THE LANCET
Psychiatry

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Gao Y, Staginnus M and the ENIGMA-Antisocial Behavior Working Group. Cortical structure and subcortical volumes in conduct disorder: a coordinated analysis of 15 international cohorts from the ENIGMA-Antisocial Behavior Working Group. *Lancet Psychiatry* 2024; **11:** 620–32.

Supplementary Appendix

Supplement to: Gao, Staginnus, and the ENIGMA-Antisocial Behavior Working Group. *Cortical structure and subcortical volumes in conduct disorder: a coordinated analysis of 15 international cohorts from the ENIGMA-Antisocial Behavior Working Group*

Table of Contents

AUTHOR AND ARTICLE INFORMATION

This study was pre-registered (https://doi.org/10.17605/OSF.IO/V6BDC) and any deviations from the pre-registration have been outlined and justified in an appendix to the pre-registration (accessible under 'Resources').

Authors and affiliations

Yidian Gao[1#], PhD, Marlene Staginnus[2#], MRes, Sophie Townend[2], MSc, Celso Arango[3], PhD, Sahil Bajaj[4], PhD, Tobias Banaschewski[5,6], MD, Edward D. Barker[7], PhD, Vivek Benegal[8], MD, Kathryn Berluti[9], PhD, Anka Bernhard[10], PhD, Robert J.R. Blair[11], PhD, Charlotte P.S. Boateng[12], MD, Arun L.W. Bokde[13], PhD, Daniel Brandeis[5,14], PhD, Jan K. Buitelaar[15,16], MD, S. Alexandra Burt[17], PhD, Elise M. Cardinale[18], PhD, Josefina Castro-Fornieles[19], PhD, Hui Chen[20], MMSc, Xianliang Chen[20], MD, Sally C. Chester[1], MSc, Olivier F. Colins[21], PhD, Harriet Cornwell[2], PhD, Michael Craig[22], PhD, Ana I. Cubillo[23], PhD, Sylvane Desrivieres[24], PhD, Dana E. Díaz[25,26], PhD, Andrea Dietrich[27,28], PhD, Daifeng Dong[29], PhD, Anouk H. Dykstra[15], MSc, Barbara Franke[15,30], PhD, Christine M. Freitag[10], MD, Jeffrey C. Glennon[31], PhD, Karen Gonzalez-Madruga[32], PhD, Cindy C. Hagan[33,34], PhD, Pieter J. Hoekstra[27,28], PhD, Bharath Holla[35], MD, Luke W. Hyde[36], PhD, Karim Ibrahim[37], PhD, Nimrah Jabeen[1], BSc, Rebecca L. Jackson[2], MRes, Yali Jiang[29,38], MD, Gregor Kohls[39], PhD, Kerstin Konrad[40,41], PhD, Alexandra Kypta-Vivanco[2], MSc, Kim Lamers[15], MSc, Ren Ma[29], PhD, Abigail A. Marsh[9,42], PhD, Anne Martinelli[10,43], PhD, Jean-Luc Martinot[44,45], PhD, Kalina J. Michalska[25], PhD, Qingsen Ming[29,46], MD, Silvia Minosse[47], PhD, Colter Mitchell[48], PhD, Christopher S. Monk[36], PhD, Declan Murphy[7], MD, Leah E. Mycue[36,49], BA, Jilly Naaijen[50], PhD, Maaike Oosterling[51], MSc, Luca Passamonti[52], MD, Ruth Pauli[1], PhD, Maria Jose Penzol Alonso[3], MSc, Harriet Phillips[53], MSc, Montana L. Ploe[9,54], BA, Nora M. Raschle[55,56], PhD, Ruth Roberts[53], PhD, Jack C. Rogers[57], PhD, Mireia Rosa-Justicia[58], PhD, Ilyas Sagar-Ouriaghli[7], PhD, Ulrike M.E. Schulze[59,60], MD, Gunter Schumann[61,62], PhD, Arjun Sethi[22], PhD, Areti Smaragdi[63], PhD, Edmund J.S. Sonuga-Barke[64], PhD, Christina Stadler[23], PhD, Michael C. Stevens[65,66], PhD, Denis G. Sukhodolsky[37], PhD, Kate Sully[67], PhD, Xiaoqiang Sun[29], BA, Nicola Toschi[68,69], PhD, Christopher D. Townsend[1], MA, Nic J.A. van der Wee[70,71], PhD, Robert Vermeiren[12], PhD, Essi Viding[53], PhD, Xiaoping Wang[20], MD, Heidi B. Westerman[36], MSc, Qiong Wu[29,72], PhD, Shuqiao Yao[29], MD, Jibiao Zhang[29,73], PhD, Jiansong Zhou[20,74], MD, Jiawei Zhou[20], BSc, Neda Jahanshad[75], PhD, Sophia I. Thomopoulos[75], BA, Christopher R.K. Ching[75], PhD, Melody J.Y. Kang[75], MSc, Paul M. Thompson[75], PhD, Eduard T. Klapwijk[76], PhD, Daniel S. Pine[77], MD, Arielle Baskin-Sommers[78], PhD, Charlotte A.M. Cecil[79–81], PhD, Moji Aghajani[82,83], PhD, Esther Walton[2], PhD, Graeme Fairchild[2#], PhD, Stephane A. De Brito[1,57,84,85#], PhD

 # contributed equally, shared first and last authorship

Affiliations

1. Centre for Human Brain Health, School of Psychology, University of Birmingham, Birmingham, United Kingdom.
2. Department of Psychology, University of Bath, Bath, United Kingdom.
3. Department of Child and Adolescent Psychiatry, Institute of Psychiatry and Mental Health, Hospital General Universitario Gregorio Marañón, IiSGM, CIBERSAM, ISCIII, School of Medicine, Universidad Complutense, Madrid, Spain.
4. Department of Cancer Systems Imaging, Division of Diagnostic Imaging, The University of Texas MD Anderson Cancer Center, Houston, TX, United States.
5. Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health Mannheim, Heidelberg University, Mannheim, Germany.
6. German Center for Mental Health (DZPG), partner site Mannheim-Heidelberg-Ulm, Germany.
7. The Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, London, United Kingdom.
8. Department of Psychiatry, National Institute of Mental Health and Neurosciences, Bengaluru, Karnataka, India.
9. Department of Psychology, Georgetown University, Washington, DC, United States.

10. Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital Frankfurt am Main, Goethe University, Frankfurt am Main, Germany.
11. Child and Adolescent Psychiatric Center, Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark.
12. LUMC Curium, Child and Adolescent Psychiatry, Leiden University Medical Center, Leiden, The Netherlands.
13. Trinity College Institute of Neuroscience and School of Medicine, Trinity College Dublin, Dublin, Ireland.
14. Department of Child and Adolescent Psychiatry and Psychotherapy, Psychiatric University Hospital, University of Zurich, Zurich, Switzerland.
15. Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition, and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands.
16. Karakter Child and Adolescent University Center, Nijmegen, The Netherlands.
17. Department of Psychology, Michigan State University, East Lansing, MI, United States.
18. Department of Psychology, The Catholic University of America, Washington, DC, United States.
19. Department of Child and Adolescent Psychiatry and Psychology, 2021SGR01319, Hospital Clínic de Barcelona, IDIBAPS, CIBERSAM, Department of Medicine, Institute of Neuroscience, University of Barcelona, Barcelona, Spain.
20. Department of Psychiatry, National Clinical Research Center for Mental Disorders, and National Center for Mental Disorders, The Second Xiangya Hospital of Central South University, Changsha, Hunan, China.
21. Department of Special Needs Education, Ghent University, Gent, Belgium.
22. Department of Forensic and Neurodevelopmental Sciences, The Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, London, United Kingdom.
23. Department of Child and Adolescent Psychiatry, University of Basel, University Psychiatric Clinics, Basel, Switzerland.
24. Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, London, United Kingdom.
25. Department of Psychology, University of California, Riverside, Riverside, CA, United States.
26. Department of Psychiatry, Columbia University Irving Medical Center, New York, NY, United States.
27. University of Groningen, University Medical Center Groningen, Department of Child and Adolescent Psychiatry, Groningen, The Netherlands.
28. Accare Child Study Center, Groningen, The Netherlands.
29. Medical Psychological Center, The Second Xiangya Hospital of Central South University, Changsha, Hunan, China.
30. Department of Human Genetics, Donders Institute for Brain, Cognition, and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands.
31. School of Medicine, Conway Institute of Biomedical and Biomolecular Research, University College Dublin, Dublin, Ireland.
32. Department of Psychology, Middlesex University London, London, United Kingdom.
33. Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, United States.
34. Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom.
35. Department of Integrative Medicine, National Institute of Mental Health and Neurosciences, Bengaluru, Karnataka, India.
36. Department of Psychology, University of Michigan, Ann Arbor, MI, United States.
37. Child Study Center, School of Medicine, Yale University, New Haven, CT, United States.
38. Department of Psychology, School of Education Science, Hunan Normal University, Changsha, China.
39. Department of Child and Adolescent Psychiatry, Faculty of Medicine, TU Dresden, Dresden, Germany.
40. Clinical Child Neuropsychology Section, Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital, RWTH Aachen, Aachen, Germany.

41. JARA-BRAIN Institute II, Molecular Neuroscience and Neuroimaging, Forschungszentrum Juelich GmbH and RWTH Aachen, Juelich, Germany.
42. Interdisciplinary Program in Neuroscience, Georgetown University, Washington, DC, United States.
43. Psychology School, Fresenius University of Applied Sciences, Frankfurt am Main, Germany.
44. INSERM U1299 'Developmental Trajectories & Psychiatry', Centre Borelli UMR 9010, ENS Paris-Saclay, University Paris-Saclay, Gif sur Yvette and EPS Barthélemy Durand, Etampes, France.
45. Université Paris-Cité, Centre Borelli UMR 9010, Paris, France.
46. Department of Psychiatry, The First Affiliated Hospital of Soochow University, Suzhou, Jiangsu, China.
47. Diagnostic Imaging Unit, Department of Biomedicine and Prevention, University of Rome 'Tor Vergata', Rome, Italy.
48. Institute for Social Research, University of Michigan, Ann Arbor, MI, United States.
49. Cognitive Neuroscience, Freie Universität Berlin, Berlin, Germany.
50. Faculty of Humanities, University of Utrecht, Utrecht, The Netherlands.
51. Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands.
52. Department of Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom.
53. Division of Psychology and Language Sciences, University College London, London, United Kingdom.
54. Department of Psychology, Washington State University, Pullman, WA, United States.
55. Jacobs Center for Productive Youth Development at the University of Zurich, Department of Psychology, Zurich, Switzerland.
56. Neuroscience Center Zurich, University and ETH Zurich, Zurich, Switzerland.
57. Institute for Mental Health, School of Psychology, University of Birmingham, Birmingham, United Kingdom.
58. Child and Adolescent Psychiatry and Psychology Department, Hospital Clinic of Barcelona, Barcelona, Spain.
59. Department of Child and Adolescent Psychiatry/Psychotherapy, University of Ulm, Ulm, Germany.
60. Center for Psychiatry (ZfP) Calw, Department of Child and Adolescent Psychiatry and Psychotherapy, Böblingen, Germany.
61. Centre for Population Neuroscience and Stratified Medicine (PONS), The Institute of Science and Technology for Brain-inspired Intelligence (ISTBI), Fudan University, Shanghai, China.
62. PONS Centre, Department of Psychiatry and Neuroscience, Charité, University Medicine Berlin, Berlin, Germany.
63. Scaling, Research and Development, Child Development Institute, Toronto, ON, Canada.
64. School of Academic Psychiatry, King's College London, London, United Kingdom.
65. School of Medicine, Yale University, New Haven, CT, United States.
66. Olin Neuropsychiatry Research Center, Hartford, CT, United States.
67. School of Psychology, University of Southampton, Southampton, United Kingdom.
68. Department of Biomedicine and Prevention, University of Rome 'Tor Vergata', Rome, Italy.
69. Martinos Center for Biomedical Imaging and Harvard Medical School, Boston, MA, United States.
70. Department of Psychiatry, Leiden University Medical Center, Leiden, The Netherlands.
71. Leiden Institute for Brain and Cognition, Leiden University Medical Center, Leiden, The Netherlands.
72. Department of Neurology, Max-Planck-Institute for Human Cognitive and Brain Sciences, Leipzig, Germany.
73. Department of Education, Jianghan University, Wuhan, Hubei, China.
74. National Technology Institute on Mental Disorders, Hunan Key Laboratory of Psychiatry and Mental Health, Changsha, Hunan, China.
75. Imaging Genetics Center, Mark and Mary Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, Marina del Rey, CA, United States.
76. Erasmus University Rotterdam, Rotterdam, The Netherlands.
77. National Institute of Mental Health Intramural Research Program (NIMH-IRP), Bethesda, MD, United States.

78. Department of Psychology, Yale University, New Haven, CT, United States.
79. Department of Child and Adolescent Psychiatry and Psychology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands.
80. Department of Epidemiology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands.
81. Molecular Epidemiology, Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands.
82. Section Forensic Family and Youth Care, Institute of Education and Child Studies, Leiden University, Leiden, The Netherlands.
83. Department of Psychiatry, Amsterdam UMC, Location Vrije Universiteit Amsterdam, Amsterdam, The Netherlands.
84. Centre for Developmental Science, School of Psychology, University of Birmingham, Birmingham, United Kingdom.
85. Birmingham Centre for Neurogenetics, School of Psychology, University of Birmingham, Birmingham, United Kingdom.

SUPPLEMENTARY METHODS

Appendix 1 - Grouping and subgrouping approach

While the specific inclusion/exclusion criteria differ by sample (see details in Appendix 2), the following exclusion criteria were applied for all analyses reported in this study: 1) a mean sample age > 18 years and individual participant age > 21 years, 2) IQ < 70 (where available) and 3) presence of neurological disorders, genetic syndromes, autism spectrum disorder as well as current diagnoses of schizophrenia or bipolar disorder (where the relevant information was available). We further excluded all participants without information on sex and age as these variables were included as covariates of no interest in the main analyses. Lastly, cohorts were only included if they comprised at least 10 participants with conduct disorder (CD) and 10 typically-developing (TD) youth.

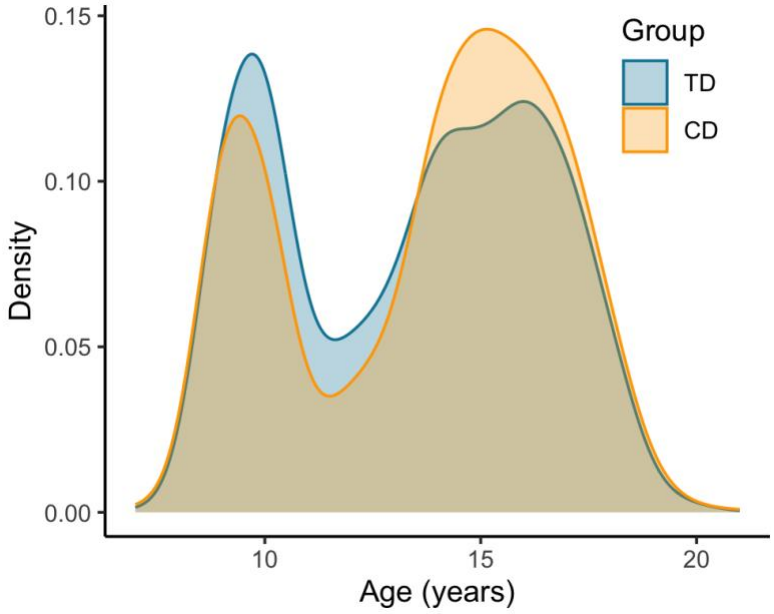
A note on autism spectrum disorder in the ABCD study: A diagnosis of 'severe' autism spectrum disorder was an exclusion criterion within the ABCD study, but participants were retained in the ABCD study if parents/caregivers reported mild/moderate autism. Hence, it is possible that some participants with mild or moderate (parent-reported) autism spectrum disorder were retained in the ABCD subsample included in this study.

A note on the inclusion of participants over the age of 18 years: Although CD is most commonly diagnosed in childhood and adolescence, in the current study, we decided to include youths up to 21 years for the following reasons:

1) Consistency with prior ENIGMA research: Our study was designed to align with previous research conducted within the ENIGMA consortium, especially the work of Hoogman et al.[1,2] on attention-deficit/hyperactivity disorder (ADHD), which adopted an age cut-off of 21 between adolescence and adulthood; as well as the work by Schmaal et al.[3] on major depressive disorder (MDD), which also set the age cut-off for adolescent versus adult analyses at ⩽21 years. This consistent/standardized approach was intended to ensure the comparability of results across disorders in youth, especially between CD and ADHD.
2) Continuity of brain development: Extensive longitudinal neuroimaging studies have demonstrated that brain maturation extends beyond adolescence, with significant structural changes observed until around age 21-25 years.[4–6] This evidence supports the notion of emerging adulthood as a post-adolescent maturation stage, rather than a distinct phase after adolescence. Given this continuum of brain development and lack of biological justification to exclude individuals aged ≥ 18 from youth samples, we opted to include youths up to age 21 years.
3) Diagnostic criteria: According to the DSM-5,[7] an individual over 18 years old can still be diagnosed with CD if they do not meet the criteria for antisocial personality disorder. As there is no strong reason to exclude those aged above 18 from a psychiatric perspective, we opted to include them.

Based on the outlined reasons, we included youth up to age 21 years. However, to avoid shifting the mean age of the sample and maintain the focus on youth, we only included samples with a mean age of ≤ 18 years. Figure S1 below shows the age distribution of the main sample (pertaining to the CD versus TD comparison). Notably, findings remained unchanged when excluding participants over the age of 18 years (n = 25, see Appendix 12).

Figure S1: **Density plot of the distribution of age by diagnostic group**



TD=typically-developing; CD=conduct disorder.

Conduct disorder versus typically-developing youth

Participants allocated to the CD group had to have a current clinical or research diagnosis of CD. The former is provided in a healthcare setting by a medical professional, whereas the latter is primarily made for the purposes of a research study. The only exception to this was the cVEDA sample, where those with a lifetime diagnosis were included as this study did not differentiate between current and past diagnoses for all disorders (leave-one-out analyses [Appendix 12] showed that excluding this sample did not impact findings). Additional exclusion criteria for the TD group beyond those listed above included a current diagnosis of ADHD or oppositional defiant disorder (ODD) or lifetime histories of any disruptive behaviour disorders. For three of the 15 CD samples, it was not possible to fully verify the absence of these disorders in the control group as ADHD and/or ODD diagnoses were not provided. However, in each case the inclusion criteria indicated that controls were free of (current) psychiatric disorders and/or had no history of antisocial behaviours, indicating that these criteria were fulfilled by all samples. Notably, in addition to the CD vs. TD analyses, we also performed analyses comparing a non-overlapping group of youth with elevated conduct problems to controls. Details on this grouping approach and the analyses can be found in Appendix 13.

Subgroup comparisons

Age-of-Onset. In line with DSM-5-TR criteria,[7] participants with CD who displayed CD symptoms before age ten were classified as having childhood-onset CD, whereas those with a later onset were classified as having adolescent-onset CD. Overall, information on age-of-onset was available for 741 CD participants (62.5% of the CD group) from seven out of 15 cohorts. Based on the described approach, 458 (61.8%) were allocated to the childhood-onset group (133 girls, age range = 9-20 years, M_{age} = 11.75, *SD* = 3.13) and 283 (38.2%) were allocated to the adolescent-onset group (86 girls, age range = 7-21 *years*, M_{age} = 15.07, *SD* = 2.29). It should be noted that 56.1% of the childhood-onset participants were derived from the baseline release of the ABCD sample when participants were aged 9 or 10. TD youth from all cohorts were included in the control group (*n* = 1,253; 446 girls, age range = 7-20 *years*, M_{age} = 13.38, *SD* = 3.01).

Low vs. high callous-unemotional (CU) traits: Youth with CD were divided into those high and low in CU traits based on the recently developed cut-offs by Kemp et al.[8] for the total scores of the Inventory of Callous-Unemotional Traits (ICU).[9] In contrast to the median split approach, which varies between studies, the use of normative cut-offs may facilitate replication in future analyses. In the current analyses, we applied the version-, gender-, and age-specific normative cut-offs, which have been shown to reduce the chance of false positives.[8] For

the self-report version of the ICU, male youth with CD were allocated to the high-CU group if they scored ≥ 34 (up to age 14) or ≥ 37 (age 15 or older), while female youth with CD were allocated to the high-CU group if they scored ≥ 29 (up to age 14) or ≥ 32 (age 15 or older). For the parent-report version of the ICU, male youth with CD were allocated to the high-CU group if they scored ≥ 34 (any age), while female youth with CD were allocated to the high-CU group if they scored ≥ 30 (regardless of age). If participants had both self- and parent-report data available, they were included in the high-CU group if they were above the respective cut-off on either version. Overall, information on CU traits based on the ICU was available for 647 CD participants (54.6% of the whole CD group) from 9 out of 15 cohorts. Based on the approach described above, 277 (42.8%) were allocated to the low-CU group (67 girls, M_{age} = 15.28, *SD* = 2.32, range = 8-20) and 370 (57.2%) were allocated to the high-CU group (140 girls, M_{age} = 15.08, *SD* = 2.07, range = 7-19). Table S2 shows which informant was available for each sample. As for the age-of-onset subgroup analyses, TD youth from all cohorts were included in the control group.

The below table depicts the overlap between the age-of-onset and CU traits subgroups.

Table S1: **Overlap of the age-of-onset and callous-unemotional traits subgroups**

	CO-CD		**AO-CD**	
total	458	100.0%	283	100.0%
low-CU	61	13.3%	73	25.8%
high-CU	111	24.2%	88	31.1%
no CU information	286	62.4%	122	43.1%
	low-CU		**high-CU**	
total	277	100.0%	375	100.0%
CO-CD	61	22.0%	111	29.6%
AO-CD	73	26.4%	88	23.5%
no onset information	143	51.6%	176	46.9%

CO-CD=childhood-onset conduct disorder. AO-CD=adolescent-onset conduct disorder. Low-CU=conduct disorder with low levels of callous-unemotional traits

Appendix 2 – Study protocols of contributing sites and inclusion flowcharts

The current study includes data from 20 cohorts who contributed to the ENIGMA-Antisocial Behavior (ASB) working group. This comprised 15 samples that contributed data on youth with a diagnosis of CD (and TD youth) as reported in the main manuscript and 11 samples that contributed data on youth with elevated conduct problems (CP; and control youth). Six cohorts contributed participants to both the CD and CP analyses (ABCD, cVEDA, FemNAT-CD, IMAGEN, MATRICS/Aggressotype, and Yale). However, the CD and CP groups did not have any overlapping participants (full information on the CP analyses and the respective grouping strategy can be found in Appendix 13). Samples were included if they comprised at least ten youth with a diagnosis of CD or elevated CP as well as ten TD youth. The working group has a rolling inclusion design, in which new groups can join at any time, but data freezes allow for analyses at fixed time points. The data-freeze for the current study was set to 31/05/2022. Below we provide information on the study protocols of the included cohorts and inclusion flowcharts, which outline how the final sample size of the contributed cohorts was achieved.

For some cohorts, particularly the included community-based and population-based cohorts, a substantially larger number of controls compared to cases were available. In these instances, to keep group sizes roughly balanced, we used propensity score matching as implemented in the R package 'MatchIt' to match each case to a control with a similar covariate profile. This is achieved by using a 'propensity score', which reflects each participant's probability of belonging to the 'treatment' group (here: CD/CP group) based on a set of covariates.[10] When matching on propensity scores, the aim is to achieve a small difference between the propensity score of the CD/CP participant and the score of the matched control participants. In the current study, control youth were always matched on sex and age, and additionally on IQ and scanner if available/applicable (please see the specific

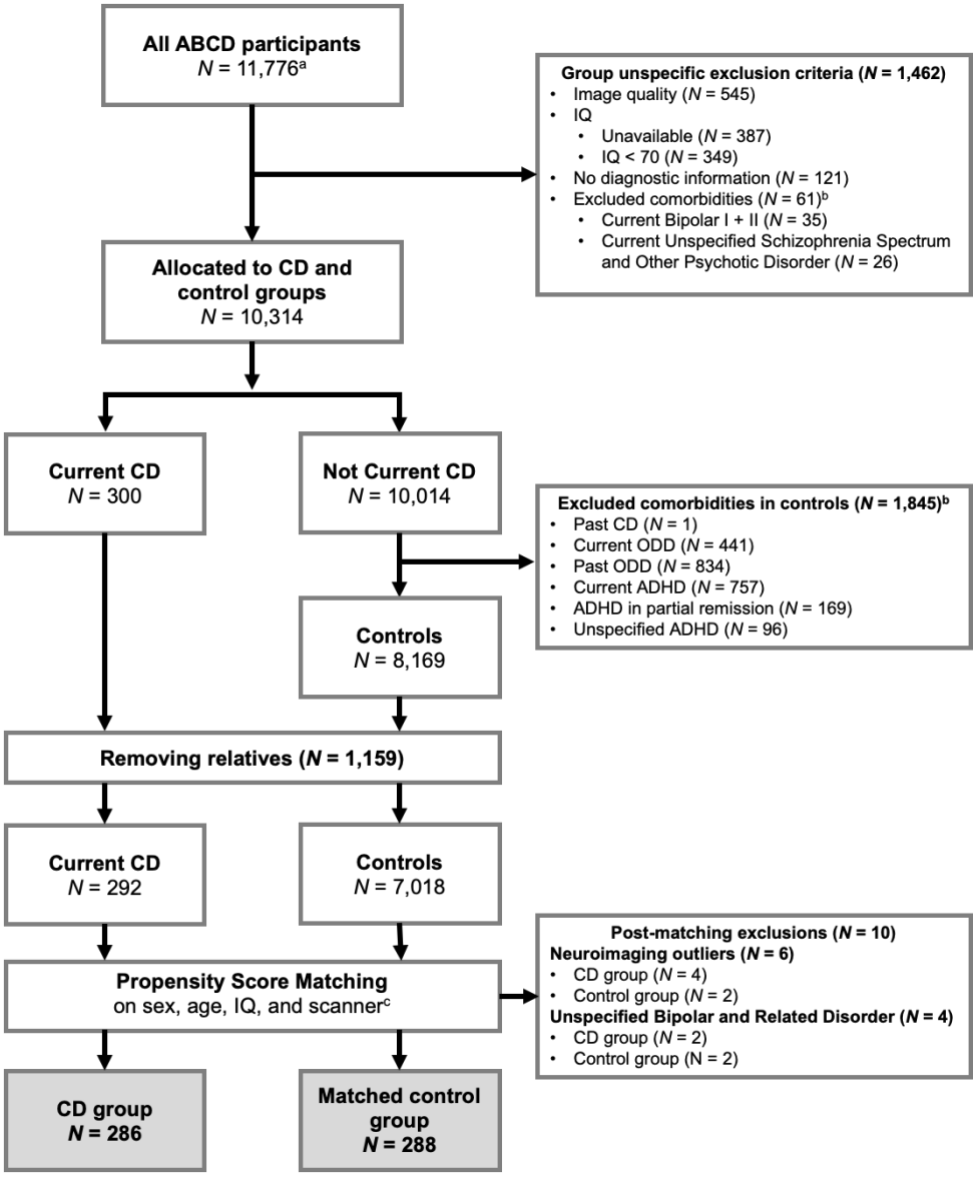
inclusion flowcharts below). We followed a 1:1 matching approach (one control participant is matched to each case), using nearest neighbour matching for continuous variables (age, IQ) and exact matching for categorical variables (sex, scanner). However, depending on the sample composition, exact matching on categorical variables was not always successful, in which case we also applied nearest neighbour matching to those variables. Nearest neighbour matching is a commonly applied approach and matches the closest eligible control unit without reference to the global balance of matches.[11]

ABCD

The Adolescent Brain Cognitive Development (ABCD) Study is a prospective longitudinal cohort study of US children born between 2006 and 2008. A total cohort of *N*=11,880 children aged 9–10 years at baseline (and their parents/guardians) was recruited from 22 sites (with one site no longer active) and are being followed for at least ten years. Eligible children were recruited from the household populations in defined catchment areas for each of the study sites during the roughly two-year period beginning in September 2016 and ending in October 2018. The ABCD Study® collects observational data to characterise US population trait distributions and to assess how biological, psychological, and environmental factors (including interpersonal, institutional, cultural, and physical environments) can relate to how individuals live and develop in today's society. From the outset, the NIH and ABCD scientific investigators were motivated to develop a baseline sample that reflected the sociodemographic variation present in the US population of 9–10 year-old children, and to follow them longitudinally through adolescence and into early adulthood. Institutional review boards at participating universities approved all study procedures. Participants provided written assent, and their legal guardians' written consent, for participation.

The data included in the current project were from the ABCD baseline wave derived from the 3.0 data release. More information on the sample can be found on the ABCD website (https://abcdstudy.org/). The information provided above was taken from Dick et al.[12] and Saragosa-Harris et al.[13]

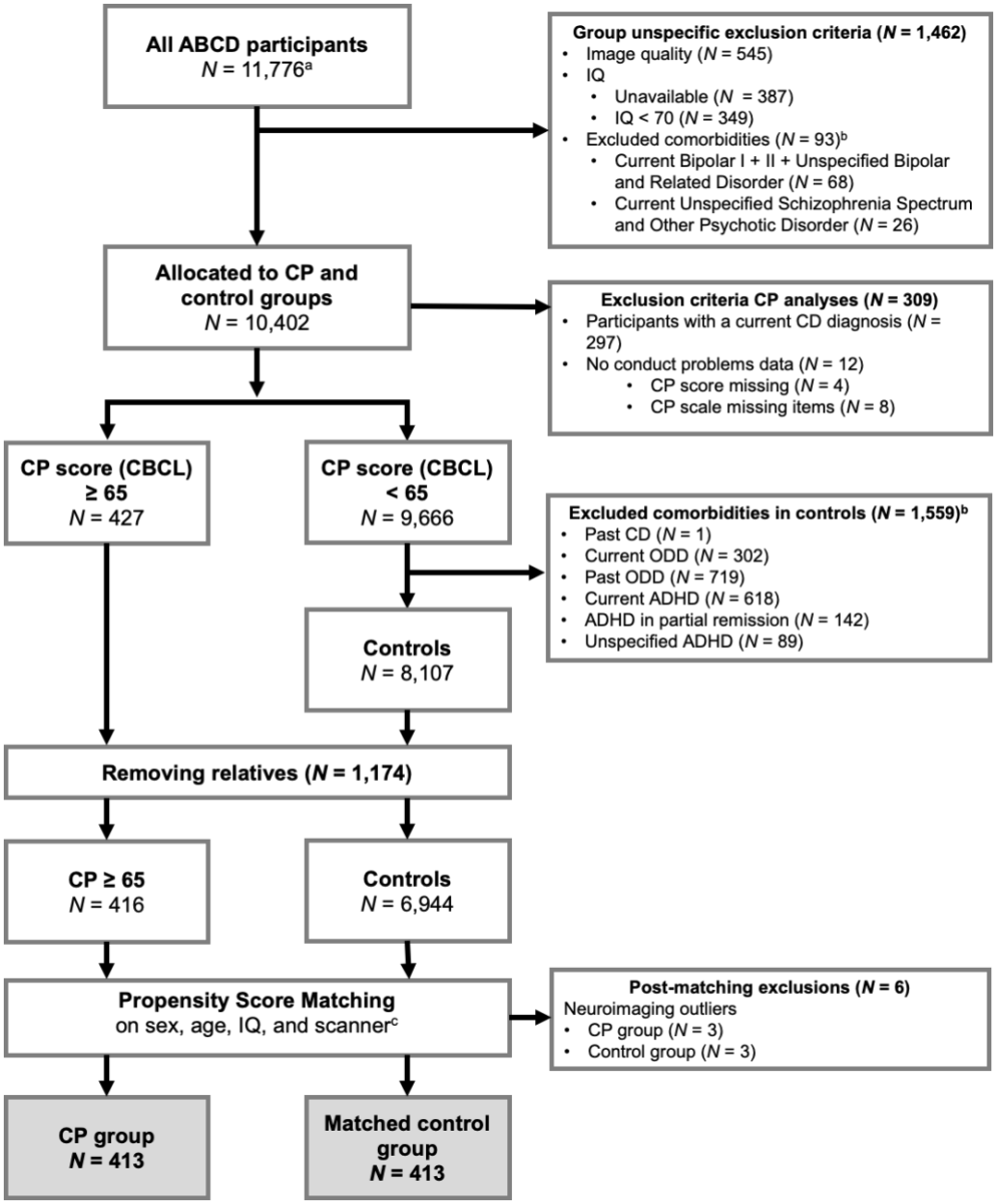
Figure S2: **ABCD inclusion flowchart for the conduct disorder analyses**



Diagnostic information was based on the K-SADS-COMP parent-report for consistency as the youth-report section was only available for a subset of modules (7/25) and not completed for conduct disorder (see Table S2 in Barch et al.[14]). Image quality assessments were based on the imaging inclusion flag provided by ABCD (see Appendix 3). The removal of relatives was performed as follows: If no participant in the family group had a diagnosis of conduct disorder, one sibling was removed at random. For the remaining relatives including youth with conduct disorder, the following was done: 1) If only one of the family group had conduct disorder, they were included and the other sibling was excluded. 2) If both siblings had conduct disorder, the case with the highest number of current symptoms (followed by lifetime if current symptom count was equal) was included. *N*=sample size. CD=conduct disorder. IQ=intelligence quotient. ODD=oppositional defiant disorder. ADHD=attention-deficit/hyperactivity disorder.

[a] Available across all used spreadsheets. [b] Total numbers may differ from individual items as participants might have had multiple comorbidities. [c] Missing scanner information was substituted with the scanner that was used most frequently at the site of the participant for whom scanner information was missing.

Figure S3: **ABCD inclusion flowchart for the conduct problems analyses**



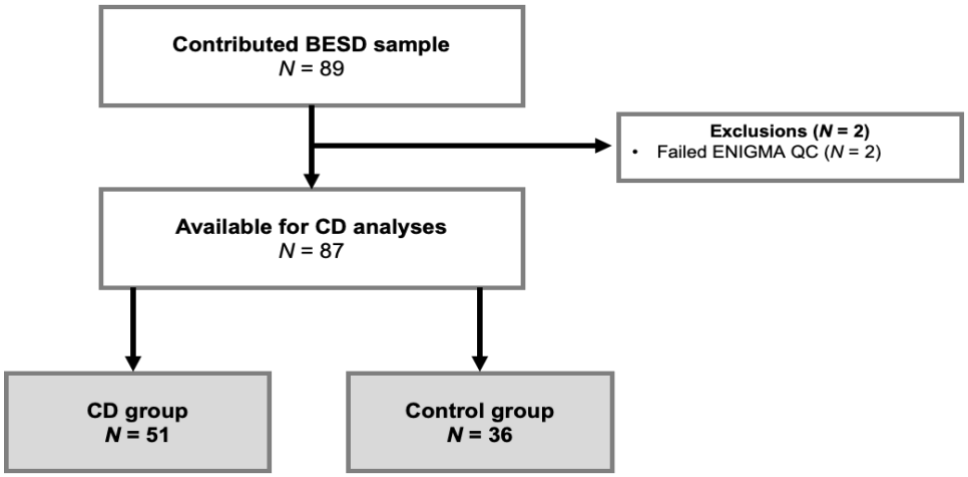
Diagnostic information was based on the K-SADS-COMP parent-report for consistency as the youth-report section was only available for a subset of modules (7/25) and not completed for conduct disorder (see Table S2 in Barch et al.[14]). Image quality assessments were based on the imaging inclusion flag provided by ABCD (see Appendix 3). The removal of relatives was performed as follows: If no participant in the family group exhibited elevated conduct problems, one sibling was removed at random. For the remaining relatives including youth with elevated conduct problems, the following was done: 1) If only one of the family group had elevated conduct problems, they were included and the other sibling was excluded. 2) If both siblings had elevated conduct problems, the case with the lower conduct problems T-score was excluded. In the case of one sibling pair, both had the same conduct problems score. The participant with the higher IQ score was retained. *N*=sample size. CP=conduct problems (based on the DSM-oriented conduct problems subscale of the Child Behavior Checklist). CD=conduct disorder. IQ=intelligence quotient. CBCL=Child Behavior Checklist. ODD=oppositional defiant disorder. ADHD=attention-deficit/hyperactivity disorder.

[a] Available across all used spreadsheets. [b] Total numbers may differ from individual items as participants might have had multiple comorbidities. [c] Missing scanner information was substituted with the scanner that was used most frequently at the site of the participant for whom scanner information was missing.

BESD

Participants were part of a larger study on the effects of juvenile antisocial, psychopathic, and autistic tendencies on socioemotional brain systems. All participants were aged 15 to 19 years old and were medication-free. Participants were scanned between February 2013 and November 2014. Juvenile offenders with CD were recruited from a juvenile detention centre and a forensic psychiatric facility, and had all been convicted for or charged with crimes such as assault, murder, and armed robbery. Healthy controls (HC) were carefully recruited through local advertisement. Inclusion criteria for the CD group were: being admitted to a forensic facility due to contact with the juvenile justice system, and having a DSM(-IV) diagnosis of CD with at least one aggressive symptom (e.g., used a weapon, has been physically cruel to people, has stolen while confronting a victim). Inclusion criteria for the HC group were: no current or past DSM-IV diagnoses of internalizing or externalizing disorders, no clinical scores on validated mood and behavioural questionnaires, and no symptoms of severe antisocial behaviour. Exclusion criteria for all participants were: primary DSM-IV diagnosis of pervasive developmental disorders, Tourette's syndrome, obsessive-compulsive disorder, bipolar disorder, and psychotic disorders, use of psychotropic medication, a history of neurological disorders, age <15 or >19 years, left-handedness, IQ score <75, as measured by either the Wechsler Intelligence Scale for Children (WISC) or the Wechsler Adult Intelligence Scale (WAIS), and general MRI contraindications (e.g., metal implants, claustrophobia). Finally, all participants were explicitly told that the slightest hint of drug intoxication on the day of scanning would result in exclusion, with juvenile offenders also being subjected to random and unannounced urine drug tests at the forensic facilities.

Figure S4: **BESD inclusion flowchart for the conduct disorder analyses**



N=sample size. CD=conduct disorder. QC=quality control.

Boys Town

The goals of this project are to determine: (i) the symptom sets associated with dysfunction in targeted neuro-cognitive systems; (ii) the neuro-cognitive variables that predict success of the Boys Town Intervention Model; and (iii) the extent to which Program response is associated with "normalisation" of pathophysiology. The project will involve both fMRI and neuropsychological assessment. This study will involve participants from the Boys Town campus as well as participants from the community. Participants will be categorised as: (i) typically-developing (TD; i.e., presenting with levels of symptoms below the clinical range as indexed by the CBCL or SDQ [see below]); (ii) clinically concerning (CC; i.e., presenting with clinically significant levels of psychiatric or behavioural symptoms for conditions as indexed by the CBCL or SDQ [see below] and/or sufficient behavioural problems to warrant residential treatment). Thus, all participants from the Boys Town campus will be considered CC participants. Participants from the community are invited to participate in six visits (one screening visit, four fMRI, and one neuropsychological assessment). Participants from the Boys Town campus are invited to participate in 12 visits (one screening visit, four fMRI, and one neuropsychological assessment as soon as possible after arrival at Boys Town and four fMRI and one neuropsychological assessment and one symptom set assessment within the three months prior to departure from Boys Town).

Inclusion criteria for TD youth:

1. 10-18 years of age
2. No current psychiatric diagnoses

Inclusion criteria for CC youth:

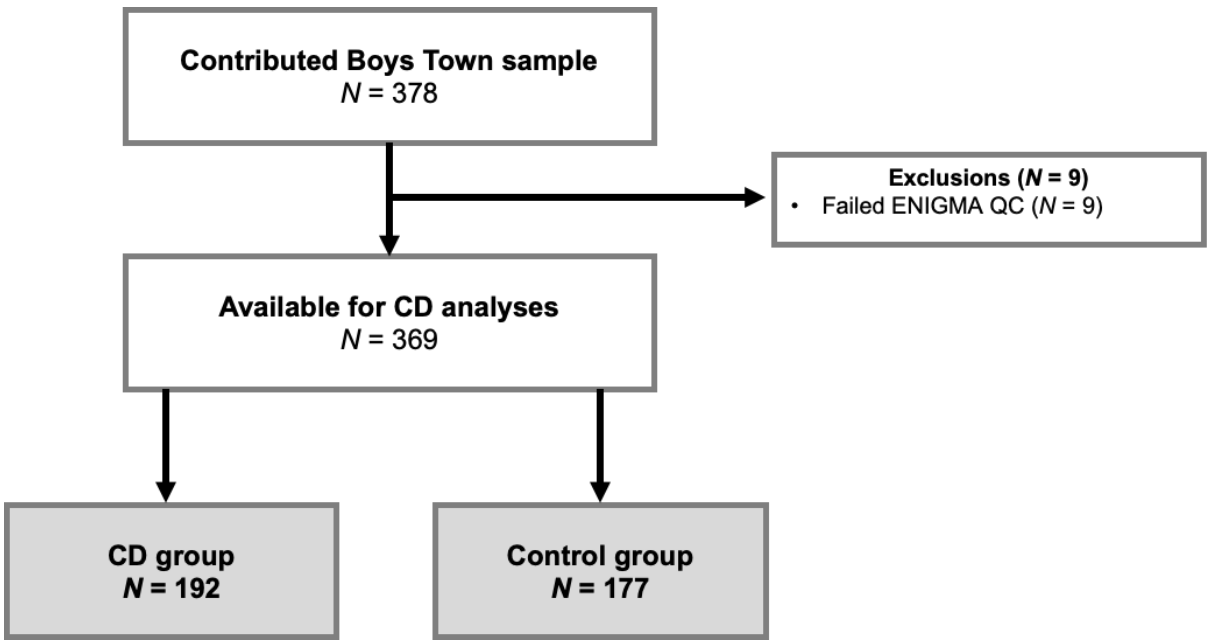
1. 10-18 years of age
2. Clinically significant levels of psychopathology as measured by the CBCL or SDQ at screening or residents at Boys Town Family Home Program or Psychiatric Residential Treatment Center

Exclusion criteria

1. IQ<75
2. Pregnancy
3. Ongoing medical illness other than those listed in the inclusion criteria for the respective groups that require use of any medication that may have psychotropic effects such as beta blockers or steroids:
    - Medications provided for psychiatric illness (specifically mood stabilising medications) will NOT be exclusory
    - Non-exclusionary medications include methylphenidate, lisdextroamphetamine, amphetamine, dexmethylphenidate hydrochloride, atomoxetine, bupropion, modafinil, or valproic acid
4. Explicit exclusions include active psychosis, Pervasive Developmental Disorders and Tourette's syndrome
5. Neurologic disorder (including seizures).
6. Any metallic objects in the body. Metal plates, certain types of dental braces, cardiac pacemakers, ext., that are sensitive to electromagnetic fields contraindicate MRI scans
7. Claustrophobia

The data contributed to the ENIGMA-ASB working group included a subgroup of participants with a diagnosis of CD and control participants.

Figure S5: **Boys Town inclusion flowchart for the conduct disorder analyses**



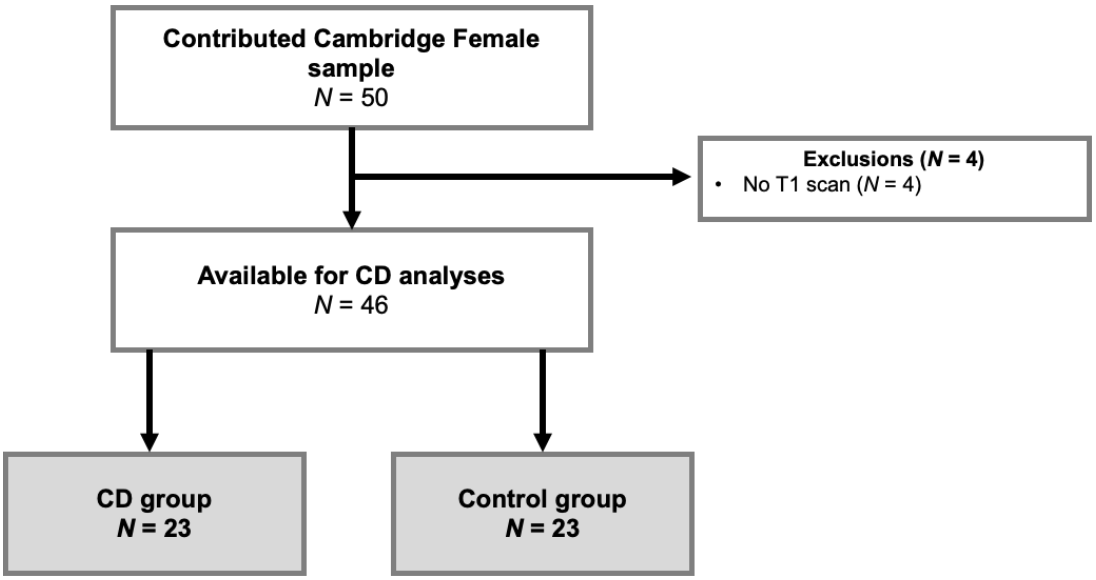
N=sample size. CD=conduct disorder. QC=quality control.

Cambridge Female

Twenty-two female adolescents with CD aged 14–20 years were recruited from schools, pupil referral units and the Cambridge Youth Offending Service. A healthy control group (HC; no history of CD/ODD and no current psychiatric illness) of 21 female adolescents, matched for age, handedness and performance IQ, was recruited from schools. All participants and their parents gave written informed consent to participate in the study, which was approved by the Suffolk NHS Research Ethics Committee. Exclusion criteria included full-scale IQ (FSIQ) <80, as estimated using the Wechsler Abbreviated Scale of Intelligence, and presence of pervasive developmental disorder (e.g. autism).

Please note that the sample contributed to the ENIGMA-ASB working group was slightly larger than that included in the original papers (see flowchart below).

Figure S6: **Cambridge Female inclusion flowchart for the conduct disorder analyses**

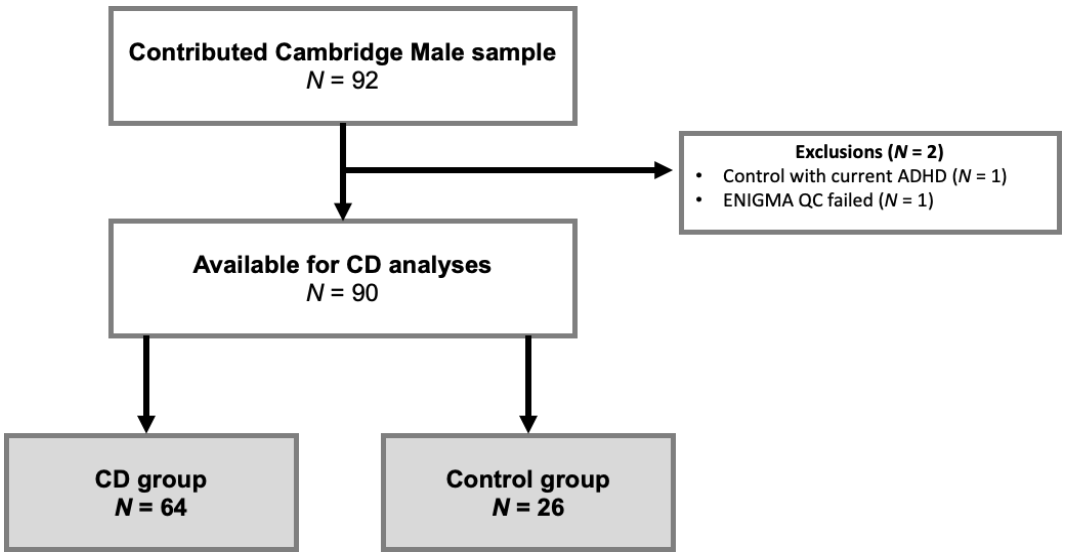


N=sample size. CD=conduct disorder.

Cambridge Male

Sixty-five male adolescents with conduct disorder were recruited from schools, pupil referral units, and the Cambridge Youth Offending Service, Cambridge, United Kingdom. All participants gave written informed consent to participate in the study, which was approved by the local research ethics committee. Exclusion criteria for the conduct disorder group included an IQ <75, as estimated using the Wechsler Abbreviated Scale of Intelligence, or presence of a pervasive developmental disorder (e.g., autism). A healthy comparison group (no history of conduct disorder/oppositional defiant disorder and no current psychiatric illness) of 27 male adolescents, matched for performance IQ, was recruited from schools and colleges.

Figure S7: **Cambridge Male inclusion flowchart for the conduct disorder analyses**



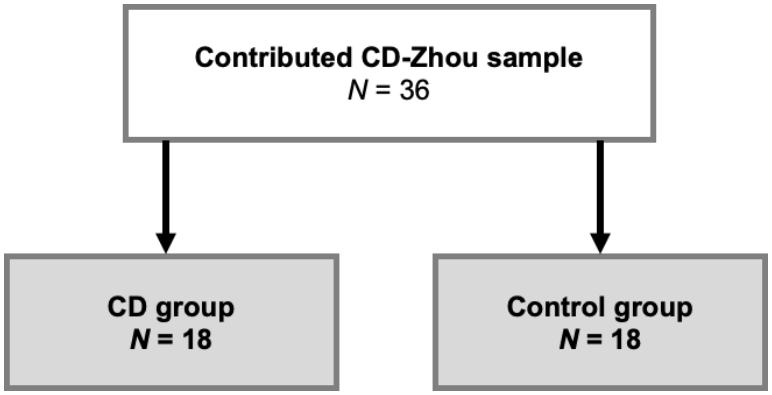
N=sample size. CD=conduct disorder. ADHD=attention-deficit/hyperactivity disorder. QC=quality control.

CD-Zhou

Eighteen boys with CD and 18 age- and education-matched typically-developing (TD) boys from a previous sample were recruited.[15] All subjects were aged between 15 and 17 years, and were right-handed with normal vision. The CD group was recruited from the Hunan province Youth Detention Centre, whereas the TD group was recruited from schools in the local community. Informed consent was provided by all participants and their parents or guardians. The study was approved by the Biomedical Ethics Board of the Second Xiangya Hospital, Central South University, People's Republic of China.

The exclusion criteria were any history of neurological disorders, including: paralysis, loss of sensation, epilepsy, muscular weakness, seizures, chronic pain, confusion, and altered levels of consciousness. In addition, any participants meeting the K-SADS-PL criteria for any other current or lifetime psychiatric disorders, or has a subthreshold or threshold levels of other disorders of symptomatology such as attention-deficit/hyperactivity disorder (ADHD), oppositional defiant disorder (ODD), mood disorder, anxiety disorder, mental retardation and substance abuse or dependence were also excluded.

Figure S8: **CD-Zhou inclusion flowchart for the conduct disorder analyses**

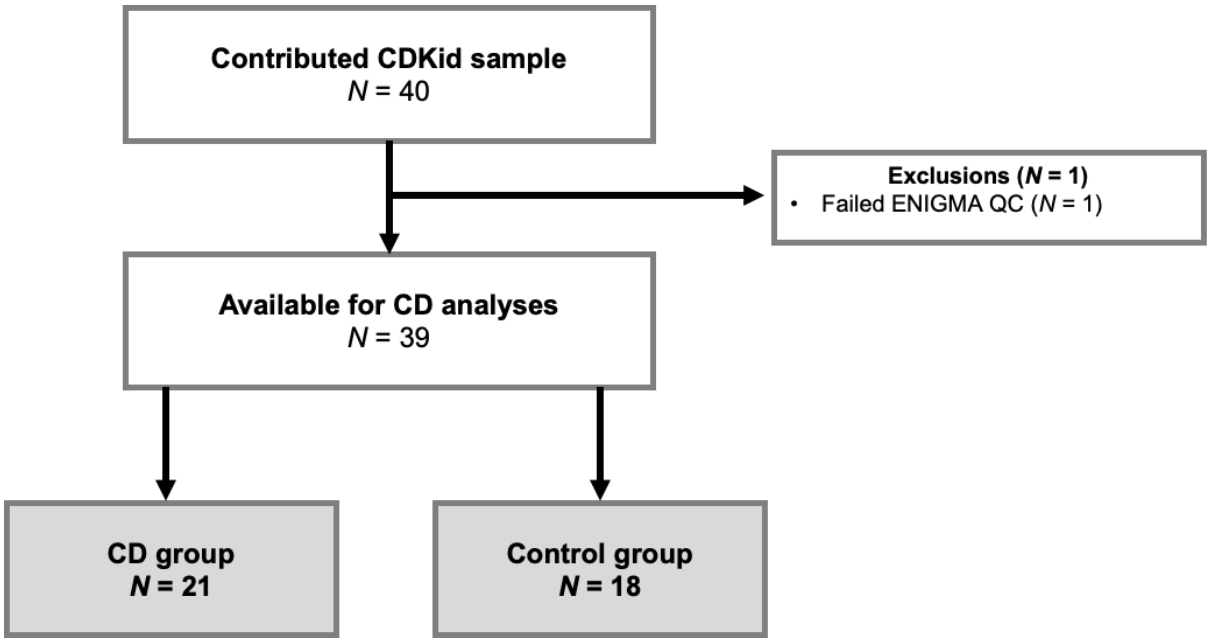


N=sample size. CD=conduct disorder.

CDKid

Twenty-one males with CD aged between 12 and 19 years were recruited as part of a larger study (see Sarkar et al.[16]) from: (1) an Institute of Psychiatry (King's College London) database of adolescents with conduct problems; (2) three Youth Offending Teams; (3) five Pupil Referral Units (facilities providing education to children who cannot attend mainstream schools, e.g., following school exclusion); (4) four youth projects; and (5) two mainstream educational institutions. A further 19 right-handed males were recruited as controls from the general public, through schools and youth services (i.e., youth clubs, 'Connexions', and several youth charities) within the same geographical areas (deprived and inner city) as the CD group. Groups did not significantly differ in age, handedness, ethnicity, and self-reported history of alcohol or cannabis use. Measures of current hyperactivity, and rates of attention-deficit/hyperactivity disorder (ADHD) did not differ significantly between groups. All study participants satisfied MRI safety requirements and were medication-free, did not have a psychiatric or substance use disorder (other than CD, ADHD, or referrals for anger management), spoke English as their first language, and were right-handed as assessed by the Edinburgh Handedness Inventory. IQ was measured using the vocabulary and matrix reasoning subtests of the Wechsler Abbreviated Scale of Intelligence.

Figure S9: **CDKid inclusion flowchart for the conduct disorder analyses**



N=sample size. CD=conduct disorder. QC=quality control.

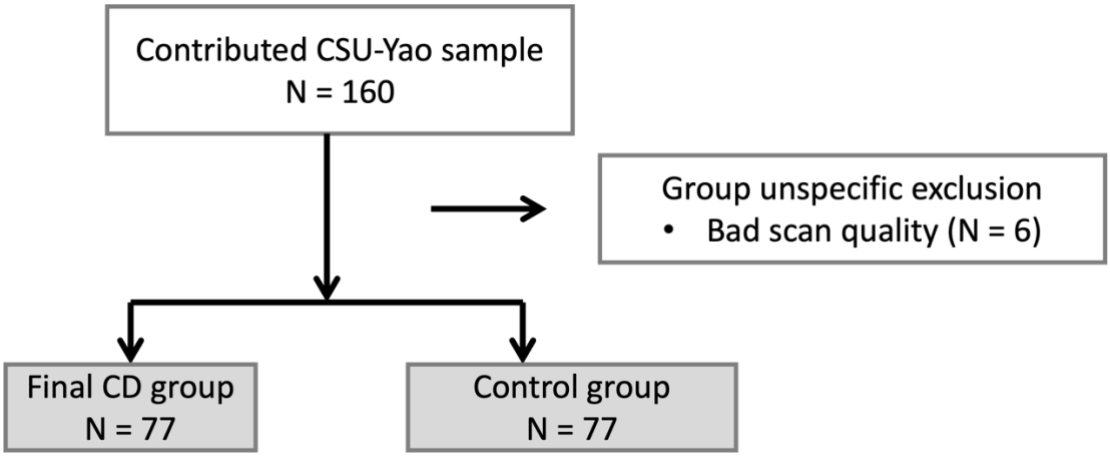
CSU-Yao

Youths with CD (aged 12–17 years old) were recruited from outpatient clinics affiliated with the Second Xiangya Hospital of Central South University in Changsha, Hunan, China. A group of age- and gender- matched typically-developing (TD) volunteers was recruited from regular secondary schools. All participants and their parents were made aware of the purpose of the study and gave written informed consent for participation. Diagnoses of CD were made based on the Structured Clinical Interview for DSM-IV-TR Axis I Disorders-Patient Edition (SCIDI/P) by two well-trained psychiatrists. To improve the reliability of the diagnostic interview, information was collected from each participant and at least one corresponding parent. The TD group was subjected to the SCIDI/P by the same group of psychiatrists that screened the CD group. None of the TD participants met the criteria for current or a history of CD.

The exclusion criteria of both groups were: current or prior history of any (other) psychiatric, behaviour or emotional disorder (including, among others, post-traumatic stress disorder and obsessive–compulsive disorder); a pervasive developmental disorder (i.e., autism); a chronic neurological disorder; Tourette's syndrome; persistent headaches; a history of head trauma; alcohol or substance abuse within the past year; any MRI contraindication;

or an IQ <80 on the C-WISC. Participants were required to be right-handed, according to the Edinburgh Handedness Inventory.

Figure S10: **CSU-Yao inclusion flowchart for the conduct disorder analyses**



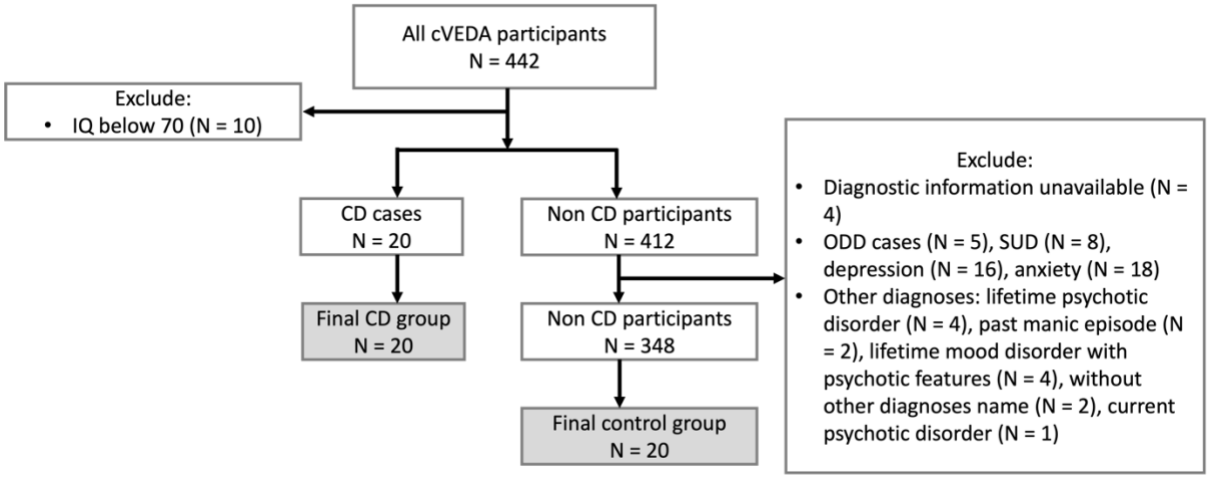
N=sample size. CD=conduct disorder.

cVEDA consortium

cVEDA, a large neurodevelopmental cohort in India, was set up to enable the long-term study of the developmental trajectories of brain and psychological functioning, and the effects of exposome and genome in modulating these trajectories, to influence vulnerability to psychiatric disorders. Participants were recruited from seven sites (Imphal (Manipur); Asansol (West Bengal); Mysore (Karnataka); National Institute of Mental Health and Neurosciences [NIMHANS] Bangalore (Karnataka); Post Graduate Institute of Medical Education and Research (PGIMER) (Chandigarh, Punjab and Haryana); Rishi Valley (Madanapelle, Andhra Pradesh); Saint John's Research institute [SJRI]) representing five different geographical locations in India, of all gender, ethnicity, socio-cultural strata and urban-rural living. The study involves a thorough assessment of behavioural, neuropsychological, clinical, and environmental exposures of each subject. Participants also undergo biological characterization with collection of T1 weighted MRI, diffusion tensor imaging, resting state functional MRI and genetics data including blood and urine for assay of neurotoxin exposures. cVEDA study was approved by the Health Ministry Screening Committee, Ministry of Health and Family Welfare, Government of India. The study protocol was approved by the Institutional Ethics Review Boards of National Institute of Mental Health and Neurosciences (NIMHANS) Bangalore, India (Item No. VII, SI. No. 7.08, Behavioural Sciences) and all other regional collaborating institutions. Written informed consent was obtained from participants over the age of 18 and from parents of participants under 18 years of age (along with assent from minor participants). The study was conducted in accordance with the Declaration of Helsinki (1964 and later versions)

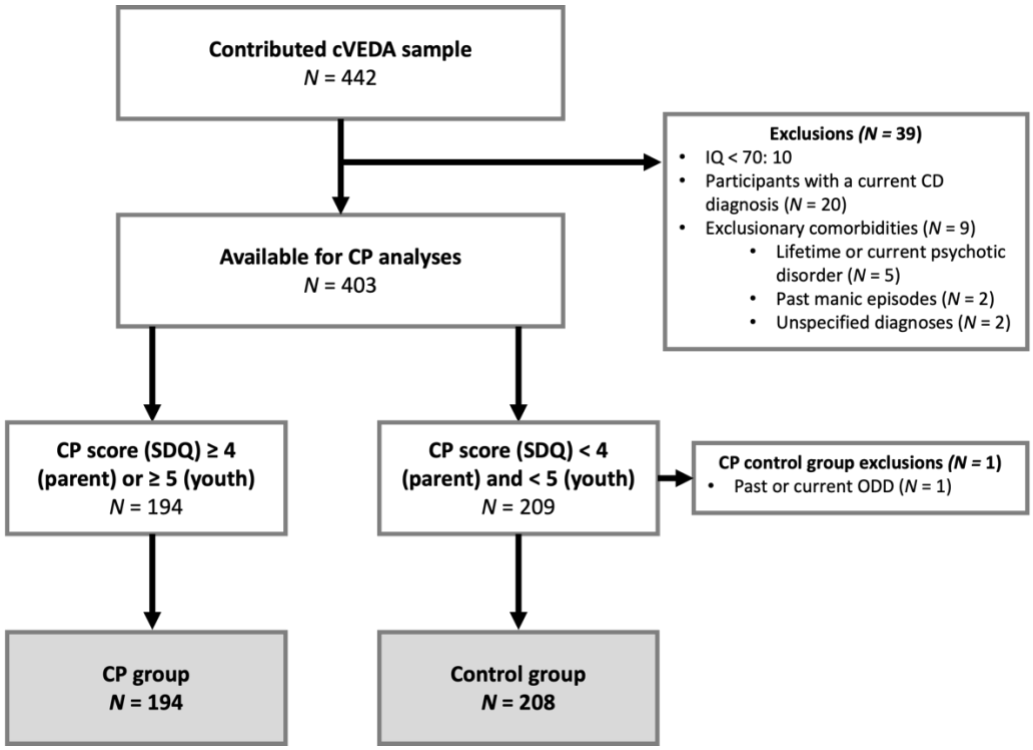
Please see the publications by Sharma et al.[17] and Zhang et al.[18] for further information.

Figure S11: **cVEDA inclusion flowchart for the conduct disorder analyses**



IQ=intelligence quotient. CD=conduct disorder. ODD=oppositional defiant disorder.

Figure S12: **cVEDA inclusion flowchart for the conduct problems analyses**



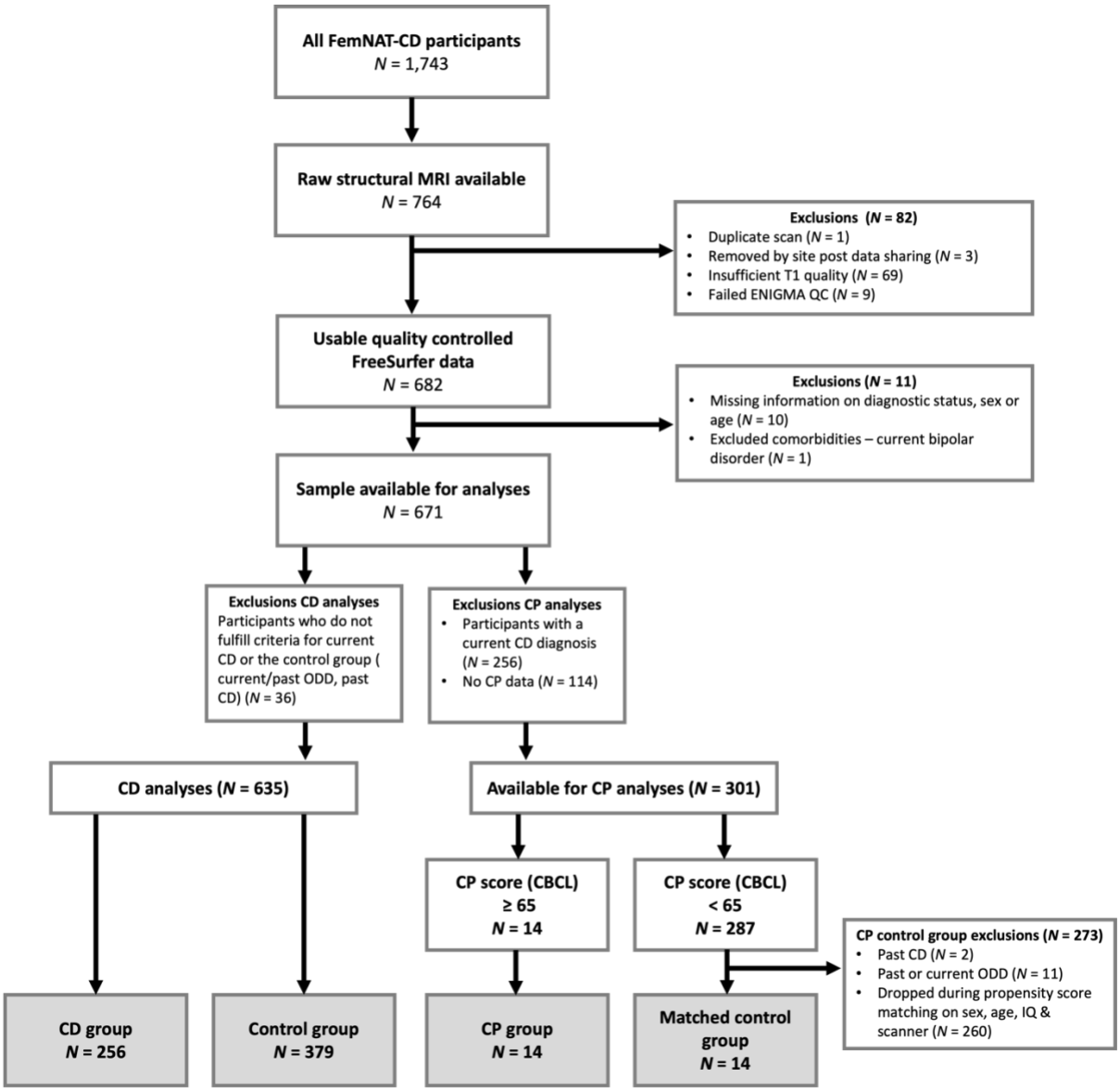
IQ=intelligence quotient. CD=conduct disorder. CP=conduct problems. SDQ=Strengths and Difficulties Questionnaire. ODD=Oppositional Defiant Disorder.

FemNAT-CD

Given the lack of studies on females with CD and the relative paucity of studies in adolescents compared to children, the aim of the European Commission-funded FemNAT-CD consortium (https://www.femnat-cd.eu/) is to study female adolescent CD using a multi-level approach including phenotypic, environmental, neurocognitive, endocrinological, psychophysiological, neuroimaging, genetic and epigenetic measures. The main objectives of the FemNAT-CD consortium are to: (1) clarify the phenomenology and neurobiology of adolescent female CD and investigate sex differences in neurocognitive and neurobiological mechanisms in CD; (2) translate knowledge of neuropsychological and neurobiological characteristics into targeted intervention; and (3) meet relevant societal and educational objectives, such as providing information, training and intervention to relevant stakeholders, and

especially young people with CD and their families. Girls were oversampled as a key aim of the consortium was to address the lack of data on female CD. Both community-based and clinically referred individuals were recruited through community outreach and from mental health clinics, welfare institutions and youth offending services. Youths with CD met diagnostic criteria for current CD according to DSM-IV-TR criteria. TD controls were free of any current psychiatric disorder (except specific learning disorders) and had no history of CD, ODD and ADHD. Exclusion criteria for both groups were IQ < 70 (based on estimates from two subtests of age-appropriate Wechsler scales), autism spectrum disorders, schizophrenia, bipolar disorder or mania, neurological disorders and genetic syndromes. Local ethics committees approved the study protocol. Written informed consent was obtained for all participants. Neuroimaging data was collected at five of the FemNAT-CD sites: Aachen (Germany), Basel (Switzerland), Birmingham (UK), Frankfurt (Germany), and Southampton (UK).

Figure S13: **FemNAT-CD inclusion flowchart for the conduct disorder and conduct problems analyses**

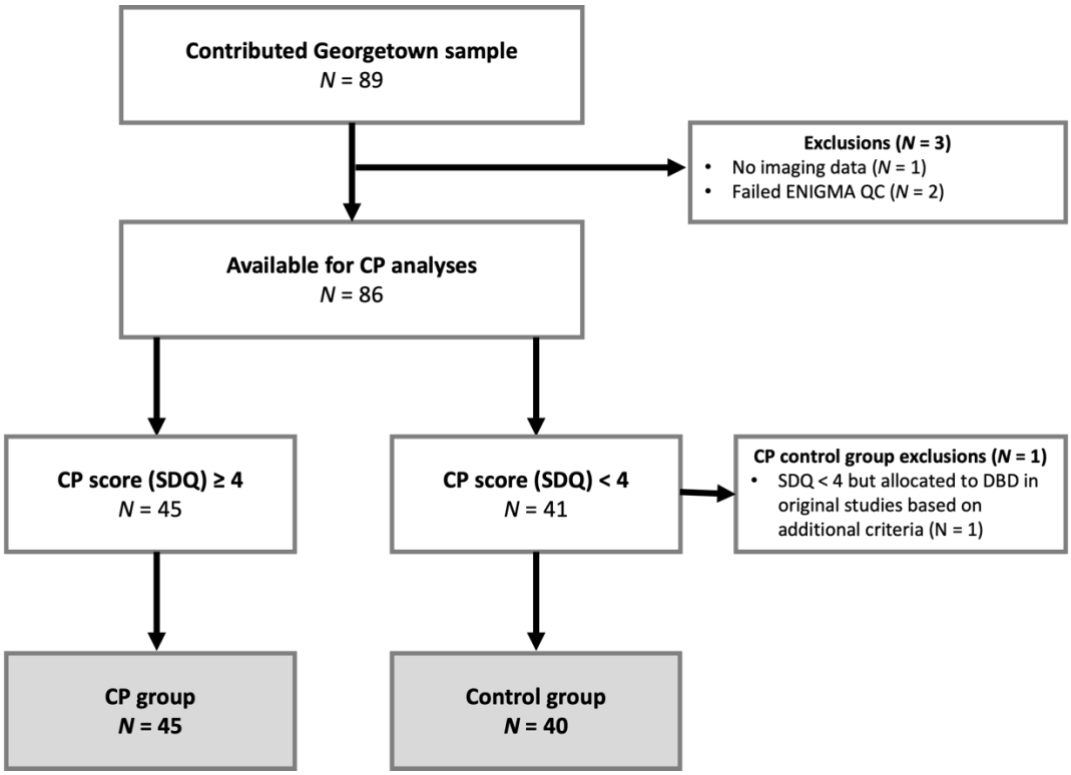


MRI=Magnetic Resonance Imaging. QC=quality control. CD=conduct disorder. CP=conduct problems (based on the DSM-oriented conduct problems subscale of the Child Behavior Checklist). CBCL=Child Behavior Checklist. ODD=oppositional defiant disorder. IQ=intelligence quotient.

Georgetown

One hundred forty-eight children, aged 9–18 (*M* = 13.96, *SD* = 2.44, % male = 59.46), were recruited from Washington, DC and surrounding regions through referrals, advertisements, and fliers seeking both healthy children and children with conduct problems. All participants and their parents first completed an initial visit during which demographic and clinical measures were completed along with IQ testing using the Kaufman Brief Intelligence Test. Participants reported a wide range of scores on our clinical measures, confirming that our sample included both healthy youth and youth with elevated conduct problems and varying CU traits, as well as psychiatric symptoms including externalising behaviours, internalising behaviours, and attentional difficulties. Consistent with our recruitment effort to specifically target both healthy children and children with elevated conduct problems, 77 participants reported clinical levels of externalising behaviour as assessed by an age and gender standardised externalising symptomology score on the Child Behavior Checklist (CBCL) that placed them above the 98th percentile. Of participants who completed the initial visit, 93 were eligible for and consented to participate in an MRI scan. Participants were excluded from MRI scanning for: history of head trauma or neurological disorder, symptoms of pervasive developmental disorder, IQ <80, or MRI contraindications such as claustrophobia or metallic implants including braces or permanent retainers. The MRI sample consisted of children aged 10–17 (M = 13.98, SD = 2.36, % male = 59.14) and varied widely in externalising behaviour, including 46 participants with clinically significant externalising scores. The MRI sample did not differ from the full sample in terms of externalising and CU scores or any other clinical or demographic measures, with the exception of a trend-level difference in age between the full sample and the scanned sample. All participants were native English speakers. Written informed assent and consent were obtained from children and parents before testing. Approval for all procedures was obtained from the Georgetown University Institutional Review Board. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Figure S14: **Georgetown inclusion flowchart for the conduct problems analyses**



QC=quality control. CP=conduct problems (based on the conduct problems subscale of the Strengths and Difficulties Questionnaire). SDQ=Strengths and Difficulties Questionnaire. Notably, the DSM-oriented conduct problems subscale of the Child Behavior Checklist was also available for this sample but only in the form of raw scores to which the cut-offs cannot be applied.

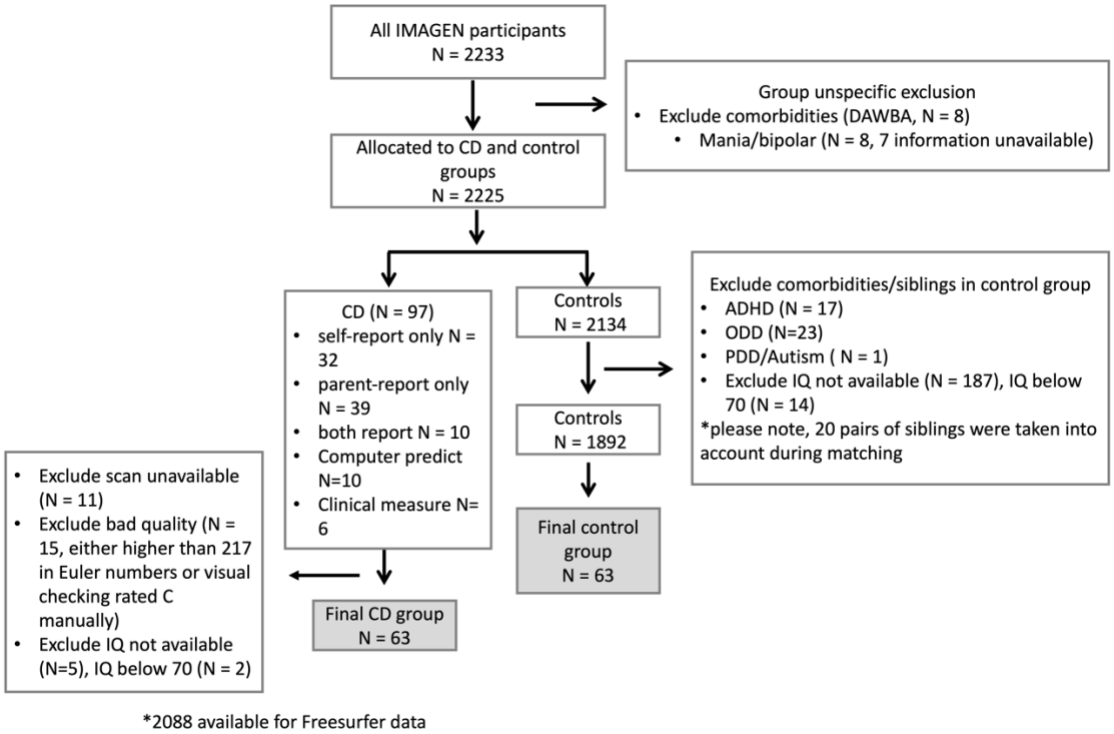
IMAGEN consortium

The European multi-centre genetic-neuroimaging IMAGEN study (https://imagen-europe.com) is a community cohort of more than 2000 participants. Adolescents were recruited from schools around the age of 14 years (age range = 12.9–15.7 years). Recruitment into the IMAGEN study targeted adolescents for whom all four grandparents were the same nationality as the participant (i.e., participants were required to have four Western European grandparents); as such, the sample is racially and ethnically homogenous. Local ethics research committees approved the study at each site. Written consent was obtained from the parent or guardian as well as verbal assent from the adolescent. A detailed description of recruitment and assessment procedures has been published elsewhere.[19] Participants completed neuropsychological, clinical, and personality assessments online and at assessments sites. Structural and fMRI is performed on 3T scanners from a range of manufacturers (Siemens, Munich, Germany; Philips, Best, The Netherlands; General Electrics, Chalfont St Giles, UK; Bruker, Ettlingen, Germany).

The IMAGEN project data included in the current study were from the IMAGEN baseline wave (T1). Please see Schumann et al.[19] for more information on the IMAGEN sample. Diagnostic information was based on the Development and Well-Being Assessment (DAWBA, https://www.dawba.info/), which is a package of interviews, questionnaires and rating techniques designed to generate ICD-10 and DSM-IV or DSM-5 psychiatric diagnoses in 2-17 years old. The participants were excluded from the final analyses if they met the following criteria: (1) absence of age, sex or IQ information, (2) IQ below 70, (3) diagnosed with bipolar disorder or lacking relevant information, (4) absence of T1 scans or failure to pass the T1 scan quality check (see Appendix 3 for details on the quality assessments). The participants meeting the CD diagnostic criteria based on the DAWBA, either by clinical rating or computer prediction, from self- or parent-report, were assigned to the CD group, resulting in 63 participants. Accordingly, 63 age-, sex- and IQ-matched TD participants were included in the final analysis. No siblings from the same family are included in the analysis.

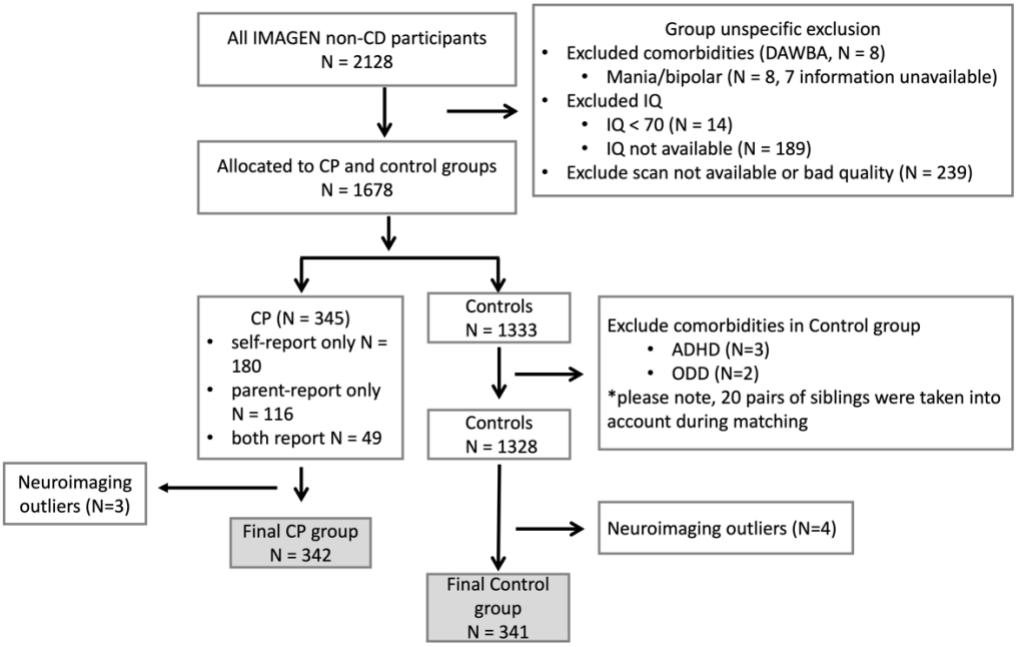
For the elevated CP vs control analysis, after excluding the participants diagnosed with CD, a similar inclusion approach was adopted. Participants who scored above 4 on the parent-report or above 5 on the self-report conduct problems subscale of the Strengths and Difficulties Questionnaire (SDQ) were included in the CP group, resulting in 342 participants. Accordingly, 341 age-, sex- and IQ-matched controls were included in the final analysis. No siblings from the same family are included in the analysis.

Figure S15: **IMAGEN consortium inclusion flowchart for the conduct disorder analyses**



DAWBA=Development and Well-Being Assessment. IQ=intelligence quotient. CD=conduct disorder. ODD=Oppositional Defiant Disorder. ADHD=attention-deficit/hyperactivity disorder. PDD=pervasive developmental disorders.

Figure S16: **IMAGEN consortium inclusion flowchart for the conduct problems analyses**



DAWBA=Development and Well-Being Assessment. IQ=intelligence quotient. CD=conduct disorder. CP=conduct problems. ODD=oppositional defiant disorder. ADHD=attention-deficit/hyperactivity disorder.

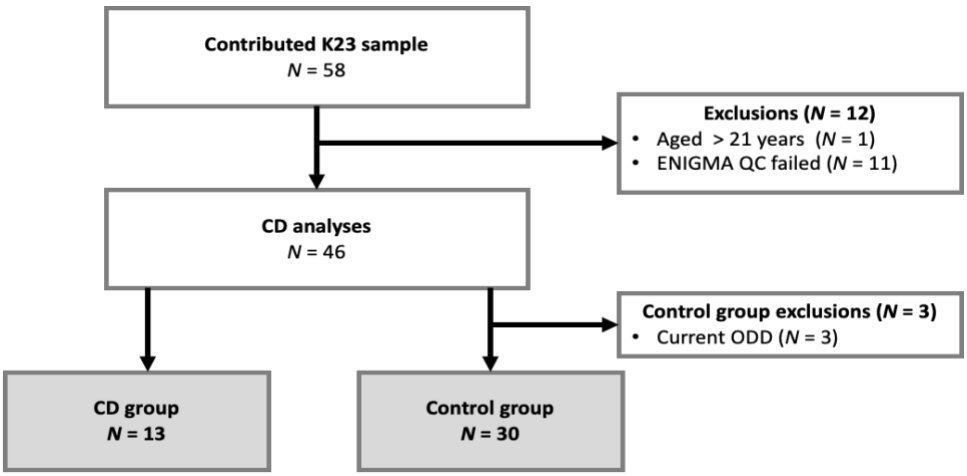
K23 (Stevens)

The CD and ADHD study groups each comprised right-handed, medically healthy adolescents (ages 12–18) recruited for a National Institute of Mental Health (NIMH)-funded project (K23 MH070036) aimed at differentiating the neural correlates of each diagnosis. Participants were recruited using a combination of community advertisements and letters sent to families of youth on probation following arrest in the Connecticut Court Support Services Division. Groups were matched by sex and mean age to control participants without diagnosable DSM-IV disorders or health problems. We obtained parental permission and informed consent to participate in the study jointly from the participants and their parent/legal guardian. All study procedures were approved by the Hartford Hospital Institutional Review Board.

Clinical diagnoses for research purposes were made using the Kiddie-Schedule for Affective Disorder and Schizophrenia — Present and Lifetime Version (K-SADS-PL) conducted by trained bachelor's- and master's-level staff working under the supervision of a licensed clinical psychologist. Interviews were performed separately for both adolescents and parents. Information was synthesized and diagnoses confirmed in weekly research group meetings. By design, we excluded youth with comorbid disorders. All ADHD participants met criteria for DSM-IV combined-subtype.

The above information was taken from Stevens & Haney-Caron.[20] Notably the sample contributed to the ENIGMA-ASB working group differed slightly from the one included in the cited publication.

Figure S17: **K23 inclusion flowchart for the conduct disorder analyses**

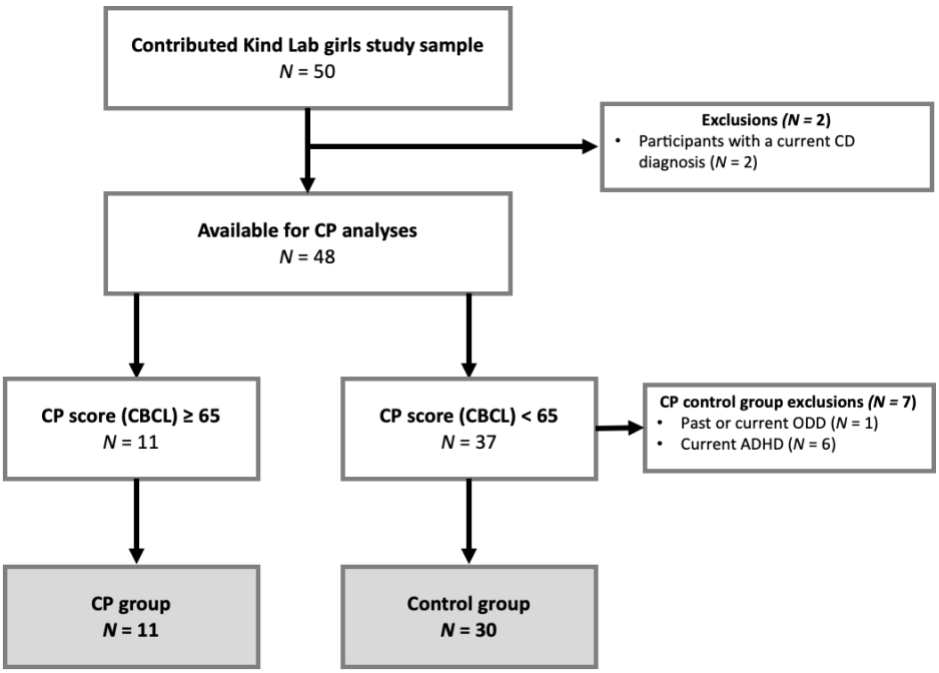


CD=conduct disorder. QC=quality control. ODD=oppositional defiant disorder.

KIND Lab girls study

The aim of the study was to examine emotional development in a community sample of Latina girls. At the first visit, children and parents completed a battery of questions about demographics, family dynamics, and mental health outcomes, as well as a child-appropriate threat conditioning paradigm. Over the course of two visits, children also participated in two fMRI tasks: 1) an implicit emotion processing task in which they labelled the genders of faces that were morphed between neutral and happy or fearful expressions, and 2) a fear generalization task with stimuli morphed between threat and safety stimuli in the previous threat conditioning paradigm. sMRI, DTI, and resting state scans were also collected. Inclusion criteria: Girls; aged 8-13 years; at least 50% Latina heritage and identified as Latina (except for one participant); medication-free. Exclusion criteria: Contraindications for neuroimaging (e.g., ferrous metal in the body, pregnant, claustrophobic); experiencing active medical problems or suicidal ideation; current psychiatric diagnosis of Tourette's syndrome or obsessive-compulsive disorder; lifetime history of mania, psychosis, or pervasive developmental disorder.

Figure S18: **KIND Lab girls study inclusion flowchart for the conduct problems analyses**



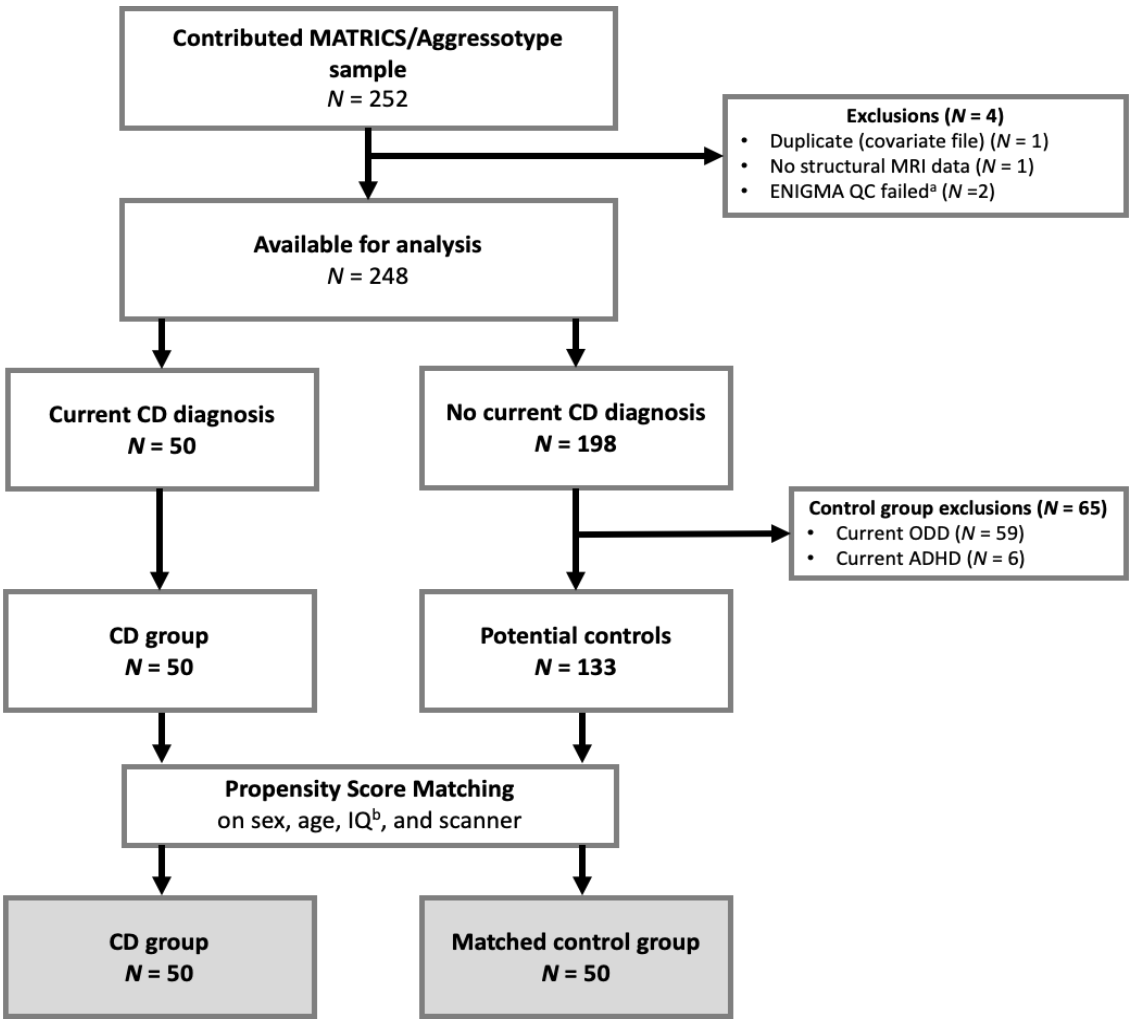
CP=conduct problems (based on the DSM-oriented conduct problems subscale of the Child Behavior Checklist). CD=conduct disorder. CBCL=Child Behavior Checklist. ODD=oppositional defiant disorder. ADHD=attention-deficit/hyperactivity disorder.

MATRICS/Aggressotype consortium

Participants aged 8–18 years were recruited across nine sites in Europe: Radboud University Medical Center and the Donders institute for Brain, Cognition and Behavior, Nijmegen, The Netherlands; Department of Neuroscience, University Medical Center Groningen, The Netherlands; Central Institute of Mental Health (CIMH), Mannheim, Germany; Department of psychiatry III and Child and Adolescent Psychiatry/Psychotherapy, University of Ulm, Ulm, Germany; Centre for Neuroimaging Sciences, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom; Department of Child Psychiatry, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom; Department of Child and Adolescent Psychiatry and Psychology, Neurosciences Institute, Hospital Clinic de Barcelona, Barcelona, Spain; Hospital Gregorio Marañón, Madrid, Spain; MR Center, Psychiatric University Hospital, Zurich, Switzerland; IRCCS Santa Lucia Foundation, Rome, Italy. Cases were recruited from child and adolescent psychiatry departments and patient associations throughout the Netherlands, Germany, Switzerland, Spain, and Italy. Controls were mainly found through elementary and high schools.

Exclusion criteria for all participants were contraindications for MRI, an IQ < 80 and a primary DSM-5 diagnosis of psychosis, bipolar disorder, major depression and/or anxiety disorder. Participants that were included as "cases" were diagnosed with conduct disorder (CD) and/or oppositional defiant disorder (ODD) and/or scored above the clinical cut-off for aggressive behaviour and/or rule-breaking behaviour as measured with the Child Behavior Checklist (CBCL) completed by parents. Within the control group no psychiatric disorders or scores within the clinical range were allowed, as determined by screening questionnaires (CBCL). Participants that were using medication were at a stable dose for at least two weeks. Ethical approval for the study was obtained for all sites separately by local ethics committees. After description of the study written informed consent was given by the participants and/or their parents. The above information was primarily taken from Naaijen et al.[21]

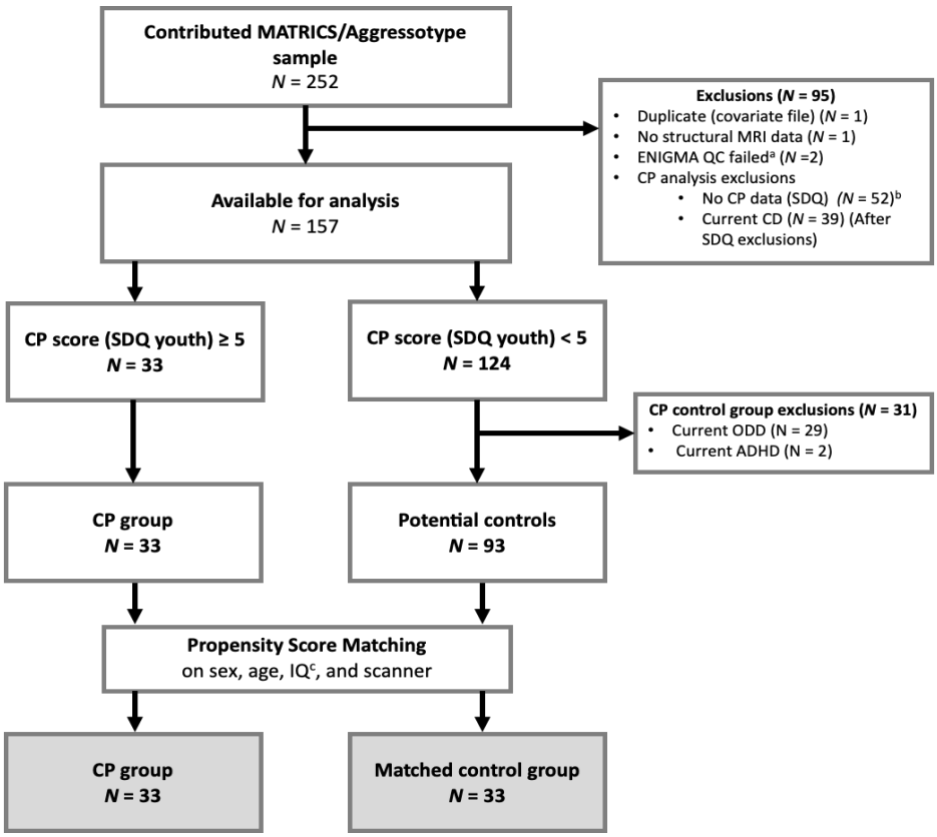
Figure S19: **MATRICS/Aggressotype inclusion flowchart for the conduct disorder analyses**



QC=quality control. CD=conduct disorder. ODD=oppositional defiant disorder. ADHD=attention-deficit/hyperactivity disorder. IQ=intelligence quotient.

[a] Due to total intracranial volume failing which precluded inclusion in the analyses. [b] IQ was group mean substituted for two cases to enable matching. Both values were subsequently set back to missing.

Figure S20: **MATRICS/Aggressotype inclusion flowchart for the conduct problems analyses**

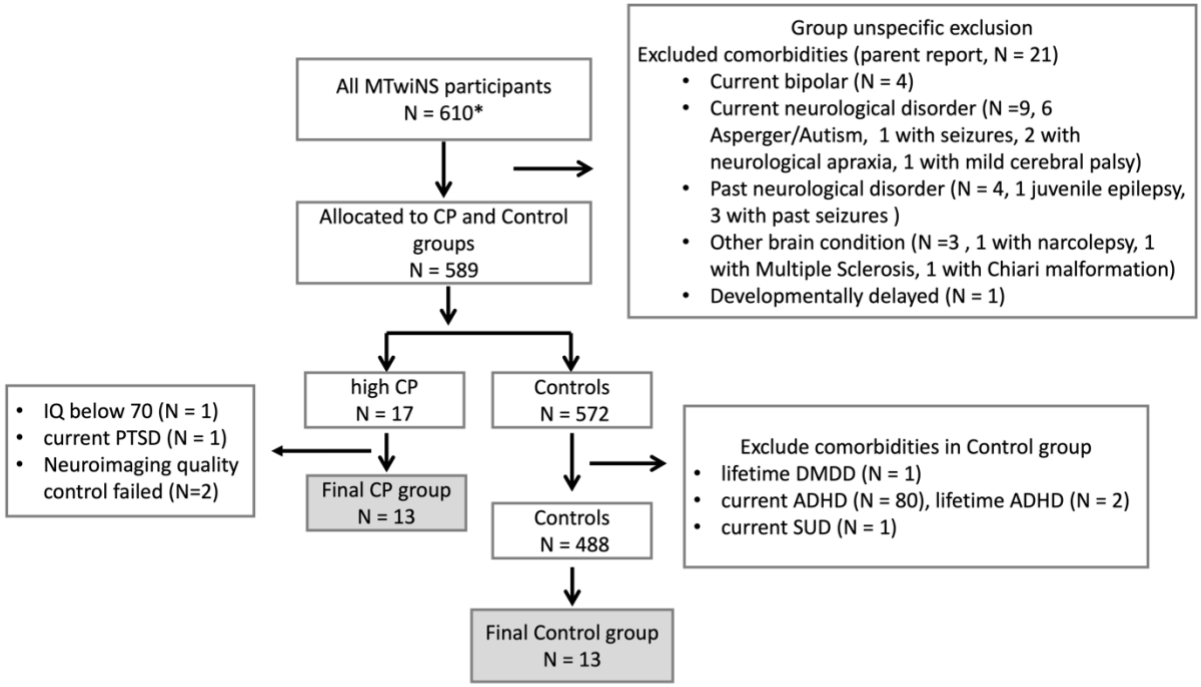


QC=quality control. CP=conduct problems (based on the conduct problems subscale of the youth-report Strengths and Difficulties Questionnaire). SDQ=Strengths and Difficulties Questionnaire. CD=conduct disorder. ODD=oppositional defiant disorder. ADHD=attention-deficit/hyperactivity disorder. IQ=intelligence quotient.
[a] Due to total intracranial volume failing which precluded inclusion in the analyses. [b] The DSM-oriented conduct problems subscale of the Child Behavior Checklist was also available for this sample but only the raw scores to which the cut-offs cannot be applied. [c] IQ was group mean substituted for two cases to enable matching. Both values were subsequently set back to missing.

MTwiNS

The primary sample included 610 youth from 354 twin pairs participating in the Michigan Twins Neurogenetics Study (MTwiNS),[22,23] an on-going study within the broader Michigan State University Twin Registry (MSUTR).[24] Youth included in this report were those who had participated up until our COVID data freeze. Youth in MTwiNS ranged in age from 7 to 19 years (mean age = 14.6 years, SD 2.2 years, 54.5% male) and were living in south-central Michigan. These youth were recruited from the Twin Study of Behavioral and Emotional Development – Child (TBED-C), which identified twins for two cohorts via birth records. The first cohort was sampled to represent families with twins living within 120 miles of Michigan State University, an area that includes Detroit, Flint, Lansing and other urban areas, as well as substantial swaths of rural Michigan. The second cohort was recruited from the same area, but only included families living in neighbourhoods with over 10.5% of families living below the poverty line (the median for the state of Michigan at the time; e.g.[25]), thus representing families living in neighbourhoods with above average poverty. For the present study, 354 twin pairs were re-recruited from the second cohort and from families in the first cohort that would have qualified for the second cohort (i.e., all families were living in a neighbourhood with above average disadvantage). This approach,, resulted in a sample that represents families living in south-central Michigan with substantial oversampling for families living in impoverished neighbourhoods. In the current MTwiNS sample, roughly 43% of twin families reported annual income below the living wage for a family of four in Michigan. Participants included in the present analyses met basic fMRI eligibility criteria, such as the absence of metal in their body and willingness to participate in the scanning session.

Figure S21: **MTwiNS inclusion flowchart for the conduct problems analyses**

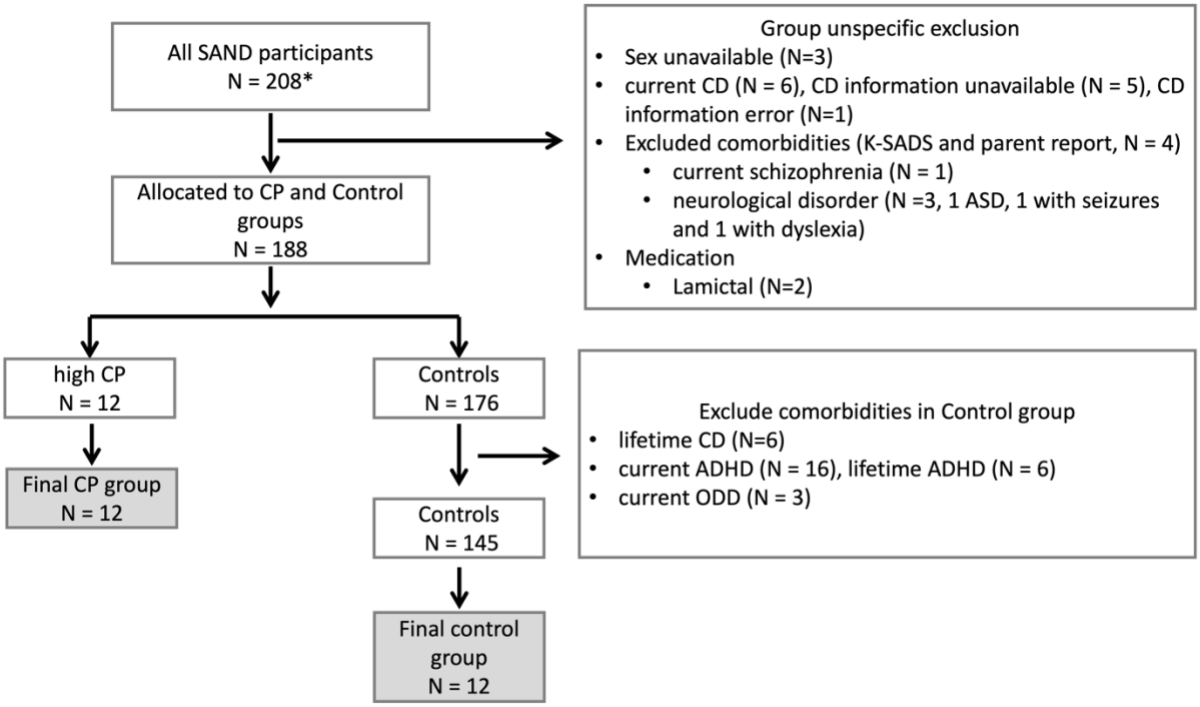


CP=Conduct problems.IQ=intelligence quotient. PTSD=post-traumatic stress disorder. DMDD=disruptive mood dysregulation disorder. ADHD=attention-deficit/hyperactivity disorder. SUD=substance use disorder.

SAND

The study sample was drawn from 237 adolescents from Detroit, Toledo, or Chicago who were part of the Study of Adolescent to Adult Neural Development (SAND),[26,27] a substudy of the Future of Families and Child Wellbeing Study (FFCWS).[28] The FFCWS is a longitudinal cohort of 4,898 (52.4% boys) children born in 20 large U.S. cities from 1998 to 2000[28] that was oversampled for nonmarital births (~3:1). Based on the initial sampling of FFCWS, along with the recruitment of SAND from 3 cities nearest to Ann Arbor, MI (and the demographics of those cities) the SAND sample contains substantial representation of African American adolescents as well as adolescents from families living in low-income contexts. Families living in Detroit, Toledo, and Chicago were invited to take part in additional data collection at the University of Michigan as part of the SAND study when the focal child was 15 years old. The complete list of measures and data for this project is publicly available from the National Institute of Mental Health data archive (https://nda.nih.gov/). The University of Michigan Medical School Institutional Review Board approved this study (UM IRBMED: HUM00074392). All adolescent participants provided written informed assent, and their primary caregivers provided written consent for both themselves and their adolescent children after the study was explained and questions were answered. Within the SAND study, the MRI component introduced sources of data loss, including some participants declining the MRI portion of the study. 208 study participants were scanned with available structural MRI data for sharing with ENIGMA.

Figure S22: **SAND inclusion flowchart for the conduct problems analyses**



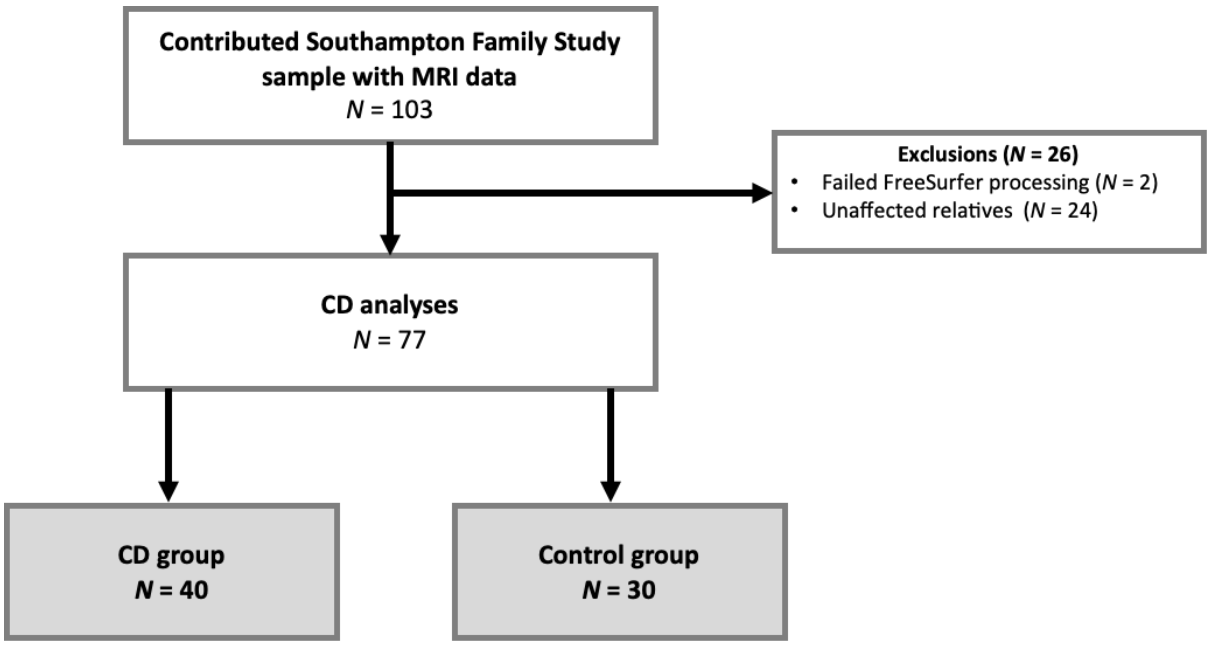
CP=conduct problems. K-SADS=The Kiddie Schedule for Affective Disorders and Schizophrenia. CD=conduct disorder. ASD=autism spectrum disorder. ADHD=attention-deficit/hyperactivity disorder. ODD=oppositional defiant disorder.

Southampton Family Study

112 adolescents aged between 11-18 years old participated in the study. There were three research groups: adolescents with conduct disorder (CD), unaffected relatives of those with CD (who did not meet criteria for the disorder themselves) and typically-developing adolescents with no history of antisocial behaviour. Control participants were recruited from mainstream schools and colleges, whereas CD participants were mainly recruited through Pupil Referral Units and local Youth Offending Teams. All participants completed a 'Family History Screen', consisting of three questions regarding: mental illness, behavioural difficulties and involvement with the law, in their first-degree relatives. The screen was designed to identify/recruit unaffected brothers or sisters of adolescents with CD who did not meet diagnostic criteria for CD themselves. In addition, it enabled us to identify the unaffected children of parents who had previously displayed CD themselves when they were aged between 11-18 years. The screen also ensured that the controls had no family history of CD or severe antisocial behaviour.

Participants were excluded if they had: (i) an IQ of <75 (as estimated using the vocabulary and matrix reasoning subtests of the Wechsler Abbreviated Scale of Intelligence; WASI); (ii) a serious psychiatric condition or pervasive developmental disorder (e.g. autism, schizophrenia, bipolar disorder) which was disclosed in the initial interview or identified using the Autism Quotient questionnaire; or (iii) metal in their body that could not be removed (e.g. head plates, cardiac stents or fixed dental braces), or claustrophobia, all of which were ascertained from the Magnetic Resonance Imaging (MRI) safety form at the initial screening and only applied to the MRI aspect of the study. In addition, female participants who were pregnant were ineligible for the MRI component of the study. All participants had to be fluent in English to understand and complete the questionnaires. Nine participants were either unable or unwilling to take part in the MRI aspect of the study, leaving a total sample of 103 participants for that part of the study.

Figure S23: **Southampton Family Study inclusion flowchart for the conduct disorder analyses**



MRI=Magnetic Resonance Imaging. CD=conduct disorder.

UCL-T1/T2

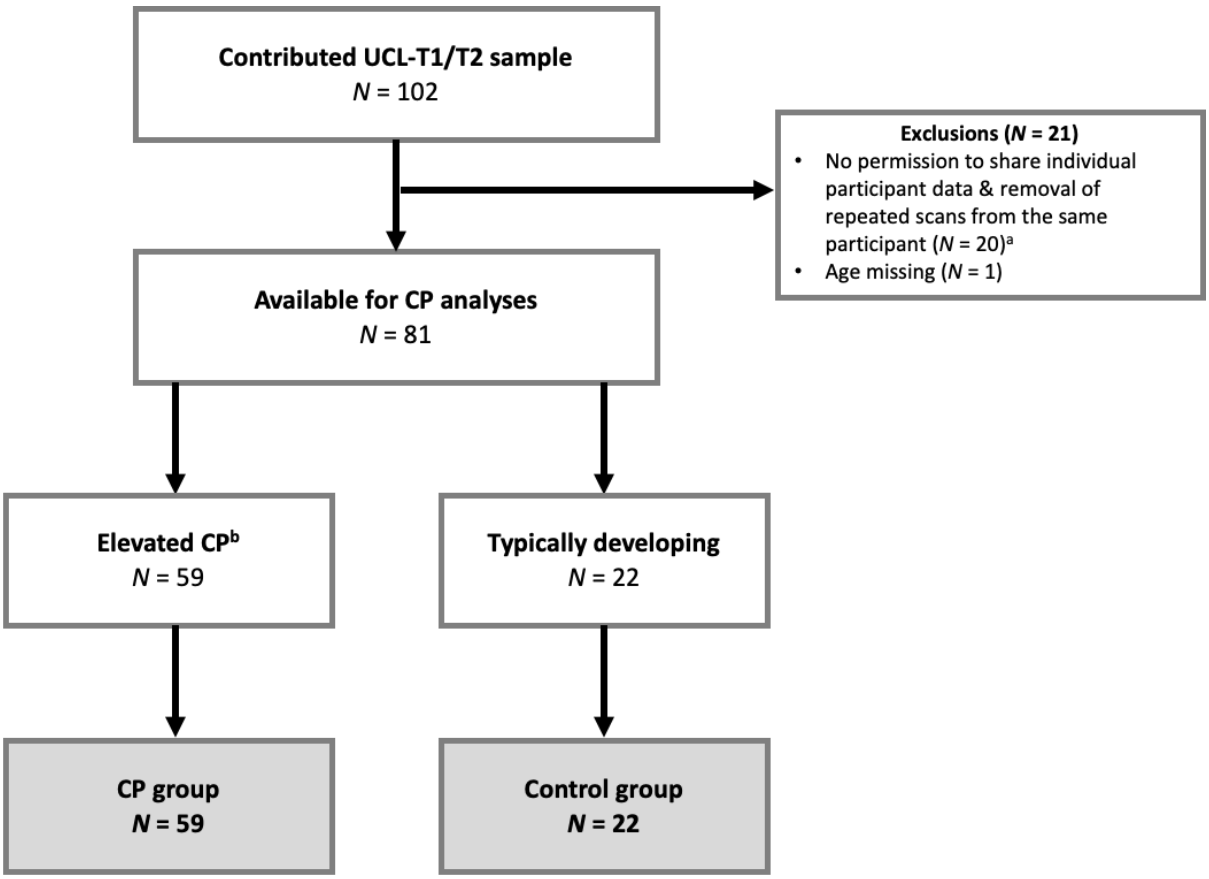
Boys aged 11-16 years were recruited from the community via newspaper advertisements, and local mainstream and specialist provision schools. Screening questionnaires were administered to parents of 360 boys and teachers of 215 boys whose families expressed an interest in taking part and provided informed consent. The screening measures yielded a research diagnosis of current conduct problems (our index of disruptive behaviour); dimensional assessment of callous-unemotional traits; an overall psychopathology screen; demographic data for group-matching purposes (i.e., socioeconomic status, parent-defined ethnicity, and handedness); and information regarding previous neurologic or psychiatric diagnoses.

Current conduct disorder symptoms were assessed using the Child and Adolescent Symptom Inventory– 4R (CASI-4R) –Conduct Disorder (CASI-CD) subscale,[29] scored by taking the highest ratings from either the parent or the teacher questionnaire for any given item.[30] For the CASI-CD scale, inclusion in the disruptive behaviour group required that the score met either parent or teacher severity cut-off (parent report: cut-off = 4+ [ages 10–12] and 3+ [ages 12–16]; teacher report: cut-off = 3+ [ages 10–12], 4+ [ages 12–14], and 6+ [ages 15–16]). These scores are associated with a clinical diagnosis of conduct disorder.[31] Typically-developing participants were required to score in the normal range for this measure, and below the atypical cut-off for total difficulties on the Strengths and Difficulties Questionnaire. Automatic exclusion criteria for both disruptive and typically-developing groups included a previous diagnosis of any neurological or psychotic disorder, or current psychiatric medication. To recruit a representative group of children with conduct problems, common comorbidities (ADHD, generalised anxiety disorder [GAD], depression, and substance/ alcohol abuse) were not used as exclusion criteria, but current parent-reported symptom counts were obtained during scanning sessions, so that their possible contribution to the findings could be systematically assessed.

Participants were provided with a complete description of the study. Informed consent was obtained from parents and written assent from participants. All aspects of the study were approved by the University College London Research Ethics Committee (Project ID number: 0622/001) and work was conducted in accordance with the Declaration of Helsinki. The above information was primarily taken from O'Nions et al.[32]

A subsample of participants which had provided consent to share de-identified individual participant data was contributed to ENIGMA-ASB.

Figure S24: **UCL-T1/T2 inclusion flowchart for the conduct problems analyses**



CP=conduct problems.
[a] Some participants were scanned twice but each participant was only included once: If one of the raw scans was of poor quality, we used the scan of better quality, if they were similar in quality both were pre-processed and we chose the final scan based on 1) having fewer fails across the cortical and subcortical quality control, 2) if the same number of regions were failed, we chose the one with fewer cortical fails; 3) if the number of cortical fails was the same, we chose the scan with the younger scan age. [b] For this sample, the grouping approach was based on age and informant-specific cut-offs on the Child and Adolescent Symptom Inventory (CASI-4R). Please see the information text on this sample above.

Yale (Sukhodolsky)

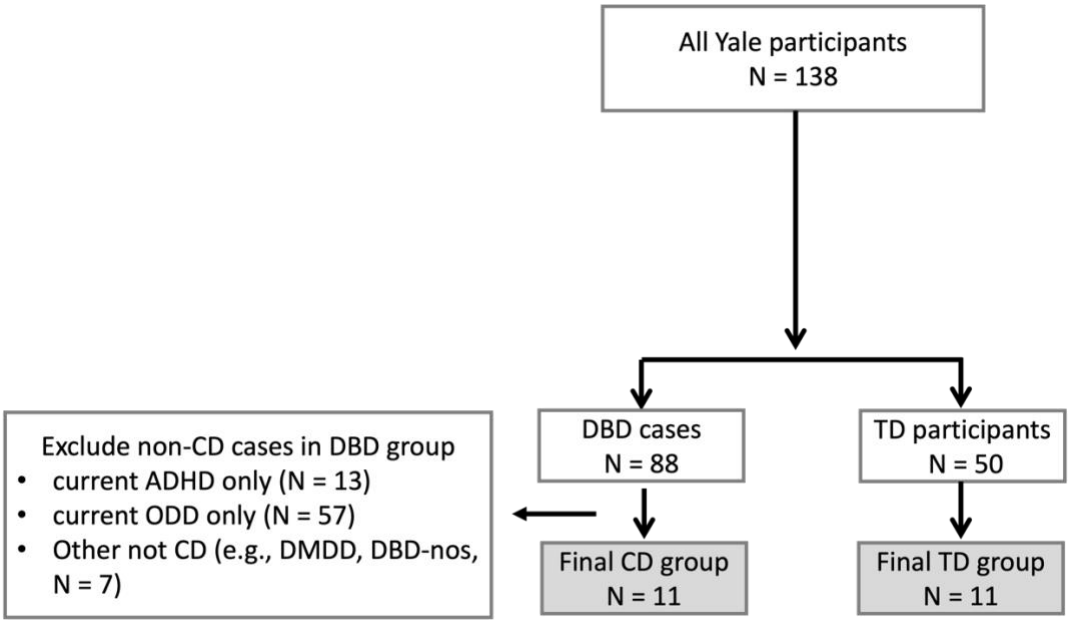
The sample included 88 children with disruptive behaviour disorders (DBD group; 30 females) and 50 typically-developing healthy controls (HC group; 20 females) matched for age and IQ. All participants were aged 8–16 years. Children with DBD participated in a treatment study of behaviour therapy for anger and aggression[33] and this paper reports structural MRI and clinical characterization data that were collected prior to initiating the treatment. Children with disruptive behaviour were recruited from the outpatient child psychiatry clinic at the Yale University Child Study Center and from outreach to the local schools, paediatricians and mental health providers. One of the inclusion criteria for the treatment study was a T-score of 65 or greater on the Aggressive Behaviour Scale of the Child Behavior Checklist (CBCL). Children were allowed to have co-occurring psychiatric disorders such as ADHD and anxiety if the presence of co-occurring disorders did not require immediate treatment. Untreated PTSD and severe depression were exclusionary criteria based on the rationale that these disorders present with pressing treatment needs. In addition to high levels of aggression on the dimensional measure (i.e. CBCL), all children met criteria for oppositional defiant disorder (ODD), conduct disorder, or disruptive mood dysregulation disorder (DMDD). All subjects who were assigned DMDD diagnoses also met criteria for ODD and following DSM-5 only one diagnosis (i.e., DMDD) was assigned. Of note, the current study was developed in response to the National Institute of Mental Health (NIMH) Research Domain Criteria (RDoC) initiative, which calls for explicating the core dimensions of psychopathology to evaluate the neural underpinnings of symptom dimensions across diagnostic boundaries.[34] Thus, children were included if they met a cut-off score

for clinically significant aggressive behaviour and these included a subgroup of children who met DSM-5 criteria for DMDD. Children were also required to be able to complete structural and functional MRI scans. Thus, this paper reports on children with disruptive behaviour disorders and high levels of aggression (indexed by CBCL aggression scale T score > 65) who were seeking treatment for disruptive behaviour. Healthy control children recruited would be matched on age, gender and IQ to children with the clinical sample. Fifty healthy control participants were recruited from the community via advertisements. Thus, 90 structural scans of participants in the DBD group and 50 structural scans of participants in the healthy control group were available for this analysis. Six structural scans from the DBD group were excluded due to high motion during scanning and two more scans were excluded after quality control assessment of reconstruction and segmentations due to artifact and segmentation errors. Thus, a total of 138 participants with high quality structural MRI data were included in the final analysis. Each participant's parent provided informed consent according to specifications by the institutional review board at the Yale University School of Medicine. Each child provided verbal and written assent.

Inclusion and exclusion criteria: Inclusion criteria for children with DBD were as follows: 1) 8 to 16 years of age at the time of consent; 2) meet criteria for a disruptive behaviour disorder according to K-SADS-PL, and confirmed by expert consensus among investigators; 3) have a Full-scale intelligence quotient (IQ) above 75, as measured by the or *DAS-II*; and 4) unmedicated or on stable medication for aggression, attention-deficit/hyperactivity disorder (ADHD), anxiety, or depression. Participants were excluded if they: 1) had a significant medical condition such as seizure disorder based on medical history; 2) were unable to meet MRI safety requirements such as absence of metal medical implants and claustrophobia; or 3) had a history of head trauma or loss of consciousness. Additional exclusionary criteria for typically-developing control subjects included any history of anxiety, ADHD, disruptive behaviour disorders, or other psychiatric, genetic, or neurological disorders.

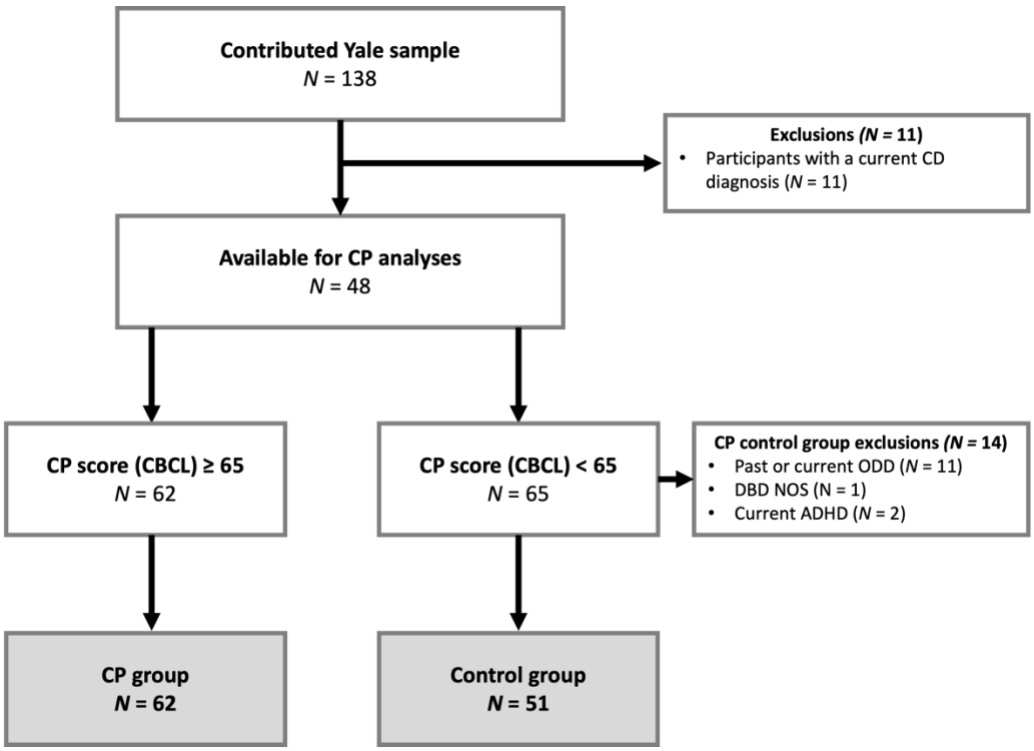
The above sample information was taken from Ibrahim et al.[35]

Figure S25: **Yale inclusion flowchart for the conduct disorder analyses**



CD=conduct disorder. TD=typically-developing. ADHD=attention-deficit/hyperactivity disorder. ODD=oppositional defiant disorder. DMDD=disruptive mood dysregulation disorder. DBD=disruptive behaviour disorder.

Figure S26: **Yale inclusion flowchart for the conduct problems analyses**



CD=conduct disorder. CP=conduct problems. CBCL=Child Behaviour Checklist. ODD=oppositional defiant disorder. DBD NOS=disruptive behaviour disorder not otherwise specified. ADHD=attention-deficit/hyperactivity disorder.

Appendix 3 - Image processing and quality control protocols

For all samples except ABCD and IMAGEN (see below), the standardised and publicly available ENIGMA protocols were followed (https://enigma.ini.usc.edu/protocols/imaging-protocols/). Each contributing site (or the main analysis sites) pre-processed previously collected structural T1-weighted MRI scans using the freely available and extensively validated FreeSurfer software,[36] version 5.3 or 6.0. This included segmentation and parcellation of the brain into subcortical and cortical regions per hemisphere (left and right), based on the Desikan-Killiany atlas[37] and FreeSurfer's subcortical *aseg* atlas. Cortical thickness and surface area were extracted for all 34 cortical regions (as well as total surface area and mean cortical thickness) and volume was extracted for seven subcortical regions (amygdala, caudate, hippocampus, nucleus accumbens, pallidum, putamen, thalamus). Estimated total intracranial volume was also obtained. Subsequently, visual quality control was performed on a region-by-region basis based on webpages showing snapshots from internal slices as well as an external 3D view of the brain segmentation from different angles. Quality assessments were aided by a visual inspection guide including pass/fail segmentation examples. Additionally, diagnostic histogram plots were generated for each site and outlying data points (2.698*SD) were flagged for further review. All regions failing quality inspection are not included in the analyses. All quality assessments were made blind to group status. Notably, training to apply the ENIGMA pre-processing and quality control protocols as well as additional supporting documentation (e.g., tutorial videos) were provided by the lead analysis sites (Birmingham and Bath).

Protocol deviations

Due to the size of the samples and means of accessing the FreeSurfer output folders, the full ENIGMA protocol was not performed for the ABCD and IMAGEN samples. Additionally, both studies also implemented their own QC procedures, which we took into account.

ABCD study

Deviations in pre-processing. As part of the ABCD pre-processing protocols, T1-weighted structural images were corrected for gradient nonlinearity distortions prior to FreeSurfer processing using scanner-specific, non-linear transformations provided by MRI scanner manufacturers.[38,39] Similarly, intensity inhomogeneity correction was performed prior to FreeSurfer processing by applying smoothly varying, estimated B1-bias fields, using a novel implementation that is similar in purpose to commonly used bias field correction methods.[40,41] Lastly, images were rigidly registered and resampled into alignment with an averaged reference brain in standard space. This pre-existing, in-house, averaged, reference brain has 1.0mm isotropic voxels and is roughly aligned with the anterior commissure/posterior commissure (AC/PC) axis. This standard reference brain was created by averaging T1w brain images from 500 adults after they had been nonlinearly registered to an initial template brain image using discrete cosine transforms (DCT).[42] The above information was taken from Hagler et al. where additional details on the ABCD imaging methods can also be found.[43]

Deviations in quality control. As part of the ABCD study protocol, all MRIs are systematically checked using automated and manual methods to assess protocol compliance and quality issues in the data. This includes a manual review of the raw data (T1w, T2w, dMRI, MRI field maps, fMRI, and fMRI field maps) for poor image quality resulting in a score of 0 (Reject) or 1 (Accept). Scores with a score of 0 (Reject) for T1w quality were excluded prior to the application of additional inclusion and exclusion criteria (please see figures Figure S2 and Figure S3 for the flowcharts). To align with the region-by-region quality control approach adopted by ENIGMA, we additionally checked for outliers (2.698*SD) for every outcome and excluded those in the final matched sample. In the case of total intracranial volume being an outlier, the whole participant was excluded as total intracranial volume was needed to apply the ComBat functions prior to data analysis and was a key covariate in models testing for group differences in surface area and subcortical volumes.

IMAGEN

Deviations in quality control. Following the IMAGEN MR protocols, two cross-site standardization and quality control procedures were regularly implemented at each site: (1) The American College of Radiology phantom was scanned to provide information about geometric distortions and signal uniformity related to hardware differences in radiofrequency coils and gradient systems, image contrast and temporal stability, and a custom phantom was scanned for diffusion-related parameters. (2) Several healthy volunteers were regularly scanned at each site to assess factors that cannot be measured using phantoms alone and at multiple sites to determine inter-site variability in structural and functional measures (for example, tissue contrast in raw MRI signal, tissue relaxation properties).[19]

The T1 scans included in this study were visually inspected in the context of previous IMAGEN studies. With a 3-tier system, T1 scans rated C (indicating bad quality) were excluded. In addition to the manual quality control, outliers were further excluded based on the Euler number automatically derived from FreeSufer: any scan with a left-right mean Euler number over 217 was excluded, following previously established recommendations.[44]

As done for the ABCD sample, to align with the region-by-region quality control approach adopted by ENIGMA, we additionally checked for outliers (2.698*SD) for every outcome and excluded those in the final matched sample. In the case of total intracranial volume being an outlier, the whole participant was excluded (see ABCD section for rationale).

Table S2: **Information on neuroimaging methods and clinical instruments used at the participating sites**

Sample	Country	Sample type	Scanner	Field Strength	FreeSurfer version	*n* (%) excluded QC	Diagnostic measure	CP measure for group allocation	CU analyses: ICU informant	IQ instrument	Key references
ABCD (3.0, baseline)	USA	population	28 scanners (Siemens: 16, GE: 9, Philips: 3)	3.0	5.3	-[a]	K-SADS-COMP	CBCL	-	NIH Toolbox Age-Corrected Standard Scores	Hawes et al., 2021, *Am J Psychiatry*
BESD	The Netherlands	forensic	Philips	3.0	6.0	2 (2.3%)	K-SADS	-	ICU youth	WISC + WAIS	Aghajani et al., 2016, *Hum Brain Mapp*
Boys Town	USA	clinical	Siemens	3.0	6.0	9 (2.4%)	psychiatric interview	-	ICU youth + parent	WASI	Zhang et al., 2021, *Psychol Med*
Cambridge Female	UK	forensic + community	Siemens	3.0	5.3	0 (0.0%)	K-SADS	-	ICU youth	WASI	Fairchild et al., 2013, *J Child Psychol Psychiatry*
Cambridge Male	UK	forensic + community	Siemens	3.0	6.0	1 (1.1%)	K-SADS	-	ICU parent	WASI	Fairchild et al., 2011, *Am J Psychiatry*
CD-Zhou	China	forensic	Siemens	3.0	6.0	0 (0.0%)	K-SADS	-	-	-	Lu et al., 2021, *Brain Imaging Behav*
CDKid	UK	forensic + community	GE	3.0	6.0	1 (2.5%)	K-SADS	-	-	WASI	Sarkar et al., 2015, *Eur Child Adolesc Psychiatry*
CSU-Yao	UK	clinical + community	Philips	3.0	6.0	0 (0.0%)	SCID	-	-	WASI	Zhang et al., 2014, *J Am Acad Child Adolesc Psychiatry*
cVEDA consortium	India	community	5 scanners (Siemens: 3, Philips: 2)	3.0	6.0	0 (0.0%)	MINI-KID (≤18) MINI-5 (≥19)	SDQ youth + parent	-	g-factor cognition score[b]	Sharma et al., 2020, *BMC Psychiatry*
FemNAT-CD	Germany, Switzerland, UK	clinical, forensic, community	5 scanners (Siemens: 4, Philips: 1)	3.0	5.3	10 (1.5%)	K-SADS	CBCL	ICU youth + parent	WASI + WISC	Freitag et al., 2018, *Eur Child Adolesc Psychiatry*
Georgetown	USA	community	Siemens	3.0	6.0	2 (2.3%)	NA	SDQ parent	-	K-BIT	Cardinale et al., 2019, *Psychol Med*
IMAGEN consortium	France, Germany, Ireland, UK	community	8 scanners (Siemens: 4, Philips: 2, GE: 1, Bruker: 1)	3.0	5.3	-[a]	DAWBA	SDQ youth + parent	-	WISC (full scale scores NA)[b]	Schumann et al. 2010, *Mol Psychiatry*
K23 (Stevens)	USA	community	Siemens	3.0	6.0	11 (19.0%)	K-SADS	-	ICU youth	WRAT3	Stevens & Haney-Caron, 2012, *J Psychiatry Neurosci*

KIND Lab girls study	USA	community	Siemens	3.0	6.0	0 (0.0%)	DISC-IV	CBCL	-	WISC	Glenn et al., 2022, *Sci Rep*
MATRICS/Aggressotype consortium	Germany, Italy, Netherlands, Spain, Switzerland, UK	clinical + community	8 scanners (Siemens: 6, Philips: 2, GE: 1)	3.0	5.3	2 (0.8%)	K-SADS	SDQ youth	ICU parent	WISC	Naaijen et al., 2020, *Neuroimage Clin*
MTwiNS	USA	twins	2 scanners (GE: 2)	3.0	6.0	2 (0.33%)	NA (K-SADS for subsample)	CBCL	-	Shipley	Burt et al., 2021, *Psychol Sci*
SAND	USA	community	GE	3.0	6.0	3 (1.44%)	K-SADS	CBCL	-	-	Gard et al., 2012, *Dev Psychopathol*
Southampton Family Study	UK	forensic + community	Siemens	1.5	5.3	0 (0.0%)	K-SADS	-	ICU youth + parent	WASI	Sully et al., 2015, *Psychol Med*
UCL-T1/T2	UK	community	2 Siemens	1.5	6.0	0 (0.0%)	NA	CASI	-	WASI	O'Nions et al., 2017, *Curr Biol*
Yale (Sukhodolsky)	USA	clinical + community	Siemens	3.0	6.0	0 (0.0%)	K-SADS	CBCL	ICU parent	WASI / Different Ability Scales-II	Ibrahim et al., 2021, *Dev Cogn Neurosci*

This table includes cohorts included in the main conduct disorder analyses, but also the five cohorts that exclusively contributed to the conduct problems analyses (Georgetown, KIND Lab girls study, MTwiNS, SAND, UCL-T1/T2, see Appendix 13). Notably, information on the measure of conduct problems or Inventory of Callous-Unemotional traits informant is only provided for samples that are included in the respective analyses. For example, some samples had measures of conduct problems available but were not included in the conduct problems analyses due to an insufficient number of participants fulfilling inclusion criteria. Similarly, some of the cohorts contributing to the conduct problems analyses had information on the Inventory of Callous-Unemotional traits but only conduct disorder samples were included in the callous-unemotional traits analyses.

[a] Full ENIGMA QC not performed. [b] Measures of cognitive abilities/IQ for cVEDA and IMAGEN were used during matching but were not used in sensitivity analyses adjusting for IQ as standardised scores comparable to full-scale IQ scores were not available.

Table S3: **Inclusion of samples across the main analyses and primary sensitivity analyses**

Sample name	CD main analysis	CD subgroup analyses		CD sensitivity analyses						Conduct Problems
		Age-of-onset[a]	CU traits[a]	IQ	ADHD	Anxiety	De-pression	SUD	Medi-cation	
ABCD	✓	✓		✓	✓	✓	✓	✓	✓	✓
BESD	✓		✓	✓	✓				✓	
Boys Town	✓		✓	✓	✓	✓	✓		✓	
Cambridge Female	✓	✓	✓	✓	✓	✓	✓	✓		
Cambridge Male	✓	✓	✓	✓	✓				✓	
CD-Zhou	✓				✓	✓	✓	✓		
CDKid	✓	✓		✓	✓				✓	
CSU-Yao	✓	✓		✓	✓	✓	✓	✓	✓	
cVEDA consortium	✓									✓
FemNAT-CD	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Georgetown										✓
IMAGEN consortium	✓				✓	✓	✓			✓
K23 (Stevens)	✓		✓		✓	✓	✓	✓		
KIND Lab girls study										✓
MATRICS/Aggressotype consortium	✓		✓	✓	✓				✓	✓
MTwiNS										✓
SAND										✓
Southampton Family Study	✓	✓	✓	✓	✓				✓	
UCL-T1/T2										✓
Yale (Sukhodolsky)	✓		✓	✓	✓	✓	✓		✓	✓
Total number of samples	**15**	**7**	**9**	**11**	**14**	**9**	**9**	**6**	**10**	**11**
Total number of participants[b]	**2438**	**1994**	**1905**	**2171**	**2378**	**2000**	**2000**	**1483**	**2131**	**2375**

[a] Controls of all samples included in the conduct disorder main analysis were included in the subgroup analyses. [b]This reflects the maximum possible sample size as sample sizes vary between individual imaging outcomes due to the applied region-by-region quality control processes and exclusions. Comorbidity analyses focused on current comorbidities. For cVEDA, only lifetime diagnoses were available and the sample was therefore not included in the sensitivity analyses adjusting for current comorbidities. CD=conduct disorder. IQ=intelligence quotient. ADHD=attention/deficit-hyperactivity disorder. SUD=substance use disorder.

Appendix 4 - Statistical approach

Mega-analysis versus meta-analysis. In the current study, we adopted a mega-analytical approach. As opposed to meta-analysis, which involves the weighted pooling of effect sizes derived from different studies/cohorts, mega-analyses involve the pooling of individual-level participant data from all included cohorts/studies. While both approaches enable a high-powered, precise estimation of the population effect, mega-analyses are (statistically) more powerful and flexible by allowing – for example – the control of confounders at the participant level.[45] In line with this, a recent ENIGMA investigation found the mega-analytical model to yield lower standard errors and narrower confidence intervals than the meta-analytical model when applied to the same data.[46] Additional advantages include that the lead analyst has greater control over the (consistent) application of inclusion criteria, better treatment of confounds and missing data, and the ability to test the assumptions underlying the used statistical models.[45] Due to the large number of working group members that were able to share de-identified individual level data, we therefore performed a pooled individual participant data (meta-)analysis (i.e., mega-analysis). For a more detailed consideration of different meta- and mega-analytical approaches within ENIGMA, please see Zugman et al.[45]

ComBat functions and statistical models. While each cohort was pre-processed and quality controlled following the ENIGMA protocols, remaining site differences (e.g., site-specific acquisition protocols and/or recruitment strategies) must be accounted for in analyses of multi-cohort individual participant data. In the current study, this was achieved by applying the modified ComBat functions developed by Radua and colleagues[47] to the imaging data prior to performing the statistical group comparisons. Using this approach, each imaging feature is modelled as a combination of variance accounted for by covariates (e.g. group status, age, sex), mean differences across sites, and error terms that follows varying normal distributions at each site. During harmonization, additive and multiplicative site effects are removed (using an approach akin to residualizing)[48] whilst retaining the effect of the pre-specified 'biological' covariates (e.g., age and sex). ComBat has multiple advantages over other site correction methods such as mixed-effect models which include 'site' as a random intercept to account for site-related effects. These include the aforementioned estimation of site-specific error terms that may follow different normal distributions (versus the assumption that error terms follow the same normal distribution within mixed models which is unlikely to be true)[47] and taking account of the fact that site effects are not independent across the different imaging features. In line with this, studies have shown that mega-analyses using ComBat harmonisation followed by standard linear models outperform mixed effects mega-analyses, providing increased statistical power and significance, thereby resulting in higher sensitivity to detect group differences and associations.[47,49]

Therefore, we adopted the following analytical strategy:

1) Step 1: **Data harmonisation** using Radua et al.'s ComBat functions[47] to remove site-related differences. ComBat models included group, sex, age, and harmonised total intracranial volume (TIV, for regional outcomes) as covariates. Please see Appendix 8 for more information on the performance of ComBat in the current sample.
2) Step 2: **Assessing group differences** using general linear models per imaging feature with diagnosis/group as the main predictor of interest, and age, sex and TIV as covariates of no interest. TIV was only included in models focused on surface area and subcortical volumes.

Model 1: $ROI_i = intercept + \beta_1(diagnosis) + \beta_2(sex) + \beta_3(age) [+ \beta_4(TIV)] + \varepsilon_i$

3) Step 3: **Sensitivity analyses to assess additional confounding**: One at a time, we added the following variables into the model to address whether group differences were robust to their inclusion: 1) IQ, 2) current ADHD (binary coded), 3) current substance use disorder (binary coded), 4) current depression (binary coded), 5) current anxiety (binary coded), and 5) psychotropic medication use (binary coded).

Models 2-7: $ROI_i = intercept + \beta_1(diagnosis) + \beta_2(sex) + \beta_3(age) + \beta_4(covariate) [+ \beta_5(TIV)] + \varepsilon_i$

4) Step 4: **Interactions with sex and age** were explored by separately adding a group-by-sex and group-by-age interaction term into the model.

Model 8: $ROI_i = intercept + \beta_1(diagnosis) + \beta_2(sex) + \beta_3(age) + \beta_4(diagnosisXsex) [+ \beta_5(TIV)] + \varepsilon_i$

Model 9: $ROI_i = intercept + \beta_1(diagnosis) + \beta_2(sex) + \beta_3(age) + \beta_4(diagnosisXage) [+ \beta_5(TIV)] + \varepsilon_i$

5) Step 5: **Subgroup analyses** exploring differences between childhood-onset and adolescent-onset CD, and between CD with low versus high CU traits were performed following a two-step approach: a) We first performed ANCOVAs per imaging feature, including group (three levels: subgroup 1, subgroup 2, typically-developing), sex, age, and TIV (where appropriate). This approach uses the same linear modelling approach as outlined in the previous steps in combination with an ANOVA wrapper to assess whether the regression coefficients associated with the three-level grouping variable were simultaneously zero. b) Significant group effects after multiple comparison correction were followed up with (uncorrected) pairwise comparisons using the emmeans R package. These were based on the same linear model used to test the overall group effect and hence included age, sex, and TIV (where appropriate) as covariates. For the specific subgrouping approach used in these analyses, please see Appendix 1.

Model 10*: $ROI_i = intercept + \beta_1(CO - CD) + \beta_2(AO - CD) + \beta_3(sex) + \beta_4(age) [+ \beta_5(TIV)] + \varepsilon_i$

Model 11*: $ROI_i = intercept + \beta_1(low - CU) + \beta_2(high - CU) + \beta_3(sex) + \beta_4(age) [+ \beta_5(TIV)] + \varepsilon_i$

*The reference category in Model 10 and 11 was the control group albeit this has no influence on the assessment of the main effect of group using the ANOVA wrapper. Models 10 and 11 with significant group effects were followed by pairwise comparisons comparing each group to the other two groups using the same covariates.

Multiple comparison correction

To increase consistency with findings of previous surface-based morphometry studies (e.g., Smaragdi et al.[50]) and the work by the ENIGMA-ADHD working group,[1] we corrected for multiple comparisons by applying a False Discovery Rate (FDR) correction with q=0.05 separately per imaging metric (i.e., surface area, cortical thickness, and subcortical volumes). Additionally, in Appendix 12 we report findings for the main analyses (CD vs TD) when adjusting across all outcomes, which largely overlapped with our main findings.

SUPPLEMENTARY RESULTS

Appendix 5 - Full results of the main analysis comparing brain structure between youth with conduct disorder and typically-developing youth

Table S4: **Case-control differences in cortical thickness between youth with conduct disorder and typically-developing controls**

	TD			CD							
Region	***n***	***M***	***SD***	***n***	***M***	***SD***	***t***	***p***	***p* FDR**	**Cohen's *d***	**95% CI (*d*)**
Banks of the Superior Temporal Sulcus	**1227**	**2.78**	**0.15**	**1161**	**2.76**	**0.15**	**-3.29**	**0.0010**	**0.0178**	**-0.13**	**-0.22, -0.05**
Caudal Anterior Cingulate Cortex	**1227**	**2.79**	**0.19**	**1159**	**2.82**	**0.20**	**3.90**	**0.0001**	**0.0034**	**0.16**	**0.08, 0.24**
Caudal Middle Frontal Gyrus	1224	2.72	0.14	1156	2.72	0.14	-0.74	0.4609	0.6824	-0.03	-0.11, 0.05
Cuneus Cortex	1210	2.01	0.13	1151	2.02	0.13	0.53	0.5978	0.6974	0.02	-0.06, 0.10
Entorhinal Cortex	1199	3.43	0.31	1143	3.46	0.32	2.31	0.0209	0.1486	0.10	0.01, 0.18
Frontal Pole	1228	3.04	0.26	1156	3.04	0.26	0.15	0.8774	0.9032	0.01	-0.07, 0.09
Fusiform Gyrus	1215	2.88	0.12	1155	2.87	0.12	-1.59	0.1111	0.3599	-0.07	-0.15, 0.02
Inferior Parietal Cortex	1205	2.67	0.12	1158	2.66	0.13	-1.36	0.1740	0.3939	-0.06	-0.14, 0.02
Inferior Temporal Gyrus	1198	2.96	0.14	1129	2.96	0.14	-0.61	0.5420	0.6974	-0.03	-0.11, 0.06
Insula	1191	3.23	0.14	1148	3.22	0.14	-0.54	0.5924	0.6974	-0.02	-0.10, 0.06
Isthmus-Cingulate Cortex	1234	2.63	0.17	1165	2.63	0.17	0.92	0.3598	0.6297	0.04	-0.04, 0.12
Lateral Occipital Cortex	1210	2.32	0.12	1151	2.32	0.12	0.02	0.9856	0.9856	0.00	-0.08, 0.08
Lateral Orbitofrontal Cortex	1230	2.89	0.14	1160	2.88	0.15	-1.88	0.0602	0.2632	-0.08	-0.16, 0.00
Lingual Gyrus	1218	2.17	0.13	1163	2.17	0.13	-1.39	0.1632	0.3939	-0.06	-0.14, 0.02
Medial Orbitofrontal Cortex	1222	2.67	0.15	1153	2.67	0.15	-1.29	0.1977	0.3939	-0.05	-0.13, 0.03
Middle Temporal Gyrus	1208	3.12	0.14	1147	3.11	0.15	-1.55	0.1215	0.3599	-0.06	-0.14, 0.02
Paracentral Lobule	1234	2.59	0.14	1170	2.58	0.14	-1.43	0.1532	0.3939	-0.06	-0.14, 0.02
Parahippocampal Gyrus	1234	2.91	0.24	1170	2.90	0.24	-0.72	0.4718	0.6824	-0.03	-0.11, 0.05
Pars Opercularis	1232	2.85	0.13	1170	2.84	0.13	-1.20	0.2305	0.4246	-0.05	-0.13, 0.03
Pars Orbitalis	1232	2.99	0.18	1168	2.99	0.19	-0.69	0.4874	0.6824	-0.03	-0.11, 0.05
Pars Triangularis	1230	2.73	0.14	1169	2.72	0.14	-1.99	0.0470	0.2350	-0.08	-0.16, 0.00
Pericalcarine Cortex	1193	1.71	0.12	1144	1.71	0.13	0.35	0.7274	0.7956	0.01	-0.07, 0.10
Postcentral Gyrus	1181	2.20	0.12	1135	2.19	0.12	-2.24	0.0255	0.1486	-0.09	-0.17, -0.01
Posterior-Cingulate Cortex	1232	2.70	0.13	1168	2.72	0.13	2.26	0.0239	0.1486	0.09	0.01, 0.17

Precentral Gyrus	1200	2.68	0.12	1145	2.67	0.13	-1.54	0.1234	0.3599	-0.06	-0.14, 0.02
Precuneus Cortex	1234	2.60	0.12	1166	2.59	0.12	-0.76	0.4450	0.6824	-0.03	-0.11, 0.05
Rostral Anterior Cingulate Cortex	1231	3.09	0.19	1161	3.09	0.19	0.30	0.7670	0.8135	0.01	-0.07, 0.09
Rostral Middle Frontal Gyrus	1224	2.58	0.13	1150	2.57	0.13	-0.61	0.5420	0.6974	-0.03	-0.11, 0.06
Superior Frontal Gyrus	1217	2.96	0.14	1154	2.96	0.14	0.73	0.4652	0.6824	0.03	-0.05, 0.11
Superior Parietal Cortex	1199	2.34	0.12	1150	2.34	0.13	0.55	0.5791	0.6974	0.02	-0.06, 0.10
Superior Temporal Gyrus	1186	3.05	0.14	1137	3.04	0.15	-1.65	0.0993	0.3599	-0.07	-0.15, 0.01
Supramarginal Gyrus	1192	2.77	0.12	1144	2.76	0.13	-1.27	0.2026	0.3939	-0.05	-0.13, 0.03
Temporal Pole	1204	3.78	0.26	1138	3.75	0.27	-2.42	0.0155	0.1486	-0.10	-0.18, -0.02
Transverse Temporal Cortex	1236	2.63	0.19	1168	2.63	0.20	-0.50	0.6179	0.6976	-0.02	-0.10, 0.06
Mean Cortical Thickness	1234	2.68	0.09	1169	2.67	0.10	-1.28	0.2019	0.3939	-0.05	-0.13, 0.03

All statistical models included group, sex and age. A negative Cohen's *d* indicates conduct disorder < controls; a positive Cohen's *d* indicates conduct disorder > controls. Regions with a significant group difference post False Discovery Rate correction are indicated in bold. Means are adjusted for covariates. TD=typically-developing controls. CD= conduct disorder group.

Table S5: **Case-control differences in surface area between youth with conduct disorder and typically-developing controls**

	TD			**CD**			***t***	***p***	***p* FDR**	**Cohen's *d***	**95% CI**
Region	***n***	***M***	***SD***	***n***	***M***	***SD***					
Banks of the Superior Temporal Sulcus	**1227**	**1055.07**	**123.56**	**1162**	**1040.40**	**127.51**	**-2.98**	**0.0029**	**0.0053**	**-0.12**	**-0.20, -0.04**
Caudal Anterior Cingulate Cortex	**1227**	**739.19**	**105.86**	**1160**	**727.01**	**109.22**	**-2.89**	**0.0039**	**0.0067**	**-0.12**	**-0.20, -0.04**
Caudal Middle Frontal Gyrus	**1225**	**2386.74**	**303.16**	**1156**	**2345.39**	**313.45**	**-3.41**	**0.0007**	**0.0016**	**-0.14**	**-0.22, -0.06**
Cuneus Cortex	**1209**	**1589.29**	**172.45**	**1149**	**1573.79**	**178.38**	**-2.24**	**0.0254**	**0.0337**	**-0.09**	**-0.17, -0.01**
Entorhinal Cortex	**1201**	**403.09**	**71.58**	**1143**	**394.82**	**73.93**	**-2.88**	**0.0040**	**0.0067**	**-0.12**	**-0.20, -0.04**
Frontal Pole	**1228**	**276.25**	**31.09**	**1158**	**270.19**	**32.06**	**-4.89**	**<0.0001**	**<0.0001**	**-0.20**	**-0.28, -0.12**
Fusiform Gyrus	**1214**	**3371.31**	**292.21**	**1157**	**3324.13**	**302.79**	**-4.03**	**0.0001**	**0.0003**	**-0.17**	**-0.25, -0.08**
Inferior Parietal Cortex	**1208**	**5445.08**	**558.01**	**1159**	**5305.42**	**577.04**	**-6.24**	**<0.0001**	**<0.0001**	**-0.26**	**-0.34, -0.18**
Inferior Temporal Gyrus	**1198**	**3504.55**	**345.84**	**1132**	**3437.10**	**356.81**	**-4.83**	**<0.0001**	**<0.0001**	**-0.20**	**-0.28, -0.12**
Insula	**1191**	**2298.22**	**180.13**	**1148**	**2274.59**	**185.96**	**-3.26**	**0.0011**	**0.0024**	**-0.13**	**-0.22, -0.05**
Isthmus-Cingulate Cortex	**1234**	**1032.07**	**123.07**	**1165**	**1015.71**	**127.18**	**-3.34**	**0.0008**	**0.0020**	**-0.14**	**-0.22, -0.06**
Lateral Occipital Cortex	**1210**	**5136.16**	**467.77**	**1150**	**5083.79**	**483.74**	**-2.79**	**0.0053**	**0.0085**	**-0.12**	**-0.20, -0.03**
Lateral Orbitofrontal Cortex	**1230**	**2756.66**	**228.39**	**1161**	**2722.52**	**235.74**	**-3.75**	**0.0002**	**0.0005**	**-0.15**	**-0.23, -0.07**

Lingual Gyrus	**1218**	**3266.20**	**370.40**	**1160**	**3230.11**	**382.53**	**-2.44**	**0.0148**	**0.0216**	**-0.10**	**-0.18, -0.02**
Medial Orbitofrontal Cortex	1221	1907.18	166.58	1154	1896.99	172.10	-1.53	0.1262	0.1472	-0.06	-0.14, 0.02
Middle Temporal Gyrus	**1211**	**3545.39**	**311.71**	**1152**	**3478.56**	**322.36**	**-5.35**	**<0.0001**	**<0.0001**	**-0.22**	**-0.30, -0.14**
Paracentral Lobule	1233	1477.71	148.41	1170	1471.27	153.42	-1.09	0.2752	0.2943	-0.04	-0.12, 0.04
Parahippocampal Gyrus	**1233**	**713.77**	**77.21**	**1170**	**701.85**	**79.77**	**-3.89**	**0.0001**	**0.0004**	**-0.16**	**-0.24, -0.08**
Pars Opercularis	1233	1630.39	201.66	1169	1622.25	208.41	-1.02	0.3099	0.3190	-0.04	-0.12, 0.04
Pars Orbitalis	**1233**	**766.58**	**72.12**	**1168**	**758.85**	**74.55**	**-2.70**	**0.0071**	**0.0107**	**-0.11**	**-0.19, -0.03**
Pars Triangularis	1232	1517.00	188.73	1168	1504.68	195.04	-1.64	0.1005	0.1256	-0.07	-0.15, 0.01
Pericalcarine Cortex	1194	1520.07	220.22	1143	1509.62	227.78	-1.17	0.2401	0.2711	-0.05	-0.13, 0.03
Postcentral Gyrus	**1185**	**4259.03**	**349.41**	**1136**	**4204.53**	**361.35**	**-3.85**	**0.0001**	**0.0004**	**-0.16**	**-0.24, -0.08**
Posterior-Cingulate Cortex	**1233**	**1274.59**	**135.34**	**1167**	**1262.61**	**139.80**	**-2.23**	**0.0260**	**0.0337**	**-0.09**	**-0.17, -0.01**
Precentral Gyrus	**1202**	**4985.18**	**377.02**	**1144**	**4927.56**	**390.21**	**-3.79**	**0.0002**	**0.0005**	**-0.16**	**-0.24, -0.08**
Precuneus Cortex	**1235**	**4151.69**	**371.02**	**1167**	**4098.34**	**383.42**	**-3.62**	**0.0003**	**0.0008**	**-0.15**	**-0.23, -0.07**
Rostral Anterior Cingulate Cortex	1230	770.51	106.21	1161	763.75	109.65	-1.60	0.1101	0.1328	-0.07	-0.15, 0.01
Rostral Middle Frontal Gyrus	**1227**	**6272.07**	**578.68**	**1152**	**6196.23**	**597.19**	**-3.28**	**0.0011**	**0.0023**	**-0.13**	**-0.22, -0.05**
Superior Frontal Gyrus	**1219**	**7526.44**	**585.95**	**1153**	**7419.71**	**605.33**	**-4.55**	**<0.0001**	**<0.0001**	**-0.19**	**-0.27, -0.11**
Superior Parietal Cortex	**1199**	**5704.91**	**520.74**	**1148**	**5657.50**	**538.83**	**-2.26**	**0.0239**	**0.0335**	**-0.09**	**-0.17, -0.01**
Superior Temporal Gyrus	**1189**	**3960.65**	**313.33**	**1138**	**3909.23**	**323.82**	**-4.07**	**<0.0001**	**0.0002**	**-0.17**	**-0.25, -0.09**
Supramarginal Gyrus	**1196**	**4072.36**	**438.76**	**1146**	**4017.38**	**453.81**	**-3.11**	**0.0019**	**0.0037**	**-0.13**	**-0.21, -0.05**
Temporal Pole	1205	460.53	48.02	1139	460.00	49.63	-0.27	0.7840	0.7840	-0.01	-0.09, 0.07
Transverse Temporal Cortex	1235	418.38	51.78	1168	416.15	53.49	-1.09	0.2775	0.2943	-0.04	-0.12, 0.04
Total Surface Area	**1234**	**89230.02**	**7425.90**	**1170**	**87463.10**	**7625.33**	**-5.95**	**<0.0001**	**<0.0001**	**-0.24**	**-0.32, -0.16**

All statistical models included group, sex, age and total intracranial volume (except in the case of total surface area). A negative Cohen's *d* indicates conduct disorder < controls; a positive Cohen's *d* indicates conduct disorder > controls. Regions with a significant group difference post False Discovery Rate correction are indicated in bold. Means are adjusted for covariates. TD=typically-developing controls. CD=conduct disorder group

***Table S6*: Case-control differences in subcortical volume and total intracranial volume between youth with conduct disorder and typically-developing controls**

	TD			CD			*t*	*p*	*p* FDR	Cohen's *d*	95% CI
Region	***n***	***M***	***SD***	***n***	***M***	***SD***					
Amygdala	**1233**	**1703.91**	**160.64**	**1173**	**1683.42**	**166.04**	**-3.21**	**0.0014**	**0.0055**	**-0.13**	**-0.21, -0.05**
Caudate	1202	4008.19	428.08	1151	3992.27	442.24	-0.93	0.3540	0.4570	-0.04	-0.12, 0.04
Hippocampus	**1248**	**4281.06**	**316.51**	**1180**	**4244.00**	**326.77**	**-2.96**	**0.0031**	**0.0082**	**-0.12**	**-0.20, -0.04**
Nucleus Accumbens	**1249**	**635.28**	**82.11**	**1180**	**626.19**	**84.78**	**-2.80**	**0.0052**	**0.0103**	**-0.11**	**-0.19, -0.03**
Pallidum	1241	1872.66	180.93	1174	1875.03	186.89	0.33	0.7407	0.7407	0.01	-0.07, 0.09
Putamen	1228	5829.56	519.78	1162	5812.13	538.12	-0.84	0.3999	0.4570	-0.03	-0.11, 0.05
Thalamus	**1228**	**7890.22**	**546.16**	**1156**	**7817.88**	**562.21**	**-3.33**	**0.0009**	**0.0055**	**-0.14**	**-0.22, -0.06**
Total Intracranial Volume	1253	1504293.15	134297.55	1185	1493411.74	137685.86	-2.04	0.0414	0.0663	-0.08	-0.16, 0.00

All statistical models included group, sex, age and total intracranial volume (except for total intracranial volume itself). A negative Cohen's *d* indicates conduct disorder < controls; a positive Cohen's *d* indicates conduct disorder > controls. Regions with a significant group difference post False Discovery Rate correction are indicated in bold. Means are adjusted for covariates. TD=typically-developing controls. CD=conduct disorder group.

Appendix 6 - Full results of the sensitivity analyses (CD vs. TD)

For the purpose of the sensitivity analyses, additional covariates were included one at time for the following reasons: First, we wanted to align our work with previous ENIGMA studies, particularly those focusing on ADHD in youth (e.g., Hoogman et al.[1,2]) to facilitate comparisons of results across studies. These studies performed sensitivity analyses in a similar way – adding covariates one at a time. Second, including all additional covariates together into one model would have made it challenging to identify which (one) of these variables had the greatest impact on the results in our sensitivity analyses – therefore to enhance interpretability we chose to include them one at a time. Third, as shown in Table S3 in the ***Supplementary Methods***, the availability of information on the covariates considered in these sensitivity analyses varied across cohorts, and not all covariates were available for all participants, ranging from ~98% (for ADHD comorbidity) and ~59% (for substance use disorder [SUD] comorbidity) of the full sample. For example, when considering participants with information available for all covariates (i.e., IQ, current comorbidity of ADHD, SUD, Depression, Anxiety, and medication use), the number of participants was reduced to n=1248, which is only half of the number retained when including each covariate one at a time (e.g., n_{ADHD}=2378). This substantial reduction in sample size would have impacted the statistical power and the resulting findings would have been difficult to interpret. Finally, the participants with comprehensive data on all covariates were mainly from two cohorts (n_{ABCD}=574, $n_{FemNAT\text{-}CD}$=633), meaning the results would predominantly be based on these two large cohorts (i.e., not 'unbiased'), thereby limiting the generalizability of our findings.

The majority of findings remained significant when adjusting (one at a time) for current comorbidities (ADHD, SUD, depression or anxiety), psychotropic medication use (all binary-coded), or IQ. It is particularly noteworthy that most group differences survived correcting for ADHD, a related disorder that is highly comorbid with CD and which is

itself associated with (overlapping) alterations in brain structure.[1,2] Supplemental Table S3 shows the sample sizes for the sensitivity analyses, which varied between cohorts. These differences in sample size may have also influenced our ability to detect significant differences after False Discovery Rate correction.

In brief, both case-control differences in cortical thickness were robust to controlling for IQ, ADHD or SUD but were affected by correcting for psychotropic medication use (caudal anterior cingulate cortex) or anxiety and depression (banks of the superior temporal sulcus). In the case of surface area, total surface area and 14 of the 26 regional surface area differences (56% of effects) remained significant across all of the sensitivity analyses. In the individual sensitivity analyses, 74-93% of case-control differences were retained, with 82% (27/33) of effects remaining significant when adjusting for ADHD. Lastly, lower amygdala volume in the CD group was significant in all of the sensitivity analyses. However, controlling for IQ (nucleus accumbens, hippocampus), ADHD (hippocampus), and SUD (nucleus accumbens, thalamus) impacted some of the subcortical volume results. Detailed findings are provided in the tables below.

Overall, these findings suggest that many of the observed group differences were not merely driven by internalising and externalising psychopathology, IQ, and psychotropic medication. However, it is notable that the fewest group differences were detected when controlling for SUD, albeit 21/33 effects remained significant. It is conceivable that substance use (and SUD comorbidity) underlie some of the brain alterations observed in CD in line with findings of brain alterations associated with substance use and dependence.[51] However, it is also notable that analyses controlling for SUD were based on only 59% of the sample and hence, lower statistical power may have also contributed to this pattern of findings. Correspondingly, we observed that many of the effects that were rendered non-significant when controlling for additional variables counted amongst the weakest effects (i.e., smallest effect sizes and largest corrected *p*-values) in the main analyses, suggesting that the smaller sample size in (some of) the sensitivity analyses and ensuing lower power likely contributed to a lack of differences in these outcomes when controlling for additional variables.

Table S7: **Case-control differences in cortical thickness when individually adjusting for current comorbidities, IQ or psychotropic medication use**

	current ADHD			**current SUD**			**current depression**			**current anxiety**			**IQ**			**psychotropic medication**		
Region	***t***	***d***	***p FDR***	***t***	***d***	***p FDR***	***t***	***d***	***p FDR***	***t***	***d***	***p FDR***	***t***	***d***	***p FDR***	***t***	***d***	***p FDR***
Banks of the Superior Temporal Sulcus*	**-3.19**	**-0.13**	**0.0493**	**-3.37**	**-0.18**	**0.0269**	-2.68	-0.12	0.0946	-2.54	-0.11	0.1247	**-3.13**	**-0.14**	**0.0310**	**-3.62**	**-0.16**	**0.0105**
Caudal Anterior Cingulate Cortex*	**2.99**	**0.12**	**0.0493**	2.92	0.15	0.0612	**4.03**	**0.18**	**0.0020**	**3.45**	**0.16**	**0.0202**	**3.51**	**0.15**	**0.0159**	2.86	0.13	0.0760
Caudal Middle Frontal Gyrus	-0.41	-0.02	0.7921	0.02	0.00	0.9820	-0.27	-0.01	0.8375	-0.04	0.00	0.9682	-0.62	-0.03	0.7235	-0.30	-0.01	0.8550
Cuneus Cortex	0.14	0.01	0.9021	0.61	0.03	0.7254	0.55	0.02	0.7739	0.49	0.02	0.8217	0.88	0.04	0.6492	0.28	0.01	0.8550
Entorhinal Cortex	2.45	0.10	0.1267	1.72	0.09	0.3421	1.96	0.09	0.2526	2.53	0.12	0.1247	1.88	0.08	0.4210	1.71	0.08	0.2871
Frontal Pole	0.27	0.01	0.8576	0.60	0.03	0.7254	0.29	0.01	0.8375	0.72	0.03	0.7222	-0.58	-0.03	0.7235	-0.54	-0.02	0.7346
Fusiform Gyrus	-1.28	-0.05	0.4416	-1.20	-0.06	0.4216	-1.45	-0.07	0.3973	-0.75	-0.03	0.7222	-1.35	-0.06	0.4969	-1.58	-0.07	0.2871
Inferior Parietal Cortex	-0.87	-0.04	0.6646	-0.54	-0.03	0.7254	-1.17	-0.05	0.4220	-0.85	-0.04	0.7222	-1.01	-0.04	0.6450	-1.63	-0.07	0.2871
Inferior Temporal Gyrus	-0.24	-0.01	0.8613	-0.51	-0.03	0.7254	-0.04	0.00	0.9719	0.19	0.01	0.9575	-0.62	-0.03	0.7235	-0.23	-0.01	0.8639
Insula	-1.12	-0.05	0.5440	-1.23	-0.06	0.4216	-0.53	-0.02	0.7739	-0.08	0.00	0.9621	-0.28	-0.01	0.8015	-0.46	-0.02	0.7561
Isthmus-Cingulate Cortex	0.79	0.03	0.6877	1.32	0.07	0.4078	1.70	0.08	0.3393	1.86	0.08	0.3163	0.43	0.02	0.7513	0.54	0.02	0.7346

Lateral Occipital Cortex	0.44	0.02	0.7921	-0.09	0.00	0.9547	-0.42	-0.02	0.7862	-0.08	0.00	0.9621	0.23	0.01	0.8196	-0.70	-0.03	0.7320
Lateral Orbitofrontal Cortex	-1.46	-0.06	0.4217	-1.94	-0.10	0.2647	-2.15	-0.10	0.1890	-1.59	-0.07	0.4927	-1.59	-0.07	0.4902	-1.87	-0.08	0.2871
Lingual Gyrus	-1.89	-0.08	0.2599	-1.29	-0.07	0.4078	-1.44	-0.06	0.3973	-1.11	-0.05	0.6470	-0.94	-0.04	0.6450	-2.01	-0.09	0.2871
Medial Orbitofrontal Cortex	-0.68	-0.03	0.7538	-1.08	-0.06	0.4696	-1.27	-0.06	0.3973	-0.62	-0.03	0.7772	-1.33	-0.06	0.4969	-1.82	-0.08	0.2871
Middle Temporal Gyrus	-1.42	-0.06	0.4217	-1.54	-0.08	0.3421	-1.25	-0.06	0.3973	-1.02	-0.05	0.6470	-1.53	-0.07	0.4902	-1.46	-0.06	0.3373
Paracentral Lobule	-0.54	-0.02	0.7921	-1.17	-0.06	0.4271	-0.95	-0.04	0.5452	-0.78	-0.04	0.7222	-1.00	-0.04	0.6450	-0.89	-0.04	0.6571
Parahippocampal Gyrus	0.35	0.01	0.8194	-0.22	-0.01	0.8854	-0.27	-0.01	0.8375	-0.40	-0.02	0.8447	-0.56	-0.02	0.7235	-0.75	-0.03	0.7233
Pars Opercularis	-1.3	-0.05	0.4416	-1.49	-0.08	0.3421	-1.33	-0.06	0.3973	-0.97	-0.04	0.6470	-1.00	-0.04	0.6450	-0.51	-0.02	0.7346
Pars Orbitalis	-0.84	-0.03	0.6646	-1.51	-0.08	0.3421	-0.97	-0.04	0.5452	-0.52	-0.02	0.8217	-0.60	-0.03	0.7235	-0.81	-0.04	0.7013
Pars Triangularis	-2.24	-0.09	0.1466	-1.95	-0.10	0.2647	-1.89	-0.09	0.2565	-1.33	-0.06	0.6470	-2.54	-0.11	0.1312	-1.89	-0.08	0.2871
Pericalcarine Cortex	0.45	0.02	0.7921	-0.60	-0.03	0.7254	0.12	0.01	0.9322	0.21	0.01	0.9575	0.47	0.02	0.7513	0.16	0.01	0.8722
Postcentral Gyrus	-1.95	-0.08	0.2553	-1.52	-0.08	0.3421	-2.14	-0.10	0.1890	-1.99	-0.09	0.2756	-1.58	-0.07	0.4902	-1.89	-0.08	0.2871
Posterior-Cingulate Cortex	2.47	0.10	0.1267	2.10	0.11	0.2647	2.55	0.12	0.0946	2.45	0.11	0.1247	1.71	0.07	0.4902	1.81	0.08	0.2871
Precentral Gyrus	-1.43	-0.06	0.4217	-2.06	-0.11	0.2647	-1.36	-0.06	0.3973	-0.97	-0.04	0.6470	-1.42	-0.06	0.4965	-1.70	-0.08	0.2871
Precuneus Cortex	-0.44	-0.02	0.7921	-0.88	-0.05	0.5800	-0.44	-0.02	0.7862	-0.39	-0.02	0.8447	-0.93	-0.04	0.6450	-0.92	-0.04	0.6571
Rostral Anterior Cingulate Cortex	-0.12	-0.01	0.9021	-0.21	-0.01	0.8854	0.84	0.04	0.5817	0.98	0.04	0.6470	0.34	0.01	0.7791	-0.53	-0.02	0.7346
Rostral Middle Frontal Gyrus	-0.59	-0.02	0.7921	-0.49	-0.03	0.7254	-0.66	-0.03	0.7124	-0.12	-0.01	0.9621	-0.77	-0.03	0.6752	-0.59	-0.03	0.7346
Superior Frontal Gyrus	1.52	0.06	0.4217	0.53	0.03	0.7254	0.86	0.04	0.5817	1.23	0.06	0.6470	0.40	0.02	0.7513	0.89	0.04	0.6571
Superior Parietal Cortex	0.47	0.02	0.7921	-0.25	-0.01	0.8854	0.45	0.02	0.7862	0.48	0.02	0.8217	0.79	0.03	0.6752	0.59	0.03	0.7346
Superior Temporal Gyrus	-1.66	-0.07	0.3787	-1.56	-0.08	0.3421	-1.54	-0.07	0.3927	-0.98	-0.05	0.6470	-1.46	-0.06	0.4965	-1.62	-0.07	0.2871
Supramarginal Gyrus	-1.29	-0.05	0.4416	-1.52	-0.08	0.3421	-1.66	-0.08	0.3393	-1.19	-0.05	0.6470	-0.86	-0.04	0.6492	-1.28	-0.06	0.4365
Temporal Pole	-2.33	-0.10	0.1399	-2.30	-0.12	0.2548	-2.62	-0.12	0.0946	-2.23	-0.10	0.1797	-2.11	-0.09	0.3044	-1.61	-0.07	0.2871
Transverse Temporal Cortex	-0.92	-0.04	0.6574	-0.98	-0.05	0.5223	-1.40	-0.06	0.3973	-1.10	-0.05	0.6470	-0.41	-0.02	0.7513	-0.20	-0.01	0.8639
Mean Cortical Thickness	-1.08	-0.04	0.5464	-1.29	-0.07	0.4078	-1.24	-0.06	0.3973	-0.74	-0.03	0.7222	-1.07	-0.05	0.6450	-1.14	-0.05	0.5226
***N* significant (out of 35)**			**2**			**1**			**1**			**2**			**2**			**1**

All statistical models included group, sex and age. The provided statistics are for the group effect. * indicates that this outcome showed a significant group difference in the main analysis (conduct disorder vs. typically-developing) after False Discovery Rate correction ($N = 2$). A negative Cohen's *d* indicates conduct disorder < controls; a positive Cohen's *d* indicates conduct disorder > controls. Regions with a significant group difference post False Discovery Rate correction are indicated in bold. *d*=Cohen's *d*. FDR=False Discovery Rate.

***Table S8*: Case-control differences in surface area when individually adjusting for current comorbidities, IQ or psychotropic medication use**

	current ADHD			current SUD			current depression			current anxiety			IQ			psychotropic medication		
Region	***t***	***d***	***p FDR***	***t***	***d***	***p FDR***	***t***	***d***	***p FDR***	***t***	***d***	***p FDR***	***t***	***d***	***p FDR***	***t***	***d***	***p FDR***
Banks of the Superior Temporal Sulcus*	-3.18	**-0.13**	**0.0038**	-1.41	-0.07	0.2397	**-2.98**	**-0.13**	**0.0057**	**-3.27**	**-0.15**	**0.0026**	**-2.52**	**-0.11**	**0.0274**	**-2.93**	**-0.13**	**0.0076**
Caudal Anterior Cingulate Cortex*	-2.66	**-0.11**	**0.0139**	**-3.00**	**-0.16**	**0.0107**	**-3.53**	**-0.16**	**0.0015**	**-3.58**	**-0.16**	**0.0011**	**-2.30**	**-0.10**	**0.0396**	**-2.84**	**-0.12**	**0.0094**
Caudal Middle Frontal Gyrus*	-3.10	**-0.13**	**0.0045**	**-2.30**	**-0.12**	**0.0442**	**-3.60**	**-0.16**	**0.0013**	**-3.68**	**-0.17**	**0.0009**	**-2.43**	**-0.11**	**0.0312**	**-2.73**	**-0.12**	**0.0125**
Cuneus Cortex*	-1.42	-0.06	0.1708	-0.28	-0.01	0.8255	-1.36	-0.06	0.2098	-1.79	-0.08	0.0947	-1.12	-0.05	0.3191	**-2.31**	**-0.10**	**0.0292**
Entorhinal Cortex*	-1.88	-0.08	0.0844	**-2.62**	**-0.14**	**0.0263**	**-2.52**	**-0.12**	**0.0189**	**-3.19**	**-0.15**	**0.0030**	**-2.75**	**-0.12**	**0.0175**	**-3.33**	**-0.15**	**0.0025**
Frontal Pole*	-4.21	**-0.17**	**0.0002**	**-3.82**	**-0.20**	**0.0017**	**-4.61**	**-0.21**	**< 0.0001**	**-4.21**	**-0.19**	**0.0002**	**-3.89**	**-0.17**	**0.0010**	**-5.08**	**-0.22**	**< 0.0001**
Fusiform Gyrus*	-3.33	**-0.14**	**0.0027**	-1.69	-0.09	0.1513	**-3.37**	**-0.15**	**0.0025**	**-3.45**	**-0.16**	**0.0016**	**-2.89**	**-0.13**	**0.0150**	**-4.12**	**-0.18**	**0.0002**
Inferior Parietal Cortex*	-5.72	**-0.24**	**< 0.0001**	**-4.51**	**-0.24**	**0.0002**	**-5.61**	**-0.26**	**< 0.0001**	**-5.84**	**-0.27**	**< 0.0001**	**-5.81**	**-0.25**	**< 0.0001**	**-5.80**	**-0.26**	**< 0.0001**
Inferior Temporal Gyrus*	-4.48	**-0.19**	**0.0001**	**-2.85**	**-0.15**	**0.0139**	**-4.56**	**-0.21**	**< 0.0001**	**-4.79**	**-0.22**	**< 0.0001**	**-3.76**	**-0.16**	**0.0012**	**-4.66**	**-0.21**	**< 0.0001**
Insula*	-3.32	**-0.14**	**0.0027**	-1.47	-0.08	0.2259	**-2.99**	**-0.14**	**0.0057**	**-2.90**	**-0.13**	**0.0067**	**-2.48**	**-0.11**	**0.0286**	**-2.49**	**-0.11**	**0.0203**
Isthmus-Cingulate Cortex*	-2.71	**-0.11**	**0.0125**	**-2.53**	**-0.13**	**0.0314**	**-3.26**	**-0.15**	**0.0030**	**-3.21**	**-0.15**	**0.0029**	**-2.79**	**-0.12**	**0.0169**	**-3.01**	**-0.13**	**0.0061**
Lateral Occipital Cortex*	-1.94	-0.08	0.0758	-0.42	-0.02	0.7585	-1.82	-0.08	0.0962	-1.75	-0.08	0.0964	-1.36	-0.06	0.2269	**-2.37**	**-0.10**	**0.0260**
Lateral Orbitofrontal Cortex*	-3.47	**-0.14**	**0.0023**	-2.08	-0.11	0.0699	**-2.77**	**-0.13**	**0.0100**	**-2.84**	**-0.13**	**0.0076**	**-2.86**	**-0.12**	**0.0151**	**-3.35**	**-0.15**	**0.0025**
Lingual Gyrus*	-1.60	-0.07	0.1284	-0.69	-0.04	0.5739	**-2.34**	**-0.11**	**0.0298**	**-2.21**	**-0.10**	**0.0400**	-1.45	-0.06	0.1989	**-2.37**	**-0.10**	**0.0260**
Medial Orbitofrontal Cortex	-1.78	-0.07	0.1015	-0.08	0.00	0.9390	-0.80	-0.04	0.4479	-0.81	-0.04	0.4409	-0.75	-0.03	0.5114	-2.01	-0.09	0.0573
Middle Temporal Gyrus*	-5.77	**-0.24**	**< 0.0001**	**-3.48**	**-0.18**	**0.0026**	**-5.11**	**-0.23**	**< 0.0001**	**-5.38**	**-0.25**	**< 0.0001**	**-4.30**	**-0.19**	**0.0003**	**-5.27**	**-0.23**	**< 0.0001**
Paracentral Lobule	-1.57	-0.07	0.1305	-1.08	-0.06	0.3954	-1.14	-0.05	0.2850	-1.78	-0.08	0.0947	-0.32	-0.01	0.7526	-0.55	-0.02	0.5818
Parahippocampal Gyrus*	-3.39	**-0.14**	**0.0025**	**-3.74**	**-0.20**	**0.0017**	**-3.24**	**-0.15**	**0.0031**	**-3.43**	**-0.16**	**0.0016**	**-3.48**	**-0.15**	**0.0022**	**-3.54**	**-0.15**	**0.0015**
Pars Opercularis	-2.33	**-0.10**	**0.0320**	-0.29	-0.02	0.8255	-1.04	-0.05	0.3283	-1.05	-0.05	0.3220	-0.48	-0.02	0.6931	-1.28	-0.06	0.2189
Pars Orbitalis*	-2.39	**-0.10**	**0.0280**	-0.85	-0.04	0.5110	-1.86	-0.08	0.0917	-1.83	-0.08	0.0906	-1.79	-0.08	0.1165	-2.05	-0.09	0.0549
Pars Triangularis	-1.65	-0.07	0.1243	-0.79	-0.04	0.5358	-1.40	-0.06	0.2038	-1.05	-0.05	0.3220	-1.15	-0.05	0.3120	-1.32	-0.06	0.2118
Pericalcarine Cortex	-0.72	-0.03	0.5000	0.98	0.05	0.4435	-0.33	-0.02	0.7599	-0.52	-0.02	0.6035	-0.38	-0.02	0.7211	-1.42	-0.06	0.1813
Postcentral Gyrus*	-3.41	**-0.14**	**0.0025**	**-3.36**	**-0.18**	**0.0035**	**-3.31**	**-0.15**	**0.0028**	**-3.61**	**-0.17**	**0.0011**	**-2.59**	**-0.11**	**0.0241**	**-2.58**	**-0.11**	**0.0176**
Posterior-Cingulate Cortex*	-1.67	-0.07	0.1230	**-2.38**	**-0.12**	**0.0403**	**-2.79**	**-0.13**	**0.0097**	**-2.92**	**-0.13**	**0.0066**	-1.75	-0.08	0.1211	-1.93	-0.08	0.0669
Precentral Gyrus*	-3.22	**-0.13**	**0.0035**	**-3.63**	**-0.19**	**0.0020**	**-3.80**	**-0.17**	**0.0007**	**-3.72**	**-0.17**	**0.0009**	-2.14	-0.09	0.0543	**-3.18**	**-0.14**	**0.0040**
Precuneus Cortex*	-2.84	**-0.12**	**0.0100**	**-2.45**	**-0.13**	**0.0362**	**-3.03**	**-0.14**	**0.0055**	**-3.34**	**-0.15**	**0.0021**	**-2.67**	**-0.12**	**0.0203**	**-3.78**	**-0.17**	**0.0006**
Rostral Anterior Cingulate Cortex	-1.61	-0.07	0.1284	-0.70	-0.04	0.5739	-1.48	-0.07	0.1786	-2.05	-0.09	0.0563	-1.46	-0.06	0.1989	-1.57	-0.07	0.1410
Rostral Middle Frontal Gyrus*	-2.77	**-0.12**	**0.0114**	-1.85	-0.10	0.1123	**-3.17**	**-0.14**	**0.0036**	**-3.18**	**-0.14**	**0.0030**	**-2.32**	**-0.10**	**0.0396**	**-3.05**	**-0.13**	**0.0058**

Superior Frontal Gyrus*	-3.85	**-0.16**	**0.0006**	**-3.77**	**-0.20**	**0.0017**	**-4.30**	**-0.20**	**0.0001**	**-4.40**	**-0.20**	**0.0001**	**-3.59**	**-0.16**	**0.0019**	**-4.36**	**-0.19**	**0.0001**
Superior Parietal Cortex*	-2.27	**-0.10**	**0.0351**	-2.21	-0.12	0.0537	-1.73	-0.08	0.1119	**-2.24**	**-0.10**	**0.0380**	-1.60	-0.07	0.1598	**-2.50**	**-0.11**	**0.0203**
Superior Temporal Gyrus*	-3.87	**-0.16**	**0.0006**	**-2.33**	**-0.12**	**0.0437**	**-3.69**	**-0.17**	**0.0010**	**-3.70**	**-0.17**	**0.0009**	**-3.55**	**-0.16**	**0.0019**	**-3.99**	**-0.18**	**0.0003**
Supramarginal Gyrus*	-2.76	**-0.12**	**0.0114**	**-2.93**	**-0.16**	**0.0122**	**-2.72**	**-0.12**	**0.0109**	**-2.57**	**-0.12**	**0.0163**	**-2.26**	**-0.10**	**0.0422**	**-2.68**	**-0.12**	**0.0138**
Temporal Pole	-0.40	-0.02	0.6872	-0.24	-0.01	0.8305	-0.18	-0.01	0.8588	-0.76	-0.03	0.4625	-0.44	-0.02	0.6969	-0.92	-0.04	0.3693
Transverse Temporal Cortex	-0.51	-0.02	0.6263	-1.14	-0.06	0.3716	-1.29	-0.06	0.2310	-1.25	-0.06	0.2454	-1.07	-0.05	0.3340	-1.04	-0.05	0.3164
Total Surface Area*	-5.54	**-0.23**	**< 0.0001**	**-3.59**	**-0.19**	**0.0020**	**-5.14**	**-0.23**	**< 0.0001**	**-5.26**	**-0.24**	**< 0.0001**	**-3.85**	**-0.17**	**0.0010**	**-5.08**	**-0.22**	**< 0.0001**
***N* significant (out of 35)**			**23**			**17**			**22**			**24**			**20**			**25**

All statistical models included group, sex, age and total intracranial volume (except for total surface area). Provided statistics are for the group effect. * indicates that this outcome showed a significant group difference in the main analysis (conduct disorder vs. typically-developing) after False Discovery Rate correction (n = 27). A negative Cohen's d indicates conduct disorder < controls; a positive Cohen's d indicates conduct disorder > controls. Regions with a significant group difference post False Discovery Rate correction are indicated in bold. d=Cohen's d. FDR=False Discovery Rate.

Table S9: **Case-control differences in subcortical volume and total intracranial volume when individually adjusting for current comorbidities, IQ or psychotropic medication use**

	current ADHD			**current SUD**			**current depression**			**current anxiety**			**IQ**			**psychotropic medication**		
Region	***t***	***d***	***p FDR***	***t***	***d***	***p FDR***	***t***	***d***	***p FDR***	***t***	***d***	***p FDR***	***t***	***d***	***p FDR***	***t***	***d***	***p FDR***
Amygdala*	**-2.88**	**-0.12**	**0.0181**	**-3.07**	**-0.16**	**0.0174**	**-3.45**	**-0.16**	**0.0029**	**-3.53**	**-0.16**	**0.0027**	**-2.53**	**-0.11**	**0.0454**	**-3.68**	**-0.16**	**0.0019**
Caudate	-0.84	-0.04	0.5341	-1.32	-0.07	0.3000	-1.22	-0.06	0.2974	-1.21	-0.06	0.3019	-1.14	-0.05	0.3984	-1.29	-0.06	0.2736
Hippocampus*	-1.73	-0.07	0.1331	**-2.81**	**-0.15**	**0.0200**	**-2.95**	**-0.13**	**0.0066**	**-3.00**	**-0.14**	**0.0054**	-2.08	-0.09	0.0751	**-3.31**	**-0.14**	**0.0025**
Nucleus Accumbens*	**-2.82**	**-0.12**	**0.0181**	**-2.31**	**-0.12**	**0.0554**	**-2.94**	**-0.13**	**0.0066**	**-3.27**	**-0.15**	**0.0029**	-2.28	-0.10	0.0601	**-3.49**	**-0.15**	**0.0019**
Pallidum	0.41	0.02	0.6841	1.03	0.05	0.4056	0.97	0.04	0.3811	0.60	0.03	0.6263	0.61	0.03	0.6186	0.17	0.01	0.8627
Putamen	-0.49	-0.02	0.6841	0.20	0.01	0.8448	0.02	0.00	0.9826	-0.43	-0.02	0.6685	-0.26	-0.01	0.7962	-1.00	-0.04	0.3601
Thalamus*	**-2.71**	**-0.11**	**0.0181**	-1.77	-0.09	0.1533	**-3.39**	**-0.15**	**0.0029**	**-3.41**	**-0.15**	**0.0027**	**-2.86**	**-0.12**	**0.0345**	**-3.08**	**-0.14**	**0.0041**
Total Intracranial Volume	-1.96	-0.08	0.0994	-0.91	-0.05	0.4125	-1.71	-0.08	0.1402	-1.51	-0.07	0.2089	-1.04	-0.04	0.3984	-1.27	-0.05	0.2736
***N* significant (out of 8)**			**3**			**3**			**4**			**4**			**2**			**4**

All statistical models included group, sex, age and total intracranial volume (except for when total intracranial volume was the outcome). Provided statistics are for the group effect. * indicates that this outcome showed a significant group difference in the main analysis (conduct disorder vs. typically-developing) after False Discovery Rate correction (N = 4). A negative Cohen's d indicates conduct disorder < controls; a positive Cohen's d indicates conduct disorder > controls. Regions with a significant group difference post False Discovery Rate correction are indicated in bold. d=Cohen's d. FDR=False Discovery Rate

Appendix 7 - Full results of hemisphere-specific analyses (CD vs. TD)

While the previous literature did not provide consistent evidence for lateralisation of brain differences in CD, we repeated the main analysis for each hemisphere separately in an exploratory fashion. Models included group (CD vs. TD), sex, age, and total intracranial volume (in the case of regional surface area and subcortical volumes). For the hemisphere-specific analyses, the False Discovery Rate correction was applied across hemispheres (i.e., 70 outcomes for surface area and cortical thickness, respectively, comprising 34 regional outcomes per hemisphere and left and right total/mean values, and 15 outcomes for (subcortical) volume, comprising 7 regional outcomes per hemisphere and total intracranial volume). Applying a more lenient correction within each hemisphere separately did not affect decisions regarding significance based on $p < 0.05$. In summary, the majority of significant group differences identified in the main analysis could also be observed in both hemispheres separately (i.e., one of the two effects on cortical thickness, 19 of the 27 surface area effects and two of the four effects on subcortical volumes). When only one of the hemispheres showed a significant difference, the coefficient in the other hemisphere was always in the same direction. For surface area, two additional regions showed a hemisphere-specific group difference: Youths with CD showed lower surface area in the right medial orbitofrontal cortex (with a very small and non-significant positive effect in the left hemisphere) and the left pars triangularis compared to controls. Detailed results can be found in Tables S10-12 (below).

Table S10: **Hemisphere-specific case-control differences in cortical thickness**

	group effect left	**group effect right**	**left hemisphere**						**right hemisphere**					
Region			***n* CD**	***n* TD**	***t***	***d***	***p***	***p FDR***	***n* CD**	***n* TD**	***t***	***d***	***p***	***p FDR***
Banks of the Superior Temporal Sulcus*	**yes**	no	**1066**	**1112**	**-3.54**	**-0.15**	**0.0004**	**0.0143**	1123	1204	-1.57	-0.07	0.1169	0.4981
Caudal Anterior Cingulate Cortex*	**yes**	**yes**	**1068**	**1145**	**3.56**	**0.15**	**0.0004**	**0.0143**	**1122**	**1174**	**3.34**	**0.14**	**0.0009**	**0.0199**
Caudal Middle Frontal Gyrus	no	no	1125	1175	-0.34	-0.01	0.7310	0.8253	1122	1182	-0.91	-0.04	0.3647	0.6382
Cuneus Cortex	no	no	1090	1143	-0.28	-0.01	0.7815	0.8548	1088	1132	1.02	0.04	0.3063	0.5795
Entorhinal Cortex	no	no	1080	1120	2.28	0.10	0.0230	0.2098	1074	1126	1.34	0.06	0.1792	0.4981
Frontal Pole	no	no	1133	1201	0.68	0.03	0.4993	0.7282	1140	1216	-0.45	-0.02	0.6547	0.7767
Fusiform Gyrus	no	no	1122	1176	-0.94	-0.04	0.3492	0.6267	1125	1190	-2.13	-0.09	0.0334	0.2340
Inferior Parietal Cortex	no	no	1093	1140	-0.86	-0.04	0.3917	0.6688	1106	1154	-1.33	-0.06	0.1850	0.4981
Inferior Temporal Gyrus	no	no	1066	1112	-0.46	-0.02	0.6452	0.7767	1102	1157	-0.77	-0.03	0.4416	0.7025
Insula	no	no	1074	1107	-1.64	-0.07	0.1004	0.4981	1041	1067	0.05	0.00	0.9576	0.9576
Isthmus-Cingulate Cortex	no	no	1153	1217	0.69	0.03	0.4901	0.7282	1145	1220	0.77	0.03	0.4395	0.7025
Lateral Occipital Cortex	no	no	1122	1171	0.46	0.02	0.6481	0.7767	1124	1175	-0.39	-0.02	0.6986	0.8150
Lateral Orbitofrontal Cortex	no	no	1148	1221	-1.03	-0.04	0.3024	0.5795	1091	1173	-2.34	-0.10	0.0193	0.2098
Lingual Gyrus	no	no	1137	1191	-1.47	-0.06	0.1410	0.4981	1110	1171	-0.68	-0.03	0.4973	0.7282
Medial Orbitofrontal Cortex	no	no	1122	1165	-0.62	-0.03	0.5385	0.7467	1106	1179	-1.35	-0.06	0.1765	0.4981
Middle Temporal Gyrus	no	no	1049	1086	-1.41	-0.06	0.1578	0.4981	1127	1160	-1.40	-0.06	0.1624	0.4981
Paracentral Lobule	no	no	1160	1221	-1.25	-0.05	0.2132	0.5330	1163	1220	-1.37	-0.06	0.1718	0.4981
Parahippocampal Gyrus	no	no	1157	1228	0.08	0.00	0.9325	0.9476	1162	1225	-1.19	-0.05	0.2344	0.5469

Pars Opercularis	no	no	1153	1210	0.08	0.00	0.9341	0.9476	1138	1194	-1.80	-0.07	0.0723	0.4600
Pars Orbitalis	no	no	1159	1223	-0.62	-0.03	0.5335	0.7467	1156	1208	-0.50	-0.02	0.6149	0.7690
Pars Triangularis	no	no	1155	1219	-0.83	-0.03	0.4079	0.6798	1141	1199	-2.21	-0.09	0.0270	0.2098
Pericalcarine Cortex	no	no	1100	1145	0.33	0.01	0.7433	0.8259	1067	1106	0.54	0.02	0.5904	0.7690
Postcentral Gyrus	no	no	1078	1124	-2.75	-0.12	0.0060	0.1043	1060	1113	-1.20	-0.05	0.2305	0.5469
Posterior-Cingulate Cortex	no	no	1139	1212	1.16	0.05	0.2470	0.5578	1149	1208	2.61	0.11	0.0091	0.1270
Precentral Gyrus	no	no	1099	1150	-1.33	-0.06	0.1838	0.4981	1095	1140	-1.61	-0.07	0.1068	0.4981
Precuneus Cortex	no	no	1153	1221	-1.06	-0.04	0.2882	0.5764	1143	1215	-0.60	-0.02	0.5507	0.7467
Rostral Anterior Cingulate Cortex	no	no	1095	1139	0.20	0.01	0.8384	0.8892	1107	1181	0.23	0.01	0.8151	0.8778
Rostral Middle Frontal Gyrus	no	no	1123	1187	-0.74	-0.03	0.4587	0.7136	1121	1189	-0.51	-0.02	0.6092	0.7690
Superior Frontal Gyrus	no	no	1120	1184	1.00	0.04	0.3164	0.5829	1134	1190	0.59	0.02	0.5547	0.7467
Superior Parietal Cortex	no	no	1108	1150	-0.17	-0.01	0.8668	0.9056	1098	1149	1.11	0.05	0.2671	0.5666
Superior Temporal Gyrus	no	no	1024	1064	-1.08	-0.05	0.2823	0.5764	1085	1138	-1.66	-0.07	0.0973	0.4981
Supramarginal Gyrus	no	no	1045	1075	-1.28	-0.06	0.2002	0.5191	1059	1104	-1.71	-0.07	0.0878	0.4981
Temporal Pole	no	no	1076	1134	-1.36	-0.06	0.1732	0.4981	1104	1180	-2.23	-0.09	0.0257	0.2098
Transverse Temporal Cortex	no	no	1153	1220	-0.50	-0.02	0.6152	0.7690	1153	1229	-0.37	-0.02	0.7119	0.8169
Mean Cortical Thickness	no	no	1169	1233	-1.11	-0.05	0.2656	0.5666	1168	1232	-1.39	-0.06	0.1651	0.4981
***N* significant (out of 35)**	**2**	**1**												

All statistical models included group, sex and age. Provided statistics are for the group effect. * indicates that this outcome showed a significant group difference in the main analysis (conduct disorder vs. typically-developing) after False Discovery Rate correction ($N = 2$). The columns referring to 'group effect right/left' indicate whether the region shows a significant difference between conduct disorder and typically-developing youth post False Discovery Rate adjustment when using the right and left hemisphere, respectively. A negative Cohen's *d* indicates conduct disorder < controls; a positive Cohen's *d* indicates conduct disorder > controls. Regions with a significant group difference post False Discovery Rate correction are indicated in bold. CD=conduct disorder. TD=typically-developing. d=Cohen's *d*. FDR=False Discovery Rate

Table S11: **Hemisphere-specific case-control differences in surface area**

			left hemisphere						right hemisphere					
Region	**group effect left**	**group effect right**	***n* CD**	***n* TD**	***t***	***d***	***p***	***p FDR***	***n* CD**	***n* TD**	***t***	***d***	***p***	***p FDR***
Banks of the Superior Temporal Sulcus*	**yes**	**yes**	**1067**	**1116**	**-2.46**	**-0.11**	**0.0139**	**0.0231**	**1122**	**1202**	**-2.68**	**-0.11**	**0.0075**	**0.0145**
Caudal Anterior Cingulate Cortex*	no	**yes**	1062	1143	-1.30	-0.06	0.1932	0.2253	**1125**	**1176**	**-3.74**	**-0.16**	**0.0002**	**0.0010**
Caudal Middle Frontal Gyrus*	**yes**	**yes**	**1126**	**1177**	**-2.81**	**-0.12**	**0.0050**	**0.0099**	**1126**	**1188**	**-3.12**	**-0.13**	**0.0018**	**0.0042**
Cuneus Cortex*	**yes**	no	**1091**	**1142**	**-2.56**	**-0.11**	**0.0106**	**0.0189**	1086	1125	-0.95	-0.04	0.3447	0.3830
Entorhinal Cortex*	**yes**	**yes**	**1078**	**1123**	**-2.44**	**-0.10**	**0.0150**	**0.0244**	**1072**	**1127**	**-2.17**	**-0.09**	**0.0303**	**0.0451**
Frontal Pole*	**yes**	**yes**	**1138**	**1199**	**-3.58**	**-0.15**	**0.0004**	**0.0014**	**1142**	**1214**	**-3.82**	**-0.16**	**0.0001**	**0.0008**
Fusiform Gyrus*	**yes**	**yes**	**1126**	**1173**	**-3.64**	**-0.15**	**0.0003**	**0.0012**	**1129**	**1191**	**-3.46**	**-0.14**	**0.0005**	**0.0017**

Inferior Parietal Cortex*	**yes**	**yes**	**1094**	**1143**	**-5.77**	**-0.24**	**< 0.0001**	**< 0.0001**	**1110**	**1158**	**-5.41**	**-0.23**	**< 0.0001**	**< 0.0001**
Inferior Temporal Gyrus*	**yes**	**yes**	**1069**	**1111**	**-3.45**	**-0.15**	**0.0006**	**0.0017**	**1107**	**1162**	**-5.00**	**-0.21**	**< 0.0001**	**< 0.0001**
Insula*	**yes**	no	**1073**	**1104**	**-3.51**	**-0.15**	**0.0005**	**0.0015**	1045	1065	-1.62	-0.07	0.1051	0.1367
Isthmus-Cingulate Cortex*	**yes**	**yes**	**1152**	**1218**	**-3.18**	**-0.13**	**0.0015**	**0.0037**	**1143**	**1220**	**-2.60**	**-0.11**	**0.0095**	**0.0180**
Lateral Occipital Cortex*	**yes**	**yes**	**1118**	**1170**	**-2.31**	**-0.10**	**0.0211**	**0.0321**	**1123**	**1176**	**-2.32**	**-0.10**	**0.0203**	**0.0321**
Lateral Orbitofrontal Cortex*	**yes**	**yes**	**1146**	**1218**	**-3.27**	**-0.13**	**0.0011**	**0.0029**	**1093**	**1172**	**-3.65**	**-0.15**	**0.0003**	**0.0012**
Lingual Gyrus*	no	**yes**	1137	1192	-1.87	-0.08	0.0615	0.0896	**1105**	**1174**	**-2.58**	**-0.11**	**0.0100**	**0.0184**
Medial Orbitofrontal Cortex	no	**yes**	1122	1163	0.56	0.02	0.5728	0.5985	**1109**	**1176**	**-3.32**	**-0.14**	**0.0009**	**0.0025**
Middle Temporal Gyrus*	**yes**	**yes**	**1053**	**1094**	**-5.63**	**-0.24**	**< 0.0001**	**< 0.0001**	**1134**	**1164**	**-4.29**	**-0.18**	**< 0.0001**	**0.0002**
Paracentral Lobule	no	no	1161	1217	-1.44	-0.06	0.1488	0.1827	1161	1220	-0.44	-0.02	0.6583	0.6775
Parahippocampal Gyrus*	**yes**	**yes**	**1159**	**1224**	**-3.13**	**-0.13**	**0.0018**	**0.0042**	**1162**	**1226**	**-3.59**	**-0.15**	**0.0003**	**0.0014**
Pars Opercularis	no	no	1150	1214	-0.74	-0.03	0.4622	0.4978	1134	1195	-1.47	-0.06	0.1419	0.1807
Pars Orbitalis*	**yes**	no	**1161**	**1225**	**-3.66**	**-0.15**	**0.0003**	**0.0012**	1157	1211	-1.43	-0.06	0.1533	0.1851
Pars Triangularis	**yes**	no	**1156**	**1222**	**-2.31**	**-0.09**	**0.0211**	**0.0321**	1149	1205	-1.05	-0.04	0.2930	0.3308
Pericalcarine Cortex	no	no	1099	1147	-1.62	-0.07	0.1055	0.1367	1064	1106	-0.60	-0.03	0.5482	0.5815
Postcentral Gyrus*	**yes**	**yes**	**1079**	**1126**	**-3.51**	**-0.15**	**0.0005**	**0.0015**	**1061**	**1115**	**-2.84**	**-0.12**	**0.0045**	**0.0094**
Posterior-Cingulate Cortex*	no	no	1135	1215	-1.72	-0.07	0.0864	0.1209	1148	1208	-1.67	-0.07	0.0957	0.1288
Precentral Gyrus*	**yes**	**yes**	**1100**	**1153**	**-3.41**	**-0.14**	**0.0007**	**0.0019**	**1096**	**1140**	**-2.91**	**-0.12**	**0.0036**	**0.0077**
Precuneus Cortex*	**yes**	**yes**	**1156**	**1221**	**-2.55**	**-0.10**	**0.0108**	**0.0189**	**1147**	**1217**	**-3.86**	**-0.16**	**0.0001**	**0.0007**
Rostral Anterior Cingulate Cortex	no	no	1094	1141	-1.69	-0.07	0.0907	0.1246	1109	1181	-1.72	-0.07	0.0852	0.1209
Rostral Middle Frontal Gyrus*	**yes**	**yes**	**1126**	**1189**	**-3.02**	**-0.13**	**0.0025**	**0.0057**	**1123**	**1192**	**-2.96**	**-0.12**	**0.0031**	**0.0068**
Superior Frontal Gyrus*	**yes**	**yes**	**1123**	**1190**	**-4.04**	**-0.17**	**0.0001**	**0.0004**	**1132**	**1190**	**-4.27**	**-0.18**	**< 0.0001**	**0.0002**
Superior Parietal Cortex*	no	**yes**	1107	1150	-1.36	-0.06	0.1731	0.2054	**1101**	**1151**	**-2.48**	**-0.10**	**0.0133**	**0.0228**
Superior Temporal Gyrus*	**yes**	**yes**	**1026**	**1067**	**-3.53**	**-0.15**	**0.0004**	**0.0015**	**1083**	**1140**	**-4.48**	**-0.19**	**< 0.0001**	**0.0001**
Supramarginal Gyrus*	**yes**	no	**1047**	**1082**	**-3.29**	**-0.14**	**0.0010**	**0.0027**	1061	1107	-1.46	-0.06	0.1454	0.1818
Temporal Pole	no	no	1079	1133	-0.43	-0.02	0.6678	0.6775	1109	1183	-0.40	-0.02	0.6893	0.6893
Transverse Temporal Cortex	**yes**	**yes**	**1169**	**1234**	**-5.99**	**-0.24**	**< 0.0001**	**< 0.0001**	**1170**	**1234**	**-5.91**	**-0.24**	**< 0.0001**	**< 0.0001**
Total Surface Area*	no	no	1156	1219	-1.20	-0.05	0.2291	0.2629	1158	1225	-0.86	-0.04	0.3875	0.4239
***N* significant (out of 35)**	**24**	**23**												

All statistical models included group, sex, age and total intracranial volume (except for total surface area). Provided statistics are for the group effect. * indicates that this outcome showed a significant group difference in the main analysis (conduct disorder vs. typically-developing) after False Discovery Rate correction (N = 27). The columns referring to 'group effect right/left' indicate whether the region shows a significant difference between conduct disorder and typically-developing youth post False Discovery Rate adjustment when using the right or left hemisphere, respectively. A negative Cohen's *d* indicates conduct disorder < controls; a positive Cohen's *d* indicates conduct disorder > controls. Regions with a significant group difference post False Discovery Rate correction are indicated in bold. CD= conduct disorder. TD=typically-developing. *d*=Cohen's *d*. FDR=False Discovery Rate.

Table S12: **Hemisphere-specific case-control differences in subcortical volumes**

Region	group effect left	group effect right	left hemisphere						right hemisphere					
			***n* CD**	***n* TD**	***t***	***d***	***p***	***p FDR***	***n* CD**	***n* TD**	***t***	***d***	***p***	***p FDR***
Amygdala*	no	**yes**	1133	1198	-1.95	-0.08	0.0511	0.0928	**1143**	**1196**	**-3.51**	**-0.15**	**0.0005**	**0.0035**
Caudate	no	no	1131	1172	-0.86	-0.04	0.3903	0.4879	1060	1091	-1.11	-0.05	0.2692	0.3671
Hippocampus*	**yes**	**yes**	**1149**	**1215**	**-2.47**	**-0.10**	**0.0135**	**0.0338**	**1166**	**1225**	**-3.08**	**-0.13**	**0.0021**	**0.0085**
Nucleus Accumbens*	**yes**	no	**1155**	**1210**	**-3.06**	**-0.13**	**0.0023**	**0.0085**	1164	1240	-1.91	-0.08	0.0557	0.0928
Pallidum	no	no	1080	1106	0.22	0.01	0.8246	0.8710	1162	1226	-0.16	-0.01	0.8710	0.8710
Putamen	no	no	1059	1096	-1.25	-0.05	0.2127	0.3190	1143	1198	-0.52	-0.02	0.6045	0.6975
Thalamus*	**yes**	**yes**	**1111**	**1175**	**-3.53**	**-0.15**	**0.0004**	**0.0035**	**1130**	**1198**	**-2.65**	**-0.11**	**0.0080**	**0.0240**
***N* significant (out of 8)**	**3**	**3**												

All statistical models included group, sex, age and total intracranial volume. Provided statistics are for the group effect. Results for total intracranial volume are not included as this value cannot be split by hemisphere. Results for this outcome can be found in Table S6 (main findings). *indicates that this outcome showed a significant group difference in the main analysis (conduct disorder vs. typically-developing) after False Discovery Rate correction ($N = 4$). The columns referring to 'group effect left/right' indicate whether the region shows a significant difference between conduct disorder and typically-developing youth post False Discovery Rate adjustment when using the right or left hemisphere, respectively. A negative Cohen's *d* indicates conduct disorder < controls; a positive Cohen's *d* indicates conduct disorder > controls. Regions with a significant group difference post False Discovery Rate correction are indicated in bold. CD=conduct disorder. TD=typically-developing. *d*=Cohen's *d*. FDR=False Discovery Rate

Appendix 8 - Additional information on ComBat and comparisons to analyses without ComBat adjustment

As evidence suggests that adjusting for site/scanner effects using Radua and colleagues' modified ComBat function is more powerful than traditional meta-analysis or mixed effect model mega-analyses that include site as a random effect,[47] we used this approach prior to performing the statistical analyses to adjust for site/scanner related heterogeneity. ComBat models included group (CD vs. TD), age, and sex for the adjustment of global outcomes and additionally included (post ComBat) total intracranial volume for regional outcomes. We adjusted across 61 scanners from 15 studies. To assess the validity of applying the modified ComBat functions to correct for scanner differences, we performed additional analyses, which are outlined below. Overall, these analyses supported the results derived from our main analyses.

Firstly, we tested for scanner effects in the data prior to and after ComBat adjustment by including scanner as an additional variable in the linear model (61 scanners were included overall). Prior to ComBat adjustment, most outcomes showed a significant scanner effect. After ComBat adjustment, no outcome showed a significant scanner effect. Relatedly, due to concerns surrounding compatibility between FreeSurfer versions,[52] we also explored whether FreeSurfer version (5.3 vs. 6.0) was a significant predictor post ComBat adjustment and whether including the FreeSurfer version in the model impacted the group differences identified in the main analyses. All previously identified group differences remained significant and FreeSurfer version was not a significant predictor for any outcome, except pallidum volume (prior to FDR correction). However, CD and TD youth did not differ in pallidum volume regardless of whether FreeSurfer version was included in the model or not. Hence, our findings suggest that differences between FreeSurfer versions seem to have been successfully adjusted by ComBat and that remaining differences did not influence our main findings.

Secondly, we performed the main analysis on the data prior to ComBat adjustment without including site/scanner in the model. The results were highly overlapping with the main analyses with only some minor deviations for cortical thickness and subcortical volumes. For example, the two regions that significantly differed between CD and TD youth in thickness (banks of the superior temporal sulcus and caudal anterior cingulate cortex) were only significant at the nominal significance level but did not survive FDR correction. For subcortical volume, an additional region became significant with the CD group additionally showing lower putamen volume compared to TD group.

Lastly, we performed the main analysis on the data prior to ComBat adjustment and added scanner as an additional (dummy-coded) variable in the linear model. The results of these analyses were 100% overlapping with the main findings reported in the manuscript (i.e., those using ComBat adjustment).

The full results of the described analyses are available on request.

Appendix 9 - Group-by-sex and group-by-age interactions

To explore whether differences between youths with CD and TD participants were moderated by sex or age, we separately added a group-by-sex or group-by-age interaction term into the main statistical model. Notably, while the group-by-sex analyses were pre-registered and accompanied by specific hypotheses (see main manuscript), the group-by-age analyses were investigated on an exploratory basis. Following multiple comparison correction, there was no significant group-by-sex or group-by-age interaction effect for any outcome, suggesting that sex and age did not moderate group differences in brain structure.

Table S13: **Group-by-sex and group-by-age interactions in cortical thickness**

Region	Group*Sex interaction								Group*Age interaction			
	n TD		*n* CD		*t*	*p*	*p* FDR	sr^2	*t*	*p*	*p* FDR	sr^2
	male	female	male	female								
Banks of the Superior Temporal Sulcus*	790	437	830	331	0.82	0.4116	0.8794	< 0.01	0.77	0.4423	0.9969	<.01
Caudal Anterior Cingulate Cortex*	788	439	829	330	-0.38	0.7037	0.8794	< 0.01	-0.63	0.5266	0.9969	<.01
Caudal Middle Frontal Gyrus	784	440	827	329	1.05	0.293	0.8794	< 0.01	0.57	0.5686	0.9969	<.01
Cuneus Cortex	773	437	823	328	1.09	0.2758	0.8794	< 0.01	-0.96	0.3370	0.9969	<.01
Entorhinal Cortex	774	425	822	321	-1.07	0.2868	0.8794	< 0.01	2.83	0.0046	0.1622	<.01
Frontal Pole	788	440	826	330	0.58	0.5603	0.8794	< 0.01	-1.35	0.1762	0.9969	<.01
Fusiform Gyrus	776	439	828	327	-0.38	0.7016	0.8794	< 0.01	1.48	0.1398	0.9969	<.01
Inferior Parietal Cortex	770	435	829	329	1.14	0.2558	0.8794	< 0.01	-0.50	0.6167	0.9969	<.01
Inferior Temporal Gyrus	770	428	802	327	-0.74	0.457	0.8794	< 0.01	0.93	0.3503	0.9969	<.01
Insula	763	428	818	330	-2.07	0.0383	0.8794	< 0.01	0.14	0.8911	0.9969	<.01
Isthmus-Cingulate Cortex	794	440	835	330	1.5	0.1342	0.8794	< 0.01	-2.03	0.0429	0.7507	<.01
Lateral Occipital Cortex	778	432	822	329	1.13	0.2596	0.8794	< 0.01	-0.01	0.9924	0.9969	<.01
Lateral Orbitofrontal Cortex	791	439	828	332	0.06	0.9533	0.9887	< 0.01	0.02	0.9826	0.9969	<.01
Lingual Gyrus	783	435	831	332	-0.01	0.9934	0.9934	< 0.01	0.34	0.7359	0.9969	<.01
Medial Orbitofrontal Cortex	785	437	824	329	0.95	0.3408	0.8794	< 0.01	-0.66	0.5093	0.9969	<.01
Middle Temporal Gyrus	778	430	821	326	-0.85	0.3978	0.8794	< 0.01	-0.21	0.8349	0.9969	<.01
Paracentral Lobule	793	441	838	332	-0.53	0.5953	0.8794	< 0.01	0.73	0.4641	0.9969	<.01
Parahippocampal Gyrus	794	440	838	332	-1.81	0.0708	0.8794	< 0.01	-1.66	0.0969	0.9969	<.01
Pars Opercularis	791	441	838	332	-0.05	0.9605	0.9887	< 0.01	0.29	0.7723	0.9969	<.01
Pars Orbitalis	791	441	836	332	0.73	0.4685	0.8794	< 0.01	-0.89	0.3714	0.9969	<.01
Pars Triangularis	790	440	837	332	0.77	0.4407	0.8794	< 0.01	-0.52	0.6051	0.9969	<.01

Pericalcarine Cortex	761	432	818	326	0.56	0.5734	0.8794	< 0.01	-0.61	0.5445	0.9969	<.01
Postcentral Gyrus	750	431	807	328	-0.28	0.7789	0.8794	< 0.01	-0.13	0.8972	0.9969	<.01
Posterior-Cingulate Cortex	793	439	836	332	0.62	0.5376	0.8794	< 0.01	-0.33	0.7384	0.9969	<.01
Precentral Gyrus	767	433	818	327	-0.63	0.5309	0.8794	< 0.01	-0.39	0.6958	0.9969	<.01
Precuneus Cortex	793	441	836	330	0.38	0.7061	0.8794	< 0.01	0.64	0.5205	0.9969	<.01
Rostral Anterior Cingulate Cortex	791	440	829	332	-0.32	0.7479	0.8794	< 0.01	-1.08	0.2821	0.9969	<.01
Rostral Middle Frontal Gyrus	784	440	821	329	0.32	0.7502	0.8794	< 0.01	-0.24	0.8067	0.9969	<.01
Superior Frontal Gyrus	777	440	823	331	0.31	0.7538	0.8794	< 0.01	0.78	0.4366	0.9969	<.01
Superior Parietal Cortex	766	433	822	328	1.35	0.177	0.8794	< 0.01	-0.07	0.9452	0.9969	<.01
Superior Temporal Gyrus	767	419	815	322	0.32	0.7509	0.8794	< 0.01	0.00	0.9969	0.9969	<.01
Supramarginal Gyrus	761	431	817	327	1.28	0.199	0.8794	< 0.01	0.78	0.4349	0.9969	<.01
Temporal Pole	779	425	818	320	-0.74	0.461	0.8794	< 0.01	0.27	0.7842	0.9969	<.01
Transverse Temporal Cortex	794	442	836	332	0.65	0.5183	0.8794	< 0.01	0.44	0.6609	0.9969	<.01
Mean Cortical Thickness	792	442	837	332	0.25	0.8057	0.8812	< 0.01	-0.10	0.9231	0.9969	<.01

All statistical models included group (conduct disorder vs. typically-developing), sex, age and a group-by-sex or group-by-age interaction term. Depicted statistics are for the respective interaction term. Effect sizes are indicated as squared semi-partial correlations. * indicates that this outcome showed a significant group difference in the main analysis (conduct disorder vs. typically-developing) after False Discovery Rate correction (N = 2). TD=typically-developing controls. CD=conduct disorder group. sr^2=squared semi-partial correlation coefficient.

***Table S14*: Group-by-sex and group-by-age interactions in surface area**

Region	**Group-by-sex interaction**								**Group*Age interaction**			
	***n* TD**		***n* CD**		***t***	***p***	***p* FDR**	***sr*²**	***t***	***p***	***p* FDR**	***sr*²**
	male	**female**	**male**	**female**								
Banks of the Superior Temporal Sulcus*	790	437	831	331	-0.08	0.935	0.9745	< 0.01	-1.69	0.0907	0.4534	<.01
Caudal Anterior Cingulate Cortex*	788	439	829	331	-0.2	0.844	0.9745	< 0.01	-0.08	0.9332	0.9697	<.01
Caudal Middle Frontal Gyrus*	785	440	827	329	0.48	0.6325	0.9609	< 0.01	-0.79	0.4308	0.9131	<.01
Cuneus Cortex*	772	437	821	328	0.65	0.5146	0.9609	< 0.01	-2.08	0.0375	0.4374	<.01
Entorhinal Cortex*	776	425	822	321	-0.16	0.8762	0.9745	< 0.01	-1.53	0.1256	0.5494	<.01
Frontal Pole*	788	440	826	332	0.07	0.9441	0.9745	< 0.01	-0.74	0.4612	0.9131	<.01
Fusiform Gyrus*	775	439	830	327	0.33	0.7413	0.9609	< 0.01	-0.37	0.7126	0.9697	<.01
Inferior Parietal Cortex*	773	435	829	330	0.7	0.486	0.9609	< 0.01	0.35	0.7266	0.9697	<.01
Inferior Temporal Gyrus*	770	428	805	327	-0.53	0.5956	0.9609	< 0.01	-0.13	0.8942	0.9697	<.01

Insula*	763	428	818	330	-0.5	0.6155	0.9609	< 0.01	-1.28	0.1990	0.6130	<.01
Isthmus-Cingulate Cortex*	794	440	835	330	-0.2	0.8444	0.9745	< 0.01	-0.69	0.4923	0.9131	<.01
Lateral Occipital Cortex*	777	433	821	329	0.85	0.3936	0.9609	< 0.01	-1.30	0.1922	0.6130	<.01
Lateral Orbitofrontal Cortex*	791	439	829	332	-0.86	0.3918	0.9609	< 0.01	0.28	0.7774	0.9697	<.01
Lingual Gyrus*	783	435	828	332	0.86	0.3875	0.9609	< 0.01	-2.86	0.0043	0.1498	<.01
Medial Orbitofrontal Cortex	784	437	825	329	-0.83	0.4082	0.9609	< 0.01	0.25	0.7998	0.9697	<.01
Middle Temporal Gyrus*	781	430	825	327	-0.53	0.5944	0.9609	< 0.01	-1.16	0.2481	0.6678	<.01
Paracentral Lobule	792	441	838	332	-0.48	0.6317	0.9609	< 0.01	-1.85	0.0648	0.4534	<.01
Parahippocampal Gyrus*	793	440	838	332	1.75	0.0798	0.9609	< 0.01	-0.38	0.7045	0.9697	<.01
Pars Opercularis	792	441	837	332	0.36	0.7182	0.9609	< 0.01	-0.68	0.4957	0.9131	<.01
Pars Orbitalis*	792	441	836	332	-1.48	0.1381	0.9609	< 0.01	-1.40	0.1605	0.6130	<.01
Pars Triangularis	792	440	836	332	-1.88	0.0604	0.9609	< 0.01	-0.84	0.4025	0.9131	<.01
Pericalcarine Cortex	762	432	817	326	-0.53	0.5936	0.9609	< 0.01	-2.23	0.0259	0.4374	<.01
Postcentral Gyrus*	754	431	808	328	0.36	0.716	0.9609	< 0.01	-0.45	0.6533	0.9697	<.01
Posterior-Cingulate Cortex*	793	440	835	332	-0.35	0.7275	0.9609	< 0.01	-0.21	0.8329	0.9697	<.01
Precentral Gyrus*	768	434	817	327	-0.03	0.9745	0.9745	< 0.01	-0.35	0.7267	0.9697	<.01
Precuneus Cortex*	794	441	836	331	0.52	0.6012	0.9609	< 0.01	-0.06	0.9552	0.9697	<.01
Rostral Anterior Cingulate Cortex	790	440	829	332	-0.81	0.4181	0.9609	< 0.01	0.08	0.9333	0.9697	<.01
Rostral Middle Frontal Gyrus*	787	440	822	330	-0.67	0.5041	0.9609	< 0.01	0.26	0.7946	0.9697	<.01
Superior Frontal Gyrus*	779	440	822	331	-1.5	0.1329	0.9609	< 0.01	-0.04	0.9697	0.9697	<.01
Superior Parietal Cortex*	766	433	820	328	0.98	0.3275	0.9609	< 0.01	-0.36	0.7223	0.9697	<.01
Superior Temporal Gyrus*	769	420	816	322	0.72	0.4738	0.9609	< 0.01	-1.93	0.0538	0.4534	<.01
Supramarginal Gyrus*	765	431	818	328	0.08	0.9368	0.9745	< 0.01	1.25	0.2102	0.6130	<.01
Temporal Pole	780	425	819	320	-1.64	0.1017	0.9609	< 0.01	-1.75	0.0799	0.4534	<.01
Transverse Temporal Cortex	793	442	836	332	0.04	0.9681	0.9745	< 0.01	-0.83	0.4049	0.9131	<.01
Total Surface Area*	793	441	838	332	-0.94	0.3489	0.9609	< 0.01	-0.05	0.9640	0.9697	<.01

All statistical models included group (conduct disorder vs. typically-developing), sex, age, total intracranial volume (except in the case of total surface area), and a group-by-sex or group-by-age interaction term. Depicted statistics are for the respective interaction term. Effect sizes are indicated as squared semi-partial correlations. * indicates that this outcome showed a significant group difference in the main analysis (conduct disorder vs. typically-developing) after False Discovery Rate correction (N = 27). TD=typically-developing controls. CD= conduct disorder group. sr^2=squared semi-partial correlation coefficient.

***Table S15*: Group-by-sex and group-by-age interactions in subcortical volume and total intracranial volume**

Region	Group*Sex interaction								Group*Age interaction			
	n TD		*n* CD		*t*	*p*	*p* FDR	sr^2	*t*	*p*	*p* FDR	sr^2
	male	female	male	female								
Amygdala*	790	443	835	338	-0.43	0.6691	0.8760	< 0.01	-2.25	0.0242	0.1776	<.01
Caudate	777	425	825	326	0.12	0.9036	0.9036	< 0.01	0.46	0.6474	0.6482	<.01
Hippocampus*	803	445	842	338	-0.98	0.3252	0.8448	< 0.01	-1.73	0.0835	0.2227	<.01
Nucleus Accumbens*	803	446	842	338	-0.63	0.5258	0.8448	< 0.01	-0.63	0.5313	0.6482	<.01
Pallidum	801	440	841	333	0.30	0.7665	0.8760	< 0.01	-1.09	0.2779	0.5557	<.01
Putamen	791	437	839	323	1.13	0.2589	0.8448	< 0.01	-2.01	0.0444	0.1776	<.01
Thalamus*	794	434	821	335	-0.63	0.528	0.8448	< 0.01	-0.86	0.3921	0.6274	<.01
Total Intracranial Volume	807	446	846	339	-1.10	0.2728	0.8448	< 0.01	0.46	0.6482	0.6482	<.01

All statistical models included group (conduct disorder vs. typically-developing), sex, age, total intracranial volume (except in the case of total intracranial volume as the outcome), and a group-by-sex or group-by-age interaction term. Depicted statistics are for the respective interaction term. Effect sizes are indicated as squared semi-partial correlations. * indicates that this outcome showed a significant group difference in the main analysis (conduct disorder vs. typically-developing) after False Discovery Rate correction (N = 4). TD=typically-developing controls. CD=conduct disorder group. sr^2=squared semi-partial correlation coefficient.

Appendix 10 - Full results of the subgroup analyses: CD age-of-onset

***Table S16*: Age-of-onset subgroup effects on cortical thickness**

	N			*F*	*p*	*p* FDR
Region	**TD**	**CO-CD**	**AO-CD**			
Banks of the Superior Temporal Sulcus*	1227	454	279	5.18	0.0057	0.1001
Caudal Anterior Cingulate Cortex*[a]	**1227**	**454**	**278**	**6.65**	**0.0013**	**0.0464**
Caudal Middle Frontal Gyrus	1224	454	278	0.01	0.9945	0.9945
Cuneus Cortex	1210	451	277	0.91	0.4012	0.6660
Entorhinal Cortex	1199	448	276	3.64	0.0263	0.3073
Frontal Pole	1228	453	279	0.91	0.4033	0.6660
Fusiform Gyrus	1215	454	277	1.25	0.2874	0.5589
Inferior Parietal Cortex	1205	452	280	0.18	0.8332	0.8837
Inferior Temporal Gyrus	1198	436	279	0.27	0.7652	0.8482
Insula	1191	455	276	1.76	0.1722	0.4892
Isthmus-Cingulate Cortex	1234	457	280	1.60	0.2018	0.4892
Lateral Occipital Cortex	1210	453	279	0.09	0.9171	0.9441
Lateral Orbitofrontal Cortex	1230	457	280	2.42	0.0891	0.4457
Lingual Gyrus	1218	456	278	1.91	0.1490	0.4892
Medial Orbitofrontal Cortex	1222	453	279	1.58	0.2067	0.4892
Middle Temporal Gyrus	1208	445	279	0.63	0.5328	0.6660
Paracentral Lobule	1234	458	282	2.54	0.0790	0.4457
Parahippocampal Gyrus	1234	458	282	0.69	0.5019	0.6660
Pars Opercularis	1232	458	282	0.68	0.5058	0.6660
Pars Orbitalis	1232	457	282	0.74	0.4787	0.6660
Pars Triangularis	1230	458	281	3.05	0.0478	0.4179
Pericalcarine Cortex	1193	449	275	2.18	0.1131	0.4892
Postcentral Gyrus	1181	444	273	1.56	0.2097	0.4892
Posterior-Cingulate Cortex	1232	458	281	2.68	0.0691	0.4457
Precentral Gyrus	1200	449	274	1.40	0.2463	0.5070
Precuneus Cortex	1234	456	281	0.25	0.7755	0.8482

Rostral Anterior Cingulate Cortex	1231	455	277	1.49	0.2253	0.4929
Rostral Middle Frontal Gyrus	1224	449	276	0.74	0.4787	0.6660
Superior Frontal Gyrus	1217	457	280	0.64	0.5297	0.6660
Superior Parietal Cortex	1199	452	277	0.47	0.6273	0.7571
Superior Temporal Gyrus	1186	444	276	0.80	0.4474	0.6660
Supramarginal Gyrus	1192	447	275	1.59	0.2034	0.4892
Temporal Pole	1204	449	273	1.95	0.1422	0.4892
Transverse Temporal Cortex	1236	458	281	0.72	0.4862	0.6660
Mean Cortical Thickness	1234	456	282	0.39	0.6750	0.7875

All statistical models included group (typically-developing vs. childhood-onset conduct disorder vs. adolescent-onset conduct disorder), sex and age. The *F* statistic indicates the overall group effect. Regions with a significant group effect post False Discovery Rate correction are indicated in bold. * indicates that this outcome showed a significant group difference in the main analysis (conduct disorder vs. typically-developing) after False Discovery Rate correction ($N = 2$). [a] The pairwise comparison for the only region (caudal anterior cingulate cortex) with a significant subgroup effect on CT indicated that the childhood-onset group showed significantly higher thickness in this region compared to both controls and the adolescent-onset group. After Bonferroni correction for multiple comparisons, only the difference to the controls remained significant. TD=typically-developing controls. CO-CD=childhood-onset conduct disorder. AO-CD=adolescent-onset conduct disorder

***Table S17*: Age-of-onset subgroup effects on surface area**

Region	*N*			*F*	*p*	*p* FDR	pairwise comparisons	pairwise comparisons Bonferroni adjusted
	TD	CO-CD	AO-CD					
Banks of the Superior Temporal Sulcus*	1227	455	279	1.98	0.1377	0.1662		
Caudal Anterior Cingulate Cortex*	1227	455	278	2.06	0.1280	0.1659		
Caudal Middle Frontal Gyrus*	1225	453	279	3.42	0.0328	0.0592		
Cuneus Cortex*	1209	449	277	2.57	0.0765	0.1162		
Entorhinal Cortex*	**1201**	**448**	**276**	**4.05**	**0.0176**	**0.0402**	**AO-CD < TD**	**AO-CD < TD**
Frontal Pole*	**1228**	**455**	**279**	**10.95**	**< 0.0001**	**0.0002**	**CO-CD & AO-CD < TD**	**CO-CD & AO-CD < TD**
Fusiform Gyrus*	1214	456	277	3.41	0.0331	0.0592		
Inferior Parietal Cortex*	**1208**	**453**	**280**	**11.94**	**< 0.0001**	**0.0002**	**CO-CD & AO-CD < TD**	**CO-CD & AO-CD < TD**
Inferior Temporal Gyrus*	**1198**	**439**	**279**	**5.79**	**0.0031**	**0.0099**	**CO-CD & AO-CD < TD**	**AO-CD < TD**
Insula*	**1191**	**456**	**276**	**5.37**	**0.0047**	**0.0138**	**AO-CD < TD & CO-CD**	**AO-CD < TD**
Isthmus-Cingulate Cortex*	**1234**	**457**	**280**	**4.85**	**0.0079**	**0.0214**	**CO-CD & AO-CD < TD**	**AO-CD < TD**
Lateral Occipital Cortex*	1210	452	279	2.28	0.1023	0.1377		
Lateral Orbitofrontal Cortex*	**1230**	**458**	**280**	**6.22**	**0.0020**	**0.0071**	**CO-CD & AO-CD < TD**	**CO-CD & AO-CD < TD**
Lingual Gyrus*	1218	454	278	1.90	0.1505	0.1755		
Medial Orbitofrontal Cortex	1221	454	279	0.87	0.4176	0.4567		
Middle Temporal Gyrus*	**1211**	**450**	**279**	**10.00**	**< 0.0001**	**0.0004**	**CO-CD & AO-CD < TD**	**CO-CD & AO-CD < TD**
Paracentral Lobule	1233	458	282	1.12	0.3281	0.3705		
Parahippocampal Gyrus*	**1233**	**458**	**282**	**6.40**	**0.0017**	**0.0071**	**CO-CD < TD**	**CO-CD < TD**
Pars Opercularis	1233	458	282	2.41	0.0905	0.1267		
Pars Orbitalis*	**1233**	**457**	**282**	**3.92**	**0.0200**	**0.0411**	**AO-CD < TD**	**AO-CD < TD**
Pars Triangularis	1232	457	281	2.53	0.0797	0.1162		
Pericalcarine Cortex	1194	448	275	0.43	0.6504	0.6899		
Postcentral Gyrus*	**1185**	**445**	**273**	**7.90**	**0.0004**	**0.0022**	**AO-CD < TD & CO-CD**	**AO-CD < TD**
Posterior-Cingulate Cortex*	1233	458	281	2.01	0.1336	0.1662		
Precentral Gyrus*	1202	448	274	3.39	0.0339	0.0592		
Precuneus Cortex*	**1235**	**457**	**281**	**6.30**	**0.0019**	**0.0071**	**AO-CD < TD & CO-CD**	**AO-CD < TD**
Rostral Anterior Cingulate Cortex	1230	455	277	0.35	0.7030	0.7030		

Rostral Middle Frontal Gyrus*	1227	451	276	2.73	0.0657	0.1046		
Superior Frontal Gyrus*	**1219**	**456**	**280**	**6.72**	**0.0012**	**0.0062**	**CO-CD & AO-CD < TD**	**CO-CD & AO-CD < TD**
Superior Parietal Cortex*	1199	451	277	3.04	0.0482	0.0804		
Superior Temporal Gyrus*	**1189**	**445**	**276**	**7.93**	**0.0004**	**0.0022**	**CO-CD & AO-CD < TD**	**AO-CD < TD**
Supramarginal Gyrus*	**1196**	**449**	**275**	**4.01**	**0.0184**	**0.0402**	**CO-CD & AO-CD < TD**	
Temporal Pole	1205	450	273	0.39	0.6743	0.6941		
Transverse Temporal Cortex	**1235**	**458**	**281**	**4.51**	**0.0111**	**0.0278**	**AO-CD < TD & CO-CD**	**AO-CD < CO-CD & TD**
Total Surface Area*	**1234**	**457**	**282**	**11.35**	**< 0.0001**	**0.0002**	**CO-CD & AO-CD < TD**	**CO-CD & AO-CD < TD**

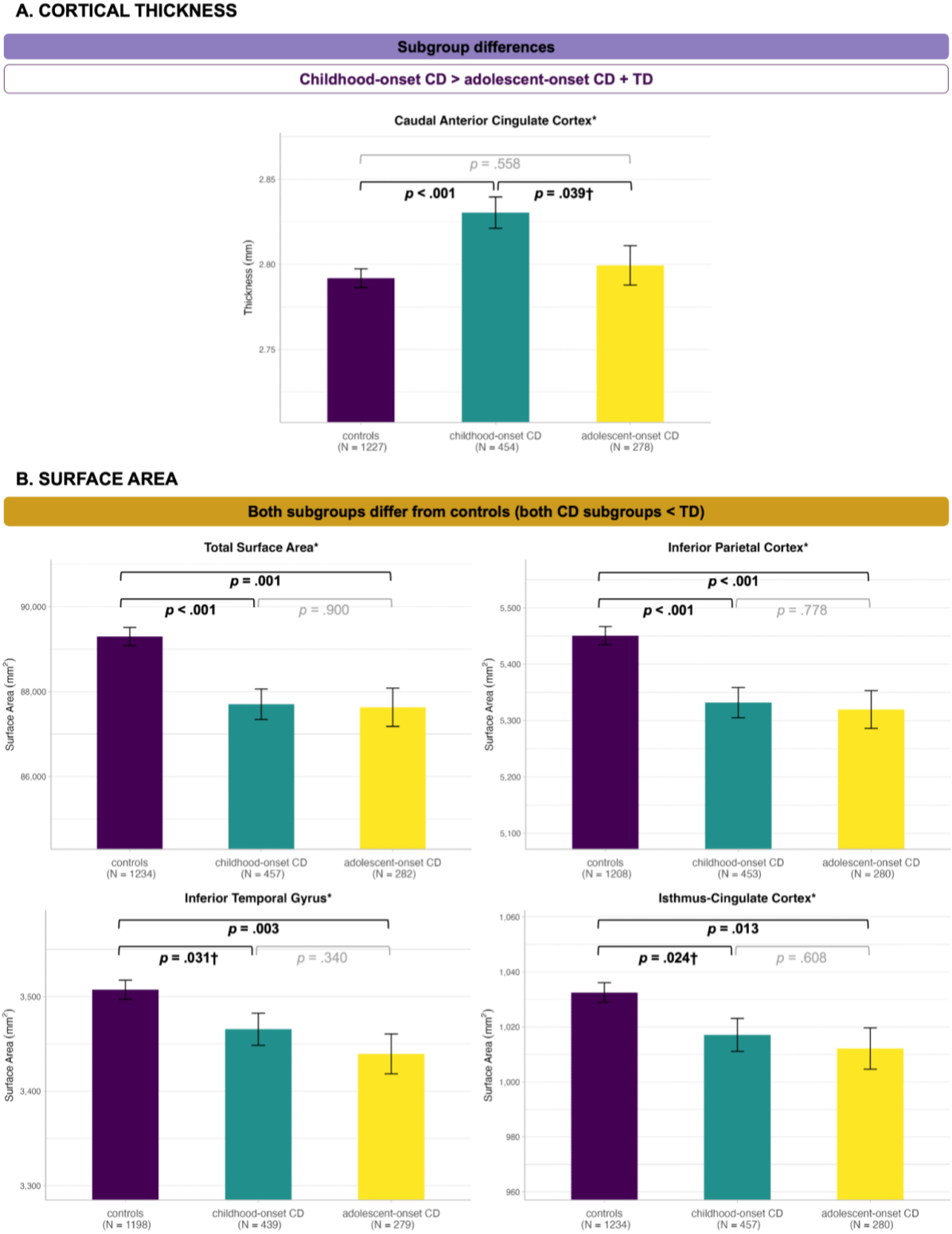
All statistical models included group (typically-developing vs. childhood-onset conduct disorder vs. adolescent-onset conduct disorder), sex, age, and total intracranial volume (except in the case of total surface area). The *F* statistic indicates the overall group effect. Regions with a significant group effect post False Discovery Rate correction are indicated in bold. * indicates that this outcome showed a significant group difference in the main analysis (conduct disorder vs. typically-developing) after False Discovery Rate correction (*N* = 27). TD=typically-developing controls. CO-CD=childhood-onset conduct disorder. AO-CD=adolescent-onset conduct disorder.

***Table S18*: Age-of-onset subgroup effects on subcortical volumes and total intracranial volume**

	N			***F***	***p***	***p* FDR**	**pairwise comparisons**	**pairwise comparisons Bonferroni adjusted**
Region	**TD**	**CO-CD**	**AO-CD**					
Amygdala*	**1233**	**457**	**278**	**4.70**	**0.0092**	**0.0368**	**AO-CD < TD**	**AO-CD < TD**
Caudate	**1202**	**445**	**272**	**4.22**	**0.0148**	**0.0395**	**CO-CD < AO-CD & TD**	**CO-CD < AO-CD**
Hippocampus*	**1248**	**455**	**282**	**6.06**	**0.0024**	**0.0190**	**AO-CD < CO-CD & TD**	**AO-CD < TD**
Nucleus Accumbens*	1249	457	282	3.68	0.0255	0.0510		
Pallidum	1241	454	281	0.21	0.8089	0.8835		
Putamen	1228	450	272	0.12	0.8835	0.8835		
Thalamus*	1228	442	272	3.44	0.0323	0.0517		
Total Intracranial Volume	1253	458	283	1.52	0.2199	0.2932		

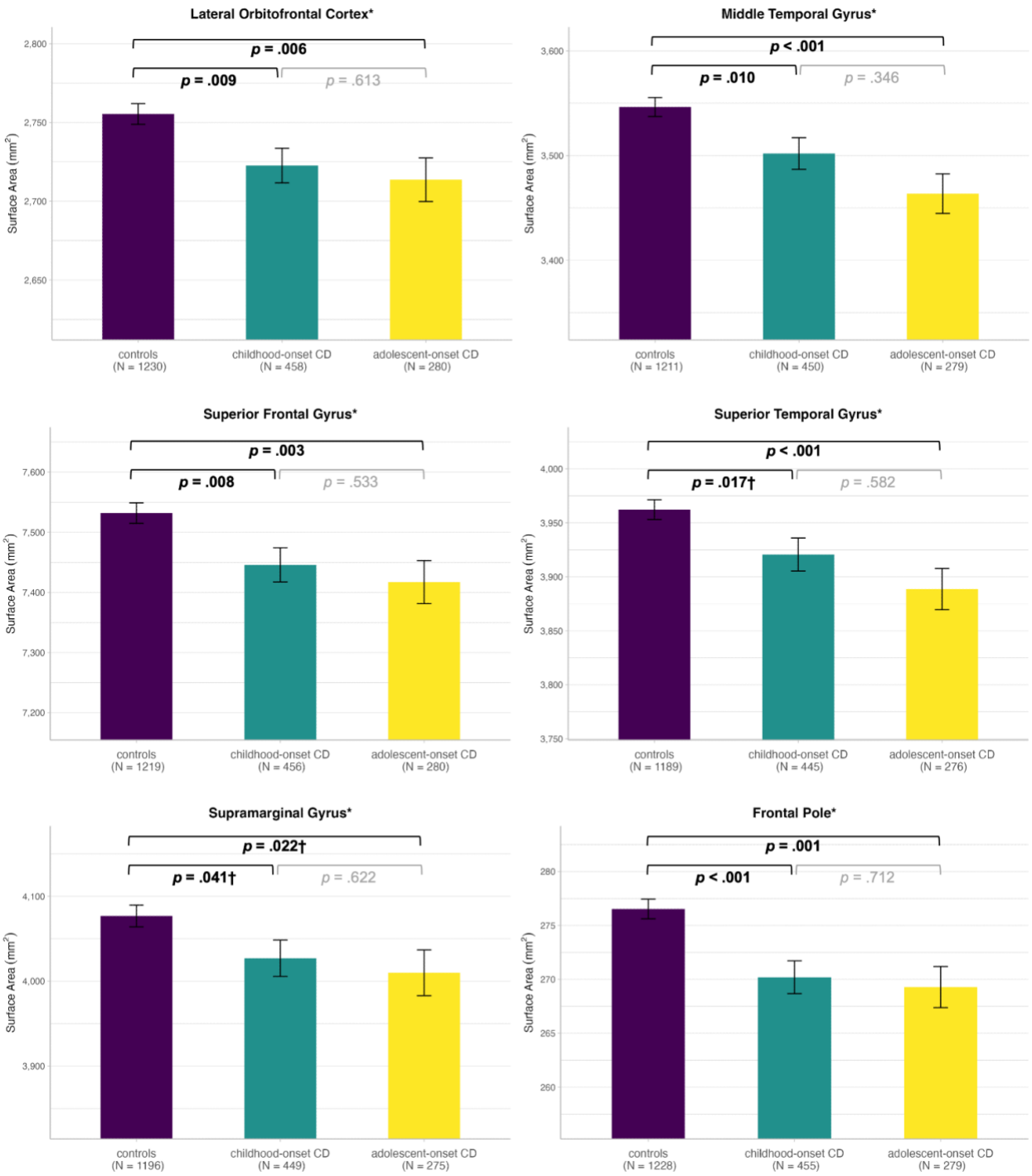
All statistical models included group (typically-developing vs. childhood-onset conduct disorder vs. adolescent-onset conduct disorder), sex, age, and total intracranial volume (except in the case of total intracranial volume as the outcome). The *F* statistic indicates the overall group effect. Regions with a significant group effect post False Discovery Rate correction are indicated in bold. * indicates that this outcome showed a significant group difference in the main analysis (conduct disorder vs. typically-developing) after False Discovery Rate correction (*N* = 4). TD=typically-developing controls. CO-CD=childhood-onset conduct disorder. AO-CD=adolescent-onset conduct disorder.

Figure S27-1: **Mean plots depicting pairwise comparisons for regions with a significant age-of-onset subgroup effect**



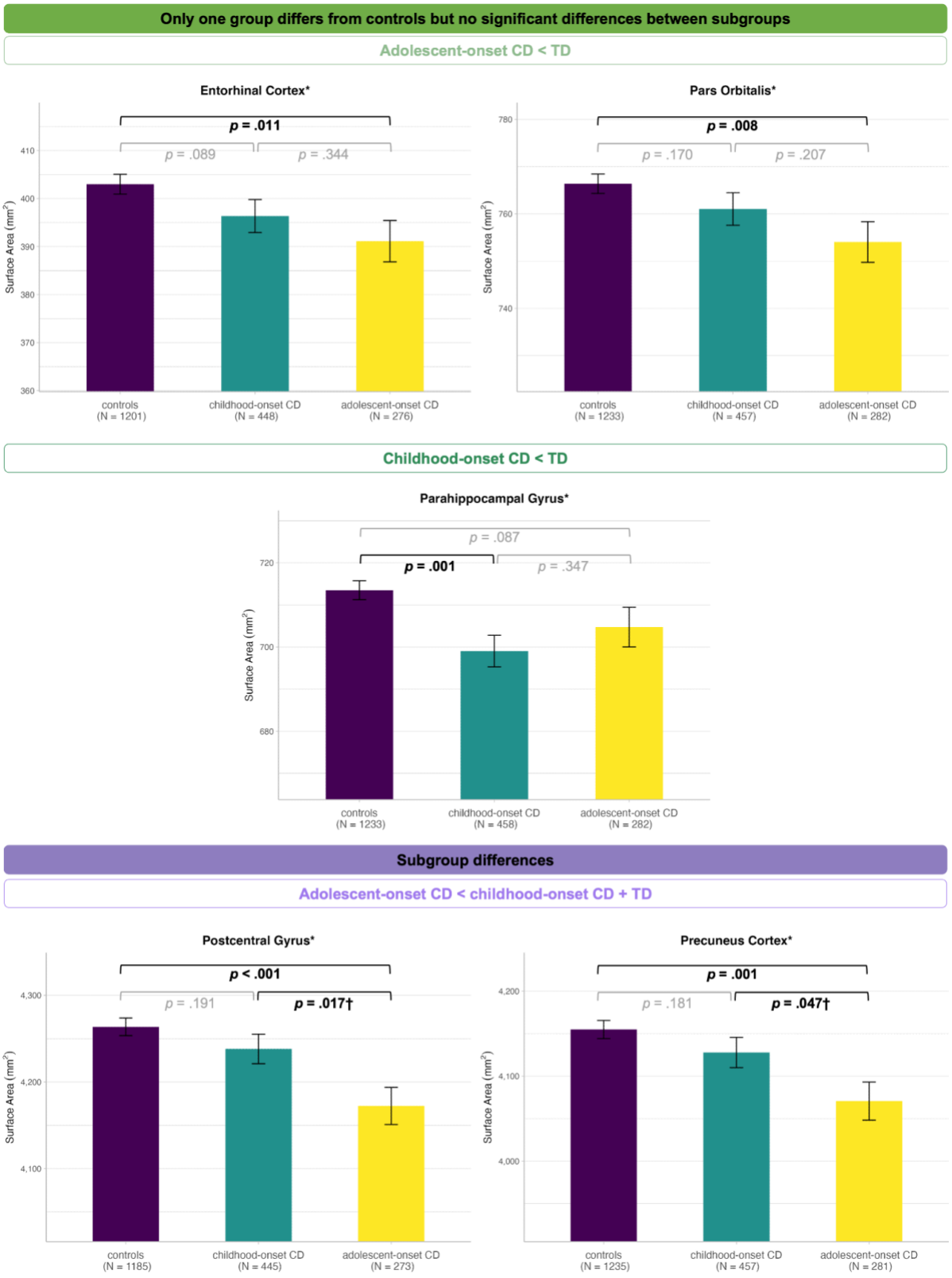
Error bars denote the standard error of the mean. Only regions with a significant *F*-test for the group effect post False Discovery Rate corrections are depicted. Uncorrected *p*-values for the pairwise comparisons are shown, † indicates that the *p*-value is no longer significant after applying a Bonferroni correction across the three comparisons. CD=conduct disorder. TD=typically-developing.

Figure S27-2: **Mean plots depicting pairwise comparisons for regions with a significant age-of-onset subgroup effect (continuation of *Figure S26-1*)**



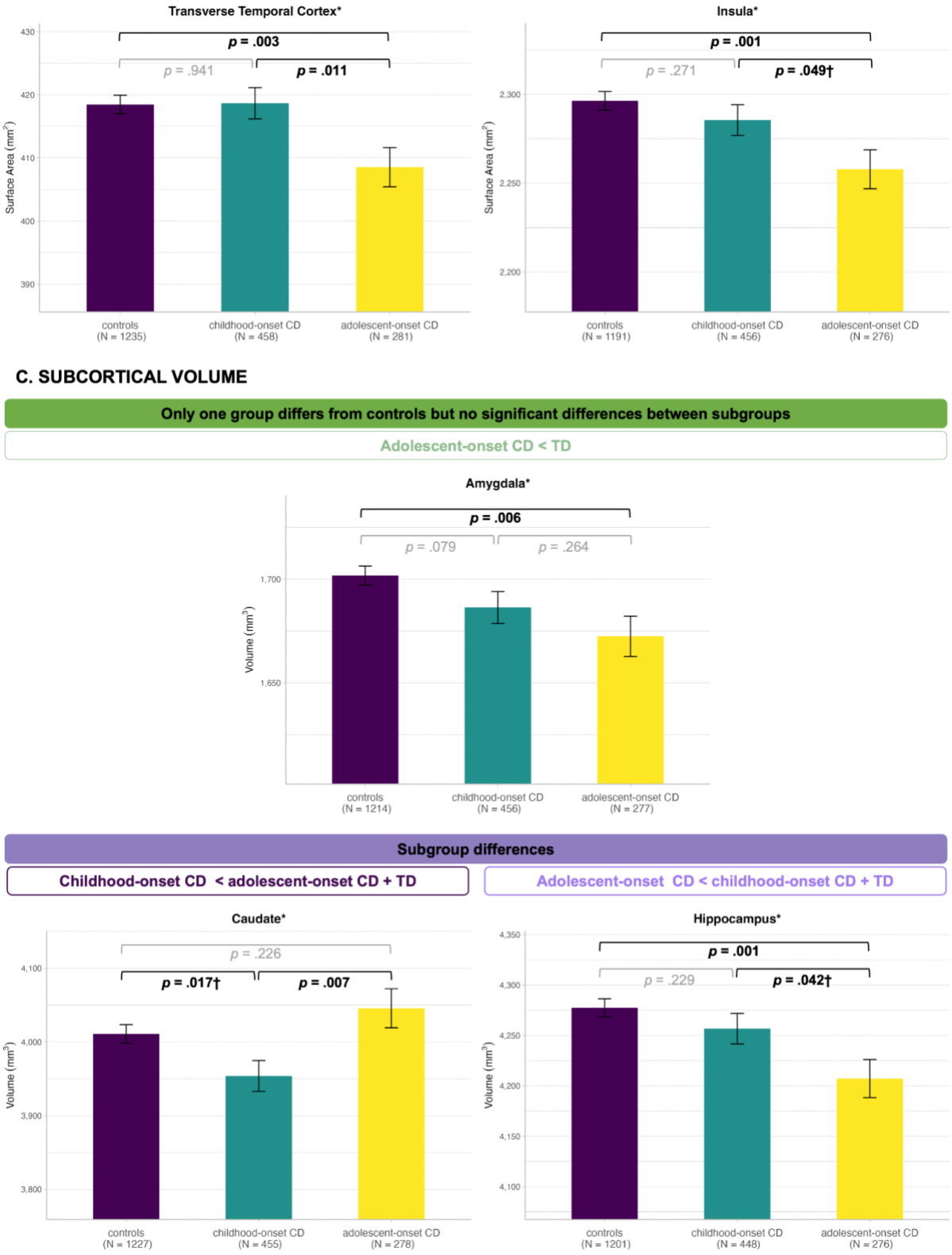
Error bars denote the standard error of the mean. Only regions with a significant *F*-test for the group effect post False Discovery Rate corrections are depicted. Uncorrected *p*-values for the pairwise comparisons are shown, † indicates that the *p*-value is no longer significant after applying a Bonferroni correction across the three comparisons. CD=conduct disorder. TD=typically-developing.

Figure S27-3: **Mean plots depicting pairwise comparisons for regions with a significant age-of-onset subgroup effect**



Error bars denote the standard error of the mean. Only regions with a significant *F*-test for the group effect post False Discovery Rate corrections are depicted. Uncorrected *p*-values for the pairwise comparisons are shown, † indicates that the *p*-value is no longer significant after applying a Bonferroni correction across the three comparisons. CD=conduct disorder. TD=typically-developing.

Figure S27-4: **Mean plots depicting pairwise comparisons for regions with a significant age-of-onset subgroup effect (continuation of *Figure S26-3*)**



Error bars denote the standard error of the mean. Only regions with a significant *F*-test for the group effect post False Discovery Rate corrections are depicted. Uncorrected *p*-values for the pairwise comparisons are shown, † indicates that the *p*-value is no longer significant after applying a Bonferroni correction across the three comparisons. CD=conduct disorder. TD=typically-developing.

Appendix 11 - Full results of the subgroup analyses: Callous-unemotional traits

***Table S19*: Callous-unemotional traits subgroup effects on cortical thickness**

	N			*F*	*p*	*p* FDR
Region	**TD**	**low-CU**	**high-CU**			
Banks of the Superior Temporal Sulcus*	1227	268	363	3.96	0.0192	0.3360
Caudal Anterior Cingulate Cortex*	1227	269	364	6.33	0.0018	0.0639
Caudal Middle Frontal Gyrus	1224	266	364	0.70	0.4984	0.7291
Cuneus Cortex	1210	269	359	0.96	0.3819	0.7182
Entorhinal Cortex	1199	260	355	2.10	0.1232	0.5345
Frontal Pole	1228	270	362	0.29	0.7497	0.9032
Fusiform Gyrus	1215	266	363	1.12	0.3272	0.7182
Inferior Parietal Cortex	1205	270	362	1.23	0.2937	0.7182
Inferior Temporal Gyrus	1198	261	358	0.17	0.8451	0.9032
Insula	1191	266	353	0.13	0.8761	0.9032
Isthmus-Cingulate Cortex	1234	269	364	0.40	0.6714	0.9032
Lateral Occipital Cortex	1210	266	357	0.13	0.8774	0.9032
Lateral Orbitofrontal Cortex	1230	271	363	2.22	0.1093	0.5345
Lingual Gyrus	1218	272	365	0.26	0.7749	0.9032
Medial Orbitofrontal Cortex	1222	266	361	0.98	0.3740	0.7182
Middle Temporal Gyrus	1208	267	355	0.76	0.4677	0.7291
Paracentral Lobule	1234	271	367	1.67	0.1877	0.5474
Parahippocampal Gyrus	1234	272	367	1.92	0.1472	0.5345
Pars Opercularis	1232	272	367	0.69	0.5000	0.7291
Pars Orbitalis	1232	271	365	0.49	0.6131	0.8583
Pars Triangularis	1230	272	366	1.79	0.1680	0.5345
Pericalcarine Cortex	1193	266	354	0.71	0.4928	0.7291
Postcentral Gyrus	1181	262	353	1.96	0.1411	0.5345
Posterior-Cingulate Cortex	1232	272	364	1.11	0.3296	0.7182
Precentral Gyrus	1200	266	355	2.04	0.1303	0.5345
Precuneus Cortex	1234	270	365	0.25	0.7763	0.9032
Rostral Anterior Cingulate Cortex	1231	269	364	1.84	0.1585	0.5345
Rostral Middle Frontal Gyrus	1224	268	358	1.02	0.3618	0.7182
Superior Frontal Gyrus	1217	270	362	1.90	0.1504	0.5345
Superior Parietal Cortex	1199	267	356	0.25	0.7822	0.9032
Superior Temporal Gyrus	1186	257	353	0.80	0.4501	0.7291
Supramarginal Gyrus	1192	263	358	0.18	0.8333	0.9032
Temporal Pole	1204	261	351	2.12	0.1203	0.5345
Transverse Temporal Cortex	1236	271	367	0.00	0.9996	0.9996
Mean Cortical Thickness	1234	272	367	0.94	0.3899	0.7182

All statistical models included group (typically-developing vs. low-CU conduct disorder vs. high-CU conduct disorder), sex and age. The F statistic indicates the overall group effect. * indicates that this outcome showed a significant group difference in the main analysis (conduct disorder vs. typically-developing) after False Discovery Rate correction (N = 2). TD=typically-developing controls. low-CU= conduct disorder youth with scores on the Inventory of Callous-Unemotional traits below the normative cut-off. high-CU= conduct disorder youth with scores on the Inventory of Callous-Unemotional traits above the normative cut-off (cut-offs based on Kemp et al.[8]).

***Table S20*: Callous-unemotional traits subgroup effects on surface area**

Region	*N*			*F*	*p*	*p* FDR	pairwise comparisons	pairwise comparisons Bonferroni adjusted
	TD	low-CU	high-CU					
Banks of the Superior Temporal Sulcus*	**1227**	**268**	**363**	**8.79**	**0.0002**	**0.0008**	**low-CU & high-CU < TD**	**low-CU & high-CU < TD**
Caudal Anterior Cingulate Cortex*	**1227**	**269**	**364**	**3.90**	**0.0205**	**0.0326**	**low-CU & high-CU < TD**	
Caudal Middle Frontal Gyrus*	**1225**	**266**	**364**	**4.96**	**0.0071**	**0.0178**	**low-CU & high-CU < TD**	**low-CU < TD**
Cuneus Cortex*	**1209**	**269**	**359**	**4.23**	**0.0147**	**0.0268**	**high-CU < TD**	**high-CU < TD**
Entorhinal Cortex*	**1201**	**260**	**355**	**4.36**	**0.0129**	**0.0250**	**low-CU & high-CU < TD**	
Frontal Pole*	**1228**	**270**	**362**	**7.93**	**0.0004**	**0.0016**	**low-CU & high-CU < TD**	**low-CU & high-CU < TD**
Fusiform Gyrus*	**1214**	**266**	**363**	**7.64**	**0.0005**	**0.0019**	**high-CU < TD**	**high-CU < TD**
Inferior Parietal Cortex*	**1208**	**270**	**362**	**14.05**	**< 0.0001**	**< 0.0001**	**low-CU & high-CU < TD**	**low-CU & high-CU < TD**
Inferior Temporal Gyrus*	**1198**	**261**	**358**	**10.90**	**< 0.0001**	**0.0001**	**low-CU & high-CU < TD**	**low-CU & high-CU < TD**
Insula*	**1191**	**266**	**353**	**4.59**	**0.0102**	**0.0215**	**high-CU < TD**	**high-CU < TD**
Isthmus-Cingulate Cortex*	**1234**	**269**	**364**	**3.45**	**0.0318**	**0.0464**	**high-CU < TD**	
Lateral Occipital Cortex*	**1210**	**266**	**357**	**4.18**	**0.0155**	**0.0268**	**high-CU < TD**	**high-CU < TD**
Lateral Orbitofrontal Cortex*	**1230**	**271**	**363**	**5.48**	**0.0043**	**0.0116**	**low-CU & high-CU < TD**	**low-CU < TD**
Lingual Gyrus*	**1218**	**272**	**365**	**5.75**	**0.0033**	**0.0104**	**high-CU < TD**	**high-CU < TD**
Medial Orbitofrontal Cortex	1221	266	361	1.52	0.2191	0.2474		
Middle Temporal Gyrus*	**1211**	**267**	**355**	**13.92**	**< 0.0001**	**< 0.0001**	**low-CU & high-CU < TD**	**low-CU & high-CU < TD**
Paracentral Lobule	1233	271	367	0.94	0.3913	0.4176		
Parahippocampal Gyrus*	**1233**	**272**	**367**	**7.20**	**0.0008**	**0.0027**	**high-CU < TD**	**high-CU < TD**
Pars Opercularis	1233	272	367	0.85	0.4271	0.4397		
Pars Orbitalis*	**1233**	**271**	**365**	**4.14**	**0.0161**	**0.0268**	**high-CU < TD**	**high-CU < TD**
Pars Triangularis	1232	272	366	2.49	0.0830	0.1037		
Pericalcarine Cortex	1194	266	354	2.49	0.0829	0.1037		
Postcentral Gyrus*	**1185**	**262**	**353**	**4.58**	**0.0104**	**0.0215**	**low-CU < TD**	**low-CU < TD**
Posterior-Cingulate Cortex*	1233	272	364	2.73	0.0655	0.0917		
Precentral Gyrus*	**1202**	**266**	**355**	**3.79**	**0.0227**	**0.0346**	**low-CU < TD**	
Precuneus Cortex*	**1235**	**270**	**365**	**4.57**	**0.0104**	**0.0215**	**high-CU < TD**	**high-CU < TD**
Rostral Anterior Cingulate Cortex	1230	269	364	2.51	0.0815	0.1037		

Rostral Middle Frontal Gyrus*	**1227**	**268**	**358**	**5.47**	**0.0043**	**0.0116**	**high-CU < TD**	**high-CU < TD**
Superior Frontal Gyrus*	**1219**	**270**	**362**	**11.57**	**< 0.0001**	**0.0001**	**high-CU < low-CU + TD**	**high-CU < TD**
Superior Parietal Cortex*	1199	267	356	2.32	0.0987	0.1176		
Superior Temporal Gyrus*	**1189**	**257**	**353**	**10.74**	**< 0.0001**	**0.0001**	**high-CU < low-CU + TD**	**high-CU < TD**
Supramarginal Gyrus*	1196	263	358	2.30	0.1008	0.1176		
Temporal Pole	1205	261	351	0.52	0.5921	0.5921		
Transverse Temporal Cortex	1235	271	367	0.93	0.3937	0.4176		
Total Surface Area*	**1234**	**272**	**367**	**12.81**	**< 0.0001**	**< 0.0001**	**low-CU & high-CU < TD**	**low-CU & high-CU < TD**

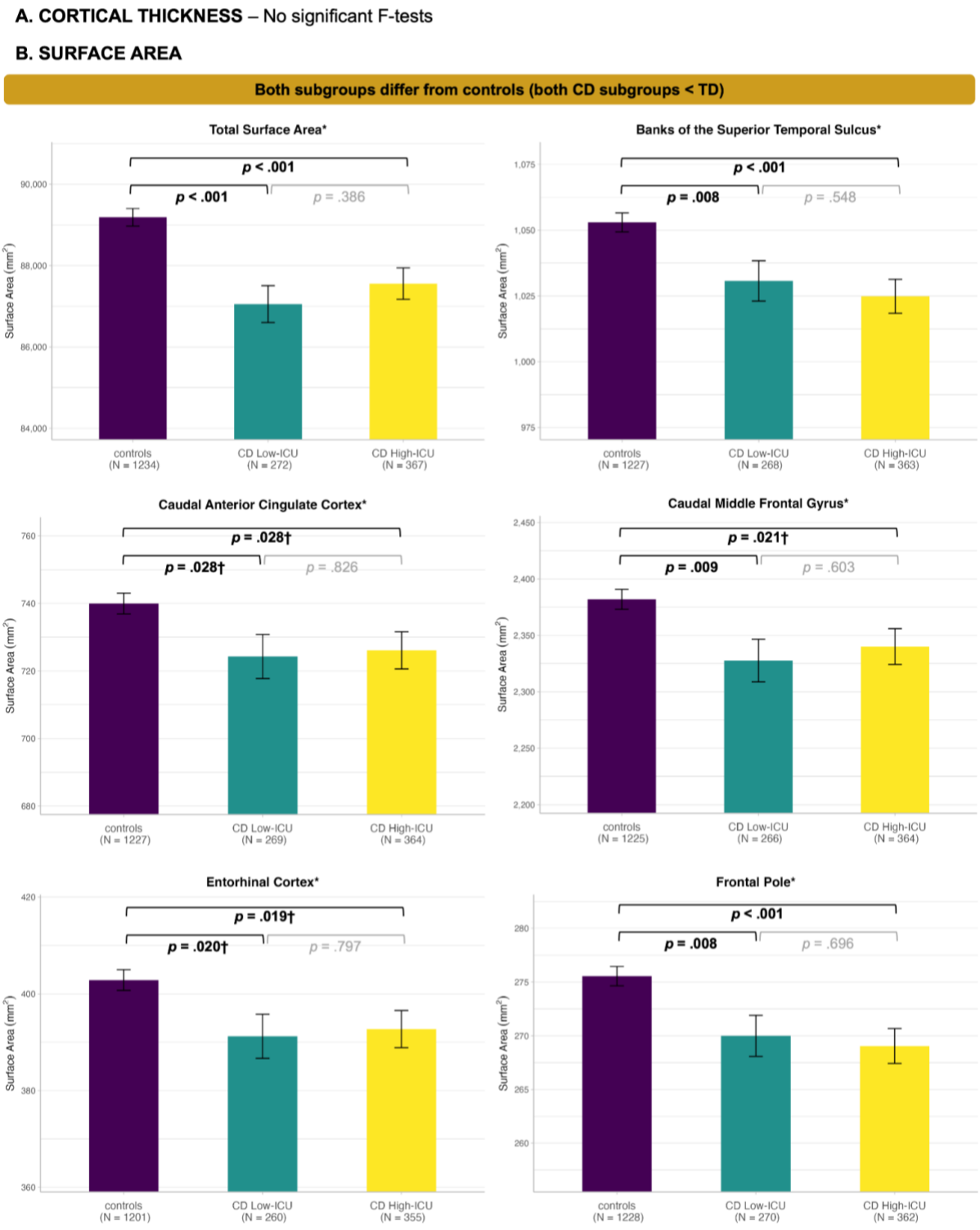
All statistical models included group (typically-developing vs. low-CU conduct disorder vs. high-CU conduct disorder), sex, age, and total intracranial volume (except in the case of total surface area). The F statistic indicates the overall group effect. Regions with a significant group effect post False Discovery Rate correction are indicated in bold. * indicates that this outcome showed a significant group difference in the main analysis (conduct disorder vs. typically-developing) after False Discovery Rate correction (N = 27). TD=typically-developing controls. low-CU= conduct disorder youth with scores on the Inventory of Callous-Unemotional traits below the normative cut-off. high-CU=conduct disorder youth with scores on the Inventory of Callous-Unemotional traits above the normative cut-off (cut-offs based on Kemp et al.[8]).

***Table S21*: Callous-unemotional traits subgroup effects on subcortical volumes and total intracranial volume**

	N			***F***	***p***	***p* FDR**	**pairwise comparisons**	**pairwise comparisons Bonferroni adjusted**
Region	**TD**	**low-CU**	**high-CU**					
Amygdala*	**1233**	**275**	**370**	**4.42**	**0.0122**	**0.0430**	**high-CU < TD**	**high-CU < TD**
Caudate	1202	268	357	1.11	0.3282	0.3750		
Hippocampus*	**1248**	**277**	**374**	**3.88**	**0.0207**	**0.0430**	**high-CU < TD**	**high-CU < TD**
Nucleus Accumbens*	**1249**	**275**	**374**	**3.85**	**0.0215**	**0.0430**	**low-CU < TD**	
Pallidum	1241	272	371	0.18	0.8362	0.8362		
Putamen	1228	269	362	1.80	0.1663	0.2218		
Thalamus*	**1228**	**268**	**359**	**4.22**	**0.0148**	**0.0430**	**low-CU < TD**	**low-CU < TD**
Total Intracranial Volume	1253	277	375	3.19	0.0412	0.0660		

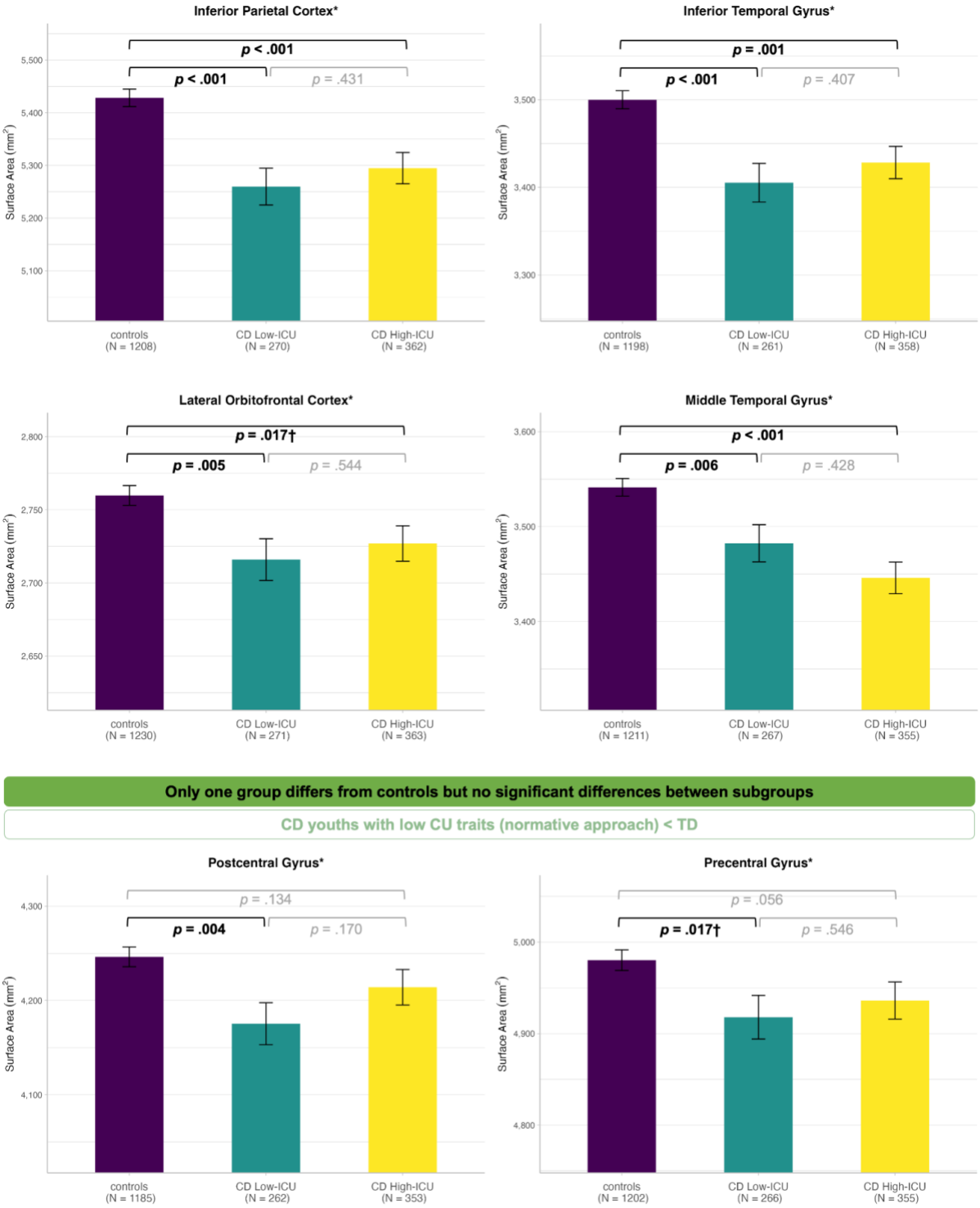
All statistical models included group (typically-developing vs. low-CU conduct disorder vs. high-CU conduct disorder), sex, age, and total intracranial volume (except in the case of total surface area). The F statistic indicates the overall group effect. Regions with a significant group effect post False Discovery Rate correction are indicated in bold. * indicates that this outcome showed a significant group difference in the main analysis (conduct disorder vs. typically-developing) after False Discovery Rate correction (N = 4). TD=typically-developing controls. low-CU= conduct disorder youth with scores on the Inventory of Callous-Unemotional traits below the normative cut-off. high-CU=conduct disorder youth with scores on the Inventory of Callous-Unemotional traits above the normative cut-off (cut-offs based on Kemp et al.[8]).

Figure S28-1: **Mean plots depicting pairwise comparisons for regions with a significant callous-unemotional traits subgroup effect (using the Inventory of Callous-Unemotional traits normative cut-off approach)**



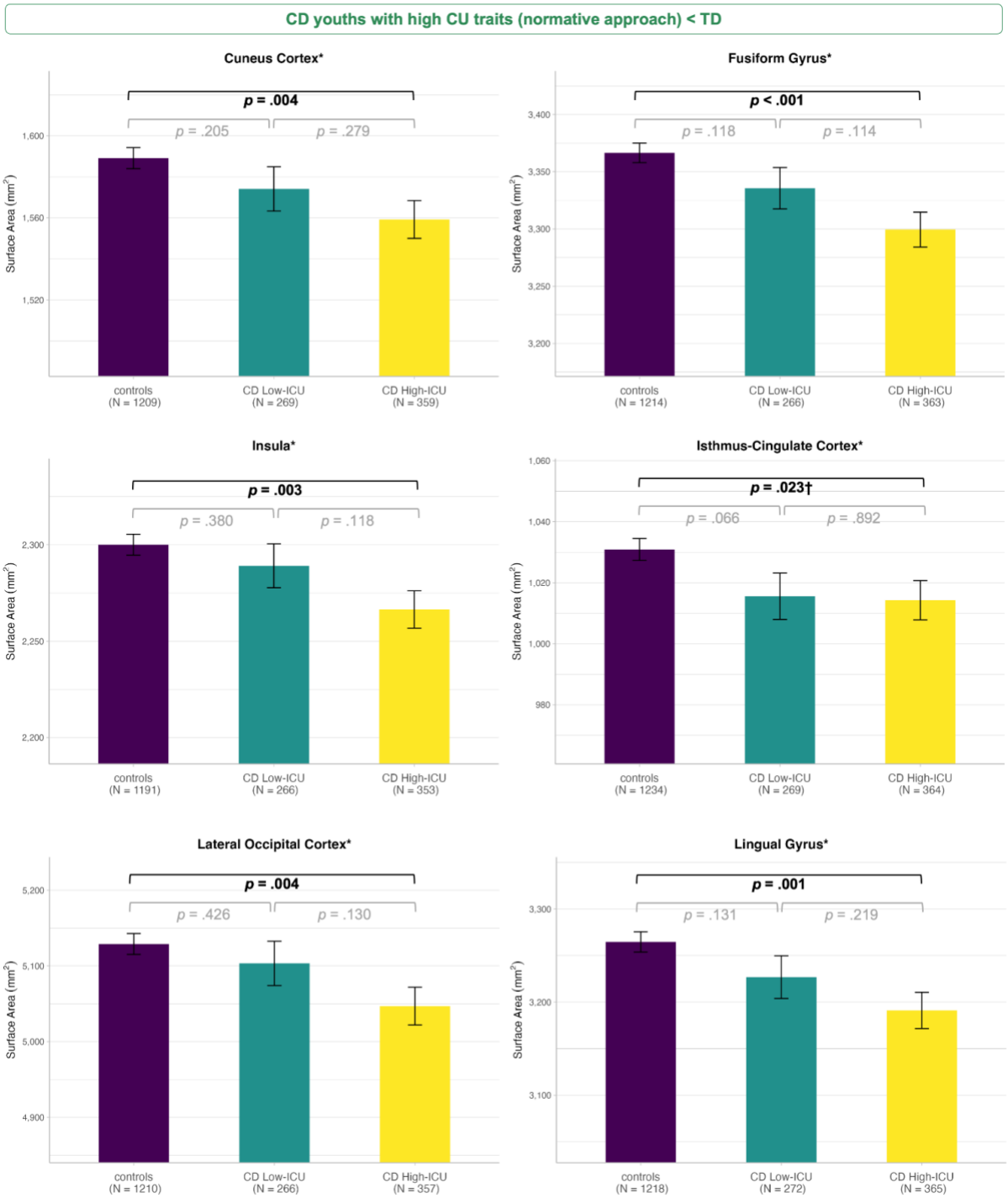
Error bars denote the standard error of the mean. Only regions with a significant *F*-test for the group effect post False Discovery Rate corrections are depicted. Uncorrected *p*-values for the pairwise comparisons are shown, † indicates that the *p*-value is no longer significant after applying a Bonferroni correction across the three comparisons. CD=conduct disorder. TD=typically-developing. CU=callous-unemotional traits. CD Low-ICU = Youths with CD below the normative cut-off on the Inventory of Callous-Unemotional traits. CD High-ICU = Youths with CD on/above the normative cut-off on the Inventory of Callous-Unemotional traits.

Figure S28-2: **Mean plots depicting pairwise comparisons for regions with a significant callous-unemotional traits subgroup effect (continuation of *Figure S27-1*)**



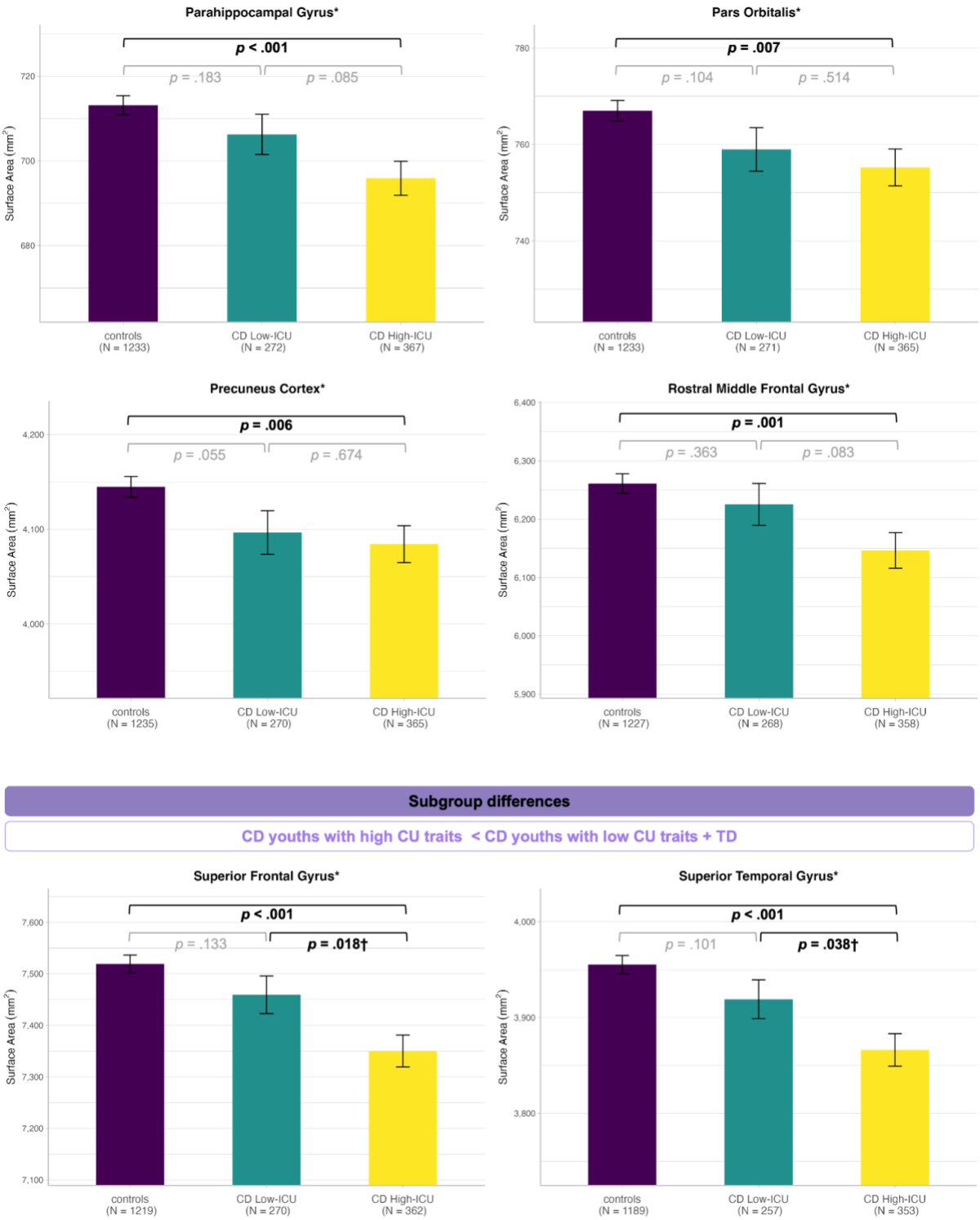
Error bars denote the standard error of the mean. Only regions with a significant *F*-test for the group effect post False Discovery Rate corrections are depicted. Uncorrected *p*-values for the pairwise comparisons are shown, † indicates that the *p*-value is no longer significant after applying a Bonferroni correction across the three comparisons. CD=conduct disorder. TD=typically-developing. CU=callous-unemotional traits. CD Low-ICU = Youths with CD below the normative cut-off on the Inventory of Callous-Unemotional traits. CD High-ICU = Youths with CD on/above the normative cut-off on the Inventory of Callous-Unemotional traits.

Figure S28-3: **Mean plots depicting pairwise comparisons for regions with a significant callous-unemotional traits subgroup effect (continuation of *Figure S27-2*)**



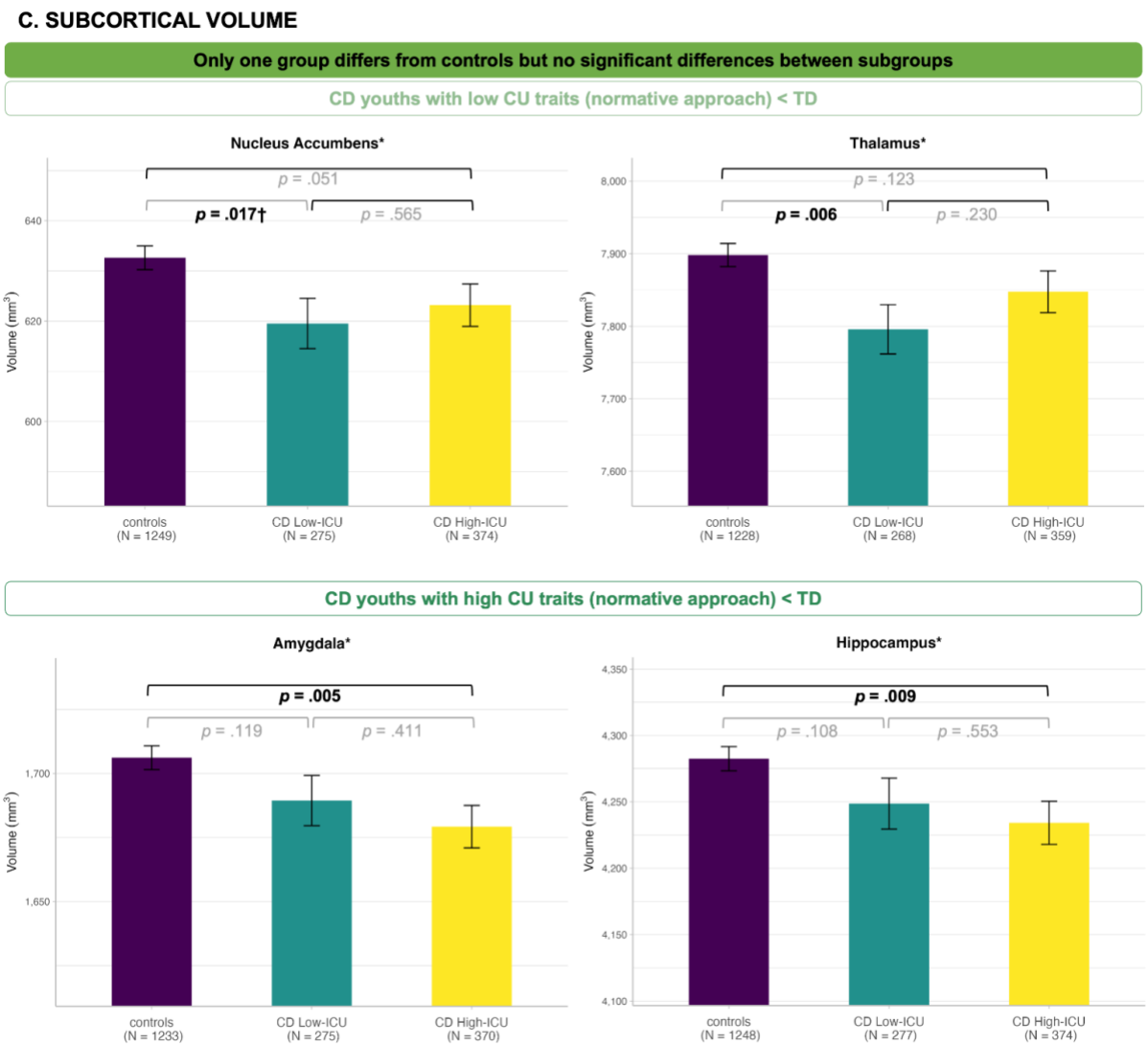
Error bars denote the standard error of the mean. Only regions with a significant *F*-test for the group effect post False Discovery Rate corrections are depicted. Uncorrected *p*-values for the pairwise comparisons are shown, † indicates that the *p*-value is no longer significant after applying a Bonferroni correction across the three comparisons. CD=conduct disorder. TD=typically-developing. CU=callous-unemotional traits. CD Low-ICU = Youths with CD below the normative cut-off on the Inventory of Callous-Unemotional traits. CD High-ICU = Youths with CD on/above the normative cut-off on the Inventory of Callous-Unemotional traits.

Figure S28-4: **Mean plots depicting pairwise comparisons for regions with a significant callous-unemotional traits subgroup effect (continuation of *Figure S27-3*)**



Error bars denote the standard error of the mean. Only regions with a significant *F*-test for the group effect post False Discovery Rate corrections are depicted. Uncorrected *p*-values for the pairwise comparisons are shown, † indicates that the *p*-value is no longer significant after applying a Bonferroni correction across the three comparisons. CD=conduct disorder. TD=typically-developing. CU=callous-unemotional traits. CD Low-ICU = Youths with CD below the normative cut-off on the Inventory of Callous-Unemotional traits. CD High-ICU = Youths with CD on/above the normative cut-off on the Inventory of Callous-Unemotional traits.

Figure S28-5: **Mean plots depicting pairwise comparisons for regions with a significant callous-unemotional traits subgroup effect (continuation of *Figure S27-4*)**



Error bars denote the standard error of the mean. Only regions with a significant *F*-test for the group effect post False Discovery Rate corrections are depicted. Uncorrected *p*-values for the pairwise comparisons are shown, † indicates that the *p*-value is no longer significant after applying a Bonferroni correction across the three comparisons. CD=conduct disorder. TD=typically-developing. CU=callous-unemotional traits. CD Low-ICU = Youths with CD below the normative cut-off on the Inventory of Callous-Unemotional traits. CD High-ICU = Youths with CD on/above the normative cut-off on the Inventory of Callous-Unemotional traits.

Appendix 12 - Assessing the robustness of the findings

1. Leave-one-out analyses

Due to varying sample sizes across cohorts (ranging from 22 to 635 participants), there were concerns that individual samples might have been driving the pattern of our findings. Therefore, to test the robustness of findings and to ensure that our findings were not unduly influenced by individual cohorts, we additionally performed leave-one-out analyses whereby we iteratively repeated the main CD vs. TD comparisons excluding one sample (of fifteen) each time. Table S22 shows the findings for all outcomes that differed significantly between youth with CD and TD participants in the main analysis. The table provides the ranges and means of the uncorrected *p*-values and Cohen's *d* values. The last two columns indicate whether the effect was observed across all leave-one-out analyses and if not, by which sample exclusion it was affected. We focus on uncorrected *p*-values in these analyses given the smaller sample sizes in these comparisons.

In summary, 23 of the 33 significant group differences are replicated in each of the fifteen leave-one-out analyses, including both differences in cortical thickness, lower surface area in 19 outcomes including total and inferior parietal surface, and lower volume in the amygdala and thalamus. In those cases where significant group differences were not observed across all leave-one-out analyses, they were usually rendered non-significant when excluding one of the largest samples, particularly FemNAT-CD (accounting for 21.6% of the CD group) and Boys Town (accounting for 16.2% of the CD group), indicating that this might be due to a reduction in statistical power. Although the ABCD sample contributed 24.1% of CD participants, the removal of this sample only rendered the smallest effect in surface area (i.e., lower posterior-cingulate cortex SA in the CD group) non-significant. Overall, the results of the leave-one-out analyses provide further support for the main findings, indicating wide-spread structural brain alterations in youth with CD, particularly in surface area.

***Table S22*: Results of the leave-one-out analyses for outcomes differed significantly between youth with conduct disorder and typically-developing participants**

	main analysis			leave-one-out analyses							
Region	***p***	***p* FDR**	***d***	***p* mean**	***p* min**	***p* max**	***d* mean**	***d* min**	***d* max**	**significant across all analyses**	**influential samples***
Cortical Thickness											
Caudal Anterior Cingulate Cortex	0.0001	0.0034	0.16	0.0003	0.0000	0.0024	0.16	0.14	0.17	yes	-
Banks of the Superior Temporal Sulcus	0.0010	0.0178	-0.13	0.0025	0.0002	0.0144	-0.13	-0.17	-0.12	yes	-
Surface Area											
Inferior Parietal Cortex	<0.0001	<0.0001	-0.26	<0.0001	<0.0001	<0.0001	-0.26	-0.27	-0.24	yes	-
Total Surface Area	<0.0001	<0.0001	-0.24	<0.0001	<0.0001	<0.0001	-0.24	-0.26	-0.23	yes	-
Middle Temporal Gyrus	<0.0001	<0.0001	-0.22	<0.0001	<0.0001	0.0001	-0.22	-0.26	-0.18	yes	-
Frontal Pole	<0.0001	<0.0001	-0.20	<0.0001	<0.0001	0.0001	-0.20	-0.21	-0.18	yes	-
Inferior Temporal Gyrus	<0.0001	<0.0001	-0.20	<0.0001	<0.0001	0.0005	-0.20	-0.23	-0.16	yes	-

Superior Frontal Gyrus	<0.0001	<0.0001	-0.19	0.0001	<0.0001	0.0004	-0.19	-0.20	-0.17	yes	-
Superior Temporal Gyrus	<0.0001	0.0002	-0.17	0.0001	<0.0001	0.0002	-0.17	-0.18	-0.16	yes	-
Fusiform Gyrus	0.0001	0.0003	-0.17	0.0003	<0.0001	0.0027	-0.17	-0.19	-0.13	yes	-
Postcentral Gyrus	0.0001	0.0004	-0.16	0.0004	<0.0001	0.0020	-0.16	-0.18	-0.14	yes	-
Parahippocampal Gyrus	0.0001	0.0004	-0.16	0.0003	<0.0001	0.0016	-0.16	-0.18	-0.14	yes	-
Precentral Gyrus	0.0002	0.0005	-0.16	0.0007	<0.0001	0.0062	-0.16	-0.18	-0.13	yes	-
Lateral Orbitofrontal Cortex	0.0002	0.0005	-0.15	0.0010	0.0001	0.0107	-0.15	-0.18	-0.12	yes	-
Precuneus Cortex	0.0003	0.0008	-0.15	0.0006	0.0002	0.0019	-0.15	-0.16	-0.13	yes	-
Caudal Middle Frontal Gyrus	0.0007	0.0016	-0.14	0.0022	0.0003	0.0198	-0.14	-0.16	-0.10	yes	-
Isthmus-Cingulate Cortex	0.0008	0.0020	-0.14	0.0020	0.0005	0.0107	-0.14	-0.14	-0.11	yes	-
Insula	0.0011	0.0024	-0.13	0.0025	0.0008	0.0146	-0.13	-0.16	-0.11	yes	-
Rostral Middle Frontal Gyrus	0.0011	0.0023	-0.13	0.0041	0.0003	0.0381	-0.13	-0.15	-0.09	yes	-
Supramarginal Gyrus	0.0019	0.0037	-0.13	0.0034	0.0008	0.0133	-0.13	-0.15	-0.12	yes	-
Banks of the Superior Temporal Sulcus	0.0029	0.0053	-0.12	0.0075	0.0001	0.0351	-0.12	-0.18	-0.10	yes	-
Entorhinal Cortex	0.0040	0.0067	-0.12	0.0192	0.0018	0.2235	-0.12	-0.13	-0.06	no	FemNAT-CD
Caudal Anterior Cingulate Cortex	0.0039	0.0067	-0.12	0.0092	0.0017	0.0568	-0.12	-0.13	-0.09	no	FemNAT-CD
Lateral Occipital Cortex	0.0053	0.0085	-0.12	0.0144	0.0007	0.1045	-0.12	-0.16	-0.07	no	Boys Town
Pars Orbitalis	0.0071	0.0107	-0.11	0.0129	0.0023	0.0608	-0.11	-0.14	-0.09	no	FemNAT-CD
Lingual Gyrus	0.0148	0.0216	-0.10	0.0302	0.0006	0.1636	-0.10	-0.16	-0.06	no	Boys Town, CSU-Yao
Superior Parietal Cortex	0.0239	0.0335	-0.09	0.0348	0.0119	0.1148	-0.09	-0.11	-0.07	no	CSU-Yao, FemNAT-CD
Cuneus Cortex	0.0254	0.0337	-0.09	0.0424	0.0030	0.1879	-0.09	-0.14	-0.06	no	Boys Town, FemNAT-CD
Posterior-Cingulate Cortex	0.0260	0.0337	-0.09	0.0370	0.0126	0.0946	-0.09	-0.10	-0.08	no	ABCD, Boys Town, CSU-Yao, FemNAT-CD
Subcortical Volume											
Thalamus	0.0009	0.0055	-0.14	0.0026	0.0004	0.0206	-0.14	-0.16	-0.10	yes	-
Amygdala	0.0014	0.0055	-0.13	0.0037	0.0006	0.0241	-0.13	-0.15	-0.09	yes	-
Hippocampus	0.0031	0.0082	-0.12	0.0105	0.0017	0.0982	-0.12	-0.14	-0.08	no	FemNAT-CD
Nucleus Accumbens	0.0052	0.0103	-0.11	0.0124	0.0029	0.0772	-0.11	-0.13	-0.08	no	Boys Town
										23/33 effects robust across all analyses	

All statistical models included group, sex and age. A negative Cohen's *d* indicates conduct disorder < controls; a positive Cohen's *d* indicates conduct disorder > controls. The depicted leave-one-out analyses show the resulting *p*-values and Cohen's *d*s after iteratively rerunning the analyses excluding one sample (out of 15) at a time. Due to the smaller sample sizes in the leave-one-out analyses, the uncorrected *p*-values are considered and a group difference was considered robust across all leave-one-out analyses if the group effect was significant prior to False Discovery Rate correction across all analyses. If the effect was not robust across all leave-one-out analyses, the column 'influential samples' indicates the samples whose exclusion rendered the group difference non-significant. Results for regions that did not show a significant group difference in the main analyses are available on request.

2. Applying a more stringent multiple comparison correction

When we applied the multiple comparison correction (False Discovery Rate approach) across all outcomes (*N*=78; rather than per metric as done in the main analyses) for the main CD vs. TD comparison, most previously significant findings remained significant. Differences in results included that an additional region showed significantly lower thickness in youth with CD versus TD youth when correcting across all metrics at once (temporal pole, Cohen's d = -0.10) and three regions no longer showed significantly lower surface area (cuneus cortex, posterior cingulate cortex, superior parietal cortex). Notably, the latter counted amongst the smallest effects in surface area (all Cohen's d = -0.09).

3. Excluding participants over the age of 18 years old

In line with previous CD and ENIGMA studies,[1,3] a lack of (neuro)biological justification for a cut-off of adolescence at age 18,[4,5] and as CD can still be diagnosed over the age of 18 (if they do not meet the criteria for Antisocial Personality Disorder),[7] we included participants up to age 21 years in the current study (albeit the mean cohort age had to be ≤18 years to be included). This resulted in the inclusion of 25 youths (14 CD, 11 TD) between 19 and 21 years from 5 different cohorts (BESD, Boys Town, Cambridge Female, Cambridge Male, and CDKid). Re-running the main analysis without these 25 participants, resulted in largely unchanged findings (i.e., the same outcomes differed significantly between youth with CD and TD participants before and after multiple comparison correction). Detailed results for these analyses are available on request.

4. Robust statistics

To further assess the robustness of findings, across all (main, sensitivity, and subgroup) analyses, we additionally assessed whether:

- bootstrapped 95% CIs around the group estimate (1000 resamples) included 0.
- heteroscedasticity-consistent standard errors (HC3) and *p*-values based on those resulted in the same conclusions.

Overall, conclusions based on analyses that were more robust to violations of normality and heteroscedasticity provided support for the main findings and interpretations. Differences were rare and minor and did not change the overall pattern of findings. Detailed tables for these findings are available on request.

In brief, for the analyses comparing youth with CD to TD youth and relevant sensitivity analyses (e.g., those adjusting for comorbidities), consideration of robust statistics produced highly overlapping results and resulted in the same conclusions. The main - albeit minor - deviation was observed in the age-of-onset analyses. In the original analyses, within the 17 surface area outcomes that showed a significant group effect (childhood-onset vs. adolescent-onset vs. TD), 31 pairwise comparisons were significant prior to applying a Bonferroni correction (23 post-correction). For two of these (i.e., the comparisons between the childhood-onset and the adolescent-onset CD subgroups in the insula and precuneus) the bootstrapped 95% CI included 0 indicating that these differences were less robust. However, these findings do not go against the overall observed pattern

or conclusion. At the most, they further weaken the support for brain structure differences between youth with childhood-onset and adolescent-onset CD as they reduce the number of regions showing subgroup differences across all three metrics from seven to five. Additionally, in the analyses exploring group-by-age interactions, one participant showed high leverage values (>0.35) for many cortical thickness outcomes. Removal of this participant did not impact the significance of findings.

Appendix 13 - Conduct problems analyses

Background

Several previous studies have not investigated youth with a CD diagnosis but instead those with elevated conduct problems (CP), which are measured via questionnaires rather than diagnostic interviews. Similar to research on youth with CD, these studies have reported brain (structural alterations) in this population.[53,54] It is likely that the elevated CP group is a combination of youth with subthreshold symptoms of CD and those with undiagnosed CD. Given this heterogeneity and previous findings of brain alterations in this closely CD-related phenotype, we therefore aimed to assess whether in a large sample of youth with questionnaire-assessed CPs but without (ascertained) CD diagnosis, we would be able to replicate previous findings in CP samples and to assess whether alterations observed in youth with CD would generalise to this more heterogeneous population. In this context, additional analyses comparing cortical and subcortical brain structure in this group with a low CP control group were performed. We predicted that compared to controls, those with elevated CP would display similar, albeit less pronounced, structural alterations than youth with CD, given that this is a more heterogeneous group comprising subthreshold and undiagnosed CD cases.

Methods

The CP analyses included data from 11 international cohorts, comprising 1,198 youth with elevated CP (491 girls) and 1,177 control youth (521 girls), aged 6 - 18 years (see Table S23 for characteristics of the individual cohorts). Importantly, six cohorts contributed participants to both the CD and CP analyses (ABCD, cVEDA, FemNAT-CD, IMAGEN, MATRICS/Aggressotype, and Yale). However, the CD and CP groups did not have any overlapping participants (see information on grouping approach). Overlap was allowed for the respective control groups.

Grouping approach

The elevated CP group included participants who scored above established cut-offs on the CP subscales of validated questionnaires but who had either not undergone a formal diagnostic assessment or did not fulfil diagnostic criteria for CD. Participants with a current diagnosis of CD were excluded to ensure that those with a known diagnosis would not be driving any of the differences that we might observe in the CP versus control comparison. However, youth with a past diagnosis of CD or a different current or past disruptive behaviour disorders (e.g., ODD) were included in the CP group if they scored above the relevant cut-offs on dimensional measures of CP.

For all but one sample (UCL-T1/2), group allocation was based on the DSM-oriented conduct problems subscale of the Child Behavior Checklist (CBCL).[55] or the conduct problems subscale of the Strengths and Difficulties Questionnaire (SDQ).[56] When both measures were available, we used the CBCL as it is a more extensive measure that captures a broader range of symptoms (17 items assessing conduct problems in the CBCL versus five items on the SDQ). In the case of the CBCL, those above the borderline clinical cut-off were included in the CP group (T-score ≥ 65). For the SDQ, we allocated participants to the elevated CP group if they fell into the 'high' category based on the four-band categorisation system (https://www.sdqinfo.org/) reflected in a score of ≥ 4 on the parent or teacher report version or a score of ≥ 5 on the self-report version. For one sample (UCL-T1/T2), group allocation was based on the CD scale of the Child and Adolescent Symptom Inventory (CASI-4R),[29] in line with the original CP studies based on this sample. Youth were included in the elevated CP group if they were above established cut-offs on the CASI-4R for either the parent- or teacher-report version (parent report = 4+ [ages 10–12] and 3+ [ages 12–16] or teacher report = 3+ [ages 10–12], 4+ [ages 12–14], and 6+ [ages 15–16]).

Information on sample specific neuroimaging methods and conduct problems measures as well as inclusion flowcharts can be found in Appendix 2 and Table S2.

Conduct problems analysis

The analyses followed the same approach as the main CD vs. TD analyses. In brief, individual participant data from all sites were pooled in one analysis. Site effects were adjusted using ComBat prior to statistical analysis. Group differences (CP vs. control) on cortical thickness, surface area, and subcortical volumes (hemisphere means) were examined using general linear models with each global and regional brain measure handled as a

separate outcome. Group was the predictor of interest, and all analyses were adjusted for sex and age (in years). Total intracranial volume was corrected for in regional SA and subcortical volume analyses. An FDR correction with q=0.05 was applied separately to cortical thickness, surface area, and subcortical outcomes. Cohen's d was calculated for all group effects based on the t-values from the general linear models.

Table S23: **Characteristics of the Conduct Problems Cohorts**

		Controls				Elevated Conduct Problems			
Sample	**Total *N***	***n***	**F:M**	**Age Mean (SD)**	**IQ Mean (SD)**	***n***	**F:M**	**Age Mean (SD)**	**IQ Mean (SD)**
ABCD[a+b]	826	413	195:218	9.48 (0.50)	95.99 (15.41)	413	188:225	9.48 (0.50)	95.38 (15.44)
cVEDA[a+b]	402	208	61:147	13.25 (2.69)	-	194	55:139	13.26 (2.32)	-
FemNAT-CD[a+b]	28	14	7:7	13.71 (2.70)	107.36 (13.61)	14	7:7	13.5 (2.68)	102.31 (12.76)
Georgetown	85	40	20:20	13.1 (2.30)	110.75 (13.96)	45	16:29	13.98 (2.33)	97.76 (10.86)
IMAGEN[a+b]	683	341	167:174	13.94 (0.48)	-	342	169:173	13.96 (0.47)	-
KIND Lab girls study	41	30	30:0	9.53 (1.20)	96.5 (9.87)	11	11:0	10.18 (1.17)	95.91 (8.24)
MATRICS/Aggressotype[a+b]	66	33	2:31	12.7 (2.62)	102.95 (11.62)	33	3:30	12.97 (2.65)	98.92 (10.77)
MTwiNS[a+b]	26	13	8:5	13.54 (2.22)	109 (1.41)	13	8:5	13.54 (2.22)	99.33 (10.60)
SAND[b]	24	12	10:2	15.42 (0.51)	-	12	10:2	15.42 (0.51)	-
UCL-T1/T2	81	22	0:22	13.91 (1.60)	95.5 (9.74)	59	0:59	14.15 (1.36)	96.54 (11.77)
Yale[b]	113	51	21:30	11.88 (1.87)	110.9 (13.35)	62	24:38	10.85 (1.94)	107.6 (14.37)
Total (11 samples)	**2375**	**1177**	**521:656**	**12.00 (2.49)**	**98.94 (15.50)**	**1198**	**491:707**	**12.10 (2.45)**	**97.2 (14.77)**

The reported values reflect n or mean (+ standard deviation). Information on sex and age were available for all participants, whereas IQ was not available for all samples or all participants within a sample. [a] Multi-site/-scanner sample. [b] Control group matched on age and sex (and IQ, if available) using propensity score matching. F:M=female:male participant sex ratio. IQ=intelligence quotient.

Results

In summary, compared to the control group, the elevated CP group showed significantly lower total intracranial volume and lower surface area in the superior temporal gyrus. We also observed greater surface area in the banks of the superior temporal sulcus region in the elevated CP group, but this effect seemed to be driven by technical issues with the ComBat adjustment for this specific region (see below). Additional differences corresponding to those observed in youth with CD were identified prior to multiple comparisons correction, including lower total and insular SA and lower amygdala volume. Please see Tables S24-26 below for the full results.

ComBat Issues with the banks of the superior temporal sulcus region

We observed that the ComBat algorithm we applied prior to performing the statistical analyses had a disproportionate impact on the thickness and surface area of the left banks of superior temporal sulcus. We therefore reran the conduct problems analyses in three ways 1) using data without ComBat adjustment and no site correction, 2) using data without ComBat adjustment and including site/batch as linear covariate, and 3) rerunning ComBat after excluding batches with fewer than 10 participants based on concerns surrounding the reliability of ComBat when small batches are included.[48] The latter resulted in the removal of 48 participants (24 with elevated CP).

In the initial analyses, the elevated CP group showed lower surface area in the banks of the superior temporal sulcus compared to the control group. However, using all three alternative approaches, the banks of the superior temporal sulcus no longer differed between groups with indications of a non-significant effect in the opposite direction. A similar pattern was observed for banks of the superior temporal sulcus thickness: The elevated CP

group showed (non-significantly) greater thickness in this region in the initial analyses compared to the control group, but significantly lower thickness prior to FDR correction across the three alternative approaches, Importantly, conclusions for the other cortical or subcortical outcomes were not affected by this issue (nor were the main CD analyses). However, group differences between the CP and control group in total intracranial volume no longer survived correction for multiple comparisons when removing small batches albeit the effect sizes were highly similar (all batches included: Cohen's d = -0.11, pFDR = 0.044; after removal of small batches: Cohen's d = -0.11, pFDR = 0.054).

Discussion

In contrast to the widespread alterations observed in youth with CD, those with elevated CP displayed few differences compared to controls, limited to lower total intracranial volume and surface area in the superior temporal gyrus. Notably, some alterations, observed prior to multiple comparison correction, overlapped with those found in youth with CD (e.g., lower total surface area). In this study, the CP group consisted of diverse samples, including studies that specifically recruited youth high in CP, participants from population-based samples with high CP, and controls from case-control cohorts that do not have a formal diagnosis but have elevated CP. As not all of the studies from which these participants were derived included diagnostic assessments, the CP group included both subclinical cases as well as those that might qualify for a diagnosis if they were assessed for CD. This sample heterogeneity might have contributed to the more limited findings. Overall, the heterogenous group of youth with elevated CP did not show the same widespread cortical and subcortical alterations that were observed in a more homogenous group of youth diagnosed with CD, underscoring the importance of specific group inclusion criteria depending on the focus of the study. It is also notable that the youth diagnosed with CD are likely to have more severe antisocial behavior/more symptoms than the elevated CP subgroup. It has been shown consistently that severity of CD matters in terms of the associated brain alterations.[57,58]

In the subgroup analyses (Figure 2 in the main manuscript), we found that youth with CD and high callous-unemotional traits showed the largest number of regions that differed significantly from TD participants (albeit those without callous-unemotional traits also showed many case-control differences). Taken together these findings might indicate a gradient of severity in brain structural alterations, with the greatest level of alterations found in youth with a diagnosis of CD and high callous-unemotional traits, followed by youth with CD without callous-unemotional traits and the fewest alterations observed in a group of youth defined by elevated conduct problems.

Table S24: **Differences in cortical thickness between youth with elevated conduct problems and controls**

	control group			elevated CP group			*t*	*p*	*p* FDR	Cohen's *d*	95% CI
Region	***n***	***M***	***SD***	***n***	***M***	***SD***					
Banks of the Superior Temporal Sulcus*	1173	2.81	0.22	1186	2.82	0.22	1.34	0.1815	0.4716	0.06	-0.03, 0.14
Banks of the Superior Temporal Sulcus (after removal of small batches)	1149	2.81	0.14	1162	2.80	0.14	-1.97	0.0493	0.3920	-0.08	-0.16, 0.00
Caudal Anterior Cingulate Cortex*	1171	2.85	0.20	1190	2.84	0.20	-1.31	0.1894	0.4716	-0.05	-0.13, 0.03
Caudal Middle Frontal Gyrus	1169	2.78	0.13	1179	2.78	0.13	-1.00	0.3188	0.5872	-0.04	-0.12, 0.04
Cuneus Cortex	1164	2.05	0.13	1181	2.05	0.13	-0.60	0.5512	0.7313	-0.02	-0.11, 0.06
Entorhinal Cortex	1157	3.47	0.30	1168	3.45	0.30	-1.55	0.1216	0.4716	-0.06	-0.15, 0.02
Frontal Pole	1172	3.08	0.24	1191	3.08	0.25	-0.66	0.5086	0.7121	-0.03	-0.11, 0.05
Fusiform Gyrus	1169	2.93	0.12	1182	2.93	0.12	-1.20	0.2313	0.4978	-0.05	-0.13, 0.03
Inferior Parietal Cortex	1155	2.73	0.12	1170	2.72	0.12	-0.69	0.4917	0.7121	-0.03	-0.11, 0.05
Inferior Temporal Gyrus	1168	3.01	0.14	1183	3.01	0.14	-0.57	0.5704	0.7313	-0.02	-0.10, 0.06
Insula	1137	3.27	0.13	1149	3.26	0.13	-1.47	0.1419	0.4716	-0.06	-0.14, 0.02
Isthmus-Cingulate Cortex	1172	2.66	0.16	1188	2.66	0.17	0.54	0.5921	0.7313	0.02	-0.06, 0.10
Lateral Occipital Cortex	1167	2.36	0.12	1181	2.35	0.12	-0.46	0.6422	0.7313	-0.02	-0.1, 0.06
Lateral Orbitofrontal Cortex	1173	2.91	0.14	1190	2.90	0.13	-1.98	0.0483	0.3379	-0.08	-0.16, 0.00
Lingual Gyrus	1172	2.22	0.11	1191	2.22	0.11	-0.73	0.4651	0.7078	-0.03	-0.11, 0.05
Medial Orbitofrontal Cortex	1162	2.70	0.14	1183	2.70	0.14	-0.85	0.3956	0.6576	-0.04	-0.12, 0.05
Middle Temporal Gyrus	1163	3.14	0.14	1184	3.13	0.14	-1.28	0.2021	0.4716	-0.05	-0.13, 0.03
Paracentral Lobule	1173	2.67	0.14	1191	2.66	0.14	-2.27	0.0231	0.3348	-0.09	-0.17, -0.01
Parahippocampal Gyrus	1173	2.95	0.22	1192	2.95	0.22	-0.15	0.8818	0.9077	-0.01	-0.09, 0.07
Pars Opercularis	1173	2.87	0.13	1190	2.87	0.13	-0.37	0.7130	0.7798	-0.02	-0.10, 0.07
Pars Orbitalis	1173	3.03	0.17	1191	3.02	0.17	-0.82	0.4134	0.6576	-0.03	-0.11, 0.05
Pars Triangularis	1172	2.76	0.13	1192	2.75	0.13	-1.62	0.1044	0.4716	-0.07	-0.15, 0.01
Pericalcarine Cortex	1162	1.73	0.13	1173	1.73	0.13	0.46	0.6477	0.7313	0.02	-0.06, 0.10
Postcentral Gyrus	1141	2.24	0.12	1161	2.23	0.12	-2.01	0.0441	0.3379	-0.08	-0.17, 0.00
Posterior-Cingulate Cortex	1173	2.76	0.13	1192	2.75	0.13	-2.19	0.0287	0.3348	-0.09	-0.17, -0.01
Precentral Gyrus	1153	2.73	0.12	1175	2.72	0.12	-3.06	0.0023	0.0796	-0.13	-0.21, -0.05
Precuneus Cortex	1170	2.66	0.12	1185	2.65	0.12	-1.36	0.1745	0.4716	-0.06	-0.14, 0.02

Rostral Anterior Cingulate Cortex	1167	3.10	0.18	1186	3.10	0.18	-1.03	0.3027	0.5872	-0.04	-0.12, 0.04
Rostral Middle Frontal Gyrus	1169	2.62	0.13	1184	2.61	0.13	-1.30	0.1940	0.4716	-0.05	-0.13, 0.03
Superior Frontal Gyrus	1169	3.04	0.14	1181	3.03	0.14	-1.17	0.2418	0.4978	-0.05	-0.13, 0.03
Superior Parietal Cortex	1156	2.41	0.12	1176	2.40	0.12	-1.41	0.1577	0.4716	-0.06	-0.14, 0.02
Superior Temporal Gyrus	1162	3.07	0.14	1181	3.07	0.14	-0.20	0.8378	0.8886	-0.01	-0.09, 0.07
Supramarginal Gyrus	1151	2.81	0.13	1155	2.81	0.13	-0.95	0.3436	0.6013	-0.04	-0.12, 0.04
Temporal Pole	1154	3.73	0.28	1163	3.73	0.28	0.10	0.9237	0.9237	0.00	-0.08, 0.09
Transverse Temporal Cortex	1173	2.68	0.18	1192	2.68	0.18	-0.51	0.6128	0.7313	-0.02	-0.10, 0.06
Mean Cortical Thickness	1172	2.72	0.09	1188	2.71	0.09	-1.83	0.0674	0.3934	-0.08	-0.16, 0.01

All statistical models included group, sex and age. A negative Cohen's *d* indicates elevated conduct problems group < controls; a positive Cohen's *d* indicates elevated conduct problems group> controls. Means are adjusted for covariates. * indicates that this outcome showed a significant group difference in the main analysis (conduct disorder vs. typically-developing) after False Discovery Rate correction (*N* = 2). CP=conduct problems.

Table S25: **Differences in surface area between youth with elevated conduct problems and controls**

	control group			**elevated CP group**			***t***	***p***	***p* FDR**	**Cohen's *d***	**95% CI**
Region	***n***	***M***	***SD***	***n***	***M***	***SD***					
Banks of the Superior Temporal Sulcus*	**1173**	**1056.47**	**185.85**	**1187**	**1087.11**	**187.62**	**4.01**	**0.0001**	**0.0022**	**0.17**	**0.08, 0.25**
Banks of the Superior Temporal Sulcus (after removal of small batches)	1149	1058.85	113.94	1163	1054.55	115.04	-0.91	0.3640	0.8477	-0.04	-0.12, 0.04
Caudal Anterior Cingulate Cortex*	1170	730.57	98.27	1189	725.93	99.19	-1.15	0.2514	0.5999	-0.05	-0.13, 0.03
Caudal Middle Frontal Gyrus*	1170	2382.32	287.63	1185	2353.07	290.36	-2.47	0.0136	0.1186	-0.10	-0.18, -0.02
Cuneus Cortex*	1166	1564.52	161.26	1181	1560.94	162.80	-0.54	0.5900	0.8260	-0.02	-0.10, 0.06
Entorhinal Cortex*	1156	389.59	65.51	1169	391.25	66.07	0.61	0.5404	0.8260	0.03	-0.06, 0.11
Frontal Pole*	1172	264.52	28.11	1192	264.41	28.37	-0.10	0.9210	0.9481	0.00	-0.08, 0.08
Fusiform Gyrus*	1169	3305.84	258.81	1186	3289.07	261.17	-1.57	0.1156	0.5059	-0.06	-0.15, 0.02
Inferior Parietal Cortex*	1156	5397.98	517.64	1176	5366.00	522.46	-1.49	0.1355	0.5270	-0.06	-0.14, 0.02
Inferior Temporal Gyrus*	1166	3393.61	321.56	1184	3398.56	324.51	0.37	0.7083	0.8549	0.02	-0.07, 0.10
Insula*	1136	2249.94	168.18	1149	2233.80	169.67	-2.30	0.0218	0.1528	-0.10	-0.18, -0.01
Isthmus-Cingulate Cortex*	1170	1010.50	114.98	1187	1015.95	115.99	1.15	0.2492	0.5999	0.05	-0.03, 0.13
Lateral Occipital Cortex*	1167	5006.25	442.19	1179	4998.52	446.30	-0.42	0.6716	0.8549	-0.02	-0.10, 0.06

Lateral Orbitofrontal Cortex*	1171	2714.04	214.03	1189	2715.17	215.97	0.13	0.8977	0.9481	0.01	-0.08, 0.09
Lingual Gyrus*	1171	3214.77	328.31	1192	3216.91	331.27	0.16	0.8741	0.9481	0.01	-0.07, 0.09
Medial Orbitofrontal Cortex	1159	1853.30	150.87	1183	1849.94	152.23	-0.54	0.5893	0.8260	-0.02	-0.10, 0.06
Middle Temporal Gyrus*	1164	3478.64	303.07	1187	3457.34	305.93	-1.71	0.0881	0.4407	-0.07	-0.15, 0.01
Paracentral Lobule	1170	1479.84	140.65	1192	1477.62	141.94	-0.38	0.7014	0.8549	-0.02	-0.10, 0.06
Parahippocampal Gyrus*	1174	703.96	68.68	1191	701.62	69.31	-0.83	0.4063	0.8260	-0.03	-0.11, 0.05
Pars Opercularis	1171	1609.44	187.82	1193	1605.95	189.56	-0.45	0.6510	0.8549	-0.02	-0.10, 0.06
Pars Orbitalis*	1173	756.68	68.23	1192	754.74	68.85	-0.69	0.4888	0.8260	-0.03	-0.11, 0.05
Pars Triangularis	1173	1513.11	175.19	1192	1504.96	176.81	-1.13	0.2571	0.5999	-0.05	-0.13, 0.03
Pericalcarine Cortex	1162	1507.17	206.16	1175	1504.57	208.05	-0.30	0.7606	0.8873	-0.01	-0.09, 0.07
Postcentral Gyrus*	1143	4239.48	325.90	1162	4221.85	329.21	-1.30	0.1940	0.5999	-0.05	-0.14, 0.03
Posterior-Cingulate Cortex*	1173	1264.27	121.95	1194	1258.11	123.08	-1.23	0.2187	0.5999	-0.05	-0.13, 0.03
Precentral Gyrus*	1155	4925.90	347.31	1179	4908.88	350.61	-1.18	0.2362	0.5999	-0.05	-0.13, 0.03
Precuneus Cortex*	1170	4100.45	348.13	1187	4108.51	351.33	0.56	0.5738	0.8260	0.02	-0.06, 0.10
Rostral Anterior Cingulate Cortex	1168	763.16	97.94	1186	760.38	98.83	-0.69	0.4909	0.8260	-0.03	-0.11, 0.05
Rostral Middle Frontal Gyrus*	1169	6233.62	534.99	1188	6232.33	539.68	-0.06	0.9530	0.9530	0.00	-0.08, 0.08
Superior Frontal Gyrus*	1168	7488.99	532.86	1181	7450.15	537.91	-1.77	0.0772	0.4407	-0.07	-0.15, 0.01
Superior Parietal Cortex*	1157	5676.20	463.37	1173	5693.57	467.77	0.91	0.3651	0.7987	0.04	-0.04, 0.12
Superior Temporal Gyrus*	**1162**	**3914.48**	**283.91**	**1180**	**3875.38**	**286.35**	**-3.34**	**0.0009**	**0.0150**	**-0.14**	**-0.22, -0.06**
Supramarginal Gyrus*	1149	4041.71	401.25	1157	4030.15	404.75	-0.69	0.4887	0.8260	-0.03	-0.11, 0.05
Temporal Pole	1155	450.93	42.93	1164	451.36	43.32	0.24	0.8095	0.9140	0.01	-0.07, 0.09
Transverse Temporal Cortex	1173	407.82	45.75	1194	406.73	46.17	-0.58	0.5618	0.8260	-0.02	-0.10, 0.06
Total Surface Area*	1172	88899.58	7174.60	1192	88067.93	7208.86	-2.83	0.0048	0.0556	-0.12	-0.20, -0.04

All statistical models included group, sex, age, and total intracranial volume (except in the case of total surface area). A negative Cohen's *d* indicates elevated conduct problems group < controls; a positive Cohen's *d* indicates elevated conduct problems group> controls. Regions with a significant group difference post False Discovery Rate correction are indicated in bold. Means are adjusted for covariates. * indicates that this outcome showed a significant group difference in the main analysis (conduct disorder vs. typically-developing) after False Discovery Rate correction (N = 27), but please note that the effect in the banks of the superior temporal sulcus was in the opposite direction to the one reported here. CP=conduct problems.

***Table S26*: Differences in subcortical volume and total intracranial volumes between youth with elevated conduct problems and controls**

	control group			elevated CP group							
Region	***n***	***M***	***SD***	***n***	***M***	***SD***	***t***	***p***	***p* FDR**	**Cohen's *d***	**95% CI**
Amygdala*	1155	1674.48	153.68	1181	1660.68	155.02	-2.17	0.0300	0.1200	-0.09	-0.17, -0.01
Caudate	1149	3967.99	399.63	1159	3978.93	403.34	0.66	0.5105	0.8096	0.03	-0.05, 0.11
Hippocampus*	1163	4125.39	294.73	1183	4107.29	297.58	-1.49	0.1367	0.3645	-0.06	-0.14, 0.02
Nucleus Accumbens*	1171	645.94	81.35	1193	642.22	82.15	-1.11	0.2657	0.5315	-0.05	-0.13, 0.03
Pallidum	1170	1781.01	162.84	1192	1782.56	164.39	0.23	0.8159	0.8159	0.01	-0.07, 0.09
Putamen	1159	5815.55	522.25	1180	5826.64	527.23	0.51	0.6072	0.8096	0.02	-0.06, 0.10
Thalamus*	1167	7456.59	470.84	1189	7462.73	475.29	0.32	0.7515	0.8159	0.01	-0.07, 0.09
Total Intracranial Volume	**1177**	**1489551.81**	**119185.32**	**1198**	**1476014.01**	**119769.06**	**-2.78**	**0.0056**	**0.0445**	**-0.11**	**-0.19, -0.03**

All statistical models included group, sex, age, and total intracranial volume (except for when total intracranial volume was the outcome). A negative Cohen's *d* indicates elevated conduct problems group < controls; a positive Cohen's *d* indicates elevated conduct problems group> controls. Regions with a significant group difference post False Discovery Rate correction are indicated in bold. Means are adjusted for covariates. * indicates that this outcome showed a significant group difference in the main analysis (conduct disorder vs. typically-developing) after False Discovery Rate correction (N = 4). CP=conduct problems.

SUPPLEMENTARY REFERENCES

1 Hoogman M, Muetzel R, Guimaraes JP, *et al.* Brain imaging of the cortex in ADHD: a coordinated analysis of large-scale clinical and population-based samples. *Am J Psychiatry* 2019; **176**: 531–42.

2 Hoogman M, Bralten J, Hibar DP, *et al.* Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: a cross-sectional mega-analysis. *Lancet Psychiatry* 2017; **4**: 310–9.

3 Schmaal L, Hibar DP, Sämann PG, *et al.* Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. *Mol Psychiatry* 2017; **22**: 900–9.

4 Blakemore S-J, Choudhury S. Development of the adolescent brain: implications for executive function and social cognition. *J Child Psychol Psychiatry* 2006; **47**: 296–312.

5 Sawyer SM, Azzopardi PS, Wickremarathne D, Patton GC. The age of adolescence. *Lancet Child Adolesc Health* 2018; **2**: 223–8.

6 Bethlehem R a. I, Seidlitz J, White SR, *et al.* Brain charts for the human lifespan. *Nature* 2022; **604**: 525–33.

7 American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 5th ed., text rev (DSM-5-TR). Washington, D.C.: American Psychiatric Publishing, 2022.

8 Kemp EC, Frick PJ, Matlasz TM, et al. Developing cutoff scores for the inventory of callous-unemotional traits (ICU) in justice-involved and community samples. *J Clin Child Adolesc Psychol* 2023; **52**: 519–32.

9 Essau CA, Sasagawa S, Frick PJ. Callous-unemotional traits in a community sample of adolescents. *Assessment* 2006; **13**: 454–69.

10 Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 2011; **46**: 399–424.

11 Thoemmes FJ, Kim ES. A systematic review of propensity score methods in the social sciences. *Multivariate Behav Res* 2011; **46**: 90–118.

12 Dick AS, Lopez DA, Watts AL, et al. Meaningful associations in the adolescent brain cognitive development study. *Neuroimage* 2021; **239**: 118262.

13 Saragosa-Harris NM, Chaku N, MacSweeney N, et al. A practical guide for researchers and reviewers using the ABCD Study and other large longitudinal datasets. *Dev Cogn Neurosci* 2022; **55**: 101115.

14 Barch DM, Albaugh MD, Baskin-Sommers A, et al. Demographic and mental health assessments in the adolescent brain and cognitive development study: Updates and age-related trajectories. *Dev Cogn Neurosci* 2021; **52**: 101031.

15 Zhou J, Chen C, Wang X, et al. Psychiatric disorders in adolescent boys in detention: a preliminary prevalence and case–control study in two Chinese provinces. *J Forens Psychiatry Psychol* 2012; **23**: 664–75.

16 Sarkar S, Craig MC, Catani M, et al. Frontotemporal white-matter microstructural abnormalities in adolescents with conduct disorder: a diffusion tensor imaging study. *Psychol Med* 2013; **43**: 401–11.

17 Sharma E, Vaidya N, Iyengar U, et al. Consortium on Vulnerability to Externalizing Disorders and Addictions (cVEDA): A developmental cohort study protocol. *BMC Psychiatry* 2020; **20**: 2.

18 Zhang Y, Vaidya N, Iyengar U, et al. The Consortium on Vulnerability to Externalizing Disorders and Addictions (c-VEDA): an accelerated longitudinal cohort of children and adolescents in India. *Mol Psychiatry* 2020; **25**: 1618–30.

19 Schumann G, Loth E, Banaschewski T, et al. The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Mol Psychiatry* 2010; **15**: 1128–39.

20 Stevens MC, Haney-Caron E. Comparison of brain volume abnormalities between ADHD and conduct disorder in adolescence. *J Psychiatry Neurosci* 2012; **37**: 389–98.

21 Naaijen J, Mulder LM, Ilbegi S, et al. Specific cortical and subcortical alterations for reactive and proactive aggression in children and adolescents with disruptive behavior. *Neuromage clin* 2020; **27**: 102344.

22 Dotterer HL, Tomlinson RC, Burt SA, Weigard AS, Klump KL, Hyde LW. Neurocognitive abilities associated with antisocial behavior with and without callous-unemotional traits in a community sample. *Neuropsychology* 2021; **35**: 374–87.

23 Michael C, Tillem S, Sripada CS, Burt SA, Klump KL, Hyde LW. Neighborhood poverty during childhood prospectively predicts adolescent functional brain network architecture. *Dev Cogn Neurosci* 2023; **64**: 101316.

24 Burt SA, Klump KL. The Michigan State University Twin Registry (MSUTR): An update. *Twin Res Hum Genet* 2013; **16**: 344–50.

25 Burt SA, Klump KL, Gorman-Smith D, Neiderhiser JM. Neighborhood disadvantage alters the origins of children's nonaggressive conduct problems. *Clin Psychol Sci* 2016; **4**: 511–26.

26 Goetschius LG, Hein TC, Mattson WI, et al. Amygdala-prefrontal cortex white matter tracts are widespread, variable and implicated in amygdala modulation in adolescents. *NeuroImage* 2019; **191**: 278–91.

27 Hein TC, Mattson WI, Dotterer HL, et al. Amygdala habituation and uncinate fasciculus connectivity in adolescence: A multi-modal approach. *Neuroimage* 2018; **183**: 617–26.

28 Reichman NE, Teitler JO, Garfinkel I, McLanahan SS. Fragile Families: sample and design. *Child Youth Serv Rev* 2001; **23**: 303–26.

29 Gadow K, Sprafkin J. The symptom inventories: an annotated bibliography. New York: Checkmate Plus, 2009.

30 Piacentini JC, Cohen P, Cohen J. Combining discrepant diagnostic information from multiple sources: are complex algorithms better than simple ones? *J Abnorm Child Psychol* 1992; **20**: 51–63.

31 Gadow KD, Sprafkin JN. Child symptom inventory 4: screening and norms manual. New York: Checkmate Plus, 2002.

32 O'Nions E, Lima CF, Scott SK, Roberts R, McCrory EJ, Viding E. Reduced laughter contagion in boys at risk for psychopathy. *Curr Biol* 2017; **27**: 3049-3055.

33 Sukhodolsky DG, Wyk BCV, Eilbott JA, et al. Neural mechanisms of cognitive-behavioral therapy for aggression in children and adolescents: design of a randomized controlled trial within the national institute for mental health research domain criteria construct of frustrative non-reward. *J Child Adolesc Psychopharmacol* 2016; **26**: 38–48.

34 Insel TR, Wang PS. Rethinking mental illness. *JAMA* 2010; **303**: 1970–1.

35 Ibrahim K, Kalvin C, Li F, et al. Sex differences in medial prefrontal and parietal cortex structure in children with disruptive behavior. *Dev Cogn Neurosci* 2021; **47**: 100884.

36 Fischl B. FreeSurfer. *Neuroimage* 2012; **62**: 774–81.

37 Desikan RS, Ségonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 2006; **31**: 968–80.

38 Jovicich J, Czanner S, Greve D, et al. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. *Neuroimage* 2006; **30**: 436–43.

39 Wald L, Schmitt F, Dale A. Systematic spatial distortion in MRI due to gradient non-linearities. *Neuroimage* 2001; **13**: 50.

40 Ashburner J, Friston KJ. Voxel-Based Morphometry–the methods. *Neuroimage* 2000; **11**: 805–21.

41 Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging* 1998; **17**: 87–97.

42 Friston KarlJ, Ashburner J, Frith CD, Poline J-B, Heather JD, Frackowiak RSJ. Spatial registration and normalization of images. *Hum Brain Mapp* 1995; **3**: 165–89.

43 Hagler DJ, Hatton SeanN, Cornejo MD, et al. Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *Neuroimage* 2019; **202**: 116091.

44 Rosen AFG, Roalf DR, Ruparel K, et al. Quantitative assessment of structural image quality. *Neuroimage* 2018; **169**: 407–18.

45 Zugman A, Harrewijn A, Cardinale EM, *et al.* Mega-analysis methods in ENIGMA: The experience of the generalized anxiety disorder working group. *Hum Brain Mapp* 2022; **43**: 255–77.

46 Boedhoe PSW, Heymans MW, Schmaal L, *et al.* An empirical comparison of meta- and mega-analysis with data from the ENIGMA Obsessive-Compulsive Disorder working group. *Front Neuroinformatics* 2019; **12**: Article 102.

47 Radua J, Vieta E, Shinohara R, et al. Increased power by harmonizing structural MRI site differences with the ComBat batch adjustment method in ENIGMA. *Neuroimage* 2020; **218**: 116956.

48 Bayer JMM, Thompson PM, Ching CRK, *et al.* Site effects how-to and when: An overview of retrospective techniques to accommodate site effects in multi-site neuroimaging analyses. *Front Neurol* 2022; **13**: 9239

49 Sun D, Rakesh G, Haswell CC, *et al.* A comparison of methods to harmonize cortical thickness measurements across scanners and sites. *NeuroImage* 2022; **261**: 119509.

50 Smaragdi A, Cornwell H, Toschi N, *et al.* Sex differences in the relationship between conduct disorder and cortical structure in adolescents. *J Am Acad Child Adolesc Psychiatry* 2017; **56**: 703–12.

51 Mackey S, Allgaier N, Chaarani B, *et al.* Mega-analysis of gray matter volume in substance dependence: general and substance-specific regional effects. *Am J Psychiatry* 2019; **176**: 119–28.

52 Haddad E, Pizzagalli F, Zhu AH, et al. Multisite test–retest reliability and compatibility of brain metrics derived from FreeSurfer versions 7.1, 6.0, and 5.3. *Hum Brain Mapp* 2023; **44**: 1515–32.

53 De Brito SA, Mechelli A, Wilke M, *et al.* Size matters: Increased grey matter in boys with conduct problems and callous–unemotional traits. *Brain* 2009; **132**: 843–52.

54 Sebastian CL, De Brito SA, McCrory EJ, *et al.* Grey matter volumes in children with conduct problems and varying levels of callous-unemotional traits. *J Abnorm Child Psychol* 2016; **44**: 639–49.

55 Achenbach TM, Ruffle TM. The Child Behavior Checklist and related forms for assessing behavioral/emotional problems and competencies. *Pediatr Rev* 2000; **21**: 265–71.

56 Goodman R. Psychometric properties of the Strengths and Difficulties Questionnaire. *J Am Acad Child Adolesc Psychiatry* 2001; **40**: 1337–45.

57 Bayer JMM, Thompson PM, Ching CRK, et al. Site effects how-to and when: An overview of retrospective techniques to accommodate site effects in multi-site neuroimaging analyses. *Front Neurol* 2022; **13**: 923988.

58 Fairchild G, Passamonti L, Hurford G, et al. Brain structure abnormalities in early-onset and adolescent-onset conduct disorder. *Am J Psychiatry* 2011; **168**: 624–33.

59 Fairchild G, Toschi N, Hagan CC, Goodyer IM, Calder AJ, Passamonti L. Cortical thickness, surface area, and folding alterations in male youths with conduct disorder and varying levels of callous–unemotional traits. *Neuroimage clin* 2015; **8**: 253–60.