Data Supplement for Muetzel et al., Tracking Brain Development and Dimensional Psychiatric Symptoms in Children: A Longitudinal Population-Based Neuroimaging Study. Am J Psychiatry (doi: 10.1176/appi.ajp.2017.16070813)

Supplemental Methods

Child psychiatric symptom assessment

The Child Behavior Checklist was used to assess a wide range of symptoms in children. At time 1 the CBCL/1½-5 was used, and at time 2 the CBCL-6-18 was used. The items utilize a three-point scale ("Not True", "Somewhat or Sometimes True", "Very or Often True"), and were filled in by the primary caregiver, which was most often the biological mother (92%). Examples of the affective problem scale items include, "Looks unhappy without good reason" and "Unhappy, sad, or depressed", and examples of the anxiety problem items are, "Worries" and "Nervous, high-strung, or tense".

Non-verbal intelligence

General intellectual functioning was assessed during the age-6 assessment wave using an abbreviated version of the Snijders-Oomen Niet-verbale Intelligentie Test – Revisie (SON-R 2½-7) (1, 2). The SON-R 2½-7 is a measure of non-verbal intelligence for children between 2.5 and 7 years of age and was selected in order to minimize language-dependent confounds that may be present in a large, ethnically diverse sample such as the Generation R Study. An intelligence quotient (IQ) was estimated from the two SON-R performance subtests that were administered (*Mosaics* and *Categories*), which is highly correlated with estimates resulting from the complete version (3).

Maternal psychiatric symptoms

Maternal psychopathology was assessed using the Brief Symptom Inventory (BSI), a 53-item self-report questionnaire (4). The full BSI was administered during pregnancy and a short version was administered again at roughly 3 months postnatally (1). A global severity index was computed by taking an average of the item scores. An average GSI was computed when data were available from both administrations, otherwise the single available GSI was used.

Image Acquisition

Structural MRI data were acquired using an IR-prepared Fast Spoiled Gradient Recalled sequence. For time 1, the following sequence parameters were used: T_R = 10.3 ms, T_E = 4.2 ms, T_I = 350 ms, flip angle = 16°, Acquisition Matrix = 256 x 256, field of view (FOV) = 230.4 mm, slice thickness = 0.9mm, Asset Acceleration Factor = 2. For time 2, similar sequence parameters were used with the GE option *BRAVO*: T_R=8.77ms, T_E=3.4ms, T_I=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.

Diffusion MRI data were collected at time 1 with 3 b=0 volumes and 35 diffusion directions using an echo planar imaging sequence (T_R = 11,000 ms, T_E = 83 ms, FOV = 256 mm x 256 mm, Acquisition Matrix = 128 x 128, slice thickness = 2 mm, number of slices = 77, b = 1000 s/mm^2). For time 2, a similar 35-direction echo planar imaging sequence was utilized (T_R = 12,500 ms, T_E = 72 ms, FOV = 240 mm x 240 mm, Acquisition Matrix = 120 x 120, slice thickness = 2 mm, number of slices = 65, Asset Acceleration Factor = 2, b = 900 s/mm^2).

Diffusion Image preprocessing

Image preprocessing was conducted using the Functional MRI of the Brain's Software Library (FSL, version 5.0.9, 5) and the Camino Diffusion Toolkit (6) via the Neuroimaging in Python Pipelines and Interfaces package (Nipype, version 0.92, 7). Details of the image processing have been described in detail elsewhere (8). Diffusion images were first corrected for eddy current-induced artifacts and translations/rotations resulting from head motion, and non-brain tissue was removed (9). In order to account for rotations applied to the diffusion data, the resulting transformation matrices were used to rotate the diffusion gradient direction table. The diffusion tensor was fit using the RESTORE method (10), and common scalar metrics (e.g., FA, MD) were subsequently computed.

Fiber tractography

Probabilistic fiber tractography was run on each subject's diffusion data using the fully automated, freely available FSL plugin, "AutoPtx"(11). The Bayesian Estimation of Diffusion Parameters Obtained using Sampling Techniques (BESTPOSTx) package from FSL was first used to estimate the diffusion parameters at each voxel, accounting for two fiber orientations (12). Next, a predefined set of seed and target masks, supplied by the AutoPtx software, were aligned to each subject's diffusion data in native space using a nonlinear registration. The FSL probabilistic fiber tracking algorithm, Probtrackx, was then used to identify connectivity distributions for a number of large, commonly reported fiber bundles, based on the predefined seed and target marks. Connectivity distributions obtained from the fiber tracking process were then normalized based on the number of successful seed-to-target attempts, and then thresholded to remove voxels that were unlikely to be part of the true distribution. For each tract, average DTI scalar metrics (e.g., FA, MD), weighted by the connectivity distribution, were computed (8). Thus, compared to voxels with a lower probability, voxels with a high probability of being part of the true white matter bundle have a higher contribution to the average DTI scalar value computed across the entire tract.

Image quality assurance

FreeSurfer reconstructions were visually inspected using a protocol similar to previously defined methods and datasets not suitable for analysis were excluded (13). The white and pial surface representations were inspected for accuracy against the brain image at a number of slices in different planes (i.e., axial/coronal/sagittal).

Diffusion image quality was assessed using two methods. First, the DTIPrep tool (https://www.nitrc.org/projects/dtiprep/) was used to automatically examine the data for slicewise variation, characteristic of artifact, in each diffusion-weighted volume. Second, the sum-of-squares error (SSE) maps from the diffusion tensor calculations were examined for structured signal that was indicative of artifact. Each SSE map was rated from 0 to 3 (0: "None",

1: “Mild”, 2: “Moderate”, 3: “Severe”). Any cases not excluded by the automated DTIPrep tool but still had a “Severe” score from the SSE rating were also excluded from analyses. Processed tractography data were also examined for problems in two ways. First, the registration of the DTI data to standard space was inspected for accuracy. Second, each tract was examined for grossly misclassified voxels in the connectivity distribution. The flow chart in Figure S1 outlines the number of datasets excluded for the above outlined quality assurance measures.

Statistical Analysis

Linear mixed-effects models were chosen to compliment analyses with cross-lagged panel models, as an explicit change term is available for interpretation. Models were fit using the LME4 package (14) to assess the association between psychiatric symptom scores and longitudinal changes in white matter microstructure. Linear mixed-effects models have numerous appealing features, including modeling of random effects, flexibility in uneven durations between time-points and handling of missing time-points. Models 1 below outlines the general modeling strategy used.

Model 1**: **Baseline psychiatric symptoms predicting changes in brain

$MRI_{ij} \sim Age\text{-}MRI_{ij} + Age\text{-}CBCL_i + Sex_i + Ethnicity_i + CBCL_i + \mathbf{Age\text{-}MRI_{ij} \times CBCL_i} + (1 \mid Subject)$

The *i* subscript denotes subject, the *j* subscript denotes time-point, and models were adjusted for age at MRI, age at CBCL behavioral assessment, sex and ethnicity. The “1 | subject” term denotes the random intercept for subject (i.e., adjusting for the baseline value in longitudinal outcomes). The interaction term in **bold** is the coefficients of interest (e.g., CBCL score predicting change in MRI). Hypothesis testing on model estimates was made using the *Anova* function from the *Car* package.

Multiple imputation of covariates

For the supplementary follow-up analyses (see below), data were missing on the covariate (child IQ and maternal psychiatric problems) in some cases. To account for missingness, multiple imputation, as implemented in the R-package “mice” was utilized. 30 imputed datasets were generated using the predicted mean matching method. Linear mixed models were run on each of the 30 imputed datasets, and effect estimates and p-values were pooled across the analyses using the mice function ‘pool’. For analyses involving the structural MRI volumes, data on IQ were missing at time 1 in 71 children and data on maternal psychopathology were missing in 71 mothers. For DTI analyses, data were missing on IQ at time 1 in 59 children and data on maternal psychopathology was missing in 63 mothers.

Non-response Analysis

A non-response analysis was conducted to compare characteristics of children with and without a time 2 assessment. A two-sample t-test showed a difference in age at time 1 between children with (M = 7.7 years) and without (M = 8.2 years) a time 2 assessment ($t(713) = 5.9$, $p < 0.05$). Chi-squared tests did not show a difference in distribution of sex ($\chi(1, 715) = 0.4$, $p = 0.53$) or ethnicity ($\chi(2, 715) = 1.3$, $p = 0.51$). The internalizing problems scores ($t(713) = 1.3$, $p = 0.21$) and externalizing problems scores ($t(713) = 1.9$, $p = 0.06$) did not differ between children with and without a time 2 MRI. Maternal education level was similar in participants with and without

a time 2 visit ($p = 0.09$*),* though family income tended to be higher in children with a time 2 visit ($p = 0.049$).

Supplemental Results

Adjustment for non-verbal IQ
When non-verbal IQ was included in significant linear mixed effects models (i.e., for total subcortical volume), associations between checklist scores and changes in subcortical volume were highly similar for both externalizing (Est. = –0.010, $p = 0.031$) and for internalizing scores (Est. = –0.012, $p = 0.012$). For the DTI global FA data, results remained virtually unchanged for both externalizing (Est = –0.05, $p = 0.036$) and internalizing (Est = –0.06, $p = 0.020$). However, though IQ is related to both DTI data and to psychiatric problems, lower IQ is often considered part of a given psychiatric disorder and thus inclusion in the model may be over-adjustment.

Adjustment of maternal psychopathology
Given that in most cases the mother rated the child's psychiatric symptoms, there is a potential for reporter bias if the mother herself has some level of psychiatric symptoms. For externalizing scores and subcortical volume, results remained similar (Est. = –0.01, $p = 0.028$), as were associations with internalizing scores (Est. = –0.011, $p = 0.012$) after adjusting models for maternal psychopathology. For the DTI global FA data, results remained highly consistent for both externalizing (Est. = –0.05, $p = 0.036$) and internalizing (Est = –0.06, $p = 0.018$).

Adjustment for motion
To rule out results from the DTI data were affected by motion, we looked at the impact of adding the total percentage of affected slices to linear mixed-effects models. This number is derived from the automated DTIprep software. Slice-wise correlations in the diffusion signal are conducted, and a drop in correlation coefficient among slices largely reflects signal attenuation in diffusion-weighted volumes due to motion and other artifacts. Slices labeled by the software as 'affected' were counted and the percentage of the total slices in the DTI series was computed. Results remained mostly consistent when the percent of affected slices was added for changes in global FA with externalizing problems (Est. = –0.052, $p = 0.037$) and with internalizing problems (Est. = –0.063, $p = 0.019$).

Adjustment for socioeconomic factors
When maternal education and income were added as covariates to models, results remained highly consistent for subcortical volume changes and changes in global FA for both internalizing and externalizing scores.

Adjustment for psychiatric medication use
When psychiatric medication status (yes/no) was added to models, results remained highly consistent for global FA and externalizing (Est. = -0.52, p = 0.04) as well as for internalizing (Est. = -0.063, p = 0.02). Similarly, results were consistent after adjusting for psychiatric medication use for externalizing and subcortical volume (Est = -0.011, p = 0.02) and for internalizing and subcortical volume (Est. = -0.011, p = 0.011).

Exclusion of clinical CBCL scores

In order to determine whether the associations between psychiatric scores at time 1 were related to changes in brain metrics purely along a continuum or if the severe end of the continuum was driving the associations, additional sensitivity analyses were run where individuals surpassing the 'clinical' cutoff on the CBCL scale were excluded (15). Table S1 outlines the percentages of participants with a score in the clinical range. For externalizing scores, 7.7% of the children were excluded, and for internalizing scores, 13.1% of the children were excluded.

Linear mixed models of externalizing symptoms at time 1 showed a similar association with changes in subcortical volume over time (Est = –0.10, $p = 0.067$) after excluding these children. This stable effect estimate and slightly increased p-value are consistent with a loss of power (7.7% of the subjects) rather than effects being solely driven by the extreme cases. Interestingly, results for internalizing scores at time 1 were even stronger after excluding cases in the clinical range, where scores predicted smaller increases in subcortical volume over time (Est. = –0.02, $p = 0.001$).

After excluding scores in the clinical range, externalizing scores at time 1 showed a similar association with changes in global FA over time (Est = –0.05, $p = 0.055$). Again, a consistent effect estimate and slightly higher p-value are consistent with reducing the sample by 13%. However, for internalizing scores, in addition to a slight decrease in effect estimate, the p-value was considerably higher (Est. = –0.045, $p = 0.20$). Thus, it is possible that for internalizing, individuals in the clinical range may have a distinct pattern of change in global FA over time.

Tables and Figures

TABLE S1. Percentage of subjects meeting borderline and clinical cutoff for the different broadband and DSM-oriented CBCL scales.

CBCL	Scale	Borderline (%)	Clinical (%)
Broadband	Externalizing	16.0	7.7
	Internalizing	22.0	13.1
DSM-Oriented	ADHD	9.7	5.1
	ODD	8.3	2.6
	AFF	15.3	6.0
	ANX	11.5	4.0

Note: Values are percentages. Cutoffs based on data from Tick et al. 2007. Data are from time 1 (baseline). CBCL = Child Behavior Checklist, ADHD = Attention deficit/hyperactivity disorder scale, ODD = oppositional defiant disorder scale, AFF = affective problems scale, ANX = anxiety problems scale.

TABLE S2. Correlation matrix of global anatomical metrics and broadband CBCL scores at both time points

		Time-1							Time-2						
		Ext	Int	TBV	Cortical	WM	Subcort	Vent	Ext	Int	TBV	Cortical	WM	Subcort	Vent
Time-1	Ext	1	-	-	-	-	-	-	-	-	-	-	-	-	-
	Int	0.72	1	-	-	-	-	-	-	-	-	-	-	-	-
	TBV	–0.09	–0.10	1	-	-	-	-	-	-	-	-	-	-	-
	Cortical	–0.12	–0.11	0.96	1	-	-	-	-	-	-	-	-	-	-
	WM	–0.07	–0.09	0.94	0.82	1	-	-	-	-	-	-	-	-	-
	Subcort	–0.08	–0.06	0.77	0.68	0.77	1	-	-	-	-	-	-	-	-
	Vent	0.07	0.04	0.29	0.24	0.26	0.27	1	-	-	-	-	-	-	-
Time-2	Ext.	0.64	0.44	–0.03	–0.07	0.00	0.03	0.07	1	-	-	-	-	-	-
	Int.	0.45	0.55	–0.08	–0.08	–0.07	–0.04	0.08	0.56	1	-	-	-	-	-
	TBV	–0.10	–0.09	0.85	0.77	0.86	0.79	0.30	0.03	–0.03	1	-	-	-	-
	Cortical	–0.10	–0.07	0.78	0.76	0.72	0.69	0.24	0.03	–0.01	0.95	1	-	-	-
	WM	–0.09	–0.09	0.82	0.69	0.91	0.76	0.25	0.02	–0.06	0.94	0.78	1	-	-
	Subcort	–0.10	–0.08	0.72	0.64	0.74	0.92	0.25	0.00	–0.05	0.79	0.70	0.76	1	-
	Vent.	0.05	0.04	0.28	0.22	0.24	0.26	0.96	0.07	0.06	0.30	0.23	0.25	0.25	1

Note: Ext = CBCL broadband externalizing, Int = CBCL broadband internalizing, TBV = total brain volume, Cortical = cortical gray matter volume, WM = white matter volume, Subcort = subcortical volume, Vent = lateral ventricle volume. Values are Spearman's rank coefficients.

TABLE S3. Correlation matrix of global DTI and broadband CBCL data at both time points

		Time-1				Time-2			
		Externalizing	Internalizing	Global FA	Global MD	Externalizing	Internalizing	Global FA	Global MD
Time-1	Externalizing	1	-	-	-	-	-	-	-
	Internalizing	0.72	1	-	-	-	-	-	-
	Global FA	0.02	0.02	1	-	-	-	-	-
	Global MD	0.03	–0.01	–0.49	1	-	-	-	-
Time-2	Externalizing	0.63	0.44	0.01	0.10	1	-	-	-
	Internalizing	0.45	0.55	0.01	0.07	0.52	1	-	-
	Global FA	–0.07	–0.10	0.64	–0.45	–0.02	0.00	1	-
	Global MD	0.05	0.05	–0.39	0.79	0.11	0.04	–0.54	1

Note: FA = fractional anisotropy, MD = mean diffusivity. Externalizing and Internalizing are broadband CBCL problem scores. Values are Spearman's rank coefficients.

TABLE S4. Linear mixed-effects models of psychiatric symptoms and global metrics of change in cortical morphology

Time 1 Predictor	Longitudinal Outcome	Est.	SE	T	χ^2	*p*
Externalizing	Total Brain Volume	0.002	0.006	0.241	0.058	0.809
	Cortical Volume	0.005	0.008	0.581	0.338	0.561
	White Matter Volume	–0.002	0.005	–0.446	0.199	0.655
	Subcortical Volume	–0.010	0.005	–2.184	4.769	0.029
	Lateral Ventricle Volume	–0.003	0.002	–1.245	1.549	0.213
Internalizing	Total Brain Volume	0.003	0.007	0.446	0.199	0.656
	Cortical Volume	0.007	0.008	0.827	0.684	0.408
	White Matter Volume	–0.001	0.005	–0.255	0.065	0.799
	Subcortical Volume	–0.012	0.005	–2.502	6.260	0.012
	Lateral Ventricle Volume	–0.001	0.002	–0.375	0.140	0.708

Note: Models adjusted for fixed effects, age at behavioral assessment, sex, and ethnicity, and random effects of subject. CBCL behavioral scores are square root transformed, and model estimates represent the association between psychiatric problems and change in MRI metric over time (i.e., the interaction between CBCL score and age).

TABLE S5. Linear mixed-effects models of associations between psychiatric symptom scores and changes in white matter microstructure

Longitudinal Outcome	Time 1 Predictor	Estimate	SE	T	χ^2	*p*
FA	Externalizing	–0.053	0.025	–2.100	4.409	0.036
	Internalizing	–0.063	0.027	–2.353	5.539	0.019
MD	Externalizing	0.000	0.002	0.210	0.044	0.834
	Internalizing	0.002	0.002	1.090	1.189	0.276

Note: Models adjusted for fixed effects, age at behavioral assessment, sex, and ethnicity, and random effects of subject. CBCL behavioral scores are square root transformed, and model estimates represent the association between psychiatric problems and change in global DTI metric (i.e., the interaction between CBCL score and age).

TABLE S6. Linear mixed-effects models of psychiatric symptoms predicting changes in FA for individual tracts

Time 1 Predictor	Tract	Hemisphere	Estimate	SE	T	χ^2	*p*	p_{FDR}
Externalizing	CB	L	–0.011	0.011	–1.022	1.045	0.307	0.460
	CB	R	–0.021	0.011	–1.934	3.739	0.053	0.160
	CST	L	–0.004	0.011	–0.369	0.136	0.712	0.855
	CST	R	0.002	0.012	0.135	0.018	0.892	0.974
	FMA	-	–0.015	0.010	–1.455	2.117	0.146	0.250
	FMI	-	0.007	0.012	0.594	0.352	0.553	0.737
	ILF	L	–0.023	0.010	–2.413	5.823	0.016	0.087
	ILF	R	–0.017	0.010	–1.706	2.909	0.088	0.209
	SLF	L	–0.024	0.011	–2.294	5.261	0.022	0.087
	SLF	**R**	**–0.036**	**0.011**	**–3.240**	**10.499**	**0.001**	**0.014**
	UF	L	0.000	0.010	0.003	0.000	0.998	0.998
	UF	R	0.017	0.010	1.622	2.632	0.105	0.209
Internalizing	CB	L	–0.022	0.012	–1.843	3.397	0.065	0.196
	CB	**R**	**–0.033**	**0.012**	**–2.850**	**8.124**	**0.004**	**0.026**
	CST	L	–0.006	0.012	–0.514	0.264	0.607	0.686
	CST	R	0.013	0.012	1.044	1.089	0.297	0.490
	FMA	-	–0.010	0.011	–0.871	0.759	0.384	0.511
	FMI	-	0.001	0.013	0.078	0.006	0.938	0.938
	ILF	L	–0.014	0.010	–1.323	1.750	0.186	0.446
	ILF	R	–0.011	0.011	–0.981	0.962	0.327	0.490
	SLF	**L**	**–0.030**	**0.011**	**–2.686**	**7.217**	**0.007**	**0.029**
	SLF	**R**	**–0.051**	**0.012**	**–4.322**	**18.679**	**0.000**	**0.000**
	UF	L	0.005	0.011	0.483	0.233	0.629	0.686
	UF	R	0.012	0.011	1.108	1.227	0.268	0.490

Note: Models adjusted for fixed effects, age at behavioral assessment, sex, and ethnicity, and random effects of subject. CBCL behavioral scores are square root transformed, and model estimates represent the association between psychiatric problems and change in MRI metric. Bold represents significant association after adjusting for multiple comparisons. CB = cingulum bundle, CST = corticospinal tract, FMA = forceps major, FMI = forceps minor, ILF = inferior longitudinal fasciculus, SLF = superior longitudinal fasciculus, UF = uncinate fasciculus.

TABLE S7. linear mixed effects models of DSM-oriented subscales predicting changes in MRI metrics

Longitudinal Outcome	Time 1 Predictor	Est	SE	T	χ^2	p
Subcortical Volume	ADHD	–0.013	0.007	–1.945	3.784	0.052
	ODD	–0.012	0.007	–1.793	3.213	0.073
	Affective	–0.008	0.008	–1.027	1.055	0.304
	Anxiety	–0.011	0.007	–1.566	2.453	0.117
Global FA	ADHD	–0.096	0.038	–2.501	6.254	0.012
	ODD	–0.080	0.037	–2.123	4.507	0.034
	Affective	–0.164	0.043	–3.834	14.697	0.000
	Anxiety	–0.050	0.039	–1.293	1.671	0.196

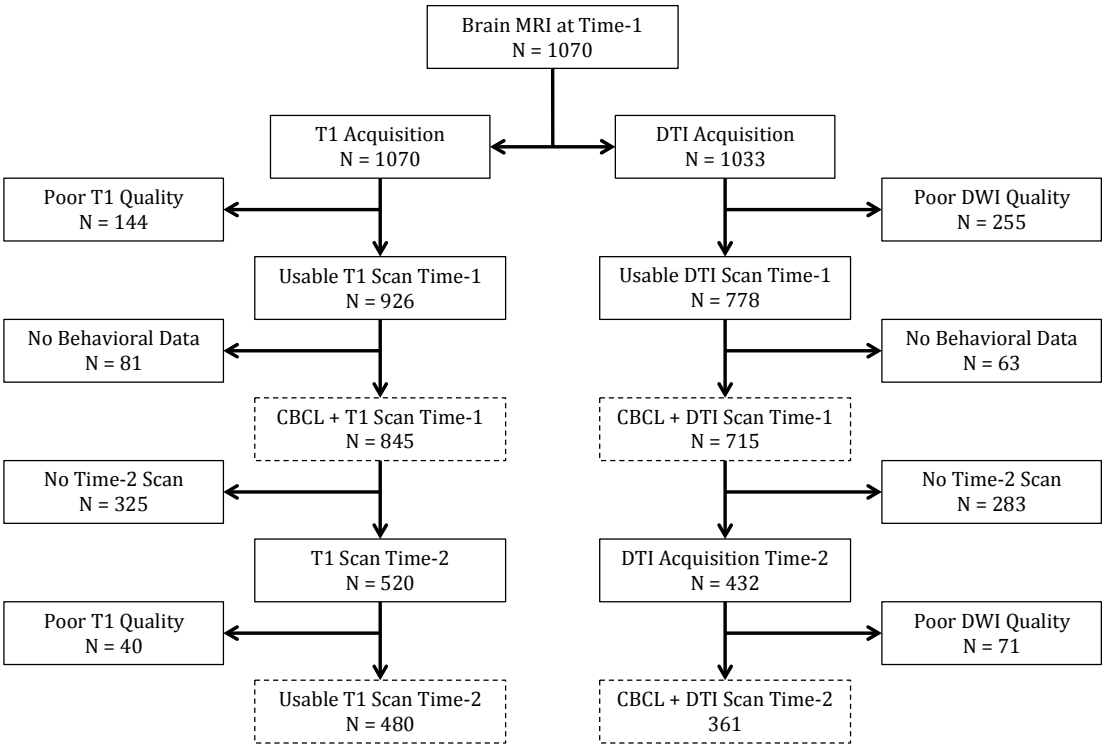
Note: Models adjusted for fixed effects, age at behavioral assessment, sex, and ethnicity, and random effects of subject. CBCL behavioral scores are square root transformed, and model estimates represent the association between psychiatric problems and change in global MRI metric (i.e., the interaction between CBCL score and age). CBCL = Child Behavior Checklist, ADHD = Attention deficit/hyperactivity disorder scale, ODD = oppositional defiant disorder scale.

TABLE S8. Main effects and model estimates for covariates from linear mixed effects models

Longitudinal Outcome	Time-1 Predictor	Model Term	Estimate	SE	T	χ^2	*p*
FA	Externalizing	Intercept	–0.45	1.17	–0.39	0.15	0.70
		Externalizing	–0.08	0.06	–1.28	1.64	0.20
		Age MRI	0.19	0.03	5.65	31.98	0.00
		Age CBCL	0.16	0.19	0.83	0.70	0.40
		Sex	–0.33	0.16	–2.04	4.14	0.04
		Ethnicity 1	–0.16	0.31	–0.51	0.26	0.61
		Ethnicity 2	–0.59	0.21	–2.86	8.16	0.00
	Internalizing	Intercept	–0.67	1.16	–0.58	0.33	0.56
		Internalizing	–0.01	0.06	–0.21	0.05	0.83
		Age MRI	0.19	0.03	5.64	31.81	0.00
		Age CBCL	0.16	0.19	0.86	0.74	0.39
		Sex	–0.30	0.16	–1.83	3.36	0.07
		Ethnicity 1	–0.17	0.31	–0.56	0.31	0.58
		Ethnicity 2	–0.61	0.21	–2.94	8.67	0.00
Subcortical Volume	Externalizing	Intercept	–0.42	0.44	–0.95	0.89	0.34
		Externalizing	–0.07	0.02	–3.09	9.57	0.00
		Age MRI	0.02	0.01	3.35	11.23	0.00
		Age CBCL	0.17	0.07	2.34	5.48	0.02
		Sex	–0.73	0.06	–11.43	130.62	0.00
		Ethnicity 1	–0.02	0.12	–0.13	0.02	0.90
		Ethnicity 2	–0.26	0.08	–3.23	10.46	0.00
	Internalizing	Intercept	–0.60	0.44	–1.37	1.87	0.17
		Internalizing	–0.03	0.02	–1.07	1.14	0.28
		Age MRI	0.02	0.01	3.34	11.13	0.00
		Age CBCL	0.17	0.07	2.42	5.87	0.02
		Sex	–0.70	0.06	–11.07	122.54	0.00
		Ethnicity 1	–0.02	0.12	–0.19	0.03	0.85
		Ethnicity 2	–0.28	0.08	–3.52	12.37	0.00

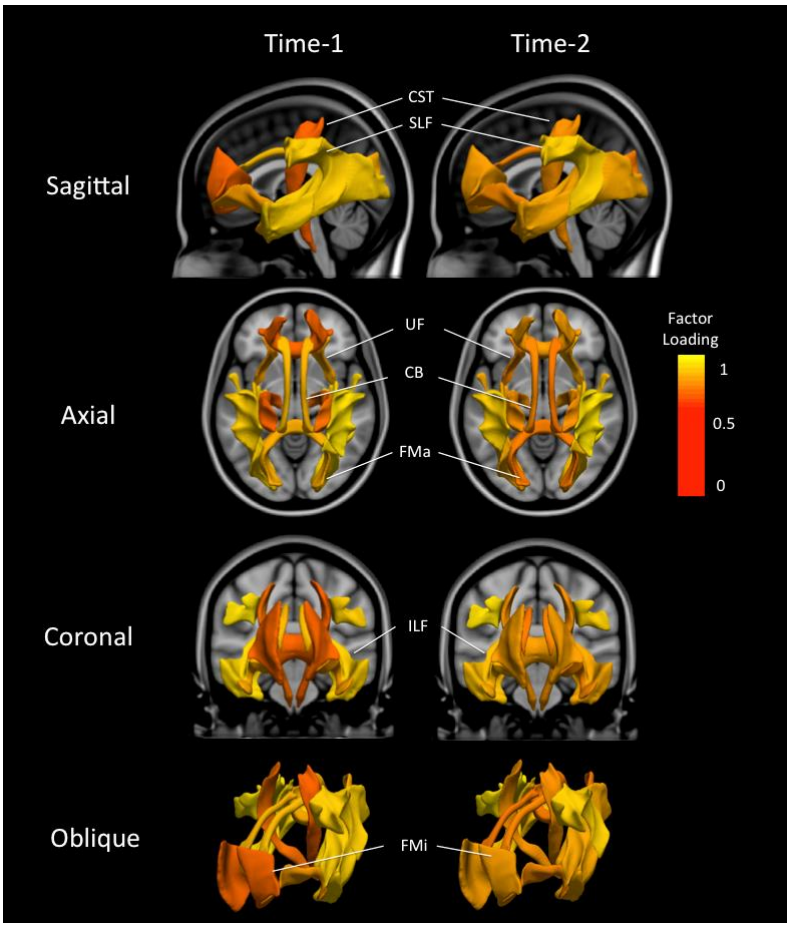
Note: CBCL behavioral scores are square root transformed. The variable for sex in the model is for females coded as 1, males as 0. The variable for ethnicity was reference coded into 3 groups, with the Dutch group being the largest and thus the reference group, and group 1 coded for other Western and group 2 coded for non-Western.

FIGURE S1. Study Inclusion Flow Chart



Note: Dotted lines represent data used in analyses. For cross-lagged analyses, only the complete Time-1/Time-2 data are used. For linear mixed models, all Time-1 data and Time-2 data were used.

FIGURE S2. Depiction of standardized factor loadings for latent global FA factor in cross-lagged models by time-point.



CB = cingulum bundle, CST = corticospinal tract, FMa = forceps major, FMi = forceps minor, ILF = inferior longitudinal fasciculus, SLF = superior longitudinal fasciculus, UF = uncinate fasciculus.

FIGURE S3. Associations between early psychiatric problems and white matter microstructural changes: clinically-relevant vs. normal range problem scores

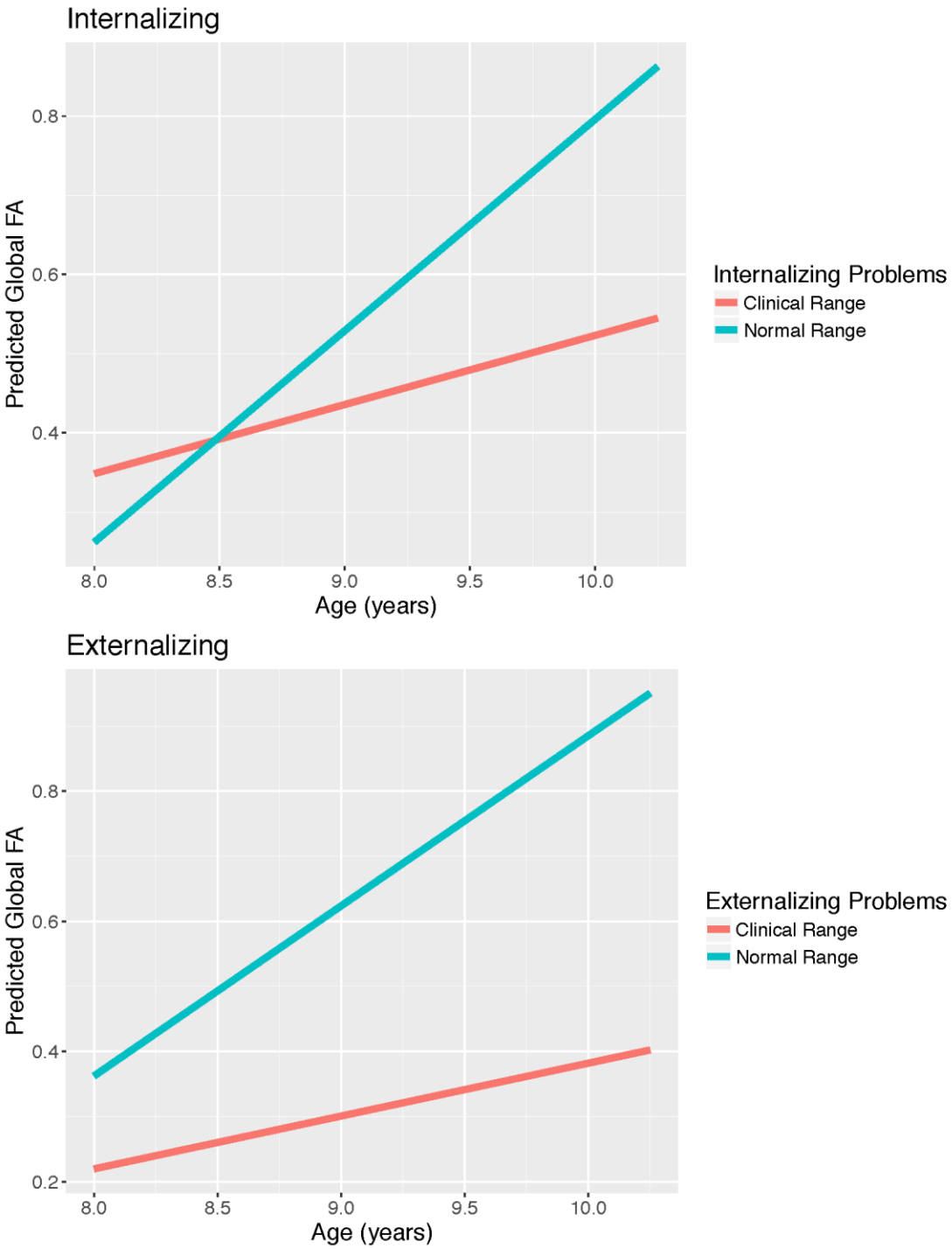


Figure represents predicted model estimates derived from linear mixed effects models. The top panel shows broadband internalizing problems and the bottom panel shows broadband externalizing problems. Separate lines for those who score in the normal and in the clinical range on any CBCL domain, with the Y-axis representing the predicted global FA value based on model estimates. Note: Fractional anisotropy is a unit-less measure.

References

1. Tiemeier H, Velders FP, Szekely E, Roza SJ, Dieleman G, Jaddoe VW, Uitterlinden AG, White TJ, Bakermans-Kranenburg MJ, Hofman A, Van Ijzendoorn MH, Hudziak JJ, Verhulst FC. The Generation R Study: A review of design, findings to date, and a study of the 5-HTTLPR by environmental interaction from fetal life onward. Journal of the American Academy of Child and Adolescent Psychiatry. 2012;51:1119-1135 e1117.
2. Tellegen PJ, Winkel M, Wijnberg-Williams B, Laros JA: Snijders-Oomen Niet-Verbale Intelligentietest: SON-R 2-1/2 -to-7. Amsterdam, Boom Testuitgevers; 2005.
3. Basten M, van der Ende J, Tiemeier H, Althoff RR, Rijlaarsdam J, Jaddoe VW, Hofman A, Hudziak JJ, Verhulst FC, White T. Nonverbal intelligence in young children with dysregulation: the Generation R Study. Eur Child Adolesc Psychiatry. 2014;23:1061-1070.
4. de Beurs E: Brief Symptom Inventory, Manual. Leiden, the Netherlands2004.
5. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. Fsl. Neuroimage. 2012;62:782-790.
6. Cook PA, Bai Y, Nedjati-Gilani S, Seunarine KK, Hall MG, Parker GJ, Alexander DC: Camino: Open-Source Diffusion-MRI Reconstruction and Processing. in 14th Scientific Meeting of the International Society for Magnetic Resonance in Medicine. Seattle, WA, USA2006. pp. 2759.
7. Gorgolewski K, Burns CD, Madison C, Clark D, Halchenko YO, Waskom ML, Ghosh SS. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. Front Neuroinform. 2011;5:13.
8. Muetzel RL, Mous SE, van der Ende J, Blanken LM, van der Lugt A, Jaddoe VW, Verhulst FC, Tiemeier H, White T. White matter integrity and cognitive performance in school-age children: A population-based neuroimaging study. Neuroimage. 2015;119:119-128.
9. Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. Med Image Anal. 2001;5:143-156.
10. Chang LC, Jones DK, Pierpaoli C. RESTORE: robust estimation of tensors by outlier rejection. Magn Reson Med. 2005;53:1088-1095.
11. de Groot M, Ikram MA, Akoudad S, Krestin GP, Hofman A, van der Lugt A, Niessen WJ, Vernooij MW. Tract-specific white matter degeneration in aging: The Rotterdam Study. Alzheimers Dement. 2015;11:321-330.
12. Behrens TE, Berg HJ, Jbabdi S, Rushworth MF, Woolrich MW. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? Neuroimage. 2007;34:144-155.
13. Hibar DP, Stein JL, Renteria ME, et al. Common genetic variants influence human subcortical brain structures. Nature. 2015;520:224-229.
14. Bates D, Maechler M, Bolker B, Walker S. lme4: Linear mixed-effects models using Eigen and S4. https://cranr-projectorg/package=lme4. 2015.
15. Tick NT, van der Ende J, Koot HM, Verhulst FC. 14-year changes in emotional and behavioral problems of very young Dutch children. Journal of the American Academy of Child and Adolescent Psychiatry. 2007;46:1333-1340.