**SUPPLEMENTARY INFORMATION**

**Testing the Ecophenotype Hypothesis: Differences in White Matter Microstructure in Youth with Conduct Disorder With Versus Without a History of Childhood Abuse**

By Townend *et al.*

**Contents**

[SUPPLEMENTARY METHODS 2](#_Toc187061626)

[Figure S1. Flowchart of the participant inclusion processa 2](#_Toc187061627)

[Inclusion and Exclusion Criteria 2](#_Toc187061628)

[TABLE S1. Number of participants per contributing sitea 3](#_Toc187061629)

[Ethical Approval 4](#_Toc187061630)

[Additional Information on Phenotypic Measures 4](#_Toc187061631)

[Diagnosis (K-SADS-PL): 4](#_Toc187061632)

Maltreatment ([Children’s Bad Experiences (CBE) interview): 5](#_Toc187061633)

[Psychopathic and callous-unemotional traits 7](#_Toc187061634)

[Attention Problems from the Child Behavior Checklist (CBCL) 8](#_Toc187061635)

[IQ measures 9](#_Toc187061636)

[Socioeconomic status (SES) 9](#_Toc187061637)

[Site Qualification Procedure 9](#_Toc187061638)

[Scanning and Diffusion-Weighted Imaging Parameters 10](#_Toc187061639)

[TABLE S2. Scanning parameters and acquisition sequences at each sitea 10](#_Toc187061640)

[Quality Assessment 11](#_Toc187061641)

[Image Processing 11](#_Toc187061642)

[Diffusion Tensor Imaging Preparation 12](#_Toc187061643)

[Statistical Analysis 12](#_Toc187061644)

[Accounting for Between-Site Heteroscedasticity 13](#_Toc187061645)

[Analytical justification 13](#_Toc187061646)

[SUPPLEMENTARY RESULTS 15](#_Toc187061647)

[TABLE S3. Demographic and clinical characteristics of the sample including post hoc comparisonsa 15](#_Toc187061648)

[TABLE S4. Significant group differences from sex-stratified analyses conducted with an all-male subsamplea 17](#_Toc187061649)

[TABLE S5. Significant group differences from sex-stratified analyses conducted with an all-female subsamplea 18](#_Toc187061650)

[TABLE S6. Significant group differences when controlling for IQa 19](#_Toc187061651)

[TABLE S7. Significant group differences when controlling for CBCL attention problemsa 20](#_Toc187061652)

[TABLE S8. Significant group differences when controlling for socioeconomic statusa 21](#_Toc187061653)

[TABLE S9. Overview of results across sensitivity analysesa 23](#_Toc187061654)

[SUPPLEMENTARY REFERENCES 24](#_Toc187061655)

# SUPPLEMENTARY METHODS

## Figure S1. Flowchart of the participant inclusion processa

Full FemNAT-CD sample

*N* = 1743

DTI data available

*n* = 376

Excluded participants: *n* = 25

* DTI data incomplete, *n* = 3
* DTI data failed quality control checks, *n* = 22

Usable DTI data available

*n* = 351

Excluded participants: *n* = 82

* No CBE data available, *n* = 68
* Healthy controls with histories of abuse, *n* = 14

Sample included in our analyses

*n* = 269

aDTI=Diffusion Tensor Imaging; CBE=Children’s Bad Experiences interview.

## Inclusion and Exclusion Criteria

Inclusion criteria of the FemNAT-CD study included being aged 9-18 years old and having normal or corrected-to-normal vision. Cases had a DSM-IV-TR diagnosis of conduct disorder (CD), or a diagnosis of Oppositional Defiant Disorder (ODD) and either 1 or 2 current symptoms of CD in younger children, depending on their age. This was the case for seven CD participants, one of whom had a history of childhood abuse. Healthy controls (HCs) were free of current Axis I disorders and had no history of externalizing disorders (i.e., CD, ODD, or Attention-Deficit/Hyperactivity Disorder [ADHD]). Exclusion criteria for both groups included an estimated IQ <70, and the presence of autism spectrum disorders, psychosis, neurological or genetic disorders or severe head trauma, as well as standard MRI exclusion criteria (e.g., metal dental braces that could not be removed). For female participants, additional exclusion criteria were being pregnant or having given birth in the last year.

Note that the current study included data that had previously been processed for a DTI study focusing on sex differences in CD by Rogers and colleagues (1). As this prior project did not examine the impact of child abuse, the availability of CBE data was not a requirement for inclusion in that study. The research team at the University of Bath (in collaboration with the authors of Rogers et al.) subsequently completed pre-processing and quality control on DTI data for an additional 78 participants. From the resulting sample of 376 participants, 25 were excluded due to low data quality, and a further 68 participants were excluded because CBE data were not available for them. As only 14 healthy control participants had a history of childhood abuse according to the CBE, this group was also excluded as the sample size was not considered sufficient for meaningful statistical analyses. This resulted in a final sample size of *n* = 269. 222 of these participants (82.5% of the present sample) were therefore included in both Rogers et al. (2019) and the current DTI study.

### TABLE S1. Number of participants per contributing sitea

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Group | Site 1  Aachen | Site 2 Southampton | Site 3  Basel | Site 4  Birmingham | **Total** |
| CD/- | 27 | 13 | 6 | 15 | **61** |
| CD/+ | 18 | 9 | 7 | 5 | **39** |
| HC | 52 | 39 | 21 | 57 | **169** |
| **Total** | **97** | **61** | **34** | **77** | **269** |

**a**CD/-=Conduct Disorder without childhood physical or sexual abuse; CD/+=Conduct Disorder with a history of childhood physical or sexual abuse; HC=healthy controls without childhood physical or sexual abuse. There was no significant association between site and group, *p* = .100, Fisher’s exact test, suggesting that the distribution of participants across the three groups did not differ between sites.

## Ethical Approval

The FemNAT-CD study was conducted in accordance with the legal regulations of the European Union and the Declaration of Helsinki. Study protocols were approved by the relevant ethical committees at each site prior to data collection. These were RWTH Aachen University Hospital (EK027/14) for the Aachen site (site 1), the Ethics Commission for Northwest and Central Switzerland (EKNZ: 336/13) for the Basel site (site 3), and the Southampton University Ethics Committee (ERGO Number: 18970) and the National Health Service Research Ethics Committee (NRES Committee West Midlands, Edgbaston; REC Reference 13/WM/0483) for the two UK sites (Birmingham [site 3] and Southampton [site 4]). Ethical approval for the current study was obtained from the University of Bath Psychology Research Ethics Committee (Ethics ref: 21-278).

## Additional Information on Phenotypic Measures

Diagnosis (K-SADS-PL): CD diagnoses were assessed with the Kiddie-Schedule for Affective Disorders and Schizophrenia-Present and Lifetime version (K-SADS-PL; (2)). Participants with CD were classified as having childhood-onset CD if at least one symptom and functional impairment were reported to have been present before the age of 10. If symptom onset occurred after age 10, the participant was classified as having adolescent-onset CD (3). The K-SADS-PL was also used to assess for the presence of comorbid disorders such as ODD, ADHD, Major Depressive Disorder (MDD), Generalized Anxiety Disorder (GAD), Post-Traumatic Stress Disorder (PTSD) and alcohol and substance abuse in line with DSM-IV-TR criteria (4). It is a semi-structured interview that was conducted with the young person and their parents/carers in separate rooms to ensure confidentiality of the information provided by each informant. Summary ratings on the presence or absence of a disorder are achieved by clinical judgement considering both sources of information (a symptom was typically considered present if it was endorsed by either informant). Inter-rater reliability (IRR) for diagnoses were assessed in a subsample of 75 participants from across all FemNAT-CD research sites. The IRR for CD diagnoses was high (Cohen’s kappa = 0.91, agreement rate = 94.7%), as well as for other comorbid disorders including ODD, ADHD, MDD and GAD (Cohen’s kappas ≥ 0.84, agreement rates ≥ 92% for all diagnostic categories).

Children’s Bad Experiences (CBE) interview: Childhood abuse/maltreatment was assessed with the CBE interview (5,6), conducted with the participant’s primary caregiver. For 94.5% of our sample, this was their biological parent (typically the mother), though other informants included adoptive, foster or step parents (2.9%), other relatives including siblings or grandparents (1.1%), or an institutional official (1.5%). The CBE is a structured interview which asks about adverse experiences the child may have been exposed to. The current study used data from items 3-5 specifically, concerning potential physical and sexual abuse (CPA/CSA) the child may have experienced, how often this may have occurred, the child’s age at the time, and whether the experience resulted in physical or psychological harm. The questions are designed to encourage honesty by remaining agnostic to the perpetrator (i.e., asking if the child has ever been harmed, but not who harmed them), thus providing a sensitive and acceptable method of assessing maltreatment. Nevertheless, informants were reminded that if serious or ongoing harm was reported that had not previously come to the attention of official agencies then the interviewer will have to pass this information on and make a decision about whether referral to such agencies is needed. The interview is scored based on the interviewer’s clinical judgement, with CPA/CSA categorised as either absent (no abuse reported), probable (abuse reported but unclear whether it definitely occurred) or definite (abuse definitely happened, for instance the police or child welfare services were involved or marks were seen on the child’s skin). Inter-rater agreement has been reported as high in previous studies (e.g., agreement rate = 90%) (5,6). As with previous studies using the CBE (7), we dichotomised this variable so that participants were classified as either having no history of childhood physical or sexual abuse or likely childhood physical or sexual abuse exposure, combining probable (n=18) and definite (n=21) categories.

In line with work by Staginnus and colleagues (8), we investigated inconsistent information in the CBE (e.g., inconsistencies between individual items and overall abuse rating). Inconsistent information was identified for 35 participants across the sample, and to ensure that correct ratings were used, these were investigated by three authors (S.T., G.F., and M.S.) using additional sources of information on childhood maltreatment or abuse. This process was completed blind to group status (i.e., CD versus controls). Note that information for these additional measures was not available for all participants (hence why the CBE was considered the main maltreatment measure). These measures included the Childhood Trauma Questionnaire (CTQ; (9)), a self-report measure for adolescents aged 12 years and older to assess five types of maltreatment: physical abuse, emotional abuse, sexual abuse, physical neglect and emotional neglect. Cut-off scores are applied to each maltreatment type to distinguish between four levels (none, low, moderate or severe), and individuals with a score of moderate or severe are considered to have been maltreated. The Childhood Experience of Care and Abuse Questionnaire (CECA-Q; (10)) is a questionnaire version of the Childhood Experience of Care and Abuse interview (11). It is a self-report measure, covering parental loss, neglect, antipathy (i.e., cold or hostile parenting), support, and physical and sexual abuse before age 17. Cut-off scores were validated to align with marked or moderate severity ratings from the CECA-Q, and only items relating to neglect, physical abuse and sexual abuse were considered in this study. Finally, we considered three items from the K-SADS-PL PTSD screen relating to whether the child had ever witnessed domestic violence, or experienced physical abuse or sexual abuse. This was completed by the parent (or caregiver) and the child. When considering inconsistent cases, the default assumption was always that the CBE rater was correct in their judgement, and ratings were only changed when convincing evidence from multiple other sources was available (e.g., both the parent and child reported experiences which may not have been picked up in the CBE measure, such as experiencing neglect or witnessing domestic abuse). Of the 35 participants’ ratings that were considered, 8 were changed from their original CBE score. One of these was changed from probable to no maltreatment/abuse, while seven were changed from no maltreatment/abuse to probable or definite maltreatment/abuse (three of whom were controls who were subsequently excluded).

Psychopathic and callous-unemotional traits: The total score of the self-report Youth Psychopathic traits Inventory (YPI; 12) was used to assess overall psychopathic traits (α=0.94), while the callous-unemotional subscale of the YPI (α=0.79), along with the parent-report Inventory of Callous-Unemotional traits (ICU; α=0.91, (13) were used to assess callous-unemotional traits. These measures were collected in order to characterize the sample and assess whether the abused and non-abused CD subgroups differed in callous-unemotional or psychopathic traits in the present study.

For the YPI, full data was available for 227 participants (84.4%), while full data for the CU subscale of the YPI was available for 247 participants (91.8%). Complete data for the ICU total score was available for 236 participants (87.7%). To handle missing data (some where data was completely missing, and some where individual items were missing), imputation was completed before the current study was planned, based on data from the whole FemNAT-CD sample. Imputation was handled at the item level (14), and scores were then calculated based on the imputed items. This was completed in SAS version 9.4 using the procedure PROC MI, and imputation by fully conditional specification (FCS) was used, as it offers a flexible method to specify the multivariate imputation model for arbitrary missing patterns including both categorical and continuous variables (15). For imputation diagnostics, distribution of the observed and imputed items and scores were checked.

Attention Problems from the Child Behavior Checklist (CBCL): The CBCL is a questionnaire that parents or caregivers complete to detect emotional and behavioural problems in children and adolescents (16). It includes 120 items, scored on a three-point Likert scale to indicate how often (if at all) their child has exhibited certain behaviours in the past six months. We used the attention problems subscale (using raw scores) as a proxy measure for ADHD symptoms in the sensitivity analyses. Data for the attention problems subscale of the CBCL was available for 237 participants (88.1%).

IQ measures: IQ was measured using the two subtest version of the Wechsler Abbreviated Scale of Intelligence (WASI-I) at the UK sites (17). At the German/Swiss sites, the Wechsler Intelligence Scale for Children (WISC-III-R/IV) was used for those aged 9-16 (18), and the Wechsler Adult Intelligence Scale (WAIS-III/IV) for those aged 17 years and above (19). IQ data was available for the entire sample.

Socioeconomic status (SES): SES scores were based on parental income, education and occupation, and assessed using the International Standard Classification of Occupations (20) and the International Classification of Education (21). To gain a measure of relative socioeconomic position accounting for potential economic variation on the country level, SES scores were centered and scaled within each country. Internal consistency of the composite SES score was acceptable (α=0.74). SES information was available for 239 participants (88.8%).

## Site Qualification Procedures

To ensure comparability of MRI data acquisition across sites, each site underwent site qualification procedures. 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 scanning stability during functional MRI sequences, providing information about scanner drift, fluctuation in signal, signal-to-noise ratio and signal-to-fluctuation noise. The data resulting from these checks underwent quality assessment by an MRI physicist at the University of Birmingham (Dr Ali Chowdhury), and scanning parameters at each site were adjusted according to the physicist’s recommendations to ensure scanning procedures were comparable. Data collection began only after successfully passing this site qualification procedure.

## Scanning and Diffusion-Weighted Imaging Parameters

Single-shot 2D spin-echo whole-brain echo planar imaging (EPI) was used. The parameters for each site are displayed in the table below. For diffusion weighted imaging, images were acquired with diffusion gradients (b-value = 1500 s/mm²) applied in 64 directions. In addition, two b-value=0 (s/mm²) volumes with reversed phase encoding (blip-up/blip-down) were acquired with distortions in opposite phase-encode directions in order to estimate susceptibility induced distortions.

### TABLE S2. Scanning parameters and acquisition sequences at each sitea

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Site 1  Aachen | Site 2 Southampton | Site 3  Basel | Site 4  Birmingham |
| Scanner make and model | Siemens Magnetom Prisma | Siemens Magnetom Tim Trio | Siemens Magnetom Prisma | Philips Achieva |
| Head coil | 20-channel | 32-channel | 20-channel | 32-channel |
| Software version | Syngo MR D13D | Syngo MR B17 | Syngo MR D13D | Version 3.2.6.1 |
| DWI acquisition parameters | TE = 71ms,  TR = 7500ms,  FOV = 256 x 256 x 124mm, voxel size = 2x2x2mm,  axial slices = 62, bandwidth = 1776 Hz/Px,  b-value = 1500s/mm2, directions = 64 | TE = 92ms,  TR = 8800ms, FOV = 256 x 256 x 124mm, voxel size = 2x2x2mm,  axial slices = 62, bandwidth = 1776 Hz/Px,  b-value = 1500s/mm2, directions = 64 | TE = 71ms,  TR = 7500ms, FOV = 256 x 256 x 124mm,  voxel size = 2x2x2mm,  axial slices = 62,  bandwidth = 1776 Hz/Px,  b-value = 1500s/mm2, directions = 64 | TE = 87ms,  TR = 8000ms, FOV = 256 x 256 x 124mm, voxel size = 2x2x2mm,  axial slices = 62, bandwidth = 1633.3 Hz/Px,  b-value = 1500s/mm2, directions = 64 |

**a**TE = echo time; TR=Repetition time; FOV = field of view; Hz/Px = Hz per pixel.

## Quality Assessment

Quality assessment was performed on all raw images prior to pre-processing to identify any excessive head motion and/or signal loss. For data included in this project from a previous study by Rogers and colleagues (*n*=222) (1), visual inspection was carried out by three independent reviewers, and any participants identified as problematic were then checked by an expert on DTI data acquisition and analysis (Dr Jesper Andersson). 35 scans were excluded at this stage. For an additional 78 participants processed for the purposes of this project, visual inspection was carried out by two independent reviewers (including S.T.), with problematic cases checked by two senior authors (G.F., J.R.) with extensive MRI experience. Of these 78 participants, 25 were excluded due to poor quality or incomplete data. A standardised visual inspection guide was created based on expert input and FSL guidance (<https://fsl.fmrib.ox.ac.uk/fsl/oldwiki/eddy(2f)UsersGuide.html>), including examples of common issues and guidance on when to pass or fail scans.

## Image Processing

Diffusion-weighted images (DWIs) were pre-processed with the FMRIB Diffusion Toolbox (FDT) (24), including the use of *TopUp* to correct for susceptibility-induced field distortions by using two non-diffusion weighted images (b-value = 0) with opposite phase-encoding directions (blip-up/blip-down) (25). The estimated susceptibility field provided by *TopUp* was then fed into *Eddy* to complete eddy-current and motion correction (26). The advanced *Eddy* option ‘repol’ (replace outliers) was also employed to assess the impact of motion-related outliers on movement and distortion corrections, by detecting slices affected by signal loss with a high level of specificity and replacing them with a non-parametric prediction (27). A second quality control assessment was conducted at this point, excluding scans with persistent head motion or strobing, or signal loss in >10% of the acquired directions.

## Diffusion Tensor Imaging Preparation

A diffusion tensor model was fit to each voxel using a least-squares minimisation approach (*dtifit*) to create fractional anisotropy (FA) images. Tract-Based Spatial Statistics (TBSS) (28) were then used to non-linearly register the individual FA maps of each participant to the FMRIB 1mm3 FA template (in the same space at the MNI152 standard space). These aligned images were then averaged to create a mean FA image, which was then fed into the FA skeletonization programme to create a mean FA skeleton. The mean FA skeleton represents the centre of all white matter tracts common to the subjects included in the study, and was thresholded at 0.2 to exclude non-white matter voxels. Each participant’s aligned FA image was projected onto the mean FA skeleton by filling the skeleton with FA values from the nearest relevant tract centre for each participant, enabling group comparison between subjects. These transformations were then applied to the other diffusion parameter maps (axial diffusivity [AD], radial diffusivity [RD] and mean diffusivity [MD]), and projected onto the same aligned skeleton.

## Statistical Analysis

To compare groups on demographic and clinical variables, we employed χ² tests and Fisher’s exact tests for categorical variables. For continuous variables, we used Welch’s t-tests, Mann-Whitney U tests, or one-way ANOVAs followed by post-hoc pairwise comparisons. These analyses were conducted in SPSS 28 (IBM Corp., Armonk, NY).

## Accounting for Between-Site Heteroscedasticity

Whole-brain voxel-wise analyses were conducted using exchangeability blocks and non-parametric permutation within FSL’s *Permutation Analysis of Linear Models* (PALM) tool (29). This is an extension of the more commonly used permutation method *Randomise*, which requires all data to be exchangeable under the null hypothesis in order to provide exact control of false positives. However, to account for across-site heteroscedasticity (related to the ratio of males and females across groups, and scanner-types), the PALM tool allowed us to model site as exchangeability blocks, which limit permutations to the site of data collection. This prevents permutations creating data without the original dependence structure to construct the reference null distribution. Instead, permutation inference is allowed by restricting permutations to only create rearrangements of data that respect true exchangeability, hence retaining the original joint distribution.

## Analytical justification

This study adopted a pairwise comparison approach to compare diffusion indices between the HC, CD/+ and CD/- groups. An alternative method would have been to conduct an omnibus F-test followed by pairwise comparisons, but this is not recommended when using threshold-free cluster enhancement (as properties of the relationship between F- and t-tests no longer hold in this situation). Additionally, conducting pairwise comparisons on significant regions identified from an F-test would not have allowed for whole-brain comparisons in individual group comparisons, and it was decided that this would be overly conservative given the exploratory nature of this study. Finally, in order to allow comparison with previous results pertaining to cortical structure (8), a similar analytical approach was judged to be preferable.

Most previous research examining WM microstructure in CD has been conducted in all-male samples (30-33), while research that has explicitly tested for sex differences in CD (including the study by Rogers et al. with an overlapping sample) has found that sex moderated the effects of CD on WM connectivity (1,34,35). Because of this, and the unequal sex distribution across our HC and CD groups, we decided to explore group differences when splitting the sample by sex. In order to allow comparison between our results and prior research in CD boys, we conducted pairwise comparisons between our CD-all and HC groups in an all-male sample, before again comparing between the three subgroups (CD/+, CD/- and HC). We then repeated this procedure in an all-female sample.

While a more parsimonious approach may have been to model a sex-by-group interaction, this was deemed inappropriate given the relatively small sample sizes in our subgroups (as it is estimated that the required sample size to detect an interaction is fourfold that required to detect a main effect (36,37)). Although stratified analyses do not allow us to reliably infer sex differences, they provide us with a preliminary indication of the differences between CD/+ and CD/- subgroups and controls within each sex separately (38). We hope that these results will provide a rationale for larger studies that are powered to detect sex interactions to explore this question further. This includes, for instance, the dataset compiled by the ENIGMA-Antisocial Behavior working group (<https://enigma.ini.usc.edu/ongoing/enigma-antisocial-behavior/>).

# SUPPLEMENTARY RESULTS

|  |
| --- |
| TABLE S3. Demographic and clinical characteristics of the sample including post hoc comparisonsa |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **HC  (*n*=169)** | |  | **CD/-**  **(*n*=61)** | |  | **CD/+**  **(*n*=39)** | |  | ***Omnibus*** | |  | ***Post hoc*** | | |
|  |  |  |  | **HC vs CD/- vs CD/+** | |  | **HC vs CD/-** | **HC vs CD/+** | **CD/- v CD/+** |
| **Characteristic** | **Mean** | **SD** |  | **Mean** | **SD** |  | **Mean** | **SD** |  | ***F*** | ***p*** |  | ***p*** | ***p*** | ***p*** |
| Age (years) | 14.15 | 2.59 |  | 14.38 | 2.11 |  | 14.62 | 2.32 |  | 0.60 | .551 |  | .832 | .557 | .883 |
| IQ | 104.15 | 11.07 |  | 95.08 | 11.20 |  | 95.46 | 13.72 |  | 18.88 | **< .001** |  | **< .001** | **< .001** | .986 |
| SESb | 0.21 | 0.83 |  | -0.26 | 0.85 |  | -0.59 | 0.78 |  | 11.83 | **< .001** |  | **.007** | **< .001** | .252 |
| CD symptoms, current | 0.07 | 0.27 |  | 4.93 | 2.22 |  | 5.51 | 2.42 |  | 411.42 | **< .001** |  | **< .001** | **< .001** | .114 |
| ADHD symptoms, current | 0.01 | 0.08 |  | 6.92 | 6.02 |  | 7.10 | 5.62 |  | 70.57 | **< .001** |  | **< .001** | **< .001** | .965 |
| Attention problems (CBCL)c | 2.09 | 2.59 |  | 8.85 | 4.05 |  | 9.52 | 3.41 |  | 142.16 | **< .001** |  | **< .001** | **< .001** | .594 |
| Psychopathic traits (YPI total) | 91.63 | 18.66 |  | 114.26 | 25.41 |  | 104.49 | 23.66 |  | 22.64 | **< .001** |  | **< .001** | **.002** | .064 |
| CU traits (YPI CU subscale) | 27.52 | 6.65 |  | 34.67 | 8.84 |  | 30.72 | 7.44 |  | 17.63 | **< .001** |  | **< .001** | **.038** | **.024** |
| CU traits (ICU total) | 16.98 | 7.65 |  | 33.70 | 12.02 |  | 37.51 | 13.03 |  | 87.25 | **< .001** |  | **< .001** | **< .001** | .135 |
|  | ***n*** | **%** |  | ***n*** | **%** |  | ***n*** | **%** |  | ***p* (Fisher’s exact)** | |  | ***p*** | ***p*** | ***p*** |
| Sex (Female) | 104 | 61.5% |  | 24 | 39.3% |  | 24 | 61.5% |  | **.009** | |  | **.004** | .999 | **.040** |
| Age of onset (CD) |  |  |  |  |  |  |  |  |  | .627d | |  |  |  |  |
| Childhood |  |  |  | 27 | 44.3% |  | 20 | 51.3% |  |  | |  |  |  |  |
| Adolescent |  |  |  | 30 | 59.2% |  | 18 | 46.2% |  |  | |  |  |  |  |
| Unspecified |  |  |  | 4 | 6.6% |  | 1 | 2.6% |  |  | |  |  |  |  |
| Psychotropic medication | 0 | 0.0% |  | 19 | 31.7% |  | 8 | 20.5% |  | **< .001** | |  | **< .001** | **< .001** | .348 |
| Lifetime diagnoses |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |
| ODD | 0 | 0.0% |  | 47 | 77.0% |  | 35 | 89.7% |  | **< .001** | |  | **< .001** | **< .001** | .120 |
| ADHD | 0 | 0.0% |  | 30 | 49.2% |  | 19 | 48.7% |  | **< .001** | |  | **< .001** | **< .001** | .999 |
| MDD | 2 | 1.2% |  | 17 | 27.9% |  | 20 | 51.3% |  | **< .001** | |  | **< .001** | **< .001** | **.021** |
| GAD | 0 | 0.0% |  | 6 | 10.0% |  | 4 | 10.3% |  | **< .001** | |  | **< .001** | **< .001** | .999 |
| PTSD | 0 | 0.0% |  | 6 | 9.8% |  | 5 | 12.8% |  | **< .001** | |  | **< .001** | **< .001** | .747 |
| Alcohol abuse | 0 | 0.0% |  | 2 | 3.3% |  | 2 | 5.1% |  | **.026** | |  | .069 | **.034** | .642 |
| Substance abuse | 0 | 0.0% |  | 7 | 11.5% |  | 4 | 10.3% |  | **< .001** | |  | **< .001** | **.001** | .999 |

aPost-hoc comparisons were based on Bonferroni-corrected Welch’s *t*-tests and Fisher’s exact tests. Significant *p*-values are marked in bold. CD participants were classified as having childhood-onset CD if at least one symptom and functional impairment occurred before the age of 10. Otherwise, participants were classified as having adolescent-onset CD. HC=Healthy Controls; CD=Conduct Disorder; CD/+ and CD/-=Conduct Disorder with and without a history of childhood physical or sexual abuse; n=sample size; SD=standard deviation; IQ=intelligence quotient; CBCL=Child Behavior Checklist; YPI=Youth Psychopathic traits Inventory; CU=callous-unemotional; ICU=Inventory of Callous-Unemotional traits; ODD=Oppositional Defiant Disorder; ADHD=Attention-Deficit/Hyperactivity Disorder; MDD=Major Depressive Disorder; GAD=Generalised Anxiety Disorder; PTSD=Post-Traumatic Stress Disorder.

bMissing for 30 participants (22 HCs, 5 CD/-, 3 CD/+)

cMissing for 32 participants (18 HCs, 8 CD/-, 6 CD/+)

dOnly the CD subgroups are compared here

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TABLE S4. Significant group differences from sex-stratified analyses conducted with an all-male subsamplea | | | | | | | | | | | | | | |
| **Comparison** |  | **Diffusion Parameter** |  | | **White matter label** | **Hemisphere**  **(L/R)** |  | **Peak MNI coordinates** | | | **Cluster size (No. of voxels)** |  | ***p*-value** | **Cohen’s *d*** |
| **X** | **Y** | **Z** |
| **CD-all vs HC** |  |  |  | |  |  |  |  |  |  |  |  |  |  |
| CD-all > HC |  | Fractional Anisotropy |  | | Genu of corpus callosum | R |  | 10 | 23 | 17 | 4 |  | .039 | 0.71 |
|  |  |  |  | |  |  |  |  |  |  |  |  |  |  |
| CD-all < HC |  | Radial Diffusivity |  | | Genu of corpus callosum | R |  | 10 | 23 | 17 | 92 |  | .038 | 0.57 |
|  |  |  |  | | Body of corpus callosum | R |  | 7 | 18 | 20 | 14 |  | .042 | 0.58 |
|  |  |  |  | | Body of corpus callosum | R |  | 1 | 17 | 19 | 2 |  | .048 | 0.60 |
| **CD/- vs HC** |  |  |  | |  |  |  |  |  |  |  |  |  |  |
| CD/- > HC |  | Fractional Anisotropy |  | | Genu of corpus callosum | R |  | 10 | 23 | 17 | 70 |  | .026 | 0.63 |
|  |  |  | |  | Genu of corpus callosum | R |  | 12 | 20 | 22 | 4 |  | .041 | 0.66 |
|  |  |  |  | |  |  |  |  |  |  |  |  |  |  |
| CD/- < HC |  | Radial Diffusivity |  | | Genu of corpus callosum | R |  | 10 | 23 | 17 | 107 |  | .025 | 0.59 |
|  |  |  |  | | Body of corpus callosum | R |  | 2 | 17 | 19 | 6 |  | .045 | 0.58 |
|  |  |  |  | | Genu of corpus callosum | R |  | 12 | 30 | 11 | 4 |  | .047 | 0.70 |
|  |  |  |  | | Genu of corpus callosum | R |  | 11 | 32 | 2 | 3 |  | .046 | 0.89 |

aAll analyses controlled for age and site. P-values are based on Threshold-free cluster enhancement, significant at *p*<.05, family wise error-corrected. Cohen’s *d* was calculated using whole-brain vertex-wise effect size brain maps. MNI=Montreal Neurological Institute. HC=Healthy Controls; CD=Conduct Disorder; CD/-=Conduct Disorder without childhood physical or sexual abuse history; CD/+=Conduct Disorder with childhood physical or sexual abuse history.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TABLE S5. Significant group differences from sex-stratified analyses conducted with an all-female subsamplea | | | | | | | | | | | | | |
| **Comparison** |  | **Diffusion Parameter** |  | **White matter label** | **Hemisphere**  **(L/R)** |  | **Peak MNI coordinates** | | | **Cluster size (No. of voxels)** |  | ***p*-value** | **Cohen’s *d*** |
| **X** | **Y** | **Z** |
| **CD-all vs HC** |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CD-all < HC |  | Radial Diffusivity |  | Fornix | L |  | -5 | -21 | 16 | 36 |  | .036 | 0.69 |
|  |  |  |  | Fornix | R |  | 5 | -14 | 17 | 31 |  | .039 | 0.69 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CD-all < HC |  | Mean Diffusivity |  | Fornix | L |  | -5 | -22 | 17 | 12 |  | .047 | 0.72 |
|  |  |  |  | Fornix | L |  | -3 | -15 | 16 | 10 |  | .045 | 0.79 |
| **CD/- vs HC** |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CD/- < HC |  | Fractional Anisotropy |  | Posterior corona radiata | R |  | 23 | -35 | 28 | 60 |  | .036 | 0.82 |
|  |  |  |  | Superior corona radiata | R |  | 26 | -19 | 26 | 14 |  | .035 | 0.91 |

aAll analyses controlled for age and site. P-values are based on Threshold-free cluster enhancement, significant at *p*<.05, family wise error-corrected. Cohen’s *d* was calculated using whole-brain vertex-wise effect size brain maps. MNI=Montreal Neurological Institute. HC=Healthy Controls; CD=Conduct Disorder; CD/-=Conduct Disorder without childhood physical or sexual abuse history; CD/+=Conduct Disorder with childhood physical or sexual abuse history.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TABLE S6. Significant group differences when controlling for IQa | | | | | | | | | | | | | | | | |  |
| **Sample** | **Comparison** |  | **Diffusion Parameter** |  | **White matter label** | **Hemisphere**  **(L/R)** |  | **Peak MNI coordinates** | | | **Cluster size (No. of voxels)** |  | ***p*-value** | **Cohen’s *d*** |  |
| **X** | **Y** | **Z** |
| **Combined sex** | **CD/+ vs CD/-** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | CD/+ > CD/- |  | Axial Diffusivity |  | Superior longitudinal fasciculus | R |  | 37 | -21 | 29 | 19 |  | .024 | 0.88 |  |
| **Male only** | **CD-all vs HC** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | CD-all < HC |  | Radial Diffusivity |  | Body of corpus callosum | R |  | 7 | 18 | 19 | 5 |  | .044 | 0.60 |  |
|  | **CD/- vs HC** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | CD/- <HC |  | Radial Diffusivity |  | Genu of corpus callosum | R |  | 7 | 18 | 19 | 59 |  | .030 | 0.68 |  |
|  |  |  | Radial Diffusivity |  | Body of corpus callosum | R |  | 2 | 17 | 19 | 3 |  | .030 | 0.69 |  |
| **Female only** | **CD-all vs HC** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | CD-all < HC |  | Radial Diffusivity |  | Fornix | L |  | -5 | -21 | 16 | 75 |  | .017 | 0.91 |  |
|  |  |  |  |  | Fornix | R |  | 4 | -16 | 17 | 104 |  | .017 | 0.86 |  |
|  | CD-all < HC |  | Mean Diffusivity |  | Fornix | L |  | -4 | -22 | 17 | 70 |  | .023 | 0.94 |  |
|  |  |  |  |  | Fornix | R |  | 5 | -14 | 17 | 31 |  | .031 | 0.94 |  |
|  | **CD/- vs HC** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | CD/- < HC |  | Fractional Anisotropy |  | Posterior corona radiata | R |  | 26 | -20 | 26 | 138 |  | .019 | 1.03 |  |
|  |  |  | Fractional Anisotropy |  | Superior corona radiata | R |  | 28 | -26 | 33 | 11 |  | .043 | 1.05 |  |

aAll analyses controlled for age and site. Sex was included as a covariate in combined sex analyses P-values are based on Threshold-free cluster enhancement, significant at *p*<.05, family wise error-corrected. Cohen’s *d* was calculated using whole-brain vertex-wise effect size brain maps. MNI=Montreal Neurological Institute. IQ=intelligence quotient. CD=Conduct Disorder; HC=Healthy Controls; CD/-=Conduct Disorder without a childhood physical or sexual abuse history; CD/+=Conduct Disorder with a childhood physical or sexual abuse history.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TABLE S7. Significant group differences when controlling for CBCL attention problemsa | | | | | | | | | | | | | | | | |  |
| **Sample** | **Comparison** |  | **Diffusion Parameter** |  | **White matter label** | **Hemisphere**  **(L/R)** |  | **Peak MNI coordinates** | | | **Cluster size (No. of voxels)** |  | ***p*-value** | **Cohen’s *d*** |  |
| **X** | **Y** | **Z** |
| **Combined sex** | **CD/+ vs CD/-** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | CD/+ > CD/- |  | Axial Diffusivity |  | Superior longitudinal fasciculus | R |  | 37 | -21 | 29 | 18 |  | .028 | 0.98 |  |
| **Male only** | **CD/- vs HC** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | CD/- < HC |  | Radial Diffusivity |  | Genu of corpus callosum | L |  | -15 | 22 | 23 | 6 |  | .047 | 1.50 |  |
| **Female only** | **CD-all vs HC** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | CD-all < HC |  | Radial Diffusivity |  | Fornix | R |  | 5 | -14 | 17 | 8 |  | .037 | 1.07 |  |

aAll analyses controlled for age and site. Sex was included as a covariate in combined sex analyses. P-values are based on Threshold-free cluster enhancement, significant at *p*<.05, family wise error-corrected. Cohen’s *d* was calculated using whole-brain vertex-wise effect size brain maps. MNI=Montreal Neurological Institute. CBCL=child behaviour checklist. CD=Conduct Disorder; HC=Healthy Controls; CD/-=Conduct Disorder without a history of childhood physical or sexual abuse; CD/+=Conduct Disorder with a history of childhood physical or sexual abuse. Note that Child Behaviour Checklist (CBCL) data were missing for 32 participants (18 HCs, 6 CD/+, and 8 CD/-), leaving an overall sample size of n = 237 for this analysis (151 HCs, 86 CD, 33 CD/+, and 53 CD/-).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TABLE S8. Significant group differences when controlling for socioeconomic statusa | | | | | | | | | | | | | | | | |  |
| **Sample** | **Comparison** |  | **Diffusion Parameter** |  | **White matter label** | **Hemisphere**  **(L/R)** |  | **Peak MNI coordinates** | | | **Cluster size (No. of voxels)** |  | ***P*-value** | **Cohen’s *d*** |  |
| **X** | **Y** | **Z** |
| **Combined sex** | **CD/+ vs CD/-** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | CD/+ > CD/- |  | Axial Diffusivity |  | Superior longitudinal fasciculus | R |  | 37 | -21 | 29 | 22 |  | .025 | 0.93 |  |
| **Female only** | **CD-all vs HC** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | CD-all < HC |  | Radial Diffusivity |  | Fornix | L |  | -4 | -22 | -17 | 74 |  | .018 | 0.85 |  |
|  |  |  |  |  | Fornix | R |  | 1 | -3 | 16 | 29 |  | .034 | 0.79 |  |
|  |  |  |  |  | Fornix | R |  | 3 | -13 | 17 | 6 |  | .042 | 0.82 |  |
|  |  |  |  |  | Fornix | R |  | 4 | -18 | 18 | 4 |  | .048 | 0.82 |  |
|  | CD-all < HC |  | Mean Diffusivity |  | Fornix | L |  | -3 | -15 | 16 | 14 |  | .038 | 0.95 |  |
|  |  |  |  |  | Fornix | L |  | -4 | -22 | 17 | 5 |  | .048 | 0.93 |  |
|  | **CD/- vs HC** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | CD/- < HC |  | Fractional Anisotropy |  | Posterior corona radiata | L |  | -30 | -46 | 16 | 205 |  | .011 | 0.93 |  |
|  |  |  |  |  | Posterior corona radiata | R |  | 27 | -37 | 24 | 130 |  | .025 | 0.91 |  |
|  |  |  |  |  | Splenium of corpus callosum | L |  | -23 | -54 | 11 | 76 |  | .011 | 1.08 |  |
|  |  |  |  |  | Splenium of corpus callosum | R |  | 26 | -52 | 15 | 28 |  | .028 | 1.03 |  |
|  |  |  |  |  | Corticospinal tract | L |  | -26 | -15 | 16 | 21 |  | .035 | 1.07 |  |
|  |  |  |  |  | Body of the corpus callosum | R |  | 31 | -43 | 17 | 20 |  | .033 | 1.02 |  |
|  |  |  |  |  | Splenium of corpus callosum | R |  | 20 | -46 | 10 | 3 |  | .048 | 1.03 |  |

aAll analyses controlled for age and site. Sex was included as a covariate in combined sex analyses. P-values are based on Threshold-free cluster enhancement, significant at *p*<.05, family wise error-corrected. Cohen’s *d* was calculated using whole-brain vertex-wise effect size brain maps. MNI=Montreal Neurological Institute. SES=socioeconomic status. CD=Conduct Disorder; HC=Healthy Controls; CD/-=Conduct Disorder without a history of childhood physical or sexual abuse; CD/+=Conduct Disorder with a history of childhood physical or sexual abuse. Note that socioeconomic status data were missing for 30 participants (22 HCs, 3 CD/+, and 5 CD/-), leaving an overall sample size of n = 239 for this analysis (147 HCs, 92 CD, 36 CD/+, and 56 CD/-).

### TABLE S9. Overview of results across sensitivity analysesa

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sample** | **Comparison** | **Diffusion Parameter** | **White matter label** | **IQ** | **ADHD** | **SES** |
| **Combined sex** | **CD-all vs HC** |  |  |  |  |  |
|  | CD-all < HC | Fractional anisotropy | Genu and body of corpus callosum | N | N | N |
|  | **CD/- vs HC** |  |  |  |  |  |
|  | CD/- > HC | Fractional Anisotropy | Genu and body of corpus callosum | N | N | N |
|  | **CD/+ vs CD/-** |  |  |  |  |  |
|  | CD/+ > CD/- | Axial Diffusivity | Superior longitudinal fasciculus | Y | Y | Y |
| **Male only** | **CD-all vs HC** |  |  |  |  |  |
|  | CD-all > HC | Fractional Anisotropy | Genu of corpus callosum | N | N | N |
|  | CD-all < HC | Radial Diffusivity | Genu and body of corpus callosum | Y | N | N |
|  | **CD/- vs HC** |  |  |  |  |  |
|  | CD/- > HC | Fractional Anisotropy | Genu of corpus callosum | N | N | N |
|  | CD/- < HC | Radial Diffusivity | Genu and body of corpus callosum | Y | Y | N |
| **Female only** | **CD-all vs HC** |  |  |  |  |  |
|  | CD-all < HC | Radial Diffusivity | Fornix | Y | Y | Y |
|  | CD-all < HC | Mean Diffusivity | Fornix | Y | N | Y |
|  | **CD/- vs HC** |  |  |  |  |  |
|  | CD/- < HC | Fractional Anisotropy | Superior and posterior corona radiata | Y | N | Y |

aIQ=intelligence quotient. ADHD=attention-deficit/hyperactivity disorder, as measured by the child behaviour checklist. SES=socioeconomic status. CD=Conduct Disorder; HC=Healthy Controls; CD/-=Conduct Disorder without a history of childhood physical or sexual abuse; CD/+=Conduct Disorder with a history of childhood physical or sexual abuse.

# SUPPLEMENTARY REFERENCES

1. Rogers JC, Gonzalez-Madruga K, Kohls G, Baker RH, Clanton RL, Pauli R, et al. (2019): White matter microstructure in youths with Conduct Disorder: Effects of sex and variation in callous traits. *J Am Acad Child Adolesc Psychiatry*. 58:1184-1196.

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 (2013): *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Washington, D.C.: American Psychiatric Association.

4. American Psychiatric Association (2000): *Diagnostic and statistical manual of mental disorders, text revision.* 4th ed. Washington, D.C.: US: American Psychiatric Association.

5. Dodge KA, Bates JE, Pettit GS (1990): Mechanisms in the cycle of violence. *Science*. 250:1678-1683.

6. 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. *Am J Psychiatr*. 168:65-72.

7. Choi JK, Wang D, Jackson AP (2019): Adverse experiences in early childhood and their longitudinal impact on later behavioral problems of children living in poverty. *Child Abuse Negl*. 98:104181.

8. Staginnus M, Cornwell H, Toschi N, Oosterling M, Paradysz M, Smaragdi A, et al. (2023): Testing the Ecophenotype Model: Cortical structure alterations in Conduct Disorder with versus without childhood maltreatment. *Biol Psychiatry Cogn Neurosci Neuroimaging*.

9. Bernstein D, Fink L (1998): Manual for the childhood trauma questionnaire. *New York: The Psychological Corporation*.

10. Bifulco A, Bernazzani O, Moran P, Jacobs C (2005): The childhood experience of care and abuse questionnaire (CECA. Q): validation in a community series. *Br J Clin Psychol*. 44:563-581.

11. Bifulco A, Brown GW, Harris TO (1994): Childhood Experience of Care and Abuse (CECA): a retrospective interview measure. *J Child Psychol Psychiatry*. 35:1419-1435.

12. Andershed HA, Kerr M, Stattin H, Levander S (2002): Psychopathic traits in non-referred youths: A new assessment tool. In: Blaauw E, Sheridan L, Haag D, editors. *Psychopaths: current international perspectives*: Elsevier, pp 131-158.

13. Essau CA, Sasagawa S, Frick PJ (2006): Callous-unemotional traits in a community sample of adolescents. *Assessment*. 13:454-469.

14. Eekhout I, De Vet HC, Twisk JW, Brand JP, 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.

15. Liu Y, De A (2015): Multiple imputation by fully conditional specification for dealing with missing data in a large epidemiologic study. *J Int Med Res*. 4:287.

16. Achenbach TM (1991): Manual for the Child Behavior Checklist/4-18 and 1991 profile. *University of Vermont, Department of Psychiatry*.

17. Wechsler D (1999): *Wechsler Abbreviated Scale of Intelligence*. Psychological Corporation.

18. Wechsler D (2003): *Wechsler Intelligence Scale for Children*. Psychological Corporation.

19. Wechsler D (1997): *Wechsler Adult Intelligence Scale*. Psychological Corporation.

20. International Labour Office (2012): *International Standard Classification of Occupations 2008 (ISCO-08): Structure, group definitions and correspondence tables*. International Labour Office.

21. UNESCO Institute for Statistics (2015): International Standard Classification of Education Fields of Education and Training 2013 (ISCED-F 2013): Detailed Field Descriptions. UNESCO Institute for Statistics Montreal, QC.

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 G (2005): FBIRN Stability phantom QA procedures.

24. Behrens TE, Woolrich MW, Jenkinson M, Johansen‐Berg H, Nunes RG, Clare S, et al. (2003): Characterization and propagation of uncertainty in diffusion‐weighted MR imaging. *Magn Reson Med*. 50:1077-1088.

25. Andersson JL, Skare S, Ashburner J (2003): How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage*. 20:870-888.

26. Andersson JL, Sotiropoulos SN (2016): An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage*. 125:1063-1078.

27. Andersson JL, Graham MS, Zsoldos E, Sotiropoulos SN (2016): Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. *NeuroImage*. 141:556-572.

28. Smith SM, Johansen-Berg H, Jenkinson M, Rueckert D, Nichols TE, Miller KL, et al. (2007): Acquisition and voxelwise analysis of multi-subject diffusion data with tract-based spatial statistics. *Nat Protoc*. 2:499-503.

29. Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE (2014): Permutation inference for the general linear model. *NeuroImage*. 92:381-397.

30. Passamonti L, Fairchild G, Fornito A, Goodyer IM, Nimmo-Smith I, Hagan CC, et al. (2012): Abnormal anatomical connectivity between the amygdala and orbitofrontal cortex in conduct disorder. *PloS one*. 7:e48789.

31. Sarkar S, Craig M, Catani M, Dell'Acqua F, Fahy T, Deeley Q, et al. (2013): Frontotemporal white-matter microstructural abnormalities in adolescents with conduct disorder: a diffusion tensor imaging study. *Psychol Med*. 43:401-411.

32. Sarkar S, Dell’Acqua F, Froudist Walsh S, Blackwood N, Scott S, Craig MC, et al. (2016): A whole-brain investigation of white matter microstructure in adolescents with conduct disorder. *PloS one*. 11:e0155475.

33. Zhang J, Zhu X, Wang X, Gao J, Shi H, Huang B, et al. (2014): Increased structural connectivity in corpus callosum in adolescent males with conduct disorder. *J Am Acad Child Adolesc Psychiatry*. 53:466-475. e461.

34. Decety J, Yoder KJ, Lahey BB (2015): Sex differences in abnormal white matter development associated with conduct disorder in children. *Psychiatry Res Neuroimaging*. 233:269-277.

35. Zhang J, Gao J, Shi H, Huang B, Wang X, Situ W, et al. (2014): Sex differences of uncinate fasciculus structural connectivity in individuals with conduct disorder. *Biomed Res Int*. 2014:673165.

36. Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ (2004): Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol*. 57:229-236.

37. Leon AC, Heo M (2009): Sample sizes required to detect interactions between two binary fixed-effects in a mixed-effects linear regression model. *Comput Stat Data Anal*. 53:603-608.

38. Rich-Edwards JW, Kaiser UB, Chen GL, Manson JE, Goldstein JM (2018): Sex and gender differences research design for basic, clinical, and population studies: essentials for investigators. *Endocr Rev*. 39:424-439.