Jongste JC et al. The Generation R Study: design and cohort update 2012. Eur J Epidemiol 2012; 27: 739–56.
2. Tiemeier H, Velders FP, Szekely E, Roza SJ, Dieleman G, Jaddoe VW V et al. 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. J Am Acad Child Adolesc Psychiatry 2012; 51: 1119–1135.e7.
3. White T, Muetzel R, Marroun H El, Blanken L, Jansen P, Bolhuis K et al. Paediatric Population Neuroimaging and the Generation R Study: The second wave. 2017; In Press.
4. Achenbach TM, Ruffle TM. The Child Behavior Checklist and Related Forms for Assessing Behavioral/Emotional Problems and Competencies. Pediatr Rev 2000; 21: 265–271.
5. Achenbach TM, Rescorla LA. Manual for the ASEBA School-Age Forms and Profiles. . Burlington, VT, 2001.
6. Arseneault L, Moffitt TE, Caspi A, Taylor A, Rijsdijk F V, Jaffee SR et al. Strong genetic effects on cross-situational antisocial behaviour among 5-year-old children according to mothers, teachers, examiner-observers, and twins' self-reports. J Child Psychol Psychiatry 2003; 44: 832–848.
7. Ringoot AP, Jansen PW, Steenweg-de Graaff J, Measelle JR, van der Ende J, Raat H et al. Young children's self-reported emotional, behavioral, and peer problems: the Berkeley Puppet Interview. Psychol Assess 2013; 25: 1273–1285.
8. Achenbach TM, McConaughy SH, Ivanova MY, Rescorla LA. Manual for the ASEBA Brief Problem Monitor (BPM). Burlington, VT ASEBA 2011.
9. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. Fsl. Neuroimage 2012; 62: 782–790.
10. Cook P a, Bai Y, Seunarine KK, Hall MG, Parker GJ, Alexander DC. Camino: Open-Source Diffusion-MRI Reconstruction and Processing. 14th Sci Meet Int Soc Magn Reson Med 2006; 14: 2759.
11. Chang L, Jones DK, Pierpaoli C. RESTORE: Robust Estimation of Tensors by Outlier Rejection. Magn Reson Med 2005; : 1088–1095.
12. Muetzel RL, Blanken LME, Ende J Van Der, Marroun H El, Shaw P, Sudre G et al. Tracking Brain Development and Dimensional Psychiatric Symptoms in Children: A Longitudinal Population-Based Neuroimaging Study. Am J Psychiatry 2017. doi:10.1176/appi.ajp.2017.16070813.
13. Groot M De, Ikram MA, Akoudad S, Krestin GP, Hofman A, Lugt A Van Der et al. Tract-specific white matter degeneration in aging: The Rotterdam Study. Alzheimer's Dement 2015; 11: 321–330.
14. Behrens TEJ, Berg HJ, Jbabdi S, Woolrich MW. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? Neuroimage 2007; : 144–155.
15. Muetzel RL, Mous SE, van der Ende J, Blanken LME, van der Lugt A, Jaddoe VW V et al. White matter integrity and cognitive performance in school-age children: A population-based neuroimaging study. Neuroimage 2015; 119: 119–128.
16. Fischl B. FreeSurfer. Neuroimage 2012; 62: 774–781.
17. Muetzel RL, Blanken LME, Thijssen S, van der Lugt A, Jaddoe VWV, Verhulst FC et al. Resting-state networks in 6-to-10 year old children. Hum Brain Mapp 2016; 37: 4286–4300.

18. Andersson J, Smith S, Jenkinson M. Fnirt-fmrib's non-linear image registration tool. Hum Brain Mapp 2008; 2008.
19. Tellegen P, Laros J. The construction and validation of a nonverbal test of intelligence: the revision of the Snijders-Oomen tests. Eur J Psychol Assess 1993; 9: 147–157.
20. Putnam SP, Rothbart MK. Development of short and very short forms of the Children's Behavior Questionnaire. J Pers Assess 2006; 87: 102–112.
21. Stichting Cito Instituut voor Toetsontwikkeling Arnhem. Cito Centrale Eindtoets. 2018.https://www.cito.nl/onderwijs/primair-onderwijs/centrale-eindtoets/ (accessed 19 Sep2018).
22. Derogatis LR, Melisaratos N. The brief symptom inventory: an introductory report. Psychol Med 1983; 13: 595–605.
23. Medina-Gomez C, Felix JF, Estrada K, Peters MJ, Herrera L, Kruithof CJ et al. Challenges in conducting genome-wide association studies in highly admixed multi-ethnic populations: the Generation R Study. Eur J Epidemiol 2015; 30: 317–330.
24. R Core Team. R: A Language and Environment for Statistical Computing. 2016.https://www.r-project.org/.
25. Rosseel Y. lavaan: An R package for structural equation modeling. J Stat Softw 2012.http://www.lce.esalq.usp.br/arquivos/aulas/2013/encontro_ppg/Lucia/paper.pdf (accessed 13 Apr2015).
26. Oberski DL. lavaan.survey: An R ackage for complex survey analysis of structural equation models. J Stat Softw 2014; 57: 1–27.
27. Buuren S, Groothuis-Oudshoorn K. MICE: Multivariate imputation by chained equations in R. J Stat Softw 2011.http://doc.utwente.nl/78938/ (accessed 13 Apr2015).
28. Neumann A, Pappa I, Lahey BB, Verhulst FC, Medina-Gomez C, Jaddoe VW et al. Single Nucleotide Polymorphism Heritability of a General Psychopathology Factor in Children. J Am Acad Child Adolesc Psychiatry 2016; 55: 1038–1045.
29. Caspi A, Houts RM, Belsky DW, Goldman-Mellor SJ, Harrington H, Israel S et al. The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? Clin Psychol Sci 2014; 2: 119–137.
30. Laceulle OM, Vollebergh W a. M, Ormel J. The Structure of Psychopathology in Adolescence: Replication of a General Psychopathology Factor in the TRAILS Study. Clin Psychol Sci 2015; : 1–11.
31. Lahey BB, Krueger RF, Rathouz PJ, Waldman ID, Zald DH. A hierarchical causal taxonomy of psychopathology across the life span. Psychol Bull 2017; 143: 142–186.
32. Takao H, Hayashi N, Inano S, Ohtomo K. Effect of head size on diffusion tensor imaging. Neuroimage 2011; 57: 958–967.
33. Smith SM, Nichols TE. Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 2009; 44: 83–98.
34. Cheung G, Rensvold R. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equ Model 2002.http://www.tandfonline.com/doi/abs/10.1207/s15328007sem0902_5 (accessed 12 Nov2015).
35. Beaujean A. Latent variable modeling using R: A step-by-step guide. 1st ed. Routledge, 2014.
36. Hirschfeld G, Von Brachel R. Multiple-Group confirmatory factor analysis in R – A tutorial in measurement invariance with continuous and ordinal indicators. Pract Assessment, Res Eval 2014; 19: 1–11.

37. Brill R V. Basic Statistics and Data Analysis. 2005.