Online Supplement to

Staginnus et al.

Testing the Ecophenotype Model: Cortical Structure Alterations in Conduct Disorder With Versus Without Childhood Maltreatment

Contents

SUPPLEMENTARY METHODS

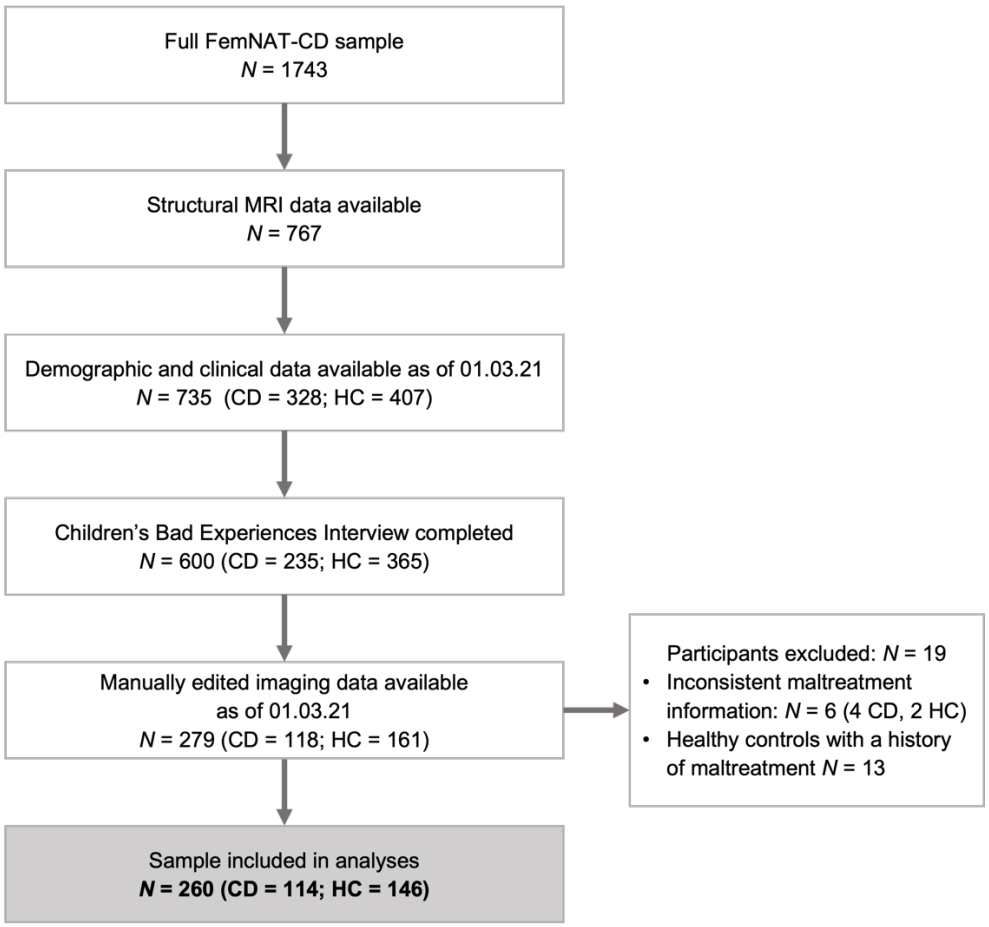

Figure S1. Flowchart of the participant inclusion process
[a] MRI=Magnetic Resonance Imaging; CD=Conduct Disorder; HC=Healthy controls

The current project drew on structural neuroimaging data that had been (partly) pre-processed and edited (blind to group status) in the context of two previous projects from our research group. The first of these projects aimed for an equal sex distribution across Conduct Disorder (CD) and healthy control (HC) groups but did not require specific maltreatment measures to be available (1). Consequently, the Children's Bad Experiences (CBE) interview, which was the focus of the current study, was not available for all participants of this subsample. Additionally, due to a data storage issue, some of the edited data from this project are no longer available. The second (ongoing) project included participants with data

on a range of psychopathology and adversity measures including the CBE. Editing for the latter project was ongoing when the present study was conceived and hence, to increase sample size for the current analyses, we prioritized CD subjects with a history of maltreatment in the editing process, whilst aiming for age- and sex-matched groups. Maltreated controls were not prioritized as according to the CBE, only 18 controls had experienced maltreatment. Thus, this group size was considered insufficient for meaningful statistical inferences. As of 01.03.2021 (the date of data freezing for the current project), 279 participants had CBE as well as manually-edited structural MRI data available. Of those, 13 participants were controls with a history of maltreatment and were thus excluded and 6 further participants were excluded due to inconsistent maltreatment information on the CBE, resulting in a final sample size of *N*=260.

Table S1. Number of participants per contributing site[a]

Group	Site 1 Frankfurt	Site 2 Aachen	Site 3 Southampton	Site 4 Basel	Site 5 Birmingham
CD/-	17	20	9	1	18
CD/+	14	17	7	3	8
HC	41	23	33	9	40
Total	**72**	**60**	**49**	**13**	**66**

[a]CD/-=Conduct Disorder without a history of childhood maltreatment; CD/+=Conduct Disorder with a history of childhood maltreatment; HC=healthy controls. There was a trend towards a significant association between site and group, but it did not reach statistical significance, p=.063, Fisher's exact test.

We note that 82 participants (32% of the sample) were included in a previous structural MRI study from our research group (1).

Ethical approvals

The study was conducted in accordance with the legal regulations of the European Union, national legislations, and the Declaration of Helsinki. The original study

protocols were approved by the relevant ethical committees at each site prior to the start of data collection: the Ethics Committee of the Medical Faculty of Goethe University Frankfurt for the Frankfurt site (site 1), RWTH Aachen University Hospital (EK027/14) for the Aachen site (site 2), the Ethics Commission Northwest and Central Switzerland (EKNZ: 336/13) for the Basel site (site 4), and the Southampton University Ethics Committee (ERGO Number: 18970) and National Health Service Research Ethics Committee (NRES Committee West Midlands, Edgbaston; REC reference 13/WM/0483) for the UK sites [Southampton (site 3) and Birmingham (site 5)]. At German-speaking sites (Basel, Frankfurt, Aachen), youths aged 18 years provided informed consent, while younger participants provided assent and their parent/caregiver informed consent. At UK sites (Southampton, Birmingham), informed consent was obtained from youths aged 16 years or over, whereas those that were younger provided assent and informed consent was obtained from their parents or caregivers. Ethical approval for the current analyses was obtained from the University of Bath Psychology Research Ethics Committee (19-297).

Additional information on phenotypic measures

Diagnostic instrument (K-SADS-PL). The Kiddie-Schedule for Affective Disorders and Schizophrenia-Present and Lifetime version (K-SADS-PL) (2) was used to assess for CD diagnoses and presence of current or lifetime comorbid disorders, including Attention-Deficit/Hyperactivity Disorder (ADHD), Major Depressive Disorder (MDD), Generalized Anxiety Disorder (GAD), Posttraumatic Stress Disorder (PTSD) and alcohol or substance abuse based on DSM-IV-TR criteria (3). The interview was conducted separately with participants and their parents/caregivers. Information from both reports was integrated to reach a clinical judgment regarding presence or absence of diagnosis (or diagnoses). Inclusion in the clinical group required a DSM-

IV-TR diagnosis of CD or in younger children fulfilling criteria for Oppositional Defiant Disorder (ODD) and having 1–2 CD symptoms. The latter was the case for 17 CD participants (14.9% of CD participants), six of whom had a history of maltreatment. Controls had to be free of current DSM-IV-TR disorders and have no history of CD, ODD or ADHD. Inter-rater reliability for diagnoses within the FemNAT-CD study was assessed using a subsample of 75 participants (5-8 participants per contributing site) and was found to be high for CD (Cohen's κ=0.91; agreement rate=94.7%) as well as for other disorders including ODD, ADHD, MDD and GAD (Cohen's κs=0.84-0.95, agreement rates=92–95%). Participants with CD were classified as having childhood-onset CD if at least one symptom and functional impairment were reported to have emerged before the age of 10. Onset of symptoms after age 10 resulted in the participant being classified as having adolescent-onset CD (4).

Children's Bad Experiences (CBE) interview. Childhood maltreatment was assessed using the Children's Bad Experiences (CBE) interview (5,6) which was administered to the participant's parent or caregiver. In the current sample, 92.7% of informants were the biological parent of the participant (75.0% were the mother). Other informants included adoptive, foster or stepparents (3.5%), other relatives (1.9%), officials (0.8%) or others (0.8%). Information on the identity of the informant was missing for two participants (0.8%). The CBE is a standardized and structured interview protocol that has been used successfully in previous studies (e.g., (7–9)]. It comprises probe questions that aim to obtain information about potential physical and sexual abuse by an adult (questions 3-5, e.g., "Has your child ever been harmed on purpose by an adult?"), including information about whether and how often it occurred (never, yes, frequent, don't know/refuse to answer), what happened, the age at which it happened and whether it resulted in physical or psychological harm.

Additional questions aim to robustly rule out accidental and peer-to-peer harm (questions 1 and 2). However, these were not used at the FemNAT-CD UK sites (Southampton and Birmingham) as other questionnaires covering overlapping constructs such as bullying were employed. The CBE interview is designed to increase honesty by phrasing questions in a way that is agnostic to the perpetrator (i.e., asking whether the child was ever harmed, but not who did this) and hence offers a sensitive and acceptable way of assessing maltreatment. However, informants were reminded that if they reported on ongoing or very serious harm (e.g., sexual abuse) that had never come to the attention of official agencies, the interviewer would have to inform their supervisor, who would take further steps if necessary. Although the interview focuses on physical and sexual abuse, informants were free to mention any experiences they regarded as relevant. Following the interview, interviewers made an overall judgment, categorizing maltreatment as absent (no maltreatment reported), probable (maltreatment reported but unclear whether it definitely happened), or definite (maltreatment definitely happened, e.g., bruises were seen, child welfare services had been involved). Previous studies reported inter-rater agreement on 90% of ratings (8,10). To facilitate group analysis and in line with previous studies [e.g., (7)], we created a dichotomous variable reflecting no maltreatment exposure versus likely exposure, whereby the latter category comprised both probable and definite maltreatment.

Accordingly, 65 participants were classified as having CD without a history of maltreatment (CD/-) and 49 were classified as having CD with a history of childhood maltreatment (CD/+; 21 rated as probable and 28 as definite). 13 HCs with edited MRI data had a history of maltreatment. Due to the small size of this group, these participants were excluded from further analyses which meant that all included HCs

were free of maltreatment exposure (according to the CBE). Six further participants were excluded due to inconsistent information on the CBE (e.g., inconsistency between individual items and overall rating potentially owing to additional information being available to the rater). Although the interview asks about age at which the abuse occurred, this information was not reported/recorded systematically and was missing for many participants and reported incidents of abuse. Hence, the following information should be interpreted with caution. Across the 49 CD participants with probable or definite maltreatment, 73 questions related to the experience of maltreatment were endorsed and age information was provided for 48 of those (65.8%). Age at maltreatment ranged from 1 to 15 years with a median of 8.00 years and a mean of 7.69 years (SD = 4.15 years).

IQ measures. IQ was measured using the Wechsler Intelligence Scales. UK sites used the two subtest version of the Wechsler Abbreviated Scale of Intelligence (WASI-I) (11) and German/Swiss sites used the vocabulary and matrix reasoning subtests of the Wechsler Scales for children (WISC-III-R/IV) (12) for those aged 9-16 and the adult version (WAIS-III/IV) (13) for those aged ≥ 17 years.

SES. SES calculations were based on parental income, education, and occupational status, and made using the International Classification of Education (14) and the International Standard Classification of Occupations (15). Human rater and computer-based ratings were combined into a standardized factor (M = 0, SD = 1) score using Principal Component Analysis. Internal consistency of the composite SES score was acceptable (α = 0.74). Due to potential economic variation on the country level, SES was centered and scaled within each country, to obtain an indicator of relative socioeconomic position.

SES information was available for 239 of the 260 participants (91.9%). However, one participant's SES value was identified as an outlier (3*IQR) and subsequently removed due to implausibility.

Psychopathic and callous-unemotional traits. The total score of the 50-item self-report Youth Psychopathic traits Inventory (YPI) (16) was used to assess psychopathic traits (α=0.94), while its callous-unemotional subscale (α=0.79) and the total score of the 24-item parent-report Inventory of Callous-Unemotional traits (ICU, α=0.91) (17) assessed callous-unemotional traits.

The amount of missing data on the employed scales from the YPI and the ICU are reported in Table S2. Missing values of the respective questionnaire scores (subscale or total) were imputed by Dr Marietta Kirchner at the Institute of Medical Biometry (IMBI), Heidelberg, based on the whole FemNAT-CD sample prior to the inception of the current project. Missing values of single items were imputed first before calculating sum scores based on the imputed items as it has been shown that missing data in a multi-item instrument is best handled by imputation at the item level (18). Imputation was performed in SAS version 9.4 using the procedure PROC MI. Imputation by fully conditional specification (FCS) was used which offers a flexible method to specify the multivariate imputation model for arbitrary missing patterns including both categorical and continuous variables (19). As all questionnaire items were measured at the ordinal level, the logistic regression method was specified in the FCS statement. The following variables were included in the imputation model: all items of the respective questionnaire, age, IQ, group (case/control), sex (male/female), site, comorbidities (PTSD, ADHD, ODD, Depression, Anxiety – present/absent, respectively), and items of other questionnaires if they showed a correlation of $r \geq 0.4$ with at least one of the to-be-imputed items. In the case of the

YPI, one item from the Alabama Parenting Questionnaire (20) and three items from the Massachusetts Youth Screening Instrument II (MAYSI-II) (21) were considered. For the ICU, no items from other questionnaires were included in the imputation model. For imputation diagnostics, distributions of the observed and imputed items and scores were checked.

Notably, in the current study, psychopathic and callous-unemotional traits are only included to provide a comprehensive clinical comparison of the CD subgroups for the purposes of sample description but are not considered in the neuroimaging analyses.

Table S2. Missingness on the Youth Psychopathic traits Inventory and the Inventory of Callous-Unemotional traits questionnaires[a]

	YPI total		YPI CU subscale		ICU total	
	N	%	*N*	%	*N*	%
Participants with complete data	233	89.6%	251	96.5%	240	92.3%
Participants missing completely	1	0.4%	1	0.4%	4	1.5%
Participants with missing items	26	10.0%	8	3.1%	16	6.2%
Overall items missing	87	0.7%	23	0.6%	123	2.0%

[a] Percentages are based on the total sample size of 260 participants for rows referring to participants and on the total number of questionnaire/subscale items for rows referring to overall items (YPI total=13,000; YPI CU subscale=3,900; ICU total=6,240). YPI=Youth Psychopathic traits Inventory; YPI CU subscale=callous-unemotional traits subscale of the Youth Psychopathic traits Inventory; ICU=Inventory of Callous-Unemotional traits.

Table S3. Scanning parameters and acquisition sequences at each site

	Site 1 Frankfurt	Site 2 Aachen	Site 3 Southampton	Site 4 Basel	Site 5 Birmingham
Scanner make and model	Siemens Magnetom Tim Trio	Siemens Magnetom Prisma	Siemens Magnetom Tim Trio	Siemens Magnetom Prisma	Philips Achieva
Software version	Syngo MR A35	Syngo MR D13D	Syngo MR B17	Syngo MR D13D	Version 3.2.6.1
Head coil	8 channel	20 channel	32 channel	20 channel	32 channel
T1-weighted MPRAGE scanning parameters	TE = 3.4 ms, TR = 1900 ms, flip angle = 9°, FHxAP field of view (FoV) = 256 mm, RL FoV = 192 mm, matrix = 256, voxel size = 1x1x1 mm, sagittal slices = 192, bandwidth = 180 Hz/pixels, total scan time = 4 minutes 26 seconds	TE = 3.4 ms, TR = 1900 ms, flip angle = 9°, FHxAP field of view (FoV) = 256 mm, RL FoV = 192 mm, matrix = 256, voxel size = 1x1x1 mm, sagittal slices = 192, bandwidth = 180 Hz/pixels, total scan time = 4 minutes 26 seconds	TE = 3.4 ms, TR = 1900 ms, flip angle = 9°, FHxAP field of view (FoV) = 256 mm, RL FoV = 192 mm, matrix = 256, voxel size = 1x1x1 mm, sagittal slices = 192, bandwidth = 180 Hz/pixels, total scan time = 4 minutes 26 seconds	TE = 3.4 ms, TR = 1900 ms, flip angle = 9°, FHxAP field of view (FoV) = 256 mm, RL FoV = 192 mm, matrix = 256, voxel size = 1x1x1 mm, sagittal slices = 192, bandwidth = 180 Hz/pixels, total scan time = 4 minutes 26 seconds	TE = 3.7 ms, TR = 1900 ms, flip angle = 9°, FHxAP field of view (FoV) = 256 mm, RL FoV = 192 mm, matrix = 256, voxel size = 1x1x1 mm, sagittal slices = 192, bandwidth = 174 Hz/pixels, total scan time = 6 minutes 5 seconds

Site qualification procedures for MRI data acquisition

To ensure comparability of MRI data across the five scanning sites, each site adopted similar scanning parameters and image acquisition sequences (see Table S3) and underwent site qualification procedures to ensure that sequences were comparable. These included scanning an American College of Radiology (ACR) phantom (22), a Functional Biomedical Informatics Research Network (FBIRN) phantom (23), and a human volunteer. The ACR phantom is designed to assess structural MRI sequences, and the FBIRN is designed to assess scanner stability during functional MRI sequences, and provide information concerning scanner drift, percent fluctuation in signal, signal-to-noise ratio, and signal-to-fluctuation-noise ratio. Once collected, the datasets were reviewed by an MRI physicist at the University of Birmingham (Dr. Ali Chowdhury), and each site adjusted the scanning parameters according to the physicist's recommendations until the sites' scanning procedures were comparable. The sites were only able to start collecting data once they had successfully passed this site qualification procedure step.

Analytical justification

Clinical/demographic data. For the demographic and clinical comparisons, analyses were repeated using non-parametric approaches (Mann-Whitney U and Kruskal Wallis tests) if issues with normality were detected. Using robust statistics had no substantial impact on the *p*-values and did not change statistical significance or any interpretations.

Neuroimaging data. In the current study, a pairwise comparison approach was adopted to compare cortical structure between the HC, CD/- and CD/+ groups. While more parsimonious, omnibus F-tests (followed by pairwise comparisons in significant

regions) would not have allowed for a whole-brain comparison of individual group pairings. This was considered too conservative given the exploratory nature of the study and would have resulted in a lack of continuity with previous studies in CD, which have mostly adopted a pairwise comparison approach.

Of note, we corrected for CBCL attention problems (using raw scores) as opposed to ADHD symptoms in sensitivity analyses, as the latter showed no variance in the control group, which meant that it was not possible to run the statistical analyses.

SUPPLEMENTARY RESULTS

Total intracranial volume

Controlling for sex, age and site, there was a trend towards smaller total intracranial volume (TIV) in the overall CD group compared to the HCs, but this difference did not reach statistical significance, $t(252)=1.76$, $p=.079$, $d=0.22$ 95% CI [-0.02,0.47]. Similarly, neither the CD/-, $t(203)=1.73$, $p=.085$, $d=0.26$ 95% CI [-0.01,0.54], nor the CD/+ group, $t(187)=1.39$, $p=.167$, $d=0.23$ 95% CI [-0.05,0.52] differed significantly from HCs, or from each other, $t(106)=0.20$, $p=.842$, $d=0.04$ 95% CI [-0.33,0.41].

Surface-Based Morphometry results obtained when adjusting for IQ

Table S4. Significant group differences in cortical thickness, surface area, volume and gyrification when controlling for IQ[a]

Comparison	Cluster	Anatomical region	Measure	H	NVtxs	Size (mm^2)	Peak MNI coordinates			CWP	Max	Cohen's *d*	SD (*d*)
							x	y	z				
CD-all versus HC													
CD-all < HC	IQ1	**Pars triangularis,** pars orbitalis, rostral middle frontal	CT	R	2130	2075.51	51.0	30.4	-4.5	<.001	-3.73	-0.35	0.06
CD-all > HC	IQ2*	**Superior temporal gyrus**, transverse temporal gyrus	IGI	L	4245	2274.63	-62.0	-18.4	-4.1	.047	2.92	0.34	0.05
CD/- versus HC													
CD/- > HC	IQ3	**Superior temporal gyrus**, transverse temporal gyrus	IGI	L	5966	3153.77	-65.9	-17.3	-2.1	.003	3.30	0.41	0.07
CD/+ versus HC													
CD/+ < HC	IQ4	**Pars triangularis,** pars orbitalis, rostral middle frontal gyrus	CT	R	2213	1829.02	47.7	24.3	3.7	<.001	-3.24	-0.42	0.06
CD/+ < HC	IQ5	**Postcentral gyrus**, precentral gyrus	CT	R	2150	1296.48	67.1	-10.1	21.5	.008	-3.56	-0.45	0.07
CD/+ < HC	IQ6*	**Middle temporal gyrus,** banks of superior temporal sulcus	CT	L	2316	1296.20	-64.7	-40.7	-12.0	.009	-2.51	-0.41	0.04
CD/+ < HC	IQ7	**Postcentral gyrus**, precentral gyrus	CV	R	3302	1867.52	64.1	-13.3	13.4	.001	-2.69	-0.40	0.04
CD/+ < HC	IQ8	**Rostral middle frontal gyrus**, superior frontal gyrus, pars opercularis	IGI	R	4695	3481.79	25.1	54.6	23.9	.002	-2.90	-0.41	0.05
CD/+ versus CD/-													
CD/+ < CD/-	IQ9	**Superior temporal gyrus**	CV	R	1329	1238.58	59.1	4.8	-13.2	.025	-2.64	-0.47	0.05
CD/+ < CD/-	IQ10	**Supramarginal,** pre- and postcentral gyrus, inferior parietal lobule, middle temporal gyrus	IGI	R	16951	8993.40	58.0	-46.1	19.1	<.001	-2.82	-0.46	0.06

CD/+ < CD/-	IQ11	**Rostral middle frontal gyrus**	IGI	R	3470	2533.01	50.6	29.6	25.5	.018	-3.56	-0.51	0.07
CD/+ < CD/-	IQ12	**Fusiform gyrus**, lateral occipital pole	IGI	L	3992	3494.93	-26.4	-78.8	-8.7	.001	-4.39	-0.54	0.12
CD/+ < CD/-	IQ13*	**Superior parietal lobule,** supramarginal & postcentral gyrus	IGI	L	6354	3347.26	-38.8	-52.6	61.3	.001	-2.43	-0.43	0.04
CD/+ < CD/-	IQ14	**Inferior temporal gyrus**, superior & middle temporal gyrus, temporal pole	IGI	L	3169	3214.76	-52.1	-12.2	-40.9	.002	-3.70	-0.50	0.08

[a] All analyses controlled for sex, age, site, and total intracranial volume (except thickness). Monte Carlo corrections for multiple comparisons were applied. Cohen's *d* was calculated using whole-brain vertex-wise effect size brain maps. Bolded regions represent the location of the peak coordinate. IQ=intelligence quotient; H=hemisphere; NVtxs=number of vertices; MNI=Montreal Neurological Institute; CWP=cluster-wise *p*-value; Max=maximum -log10(*p*-value) in the cluster; CD=Conduct Disorder; HC=healthy controls; CD/-=Conduct Disorder without maltreatment; CD/+=Conduct Disorder with maltreatment; CT=cortical thickness, SA=surface area; CV=cortical volume; IGI=local gyrification index. *indicates clusters that were not identified in the main analyses not adjusting for IQ.

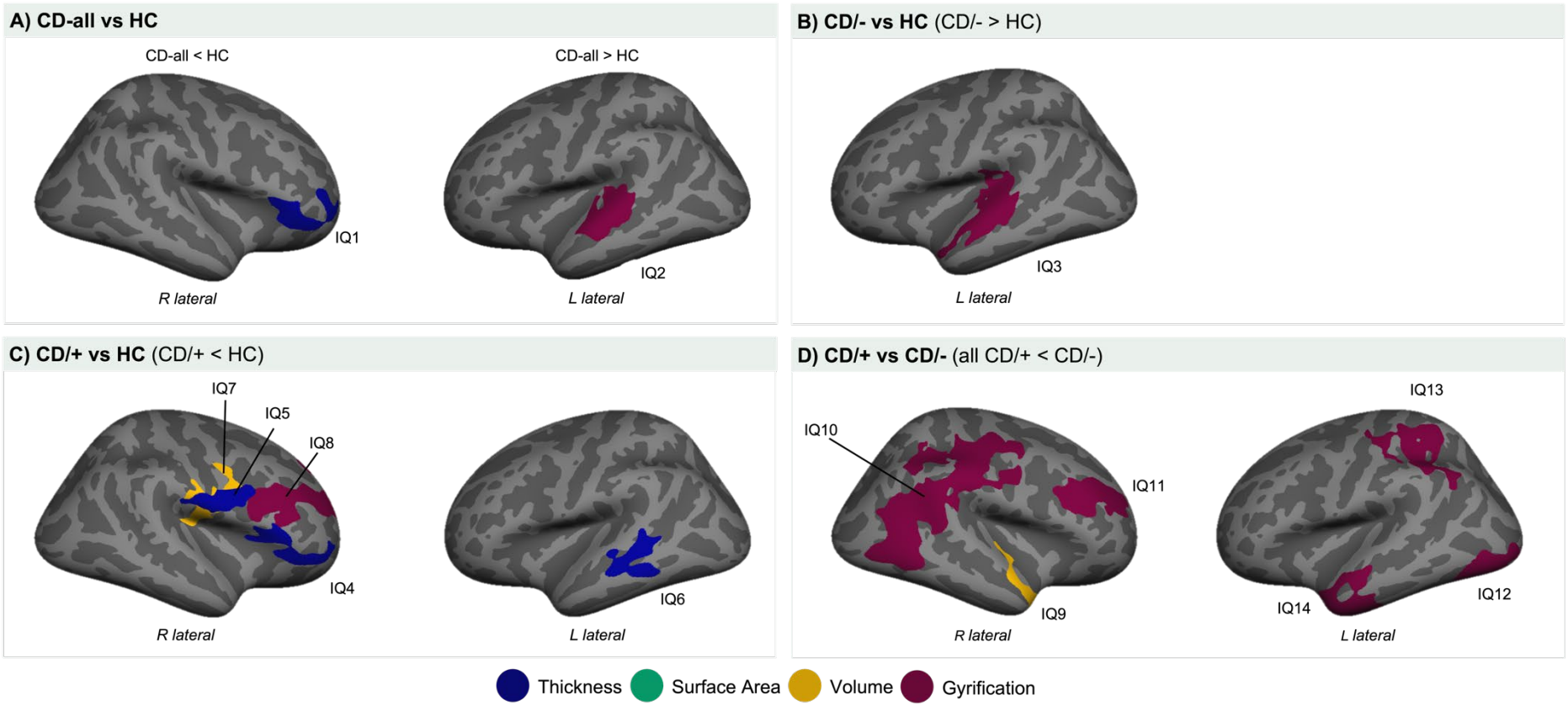

Figure S2. Group differences in cortical thickness, surface area, volume and gyrification when controlling for sex, age, site, total intracranial volume, and IQ[a]

[a] Total intracranial volume was not covaried in the cortical thickness analyses. CD=Conduct Disorder; HC=Healthy Controls. CD/-=Conduct Disorder without maltreatment history; CD/+=Conduct Disorder with maltreatment history.

Surface-Based Morphometry results obtained when adjusting for CBCL attention problems

Childhood Behavior Checklist (CBCL) data were missing for 13 participants (5 HC, 5 CD/-, 3 CD/+), resulting in an overall sample size of 247 participants (HC=141, CD-all=108; CD/-=60, CD/+=46).

Table S5. Significant group differences in cortical thickness, surface area, volume and gyrification when controlling for CBCL attention problems[a]

Comparison	Cluster	Anatomical region	Measure	H	NVtxs	Size (mm^2)	Peak MNI coordinates			CWP	Max	Cohen's *d*	SD (*d*)
							x	y	z				
All CD versus HC													
All CD < HC	AP1	**Pars triangularis,** pars orbitalis	CT	R	1340	1120.55	50.7	28.1	-2.1	.026	-2.87	-0.45	0.06
CD/- versus HC													
CD/- > HC	AP2	**Superior temporal gyrus**, transverse temporal gyrus	IGI	L	4963	2690.92	-66.5	-16.9	0.2	.016	2.82	0.52	0.07
CD/+ versus HC													
CD/+ < HC	AP3*	**Middle temporal gyrus**	CT	L	1370	1200.03	-65.0	-38.8	-13.5	.015	-3.83	-0.66	0.11
CD/+ < HC	AP4	**Rostral middle frontal gyrus**	CV	R	1743	1441.01	44.8	26.0	34.9	.006	-3.21	-0.61	0.10
CD/+ < HC	AP5	**Postcentral gyrus**	CV	R	2055	1159.23	66.7	-10.5	12.3	.039	-4.90	-0.70	0.17
CD/+ versus CD/-													
CD/+ < CD/-	AP6	**Superior frontal gyrus**	SA	R	2852	1998.31	3.8	30.6	51.5	.012	-2.48	-0.47	0.05
CD/+ < CD/-	AP7	**Supramarginal gyrus**, pre- and postcentral gyrus, banks of superior temporal sulcus	IGI	R	10756	5330.16	61.2	-25.0	42.0	<.001	-2.60	-0.48	0.05
CD/+ < CD/-	AP8	**Inferior temporal gyrus**, middle and superior temporal gyrus	IGI	L	3219	3372.38	-51.5	-11.7	-41.5	.001	-4.79	-0.56	0.12

[a] All analyses controlled for sex, age, site, and total intracranial volume (except thickness). Monte Carlo corrections for multiple comparisons were applied. Cohen's *d* was calculated using whole-brain vertex-wise effect size brain maps. Bolded regions represent the location of the peak coordinate. CBCL=Child Behavior Checklist; H=hemisphere;

NVtxs=number of vertices; MNI=Montreal Neurological Institute; CWP=cluster-wise *p*-value; Max=maximum -log10(*p*-value) in the cluster; CD=Conduct Disorder; HC=healthy controls; CD/-=Conduct Disorder without maltreatment; CD/+=Conduct Disorder with maltreatment; CT=cortical thickness, SA=surface area; CV=cortical volume; lGI=local gyrification index. *indicates clusters that were not identified in the main analyses not adjusting for attention problems.

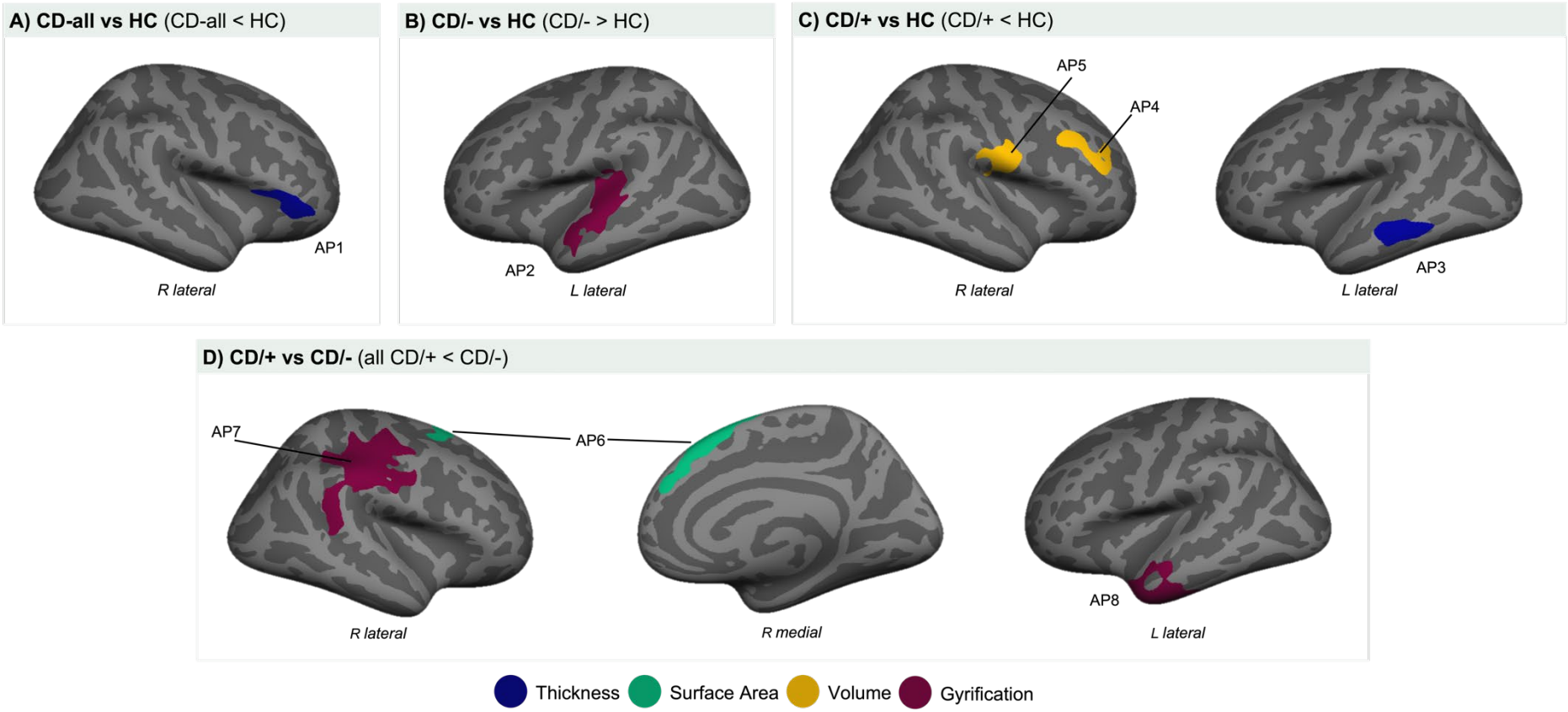

FIGURE S3. Group differences in cortical thickness, surface area, volume and gyrification when controlling for sex, age, site, total intracranial volume, and CBCL attention problems[a]

[a] Total intracranial volume was not covaried in analyses of cortical thickness. CD=Conduct Disorder; HC=Healthy Controls. CD/-=Conduct Disorder without maltreatment history; CD/+=Conduct Disorder with maltreatment history.

Surface-Based Morphometry results obtained when adjusting for socioeconomic status (SES)

Information on SES was missing for 22 participants (9 HC, 8 CD/-, 5 CD/+), resulting in an overall sample size of 238 participants (HC=137, CD-all=101; CD/-=57, CD/+=44) (please see supplemental methods for more information).

Table S6. Significant group differences in cortical thickness, surface area, volume and gyrification when controlling for SES[a]

Comparison	Cluster	Anatomical region	Measure	H	NVtxs	Size (mm^2)	Peak MNI coordinates			CWP	Max	Cohen's *d*	SD (*d*)
							x	y	z				
CD-all versus HC													
CD-all < HC	SES1	**Rostral middle frontal gyrus**, pars orbitalis, pars triangularis	CT	R	2929	2781.70	32.4	53.5	3.3	<.001	-3.59	-0.36	0.05
	SES2*	**Frontal pole**, rostral middle frontal gyrus, lateral orbitofrontal cortex	CT	L	1149	1165.60	-3.7	62.0	-19.5	.019	-2.57	-0.33	0.04
	SES3*	**Superior frontal gyrus**	SA	R	1966	1669.72	3.9	40.8	50.1	.037	-3.46	-0.36	0.06
	SES4*	**Postcentral gyrus**, precentral gyrus	CV	R	4000	1969.29	65.2	-12.3	12.2	<.001	-3.90	-0.37	0.06
CD/- versus HC													
CD/- > HC	SES5	**Superior temporal gyrus**, transverse temporal and supramarginal gyrus	IGI	L	7463	4285.37	-65.9	-17.3	-2.1	<.001	3.58	0.44	0.07
CD/+ versus HC													
CD/+ < HC	SES6	**Pars orbitalis**, rostral middle frontal gyrus	CT	R	1355	1292.30	43.7	53.2	-6.0	.008	-2.47	-0.44	0.05
	SES7	**Frontal pole**, rostral middle frontal gyrus, lateral orbitofrontal cortex	CT	L	1403	1498.32	-3.5	63.2	-17.1	.003	-3.17	-0.45	0.06
	SES8*	**Superior frontal gyrus**	SA	R	2313	1913.72	4.0	36.9	51.5	.017	-3.62	-0.46	0.08
	SES9	**Postcentral gyrus**, precentral gyrus	CV	R	3569	2102.15	63.3	-13.6	14.2	<.001	-4.21	-0.46	0.07

	SES10*	**Rostral middle frontal gyrus**, pars orbitalis, lateral orbitofrontal cortex	CV	R	1403	1578.69	41.2	56.5	-6.3	.004	-2.68	-0.43	0.05
	SES11	**Caudal middle frontal gyrus**, rostral middle frontal gyrus	CV	L	1627	1426.23	-44.2	11.7	46.6	.009	-2.86	-0.43	0.05
	SES12*	**Superior parietal lobule**, postcentral and supramarginal gyrus, inferior parietal lobule	IGI	R	4932	2257.54	36.8	-44.8	52.2	.038	-1.94	-0.40	0.03
CD/+ > HC	SES13*	**Pars opercularis**, pars triangularis	SA	R	2703	2103.63	54.8	13.6	2.0	.009	3.20	0.44	0.06
CD/+ versus CD/-													
CD/+ < CD/-	SES14	**Superior frontal gyrus**, precentral gyrus	SA	R	5606	3507.85	8.3	22.6	63.0	<.001	-2.42	-0.51	0.06
	SES15	**Superior temporal gyrus**	CV	R	1550	1195.60	46.0	-13.5	-6.5	.033	-3.01	-0.52	0.07
	SES16*	**Superior frontal gyrus**	CV	R	1592	1193.47	21.3	17.6	60.7	.033	-4.14	-0.53	0.11
	SES17	**Supramarginal gyrus**, pre- and postcentral gyrus, superior and inferior parietal lobule, middle temporal gyrus	IGI	R	23011	12388.39	62.6	-29.8	38.7	<.001	-3.54	-0.55	0.08
	SES18	**Rostral middle frontal gyrus**	IGI	R	4704	3098.92	50.6	29.6	25.5	.004	-3.99	-0.58	0.11
	SES19	**Fusiform gyrus**, lateral occipital pole, inferior and superior parietal lobule, supramarginal and postcentral gyrus	IGI	L	18499	10785.30	-38.2	-72.9	-19.5	.001	-3.69	-0.51	0.07
	SES20	**Middle temporal gyrus**, superior and inferior temporal gyrus	IGI	L	4157	3287.50	-48.1	6.0	-37.7	.001	-3.10	-0.50	0.06

[a] All analyses controlled for sex, age, site, and total intracranial volume (except thickness). Monte Carlo corrections for multiple comparisons were applied. Cohen's *d* was calculated using whole-brain vertex-wise effect size brain maps. Bolded regions represent the location of the peak coordinate. SES=socioeconomic status; H=hemisphere; NVtxs=number of vertices; MNI=Montreal Neurological Institute; CWP=cluster-wise *p*-value; Max=maximum -log10(*p*-value) in the cluster; CD=Conduct Disorder; HC=healthy controls; CD/-=Conduct Disorder without maltreatment; CD/+=Conduct Disorder with maltreatment; CT=cortical thickness, SA=surface area; CV=cortical volume; IGI=local gyrification index. *indicates clusters that were not identified in the main analyses not adjusting for SES.

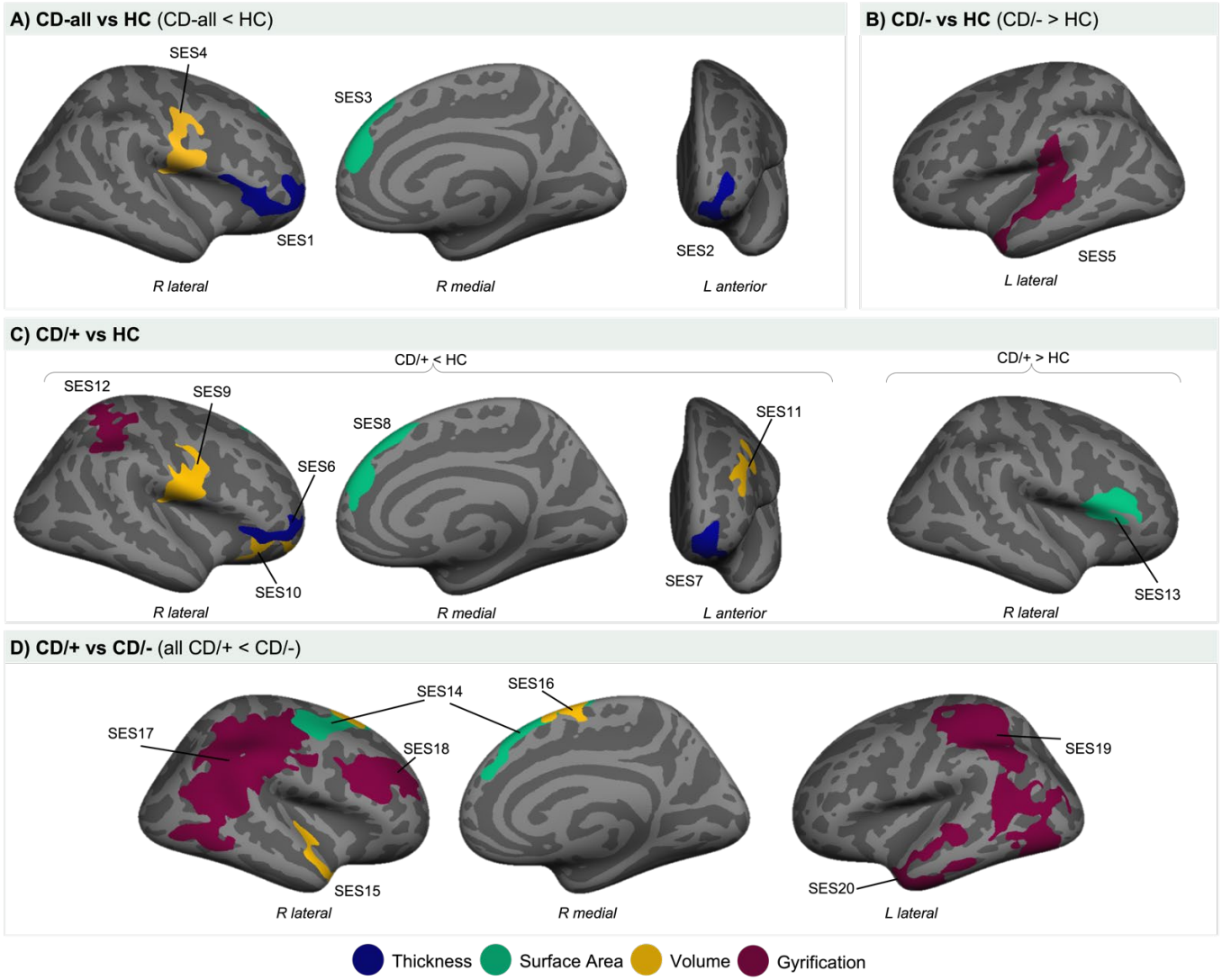

Figure S4. Group differences in cortical thickness, surface area, volume and gyrification when controlling for sex, age, site, total intracranial volume, and socioeconomic status[a]

[a] Total intracranial volume was not covaried in the cortical thickness analyses. CD=Conduct Disorder; HC=Healthy Controls. CD/-=Conduct Disorder without maltreatment history; CD/+=Conduct Disorder with maltreatment history.

Surface-Based Morphometry results obtained when focusing on a male-only subsample

The following analyses were performed in the male participants only (HC=72, CD-all=77; CD/-=48, CD/+=29). IQ, attention problems, and SES adjusted results are available on request.

Table S7. Significant group differences in surface area, volume and gyrification in male participants only[a]

Comparison	Cluster	Anatomical region	Measure	H	NVtxs	Size (mm^2)	Peak MNI coordinates			CWP	Max	Cohen's *d*	SD (*d*)
							x	y	z				
CD-all versus HC													
CD-all < HC	M1*	**Superior frontal gyrus**	SA	R	2056	1802.14	5.1	48.4	27.4	.022	-3.38	-0.46	0.08
CD/- versus HC	*None*												
CD/+ versus HC													
CD/+ < HC	M2*	**Superior frontal gyrus**	CV	R	1609	1161.38	5.3	39.7	27.2	.038	-3.63	-0.63	0.11
	M3*	**Inferior temporal gyrus**, middle temporal gyrus, banks of superior temporal sulcus, lateral occipital pole	IGI	R	5418	3581.11	52.3	-59.8	-2.4	.001	-3.18	-0.61	0.09
	M4*	**Fusiform gyrus**, lateral occipital pole	IGI	L	2315	2315.33	-29.9	-74.3	-16.8	.033	-2.83	-0.57	0.07
CD/+ versus CD/-													
CD/+ < CD/-	M5*	**Caudal middle frontal gyrus**, rostral middle frontal & precentral gyrus	CV	R	3108	1964.29	38.3	26.4	48.4	<.001	-3.73	-0.58	0.09
	M6*	**Inferior temporal gyrus**, middle temporal gyrus	CV	R	1285	1179.33	48.0	-48.0	-17.2	.035	-3.18	-0.60	0.09
	M7	**Middle temporal gyrus**, inferior temporal gyrus, inferior parietal lobule, lateral occipital pole	IGI	R	5299	3015.11	62.2	-59.1	5.4	.005	-2.88	-0.58	0.07
	M8	**Fusiform gyrus**, lateral occipital pole	IGI	L	3291	2648.14	-30.1	-77.4	-13.5	.012	-3.85	-0.67	0.14
CD/+ > CD/-	M9*	**Precuneus**, paracentral lobule, posterior & isthmus cingulate cortex	IGI	R	5761	2574.65	8.1	-43.1	37.0	.016	2.36	0.58	0.06

[a] All analyses controlled for age, site, and total intracranial volume (except thickness). Monte Carlo corrections for multiple comparisons were applied. Cohen's *d* was calculated using whole-brain vertex-wise effect size brain maps. Bolded regions represent the location of the peak coordinate. H=hemisphere; NVtxs=number of vertices; MNI=Montreal Neurological Institute; CWP=cluster-wise *p*-value; Max=maximum -log10(*p*-value) in the cluster; CD=Conduct Disorder; HC=healthy controls; CD/-=Conduct Disorder without maltreatment; CD/+=Conduct Disorder with maltreatment; SA=surface area; CV=cortical volume; lGI=local gyrification index. *indicates clusters that were not identified in the main analyses with the whole sample. There were no group differences in cortical thickness.

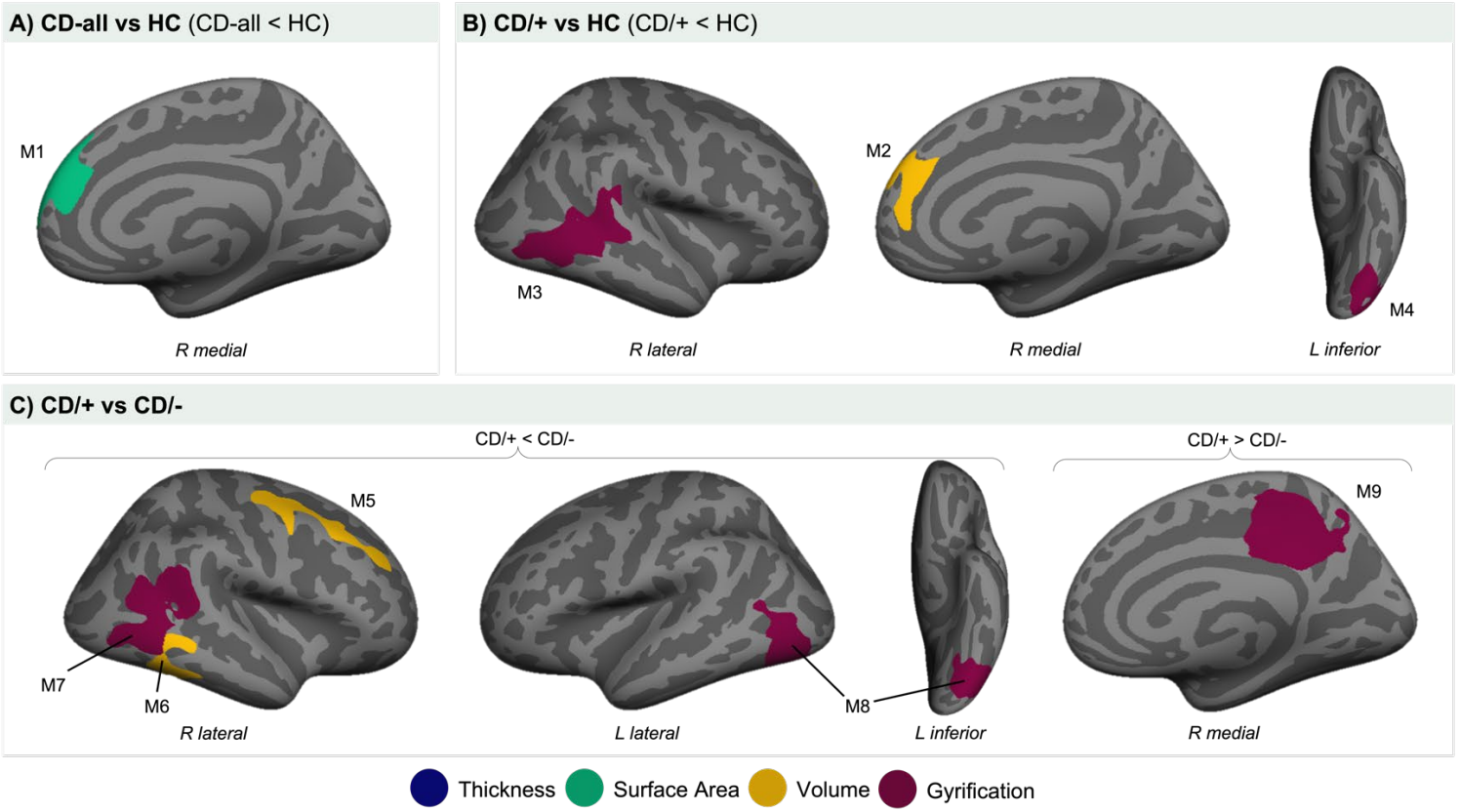

Figure S5. Group differences in surface area, volume and gyrification when controlling for age, site, and total intracranial volume in male participants only

CD=Conduct Disorder; HC=Healthy Controls. CD/-=Conduct Disorder without maltreatment history; CD/+=Conduct Disorder with maltreatment history. There were no group differences in cortical thickness.

Statistical Power Considerations

Recently published evidence by Marek and colleagues (43) highlighted that sample sizes in the thousands are necessary to robustly identify brain-behavior associations, particularly if those are small in (effect) size. This suggests that the current analyses, albeit based on a sample larger than most CD studies, might have still been underpowered. With regard to SBM studies, Liem et al. (44) found that to achieve >80% power to detect between-group differences of 10% in vertex-wise whole-brain analyses, 39 participants per group are required for thickness analyses, 21 for surface area and 81 for cortical volume (assuming α=0.05 and 10mm smoothing). Similarly, Pardoe et al. (45) calculated that to detect a 0.25mm thickness difference between two groups, approximately 50 subjects per group are required (assuming α=0.05 and 10mm smoothing).

We note that our group differences were smaller than those focused on in the above papers. For example, the differences in right rostral middle frontal gyrus gyrification between the maltreated (CD/+) and non-maltreated CD groups (CD/-) had an effect size of d=-0.50, reflecting a difference of 4% (CD/+ < CD/-). Using the calculations provided by Pardoe et al. (45), we estimated that 386 participants per group would be required to achieve 80% power to detect an effect of this size (though we note that these calculations pertain to thickness and their applicability to gyrification may be limited). This highlights that replications of the current findings in larger samples are needed, including for example, in the dataset compiled by the ENIGMA-Antisocial Behavior working group.

Subcortical analyses: Statistical analysis strategy and results

Hypotheses. Based on evidence of reduced amygdala and striatal volumes in youths with CD or Disruptive Behavior Disorders (including CD and ODD) (24–26), we expected to observe main effects of group in these regions. Although not consistently identified in youth samples, studies have reported maltreatment-related gray matter volume reductions in the amygdala (27–29). Hence, we tentatively hypothesized that lower amygdala volumes in the CD group might be driven by maltreatment, reflected in alterations being limited to (or more pronounced) in the CD group with maltreatment and/or differences between the CD subgroups in this region. Additionally, previous studies have supported maltreatment-related volumetric decreases in the hippocampus [e.g., (29–31)] albeit more consistently in adult as opposed to youth samples (28). Hence, we tentatively hypothesized lower hippocampal volume in the CD group with maltreatment history. Gao and colleagues' (32) study was published after these predictions were formulated but based on their findings further differences between the CD subgroups in the putamen can be hypothesized (CD/+ > CD/-).

Extraction of subcortical volumes. Subcortical volumes were estimated using FreeSurfer's automatic segmentation pipeline for subcortical regions (33). In short, each voxel in the normalized brain is assigned to one of 40 labels including cerebral white matter, cerebral cortex, ventricles, and subcortical structures, based on FreeSurfer's Aseg Atlas. Subsequently, volume and intensity statistics are extracted for each segmentation. For the current analyses, we used the estimated volumes of the amygdala, hippocampus, caudate, pallidum, putamen, thalamus, and nucleus accumbens per hemisphere.

Statistical analyses. Corresponding to the analytic approach in the vertex-wise analyses of cortical structure, we tested for group effects on these volumes using General Linear Models (GLMs) adjusted for sex, age, site (dummy coded), and total intracranial volume (TIV; orthogonalized to sex). GLMs were fitted separately per subcortical volume and hemisphere. First, the overall CD group (CD-all) was compared with the healthy controls (HCs), before each CD subgroup (CD/- and CD/+) was compared to the control group and then to each other. Analyses were performed in R (v4.0.3) (34). We applied a False-Discovery-Rate multiple comparison correction at q=0.05 across all included regions of both hemispheres (i.e., across 14 regions). Sensitivity analyses additionally included IQ, CBCL attention problems (missing for 13 participants: 5 HC, 5 CD/-, 3 CD/+) or SES (missing for 22 participants: 9 HC, 8 CD/-, 5 CD/+). For each analysis, outliers defined as subjects with a standardized residual >|3.29| (35) and/or a Cook's distance >1 (36) were identified, and heteroscedasticity was assessed visually and statistically. If outliers were identified, analyses were repeated without them to gauge their influence (37). Similarly, analyses were repeated using heteroscedasticity-consistent standard errors (type "HC3") when heteroscedasticity was detected (38).

Results. No group differences in subcortical volumes survived False-Discovery-Rate correction, regardless of IQ, attention problems or SES adjustment. Group differences obtained in the main analyses when controlling for sex, age, site and TIV, but not correcting for multiple comparisons, are presented in Table S8. The uncorrected findings indicated higher left accumbens volumes in the overall CD group and both CD subgroups relative to controls (with small to medium effect sizes, see Table S8). Relative to the HCs, the CD-all group further showed greater volume in the bilateral pallidum at the nominal significance level. A similar trend could be

observed for both CD subgroups, but these effects did not reach nominal significance (0.05<*p*s<0.1). For right pallidum volume, when adjusting for IQ or attention problems, the differences between the CD subgroups and the control group reached nominal significance, respectively. For left pallidum volume, when correcting for IQ, the CD/+ group also differed significantly from controls. None of the subcortical group differences survived adjustment for SES. Regardless of adjustment for IQ, attention problems or SES, there were no differences between the CD subgroups in any subcortical region. Excluding outliers or using heteroscedasticity-consistent standard errors did not alter the results to a notable extent in most cases (see Table S8).

Table S8. Significant group differences in subcortical volumes (not corrected for multiple comparisons)[a]

Subcortical Region	Comparison	*t*	*df*	*p*	*d*	95% CI		Sensitivity		
						lower	upper	IQ	AP	SES
Left Nucleus Accumbens	CD-all > HC	-2.55	251	0.011	-0.32	-0.57	-0.08	yes	yes	no
	CD/- > HC*	-2.05	202	0.041	-0.31	-0.59	-0.04	no	yes	no
	CD/+ > HC	-2.20	186	0.029	-0.37	-0.66	-0.09	yes	yes	no
Right Pallidum	CD-all > HC	-2.21	251	0.042	-0.28	-0.53	-0.04	yes	yes	no
Left Pallidum	CD-all > HC*	-2.04	251	0.028	-0.26	-0.5	-0.01	yes	yes	no

[a]All analyses controlled for sex, age, site, and total intracranial volume. The *p*-values presented here are not corrected for multiple comparisons. None of the displayed group differences survived False-Discovery-Rate correction for multiple comparisons. *d*=Cohen's d; CI=confidence interval; IQ=intelligence quotient; AP=attention problems (Child Behavior Checklist); SES=socioeconomic status; CD=Conduct Disorder; HC=healthy controls; CD/-=Conduct Disorder without maltreatment; CD/+=Conduct Disorder with maltreatment.
*Differences in this region were no longer significant when removing one outlier and/or using heteroscedasticity-consistent standard errors.

Brief discussion. In contrast to our hypotheses, we found no robust evidence of subcortical alterations in the CD group or maltreatment-related effects. The lack of significant effects of group on subcortical volumes when applying a False-Discovery-Rate multiple comparison correction is in line with previous findings based on an overlapping sample (1). Considering the uncorrected findings, the current results

tentatively indicate larger volumes in the left accumbens and bilateral pallidum in CD (with small to medium effect sizes in each case). Both structures are part of the striatum which is implicated in a variety of functions including reinforcement-learning and altering one's behavior in the face of changing situational demands (39). Greater striatal volume has been reported in adults with psychopathy (40) and was found to be positively correlated with self-reported callous-unemotional traits in boys with CD (41). Relatedly, in meta-regression analyses, callous-unemotional traits were associated with a *lower* reduction in gray matter volume in the left putamen in youths with conduct problems (25). Hence, it seems possible that the observed moderate increases in accumbens and pallidum volumes might be driven by psychopathic and callous-unemotional traits in the CD (sub)groups. However, additional exploratory regression analyses did not support this assumption. Adjusting for age, sex, site and TIV, there were no significant associations between measures of callous-unemotional and psychopathic traits (YPI total score, YPI callous-unemotional traits subscale score, ICU total score) and pallidum and accumbens volumes in the whole CD group or the CD subgroups, respectively (all *p*s > 0.05).

Returning to the group comparisons, there was no evidence of specific effects of maltreatment, that is, the CD subgroups did not differ from each other, and where there were indications of volumetric increases relative to the control group, both CD subgroups showed similar alterations in the same direction (e.g., higher accumbens volume). This suggests that having a history of maltreatment does not influence alterations in subcortical volumes in CD. This contrasts with the findings of Gao et al. (32) who reported greater putamen volume in CD youth with a history of maltreatment compared to those without.

Alternatively, the lack of (maltreatment-related) subcortical alterations may also be explained by the young age of the sample and the investigated modality (i.e., brain *structure*). For example, corresponding to their susceptibility to stress, both the amygdala and hippocampus have previously been implicated in maltreatment (28). However, neither of these structures differed between the CD/+ group and the controls or between the CD subgroups. Critically, previous studies suggest that maltreatment-related alterations in the hippocampus are more consistently observed in adult as opposed to youth samples (28). This may indicate a 'silent' period between the experience of maltreatment and observable effects on the hippocampus, potentially explaining why maltreatment-related hippocampus alterations were not observed in the current sample, which had a mean age of ~14 years. With regard to the amygdala, previous findings have been mixed with many studies failing to find maltreatment-related volumetric differences in this region (28). Reviews have indicated that functional (as opposed to structural) alterations are more consistently identified in maltreated samples and that factors such as timing of maltreatment, the specific type of maltreatment, and presence of concurrent psychopathology may explain between-study variability with regards to structural amygdala alterations (28,42). This might also apply to the current study where we were unable to systematically explore the contributions of these factors due to the sample size and limitations of the chosen maltreatment measure (Children's Bad Experiences interview). The role of these factors could be explored in future studies with larger samples and more detailed (and time-sensitive) measures of maltreatment. Lastly, we note that FreeSurfer's subcortical stream only provides an aggregate index of volume, which – unlike measures derived using voxel-based morphometry (VBM) methods – is not specific to gray matter volume. This might

constitute a further reason why group differences that have previously been identified in VBM studies (e.g., lower amygdala gray matter volume) (25) were not detected in the current study and may explain the differences between our findings and those reported by Gao et al. (32).

Overall, the current data provide no robust evidence of alterations in subcortical volumes in youths with CD – or differences between CD subgroups with versus without a history of maltreatment. The uncorrected findings indicated greater volume in regions of the striatum in youths with CD, irrespective of maltreatment history. Hence, these data do not support the view that maltreatment contributes to alterations in subcortical brain structure in CD.

REFERENCES

1. Smaragdi A, Cornwell H, Toschi N, Riccelli R, Gonzalez-Madruga K, Wells A, *et al.* (2017): Sex differences in the relationship between conduct disorder and cortical structure in adolescents. *J Am Acad Child Adolesc Psychiatry* 56: 703–712.
2. Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, *et al.* (1997): Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry* 36: 980–988.
3. American Psychiatric Association (2000): *Diagnostic and Statistical Manual of Mental Disorders,* 4th ed., text revision. Washington, DC: American Psychiatric Publishing.
4. American Psychiatric Association (2013): *Diagnostic and Statistical Manual of Mental Disorders,* 5th ed. Washington, DC: American Psychiatric Publishing.

5. Dodge KA, Bates JE, Pettit GS (1990): Mechanisms in the cycle of violence. *Science* 250: 1678–1683.
6. Lansford JE, Criss MM, Laird RD, Shaw DS, Pettit GS, Bates JE, Dodge KA (2011): Reciprocal relations between parents' physical discipline and children's externalizing behavior during middle childhood and adolescence. *Dev Psychopathol* 23: 225–238.
7. Choi KW, Houts R, Arseneault L, Pariante C, Sikkema KJ, Moffitt TE (2019): Maternal depression in the intergenerational transmission of childhood maltreatment and its sequelae: Testing postpartum effects in a longitudinal birth cohort. *Dev Psychopathol* 31: 143–156.
8. Arseneault L, Cannon M, Fisher HL, Polanczyk G, Moffitt TE, Caspi A (2011): Childhood trauma and children's emerging psychotic symptoms: a genetically sensitive longitudinal cohort study. *AJP* 168: 65–72.
9. Jaffee SR, Bowes L, Ouellet-Morin I, Fisher HL, Moffitt TE, Merrick MT, Arseneault L (2013): Safe, stable, nurturing relationships break the intergenerational cycle of abuse: a prospective nationally representative cohort of children in the United Kingdom. *J Adolesc Health* 53: S4–S10.
10. Dodge KA, Pettit GS, Bates JE, Valente E (1995): Social information-processing patterns partially mediate the effect of early physical abuse on later conduct problems. *J Abnorm Psychol* 104: 632–643.
11. Wechsler D (1999): *Wechsler Abbreviated Scale of Intelligence.* Psychological Corporation.
12. Wechsler D (2003): *Wechsler Intelligence Scale for Children*. Psychological Corporation.

13. Wechsler D (1997): *Wechsler Adult Intelligence Scale*. Psychological Corporation.
14. UNESCO Institute for Statistics (2015). *International standard classification of education field of education and training 2013 (ISCED-F 2013): Detailed field descriptions.* Montreal, QC: UNESCO Institute for Statistics.
15. International Labour Organization (2012). *International standard classification of occupations, structure, group definitions and correspondence tables*. Geneva: International Labour Organization.
16. Andershed HA, Kerr M, Stattin H, Levander S (2002): Psychopathic traits in non-referred youths: a new assessment tool. Elsevier, pp 131–158. In Blaauw E, Sheridan L, editors. *Psychopaths: New international perspectives*. The Hague, Netherlands: Elsevier, pp 131–58.
17. Essau CA, Sasagawa S, Frick PJ (2006): Callous-unemotional traits in a community sample of adolescents. *Assessment* 13: 454–469.
18. Eekhout I, de Vet HCW, Twisk JWR, Brand JPL, de Boer MR, Heymans MW (2014): Missing data in a multi-item instrument were best handled by multiple imputation at the item score level. *J Clin Epidemiol* 67: 335–342.
19. Liu Y, De A (2015): Multiple imputation by fully conditional specification for dealing with missing data in a large epidemiologic study. *Int J Stat Med Res* 4: 287–295.
20. Essau CA, Sasagawa S, Frick PJ (2006): Psychometric properties of the alabama parenting questionnaire. *J Child Fam Stud* 15: 595–614.
21. Grisso T, Barnum R, Fletcher KE, Cauffman E, Peuschold D (2001): massachusetts youth screening instrument for mental health needs of juvenile justice youths. *J Am Acad Child Adolesc Psychiatry* 40: 541–548.

22. Chen C-C, Wan Y-L, Wai Y-Y, Liu H-L (2004): quality assurance of clinical MRI scanners using ACR MRI phantom: preliminary results. *J Digit Imaging* 17: 279–284.
23. Glover GH (2005): *FBIRN Stability phantom QA procedures*. Stanford University and FBIRN.
24. Marek S, Tervo-Clemmens B, Calabro FJ, Montez DF, Kay BP, Hatoum AS, *et al.* (2022): Reproducible brain-wide association studies require thousands of individuals. *Nature* 603: 654–660.
25. Liem F, Mérillat S, Bezzola L, Hirsiger S, Philipp M, Madhyastha T, Jäncke L (2015): Reliability and statistical power analysis of cortical and subcortical FreeSurfer metrics in a large sample of healthy elderly. *Neuroimage* 108: 95–109.
26. Pardoe HR, Abbott DF, Jackson GD (2012): Sample size estimates for well-powered cross-sectional cortical thickness studies. *Hum Brain Mapp* 34: 3000–3009.
27. Noordermeer SDS, Luman M, Oosterlaan J (2016): A systematic review and meta-analysis of neuroimaging in oppositional defiant disorder (ODD) and conduct disorder (CD) taking attention-deficit hyperactivity disorder (ADHD) into account. *Neuropsychol Rev* 26: 44–72.
28. Rogers JC, De Brito SA (2016): Cortical and subcortical gray matter volume in youths with conduct problems: a meta-analysis. *JAMA Psychiatry* 73: 64–72.
29. Waller R, Hawes SW, Byrd AL, Dick AS, Sutherland MT, Riedel MC, *et al.* (2020): Disruptive behavior problems, callous-unemotional traits, and regional gray matter volume in the adolescent brain and cognitive development study. *Biol Psychiatry Cogn Neurosci Neuroimaging* 5: 481–489.

30. Lim L, Radua J, Rubia K (2014): Gray matter abnormalities in childhood maltreatment: a voxel-wise meta-analysis. *AJP* 171: 854–863.
31. Teicher MH, Samson JA (2016): Annual Research Review: Enduring neurobiological effects of childhood abuse and neglect. *J Child Psychol Psychiatry* 57: 241–266.
32. Pollok TM, Kaiser A, Kraaijenvanger EJ, Monninger M, Brandeis D, Banaschewski T, *et al.* (2022): Neurostructural traces of early life adversities: A meta-analysis exploring age- and adversity-specific effects. *Neurosci Biobehav Rev* 135: 104589.
33. Dannlowski U, Stuhrmann A, Beutelmann V, Zwanzger P, Lenzen T, Grotegerd D, *et al.* (2012): Limbic scars: long-term consequences of childhood maltreatment revealed by functional and structural magnetic resonance imaging. *Biol Psychiatry* 71: 286–293.
34. Dahmen B, Puetz VB, Scharke W, Polier GG von, Herpertz-Dahlmann B, Konrad K (2018): Effects of early-life adversity on hippocampal structures and associated HPA axis functions. *Dev Neurosci* 40: 13–22.
35. Gao Y, Jiang Y, Ming Q, Zhang J, Ma R, Wu Q, *et al.* (2021): Neuroanatomical changes associated with conduct disorder in boys: influence of childhood maltreatment. *Eur Child Adolesc Psychiatry:* 1-13
36. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, *et al.* (2002): Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33: 341–355.
37. R Core Team (2020): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

38. Field AP, Miles J, Field Z (2012): *Discovering statistics using R*. London, England: Sage.
39. Cook, RD, Weisberg S (1982): *Residuals and influence in Regression*. New York, US: Chapman and Hall.
40. Aguinis H, Gottfredson RK, Joo H (2013): Best-practice recommendations for defining, identifying, and handling outliers. *Organ Res Methods* 16: 270–301.
41. Hayes AF, Cai L (2007): Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behav Res Methods* 39: 709–722.
42. Glenn AL, Yang Y (2012): The potential role of the striatum in antisocial behavior and psychopathy. *Biol Psychiatry* 72: 817–822.
43. Glenn AL, Raine A, Yaralian PS, Yang Y (2010): Increased volume of the striatum in psychopathic individuals. *Biol Psychiatry* 67: 52–58.
44. Fairchild G, Passamonti L, Hurford G, Hagan CC, von dem Hagen EAH, van Goozen SHM, *et al.* (2011): Brain structure abnormalities in early-onset and adolescent-onset conduct disorder. *AJP* 168: 624–633.
45. McLaughlin KA, Weissman D, Bitrán D (2019): Childhood adversity and neural development: a systematic review. *Annu Rev Dev Psychol* 1: 277–312.