



ARTICLE

Integration of brain and behavior measures for identification of data-driven groups cutting across children with ASD, ADHD, or OCD

Grace R. Jacobs^{1,2}, Aristotle N. Voineskos^{1,2,4}, Colin Hawco^{2,4}, Laura Stefanik^{1,3}, Natalie J. Forde², Erin W. Dickie^{1,2,4}, Meng-Chuan Lai^{1,4,5,6,7,8}, Peter Szatmari^{4,5,6}, Russell Schachar^{4,6}, Jennifer Crosbie^{4,6}, Paul D. Arnold^{9,10,11}, Anna Goldenberg^{9,12,13}, Lauren Erdman^{9,12} and Stephanie H. Ameis^{4,5,6}

Autism spectrum disorder (ASD), obsessive-compulsive disorder (OCD) and attention-deficit/hyperactivity disorder (ADHD) are clinically and biologically heterogeneous neurodevelopmental disorders (NDDs). The objective of the present study was to integrate brain imaging and behavioral measures to identify new brain-behavior subgroups cutting across these disorders. A subset of the data from the Province of Ontario Neurodevelopmental Disorder (POND) Network was used including participants with different NDDs (aged 6–16 years) that underwent cross-sectional T1-weighted and diffusion-weighted magnetic resonance imaging (MRI) scanning on the same 3T scanner, and behavioral/cognitive assessments. Similarity Network Fusion was applied to integrate cortical thickness, subcortical volume, white matter fractional anisotropy (FA), and behavioral measures in 176 children with ASD, ADHD or OCD with complete data that passed quality control. Normalized mutual information was used to determine top contributing model features. Bootstrapping, out-of-model outcome measures and supervised machine learning were each used to examine stability and evaluate the new groups. Cortical thickness in socio-emotional and attention/executive networks and inattention symptoms comprised the top ten features driving participant similarity and differences between four transdiagnostic groups. Subcortical volumes (pallidum, nucleus accumbens, thalamus) were also different among groups, although white matter FA showed limited differences. Features driving participant similarity remained stable across resampling, and the new groups showed significantly different scores on everyday adaptive functioning. Our findings open the possibility of studying new data-driven groups that represent children with NDDs more similar to each other than others within their own diagnostic group. Future work is needed to build on this early attempt through replication of the current findings in independent samples and testing longitudinally for prognostic value.

Neuropsychopharmacology (2021) 46:643–653; <https://doi.org/10.1038/s41386-020-00902-6>

INTRODUCTION

Neurodevelopmental disorders (NDDs), such as autism spectrum disorder (ASD), pediatric obsessive-compulsive disorder (OCD) and attention-deficit/hyperactivity disorder (ADHD), are often associated with poor cognitive and functional outcomes, although long-term trajectories vary considerably [1–4]. There are high rates of co-occurrence between different NDDs [5], as well as similarities in functional impairment [6] and clinical features (e.g., inattention [7], repetitive behaviors [5]). Together with similarity in genetic variants implicated in risk across NDDs [8], these convergences suggest that some children with different NDD diagnoses may be more similar to each other at the biological and behavioral level

despite current distinct categorical (i.e., DSM-5/ICD-10-based) classifications.

Recent transdiagnostic neuroimaging studies highlight the heterogeneity within and across different NDDs and emphasize the need for new research models to move the field forward [9–11]. For example, a prior study from our group using the transdiagnostic Province of Ontario Neurodevelopmental Disorders (POND) dataset showed that children with ASD, ADHD, or OCD all featured non-distinct corpus callosum alterations compared to typically developing controls [12]. A continuous positive association between white matter microstructure and adaptive (everyday) functioning across children, irrespective of NDD

¹Institute of Medical Science, University of Toronto, Toronto, ON, Canada; ²Kimel Family Translational Imaging Genetics Research Laboratory, Campbell Family Mental Health Research Institute, The Centre for Addiction and Mental Health, University of Toronto, Toronto, ON, Canada; ³Targeted Neonatal Echocardiography and Hemodynamic Program, Department of Paediatrics, Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, Canada; ⁴Department of Psychiatry, University of Toronto, Toronto, ON, Canada; ⁵The Margaret and Wallace McCain Centre for Child, Youth & Family Mental Health and Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada; ⁶Program in Neurosciences and Mental Health, The Hospital for Sick Children, Toronto, ON, Canada; ⁷Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge, UK; ⁸Department of Psychiatry, National Taiwan University Hospital and College of Medicine, Taipei, Taiwan; ⁹Genetics & Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada; ¹⁰The Mathison Centre for Mental Health Research & Education, Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada; ¹¹Departments of Psychiatry and Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada; ¹²Department of Computer Science, University of Toronto, Toronto, ON, Canada and ¹³Vector Institute, Toronto, ON, Canada
Correspondence: Stephanie H. Ameis (Stephanie.ameis@camh.ca)

Received: 13 February 2020 Revised: 25 September 2020 Accepted: 26 October 2020
Published online: 9 November 2020

Table 1. Sample characteristics presented by diagnostic group.

	1. ADHD <i>n</i> = 56 M (SD)	2. ASD <i>n</i> = 81 M (SD)	3. OCD <i>n</i> = 39 M (SD)	X ² df = 3	<i>p</i> value	Post-hoc
Sex (F)	9 (16%)	15 (19%)	14 (36%)	6.2	n.s	
Age (years)	10.1 (2.0)	11.6 (2.9)	12.1 (2.2)	17.5	<i>p</i> < 0.001	2, 3 > 1
FSIQ	102.1 (15.0) <i>n</i> = 47	97.1 (17.0) <i>n</i> = 72	115.0 (16.5) <i>n</i> = 17	13.8	<i>p</i> < 0.001	3 > 1, 2
CBCL Internalizing	62.3 (9.3)	64.4 (9.8)	65.5 (9.3)	4.2	n.s	
CBCL Externalizing	62.7 (9.4)	58.6 (9.3)	54.6 (11.5)	13.6	<i>p</i> = 0.001	1 > 2 > 3
TOCS	−18.6 (22.9)	−4.3 (21.5)	16.3 (19.3)	49.7	<i>p</i> < 0.0001	3 > 2 > 1
SWAN I	5.9 (2.2)	4.6 (2.9)	2.3 (3.0)	30.3	<i>p</i> < 0.0001	1 > 2 > 3
SWAN H/I	4.4 (3.0)	3.4 (2.9)	1.5 (2.4)	25.4	<i>p</i> < 0.0001	1 > 2 > 3
RBS-R	14.4 (14.3)	31.5(19.5)	27.1 (22.3)	35.6	<i>p</i> < 0.0001	2, 3 > 1
SCQ	8.1 (5.6)	19.5 (7.2)	5.9 (7.2)	86.2	<i>p</i> < 0.0001	2 > 1 > 3
GAC	79.2 (15.6) <i>n</i> = 48	68.9 (15.1) <i>n</i> = 85	93.7 (21.6) <i>n</i> = 34	33.7	<i>p</i> < 0.0001	3 > 1 > 2

All descriptive statistics are represented as means with standard deviations indicated in brackets except for self-reported sex, where percentage of females is included in brackets. Chi-squared (sex) and Kruskal–Wallis tests followed by Dunn’s tests (all other measures) were used to test group differences. All participants were included in statistics except in the case of IQ and ABAS-II where a subset of the sample was used with available data (as indicated). Raw scores for Toronto Obsessive-Compulsive Scale (TOCS), Strengths and Weaknesses of Attention-Deficit/Hyperactivity Symptoms and Normal Behaviors (SWAN) inattention (I) and hyperactivity/impulsivity (H/I), Repetitive Behaviors-Revised (RBS-R), and Social Communication Questionnaire (SCQ) are presented. The General Adaptive Composite (GAC) Scaled Score for the Adaptive Behavior Assessment System-II (ABAS-II) is also presented. ASD Autism Spectrum Disorder, OCD Obsessive-Compulsive Disorder, ADHD Attention-Deficit/Hyperactivity Disorder, F female, FSIQ Full-Scale Intellectual Quotient (estimated using WASI Wechsler Abbreviated Scale of Intelligence or WISC Wechsler Intelligence Scale for Children), CBCL Internalizing/Externalizing Child Behavior Checklist Internalizing or Externalizing T-score.

category was also found. Others have additionally reported on the absence of clear biological distinctions on structural or functional neuroimaging measures when comparing different NDD diagnostic groups [13–18].

Data-driven clustering approaches offer a methodological alternative to conventional comparisons between clinically defined groups. This alternative approach may better disentangle heterogeneity within and across current diagnostic categories to identify participant subgroups that may be more similar to each other in brain or behavior [10]. Some of these approaches use data integration techniques to identify data-driven subgroups beyond using neuroimaging [19] or behavioral features alone [13]. Different clustering techniques can identify subgroups of participants across disorders with more similar brain-behavior profiles than those within a disorder [20, 21], including a recent effort in the POND sample that integrated cortical thickness and behavioral measures, showing that identified clusters did not divide along diagnostic boundaries [22].

The present study aims to build on these efforts by simultaneously integrating different brain imaging phenotypes (regional cortical thickness, subcortical volume, and white matter tract fractional anisotropy, FA) with behavioral measures in children with primary ASD, ADHD or OCD clinical diagnoses using Similarity Network Fusion (SNF), a data integration approach [23]. SNF identifies participant similarity networks by integrating within and across data types, allowing participants who are most similar to each other to be grouped together. We hypothesized that we would find new groups, each comprised of children with different NDDs who would show similar brain imaging and behavioral features to each other (i.e., within group) but different from other participants (i.e., between group); further, these differences would be of larger effect size than those found using categorical NDD diagnoses. We then examined whether differences between new groups would extend to out-of-model measures (e.g., everyday functioning, structural covariance network indices, surface area), hypothesizing again that a similar pattern would emerge. Finally, we examined the stability of our model, and explored whether supervised machine learning could be used to compare accuracy of subgroup identification using different sets of top contributing model features.

MATERIALS AND METHODS

Participants

Participants included children recruited through the POND Network between June 2012 to June 2016 from the Hospital for Sick Children and Holland Bloorview Kids Rehabilitation Hospital that were all scanned on the same Siemens Tim Trio (Malvern, Pa.) 3T magnetic resonance imaging (MRI) scanner located at the Hospital for Sick Children (Toronto, Canada). Additional data collection through POND is ongoing post scanner upgrade to the PrismaFIT, which was not analyzed for this report. Each institution received approval for this study from their respective research ethics boards. Following a complete description of the study, written informed consent/assent from primary caregivers/participants was obtained. Inclusion criteria included: age < 18 years, presence of a primary clinical diagnosis of ASD, ADHD or OCD, confirmed using the Autism Diagnostic Interview-Revised [24] and Autism Diagnostic Observation Schedule-2 [25] for ASD, Parent Interview for Child Symptoms [26] for ADHD, and the Schedule for Affective Disorders–Children’s Version (Kiddie-SADS) [27] and the Children’s Yale-Brown Obsessive Compulsive Scale [28] for OCD. Full-scale IQ was estimated using age-appropriate Wechsler scales in all participants. After quality control of MRI data (*n* = 57 removed), removal of participants with missing behavioral data (*n* = 26) and those older than 16 years (*n* = 7) to ensure similar age variance across diagnostic groups (Fig. S1), data from a total of 176 participants were used for the main analyses (Table 1; see Supplementary Materials and Methods).

Clinical/Behavioral assessments

Behavioral measures were selected to capture clinical features of each NDD that varied dimensionally across participants. Seven total raw parent-report scores from five behavioral scales were selected as input features for the SNF analysis: the Child Behavior Checklist (CBCL 6-18) externalizing and internalizing broad-band scores [29]; Toronto Obsessive-Compulsive Scale (TOCS) total score [30]; Social Communication Questionnaire (SCQ) total score [31]; Repetitive Behaviors Scale-Revised (RBS-R) total score [32]; and total Strengths and Weaknesses of ADHD Symptoms and Normal Behavior (SWAN) inattention and hyperactivity/impulsivity item scores [33]. The parent-reported general adaptive composite score from the Adaptive Behavior Assessment System-II (ABAS-II)

[34] capturing cross-disorder impairments in adaptive functioning was also assessed in participants.

MRI acquisition parameters

All MRI data was acquired on the same 3T scanner using a 12-channel head coil. T1-weighted and diffusion-weighted acquisitions are detailed in Supplementary Materials and Methods [12, 17, 35].

Image analysis

Cortical thickness and surface area values (for 68 regions) and subcortical volumes (for 14 regions) were derived from T1-weighted images using the Desikan-Killiany Atlas in FreeSurfer (v.5.3) [36] for regional parcellation. Diffusion MRI data were preprocessed using a combination of FSL (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>) and MRtrix (<http://www.mrtrix.org/>) to denoise, upsample the data, and correct for motion and eddy currents (Supplementary Materials and Methods). Tract-based spatial statistics (v1.2) [37] and the ENIGMA DTI pipeline were used to estimate white matter FA as an indirect index of white matter microstructure (for 46 regions) [38]. Imaging metrics for subcortical volume, cortical thickness and white matter FA were included as brain-based input features in the SNF analysis. Subcortical volumes were divided by intracranial volume. Surface area was used subsequently as an out-of-model measure to test data-driven group differences.

Quality control

Please see Supplementary Materials and Methods for quantitative and qualitative MRI quality control details.

SNF data integration and cluster determination

SNF(v2.3.0 R package) was used to integrate structural imaging (cortical thickness, subcortical volume, white matter FA) and behavioral (CBCL, SCQ, RBS-R, SWAN, TOCS scores) data types (see Table S3 listing all SNF input features). Separate networks describing participant similarity for each data type were first created, followed by the use of a nonlinear combination method to iteratively fuse networks for each data type into a single participant similarity network representing the full spectrum of included features [23]. Similarity matrices for each of the four data types (i.e., cortical thickness, subcortical volume, white matter FA and behavioral data) were calculated using Euclidean distance with a nearest neighbors value of 18 and normalization parameter of 0.8, based on consultation with SNF developers and suggested nearest neighbor value of sample size/10. Normalized mutual information (NMI) was used as a metric to describe the overlap in a similarity matrix created using any single model feature compared to the fused matrix created using all model features (NMI range 0-1). Features with higher NMI scores indicate greater contribution to participant similarity. Spectral clustering (SNF *spectralClustering* function) was then applied to delineate groups based on participant similarity matrices determined using 135 model features across 1000 iterations of resampling 80% of participants. A cluster number of 4 was chosen as it was consistently optimal across parameters based on SNF's *estimateNumberOfClustersGivenGraph* function. A silhouette plot quantified the similarity between participants within a given group compared to participants in all other groups. The R package *qgraph* (v.1.6.1) was used for visualization of relative similarity among participants. See Supplementary Materials and Methods for further details describing the SNF method and analysis.

COMPARISONS AMONG IDENTIFIED DATA-DRIVEN GROUPS ON DEMOGRAPHIC, COGNITIVE AND TOP CONTRIBUTING MODEL FEATURES

Separate one-way ANCOVAs were conducted to examine whether data-driven groups differed on age, sex and IQ measures. Based

on the results, all subsequent analyses used to evaluate data-driven group distinctions on model features covaried for age, sex and IQ. Separate one-way ANCOVAs were conducted to provide a standardized effect size estimate (using eta squared) of data-driven group distinctions for model features contributing to participant similarity groups, as well as for diagnostic groups. Correction for multiple comparisons was applied to all 135 features using a false discovery rate (FDR) of 5%. Where ANCOVAs were significant, follow-up Tukey comparison tests were run to determine distinctions between specified groupings.

Evaluation of clusters

Cluster stability. Internal cluster reproducibility was evaluated via bootstrap resampling across 1000 iterations of 80% of participants to calculate stability measures. We used top feature proportion of resampling (i.e., would top features consistently remain top features), the percentage of time that each participant clustered with each other participant, and an Adjusted Rand Index measure of the overlap between clusters to determine stability of the data-driven model.

Extension of data-driven group differences to out-of-model features. Brain and behavioral measures, ABAS-II General Adaptive Composite and surface area, that were excluded from the SNF analysis were compared using ANCOVAs to evaluate whether distinctions found between data-driven groups extended to out-of-model features. Similarly, cortical thickness covariance network measures of global efficiency, network strength, and density were compared using permutation testing across a range of thresholds. All analyses accounted for the effects of age, sex, and IQ. Please see Supplementary Materials and Methods for a full description of methods used for out-of-model testing.

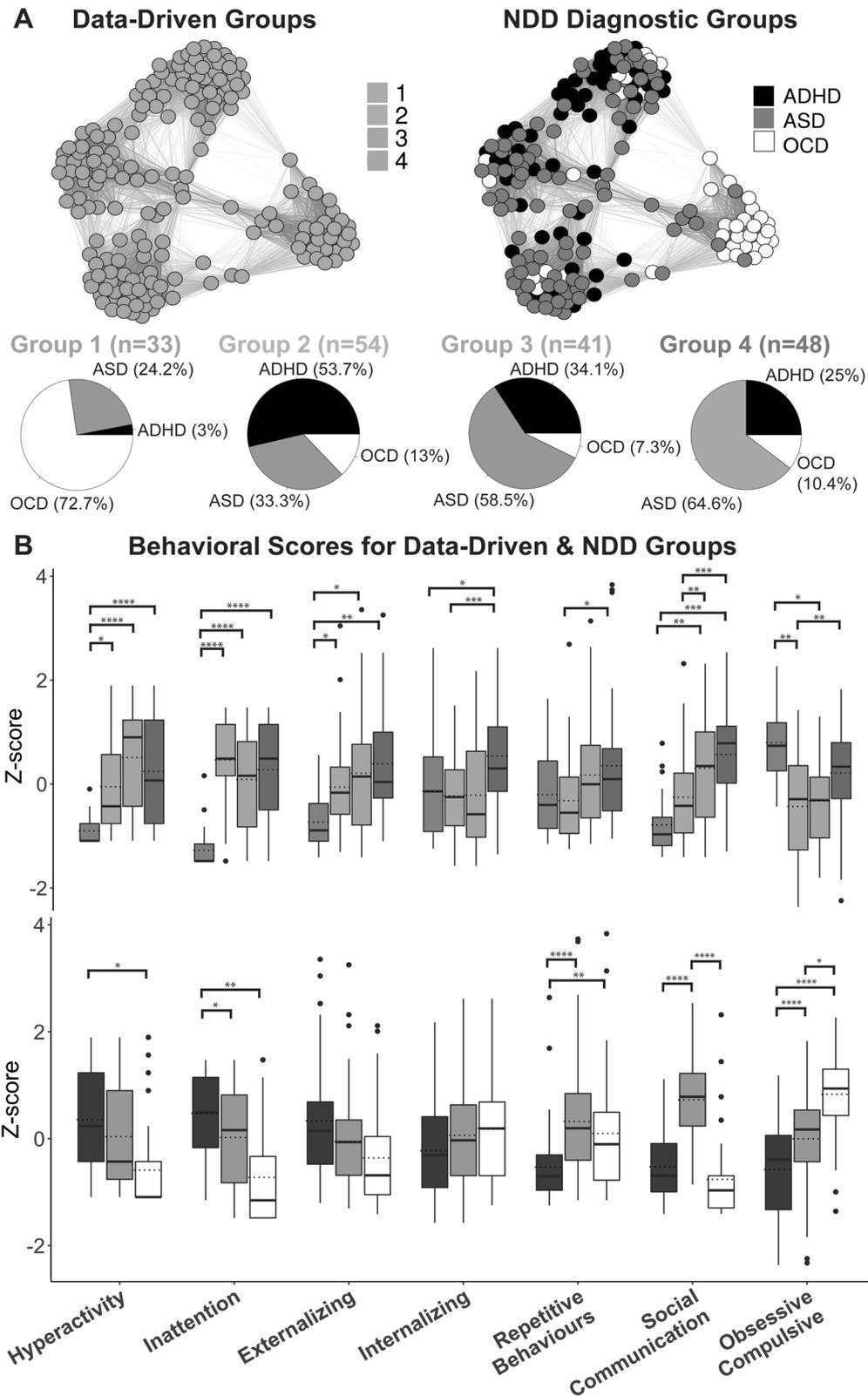
Comparison of classification accuracy based on different selections of top contributing features. A random forest machine learning algorithm was applied across 100 permutations of randomly sampled participants with an 80/20 train-test split to evaluate the reliability of group prediction using different sets of top contributing model features. Due to the limitations of testing classification within the sample used to identify initial groups (versus out-of-sample testing), this approach was considered exploratory to help understand the reliability of group identification and to determine which features might be more likely to accurately classify groups (see Supplementary Materials and Methods for further detail).

Scripts used for analyses have been made available at <https://github.com/gracejacobs/SNF-NDD-Clustering>. The POND Network has made a commitment to release the data. Data release is controlled and managed by the funding agency, known as the Ontario Brain Institute (OBI). The OBI has estimated that POND data will be released in 2020 via the Brain-CODE portal. Please see <https://www.braincode.ca/>.

RESULTS

Top ranking features contributing to formation of four transdiagnostic data-driven groups

The four transdiagnostic data-driven participant similarity groups identified using SNF and spectral clustering (Fig. 1, Table S2) featured an average silhouette width of 0.69, indicating a good-to-strong cluster structure (Fig. S2). Model features with the highest NMI scores (top 10 features) included cortical thickness of the pars triangularis, insula, middle temporal, supramarginal, superior and middle frontal gyrus regions and the SWAN inattention score (Table 2). The SWAN hyperactivity/impulsivity score and right pallidum volume were the only model features besides additional cortical thickness regions ranked among the top 35 features contributing to participant similarity. Patterns of group differences



between data types differed, but were consistent within type (e.g., across cortical thickness features) as can be seen in plots of representative top ranking features (Fig. 2). White matter FA measures did not prominently drive clustering (NMI scores were in the bottom half of all included features).

Influence of demographic and IQ variables on data-driven groups
 There was a significant effect of age on data-driven groups ($F_{3,172} = 13.1, p < 0.0001$), due to a younger age, on average, among Group 3 participants compared to all other groups (Fig. S3a). There was also an effect of IQ ($F_{3,132} = 4.5, p = 0.005$) and sex ($X^2 = 17.0$,

Fig. 1 Relative participant similarities, diagnostic label break-down, and behavior scores for data-driven and NDD groups. A Representations of relative participant similarities derived from the final SNF similarity matrix labeled by data-driven group (upper left) and by diagnostic label (upper right), and break-down of each data-driven group by diagnosis (lower **A**). **B** Bar graphs displaying symptom scores and between-group differences for data-driven groups (upper **B**) and neurodevelopmental disorder (NDD) diagnostic groups (lower **B**) (i.e., ASD Autism Spectrum Disorder, OCD Obsessive-Compulsive Disorder, ADHD Attention-Deficit/Hyperactivity Disorder) after the effects of age, sex and IQ have been regressed out. Hyperactivity = Strengths and Weaknesses of ADHD Symptoms and Normal Behavior (SWAN) hyperactivity/impulsivity raw scores, Inattention = SWAN inattention raw scores, Externalizing or Internalizing = Child Behavior Checklist (CBCL) internalizing or externalizing broad-band raw score, Repetitive Behaviors = Repetitive Behaviors-Revised raw score, Social Communication = Social Communication Questionnaire raw score, Obsessive Compulsive = Toronto Obsessive-Compulsive Scale (TOCS) raw score. Boxplots show the first and third quartile with extensions representing 1.5 * the inter-quartile range. Dashed lines indicate group means and solid lines indicate medians. Error bars show standard errors. Stars indicate significance on follow-up Tukey tests (**** $p < 0.00005$, *** $p < 0.0005$, ** $p < 0.005$, * $p < 0.05$).

$p = 0.0007$) on data-driven groups (Table S2), due to higher IQ in Group 1 compared to all other groups and proportionally more females in Groups 1 and 2 compared to Groups 3 and 4 (Fig. S3b, c). However, differences between groups on top ranking features were consistent across age and IQ when the effects of the other and sex were regressed out as presented in Fig. 2, and no group-by-age, -sex or -IQ interaction effects were shown after FDR correction for any feature. There were two significant group-by-IQ interaction effects among top contributing features before correction in left middle temporal cortical thickness ($F_{3,126} = 4.9$, $p = 0.003$, $\eta^2 = 0.06$) and left pars orbitalis cortical thickness ($F_{3,126} = 2.9$, $p = 0.04$, $\eta^2 = 0.05$) (Fig. S4).

Comparison of top contributing features between data-driven vs. NDD groups

One-way ANCOVAs showed a significant main effect of group (FDR-corrected $p < 0.05$) across top ranking model features (i.e., top 35), after accounting for the effects of age, sex and IQ (see Table 2, Fig. 2). Top features included 32 measures of cortical thickness, 2 behavioral measures and 1 subcortical volume measure (right pallidum), and with the exception of the SWAN inattention score (where effect size was much smaller), top ranking SNF features were not different between NDD groups. Further, there were no significant differences between NDD groups on any of the brain features included in the SNF analysis; effect sizes were typically smaller in NDD relative to data-driven groups by ten-fold (or more) (Table 2).

Follow-up Tukey comparison tests (Table S3) examining data-driven group differences on top contributing model features indicated that compared to other groups, Group 2 consistently featured significantly lower cortical thickness across regions with top NMI scores (e.g., right pars triangularis, $F_{3,129} = 20.3$, $p = 1.06e-9$, $\eta^2 = 0.32$) compared to Groups 1, 3 and 4 (Fig. 2 and Fig. S5). Group 2 ($n = 54$), consisted of more children with ADHD ($n = 29$) than ASD ($n = 18$), and fewer OCD ($n = 7$). In contrast, Group 3 ($n = 41$) had the highest cortical thickness relative to other groups and was comprised mainly of children with ASD ($n = 24$), a sizeable minority with ADHD ($n = 14$), and few with OCD ($n = 3$), and was male dominant. Group 4 ($n = 48$) included predominantly children with ASD ($n = 31$), with a sizeable minority of ADHD ($n = 12$), and fewer children with OCD ($n = 5$), and was male dominant. Group 4 featured significant decreases in right pallidum volume compared to Groups 2 and 3 ($F_{3,129} = 8.02$, $p = 0.002$, $\eta^2 = 0.15$). Group 1 ($n = 33$) was comprised mainly of children with OCD ($n = 24$), fewer with ASD ($n = 8$) and one with ADHD, and featured lower SWAN inattention ($F_{3,129} = 16.0$, $p = 3.46e-8$, $\eta^2 = 0.27$) and hyperactivity/impulsivity scores ($F_{3,129} = 7.8$, $p = 2.35e-4$, $\eta^2 = 0.15$) compared to all other groups. Further examination of symptom scores indicated that ASD participants that clustered into Group 1 tended to have lower social communication impairment scores on the SCQ than ASD participants that clustered into Groups 2, 3 and 4 and higher obsessive-compulsive symptoms than ASD participants that clustered into Groups 2 and 3 (see Fig. S5).

Notably, cortical thickness and subcortical measures in Group 1 were neither highest nor lowest.

In comparing data-driven groups across other behavioral features included in the SNF analysis (Fig. 1B), Group 1 generally featured lower impairment across behavioral measures, with the exception of obsessive-compulsive symptoms. Group 2 featured similar symptom scores relative to Group 4 on hyperactivity/impulsivity, inattention and externalizing symptoms but lower symptom scores on repetitive behavior, social communication and obsessive-compulsive measures relative to this group. Group 4 was also generally more impaired than Group 3 based on mean scores (though these groups did not differ significantly on any behavioral measure examined). Among white matter regions, FA within the right posterior limb of the internal capsule ($F_{3,129} = 3.73$, $p = 0.02$, $\eta^2 = 0.08$) and right cingulate gyrus ($F_{3,129} = 3.2$, $p = 0.04$, $\eta^2 = 0.06$) featured the largest between group effects (Table S3), driven by increased FA in these regions in Group 4 versus Group 2.

Cluster evaluation

Cluster stability. Top ranking features based on NMI score remained relatively consistent across resampling, particularly for those features ranking within the top 35. As can be seen in Fig. 2C (Fig. S7), only those features within the top 35 had a substantial number that remained among the top 35 more than 70% of the time across permutations. None in the middle third were ever present more than 70% of the time, and none in the bottom third were present more than 25% of the time. Across resampling, any given participant clustered with each other participant within their group, on average, 67% of the time, as compared to across them (9.4% of the time, Fig. S8). An Adjusted Rand Index of 0.46 across 1000 iterations was found, indicating over 70% agreement across clusters [39].

Extension of data-driven group distinctions to out-of-model features

In the three out-of-model 'phenotypes' [adaptive (everyday) functioning, brain structural covariance network indices, surface area] shown in Fig. 3, we found that effect sizes for differences between data-driven groups were larger than for NDD groups. For data-driven groups, the effect size of the between group difference on the General Adaptive Composite score ($F_{3,126} = 9.1$, $p = 1.8E-5$, $\eta^2 = 0.16$), where Group 1 had increased functioning compared to all other groups, was larger compared to NDD groups ($F_{2,127} = 9.9$, $p = 0.0001$, $\eta^2 = 0.12$), when covarying for sex, age and IQ (Fig. 3A). For structural covariance network indices, effects were larger among data-driven groups compared to NDD groups for network strength and density across thresholds (Fig. 3B). Group 2 had reduced network strength compared to all other groups and reduced network density compared to Group 3, as compared to no significant differences between NDD groups. When surface area was examined in the top ten contributing cortical thickness regions, left insula surface area was the only region that was significantly different among data-driven groups,

Table 2. The top 35 ranking model features contributing to participant similarity as determined by normalized mutual information and corresponding statistical summaries of between group differences for data-driven and diagnostic groups.

Rank	Top contributing features	Normalized mutual information	Data-driven groups			Diagnostic groups		
			Statistic	pFDR	Eta squared	Statistic	pFDR	Eta squared
1	Right pars triangularis CT	0.182	20.3	1.06E-09	0.315	0.685	n.s	0.0098
2	Right insula CT	0.182	24.9	4.05E-11	0.355	1.35	n.s	0.018
3	SWAN Inattention Score	0.181	16.0	3.46E-08	0.268	7.68	0.0238	0.099
4	Right middle temporal CT	0.180	28.2	3.01E-12	0.392	0.687	n.s	0.0097
5	Left supramarginal CT	0.172	23.7	9.67E-11	0.343	1.61	n.s	0.022
6	Right inferior temporal CT	0.169	17.5	9.36E-09	0.285	1.7	n.s	0.024
7	Left middle temporal CT	0.165	29.9	1.28E-12	0.404	2.34	n.s	0.033
8	Right superior frontal CT	0.165	19.7	1.60E-09	0.308	0.503	n.s	0.0072
9	Left rostral middle frontal CT	0.162	21.7	3.68E-10	0.321	1.39	n.s	0.019
10	Left insula CT	0.161	21.2	5.12E-10	0.307	1.91	n.s	0.025
11	Left fusiform CT	0.159	19.3	2E-09	0.308	3.56	n.s	0.049
12	Left pars opercularis CT	0.158	23.3	1.12E-10	0.350	1.35	n.s	0.02
13	Right inferior parietal CT	0.158	20.2	1.09E-09	0.306	2.22	n.s	0.029
14	Right rostral middle frontal CT	0.157	18.0	6.33E-09	0.292	1.07	n.s	0.016
15	Left inferior parietal CT	0.157	20.4	1.0E-09	0.304	1.89	n.s	0.025
16	Left inferior temporal CT	0.153	21.8	3.68E-10	0.335	1.65	n.s	0.024
17	Right precentral CT	0.149	13.2	6.66E-07	0.228	0.327	n.s	0.0047
18	Left lateral orbitofrontal CT	0.148	16.6	1.88E-08	0.261	1.17	n.s	0.016
19	Left lateral occipital CT	0.148	15.1	8.89E-08	0.253	2.48	n.s	0.033
20	Left lingual CT	0.145	10.2	1.32E-05	0.181	1.93	n.s	0.024
21	Right supramarginal CT	0.145	17.2	1.24E-08	0.272	1.12	n.s	0.015
22	Left pars triangularis CT	0.137	17.5	9.36E-09	0.281	0.402	n.s	0.0058
23	Right fusiform CT	0.136	16.9	1.43E-08	0.278	1.27	n.s	0.018
24	Right lateral orbitofrontal CT	0.134	13.1	6.94E-07	0.224	0.835	n.s	0.012
25	Left postcentral CT	0.134	15.2	7.8E-08	0.259	3.9	n.s	0.053
26	SWAN Hyperactivity Score	0.132	7.59	2.35E-04	0.146	4.48	n.s	0.061
27	Right lateral occipital CT	0.132	19.2	2.18E-09	0.299	3.67	n.s	0.048
28	Left superior frontal CT	0.131	13.9	2.91E-07	0.239	0.728	n.s	0.01
29	Right medial orbitofrontal CT	0.127	12.9	8.3E-07	0.217	0.405	n.s	0.0055
30	Left pericalcarine CT	0.126	7.49	0.000258	0.143	0.531	n.s	0.0073
31	Left caudal middle frontal CT	0.123	19.6	1.69E-09	0.310	0.62	n.s	0.0093
32	Right caudal middle frontal CT	0.119	18.6	3.71E-09	0.296	0.173	n.s	0.0026
33	Right postcentral CT	0.118	10.4	1.01E-05	0.194	4.81	n.s	0.067
34	Left pars orbitalis CT	0.115	15.0	9.18E-08	0.253	1.23	n.s	0.018
35	Right pallidum	0.115	8.02	0.000155	0.154	0.248	n.s	0.0037

Group differences were determined using ANCOVAs while including age, sex and IQ as covariates.

though this finding did not survive FDR correction (Fig. 3C). There was a larger effect size found for the difference in left insula surface area among data-driven groups ($F_{3,126} = 3.7, p = 0.01, \eta^2 = 0.07$), driven by greater surface area in Group 4 compared to Group 2, in contrast to NDD groups where no significant differences were found ($F_{2,127} = 0.1, p = 0.91, \eta^2 = 0.001$). See Supplementary Materials and Methods for detailed out-of-model results.

Comparison of classification accuracy based on different selections of top ranking features. Model performance was highest when the top 2 ranking features (i.e., highest NMI scores) from each of the four data types (i.e., right pars triangularis and right insula thickness, SWAN inattention and hyperactivity/impulsivity, right pallidum and right putamen volume, left anterior limb internal capsule and right retrolenticular internal capsule FA), or all 135

input features were included in the classifier (Fig. S9), compared to when only top ranking features were included in the classifier. Mean sensitivity performance for detecting data-driven groups when including the top 2 features across data types in the classifier ranged from 62–75%, exceeding chance. Mean specificity performance was >80%. When the top 10 or 35 ranking features were included in the classifier, performance remained stable for prediction of Groups 1 and 2, but comparatively declined for Groups 3 and 4. When features from only one of the four data types were included, mean sensitivity dropped to below 50% for at least two of Groups 1, 3 and 4.

DISCUSSION

By fusing data across multiple brain imaging phenotypes and behavioral measures, we identified novel transdiagnostic data-

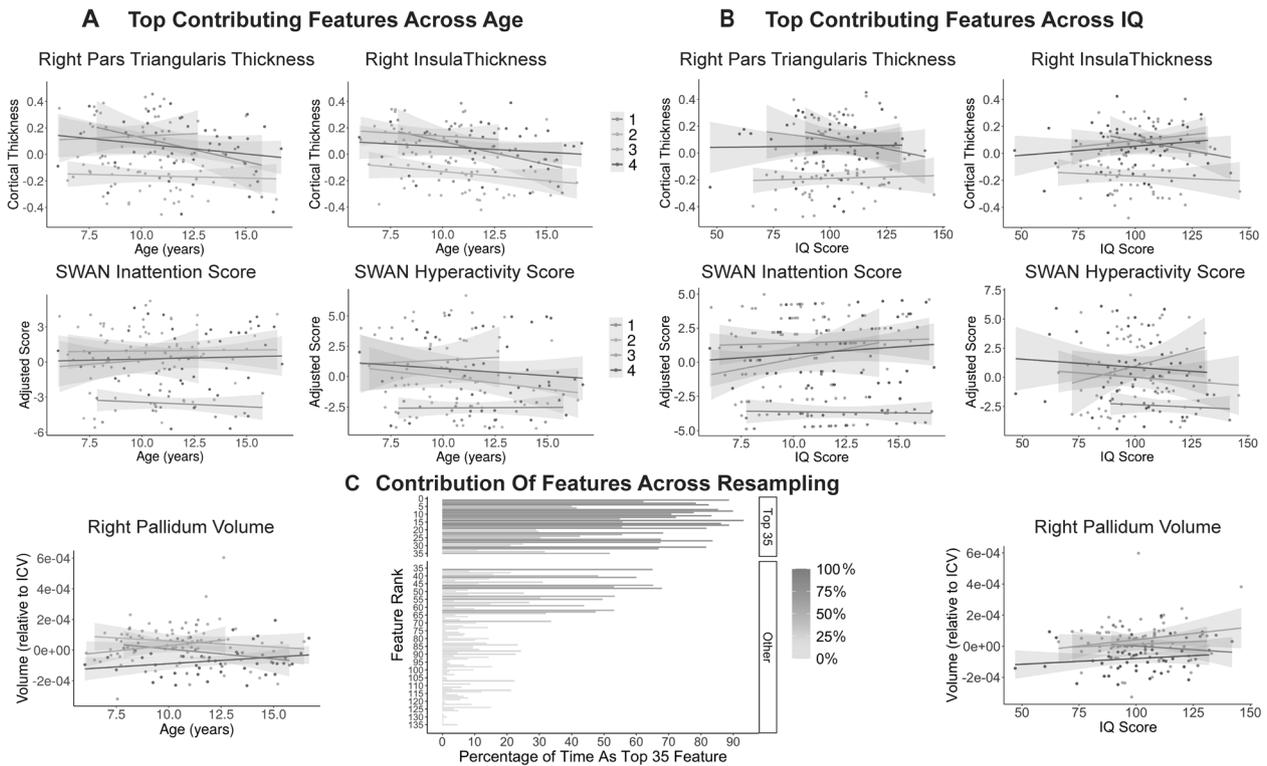


Fig. 2 Data-driven group differences for top contributing features across age and IQ as well as their stability across resampling. Figure panels depict top ranking SNF input features across different data types contributing to participant similarity (i.e., in top 35), including the top two cortical thickness regions (right pars triangularis = rank#1, right insula = rank#2), top two behavioral measures (Strengths and Weaknesses of Attention-Deficit/Hyperactivity Symptoms and Normal Behaviors (SWAN) Inattention = rank#3, SWAN Hyperactivity = rank#26), and right pallidum subcortical volume (rank = #35). Between group differences are plotted across (A) age and (B) IQ. Graphs show differences between groups on top ranking features were consistent across age and IQ when the effects of the other and sex were regressed out. Shaded areas represent 95% confidence intervals. Similar group difference patterns were found across other features (including additional cortical thickness regions) (see Fig. S6). C The percentage with which any given top contributing feature remains among the top 35 features across 1000 permutations (see Fig. S7 for detailed feature labels).

driven groups, which feature more homogeneous characteristics within groups in both brain and behavioral measures compared to current DSM-5 categories (ASD, ADHD, OCD). In particular, we found that cortical thickness in regions important for social or language-related behavior (inferior frontal gyrus, insula, inferior parietal cortex, temporal cortex) and executive function (superior and middle frontal gyrus) along with inattention scores were the top contributors to the model. These differences were consistent across the sample age range, a period of dynamic brain growth and change [40, 41]. Data-driven participant similarity groups displayed internal stability of clustering, top contributing model features and pairwise participant grouping across resampling. Stronger differences between data-driven groups compared to DSM-5 diagnostic groups extended to aspects of clinically (such as everyday functioning) and biologically relevant features excluded from the SNF analysis. Although white matter FA was not among the top contributing features, classification accuracy was best when this data type was included in the model.

Across the four data-driven groups, the pattern of broad clinical symptom severity ranged from more circumscribed symptom scores in Groups 1 and 2 to increases across a fuller range of clinical measures in Group 3 and Group 4, with Group 4 generally featuring the highest symptom load across measures. However, among these groups, there were notable differences, suggesting different neurobiological features may relate to different behavioral profiles. Group 2, comprised mainly of children with ADHD or ASD, had higher inattention scores and a distinct pattern of decreased cortical thickness compared to all other groups. The Group 2 profile found in the current study may be most consistent

with prior work showing delayed and deviated cortical and neural network maturation, particularly in frontal regions in large-scale studies of children with ADHD, including reduced thickness [42, 43]. In direct contrast to Group 2, Group 3 showed elevated hyperactivity and higher cortical thickness in top ranking regions, but was also the youngest group and in earlier stages of normative behavioral and cortical development, perhaps accounting for some of these differences [40]. Nevertheless, plotting cortical thickness across age showed that these findings were sustained and age-independent. These contrasting cortical thickness phenotypes were not elicited through diagnostic comparisons (for which there were no significant differences in thickness). By contrasting data-driven versus diagnostic groups and via inclusion of multiple imaging phenotypes (i.e., cortical thickness, subcortical volumes, white matter FA) our findings build on a previous analysis of the POND sample [22]. In addition, the novelty of our findings is also notable because of out-of-model differences among the data-driven groups in adaptive functioning and brain network structure. Similar frontal and temporal cortical regions have also been implicated in a mega-analysis from the ENIGMA group comparing ASD to typically developing controls, with both increased and decreased thickness found in ASD [44]. Our work suggests that these same regions contribute to biological differences among data-driven groups. It is possible that reduced cortical thickness in some cases versus increased thickness in others (which may reflect delayed maturation) exists in subgroups of children with the same NDD diagnosis, but is associated with different behavioral phenotypes (e.g., more inattention vs. more hyperactivity).

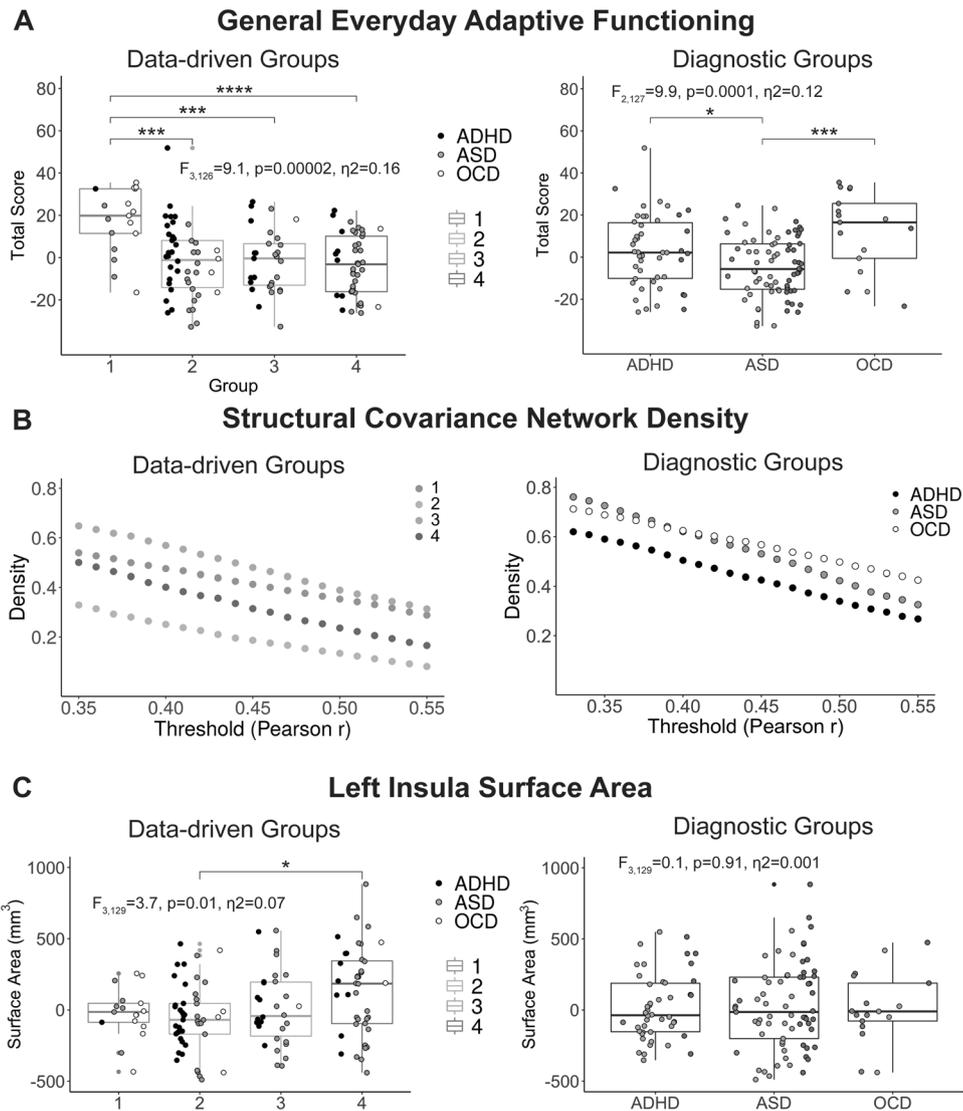


Fig. 3 Data-driven and NDD group differences on out-of-model brain and behavior measures. A ABAS-II General Adaptive Composite scores for both data-driven and diagnostic groups after regressing out the effects of age, sex and IQ. **B** Structural covariance network densities across a range of Pearson's *r* thresholds for both data-driven and diagnostic groups. Networks were created by correlating cortical thickness across regions after effects of age, sex and IQ were regressed out. **C** Left insula surface area in data-driven groups and diagnostic groups after regressing out the effects of age, sex and IQ (*****p* < 0.0001, ****p* < 0.001, ***p* < 0.01, **p* < 0.05). Boxplots show the first and third quartile with extensions representing 1.5 * the inter-quartile range.

Group 4 was characterized biologically by decreases in striatal and thalamic subcortical volumes. Reductions in similar regions were shown in ASD in the recent case-control ENIGMA ASD mega-analysis [44]. When these features are considered alongside top contributing cortical thickness features, involvement of the cortico-striatal-thalamic-cortical (CSTC) circuit also emerges as a notable pattern. The CSTC is a network widely implicated as vulnerable in ASD, ADHD and OCD [45–49]. Children in this group may represent those with this shared vulnerability pathway cutting across diagnoses.

In contrast to Groups 2–4, Group 1 was largely comprised of children with OCD. This group featured lower clinical symptom scores across included behavioral features (except for OCD symptoms), lacked distinguishable biological impairments, and had higher IQ scores than Groups 3 and 4. Some prior studies using POND data have found, on average, children with OCD have milder alterations at the brain and behavioral level compared to children with ASD or ADHD [12, 22, 35]. Our results identified

some children with ASD and one child with ADHD that also fit into this group. Group 1 also featured the highest adaptive (i.e., improved everyday) functioning compared to Groups 2, 3 and 4 on out-of-model evaluation. Longitudinal analyses are needed to track whether outcomes differ in this group compared to other groups over time. If higher adaptive functioning in this group remains stable over time, features such as behavior and functioning scores differentiating this group could potentially have clinical utility for identifying children with NDDs that may have favorable outcomes or respond differently to available treatment or clinical management approaches.

Although modest, brain network comparisons among the groups provided support for generalizability of distinctions between data-driven groups to features that were not utilized to delineate groups. In particular, the lower network density in Group 2 among cortical thickness regions supported the impaired cortical thickness phenotype found in this group using the data-driven model. Lower network density in Group 2 (and 4) may

indicate broader, more wide-ranging network-based alterations associated with their respective behavioral alteration profiles, and perhaps earlier developmental insults affecting more of the brain [50].

It is notable that females with a diagnosis of OCD or ADHD mainly clustered into Groups 1 and 2, while females with ASD clustered across data-driven groups. Biological sex is an important source of heterogeneity in NDDs [11] and aspects of sex-specific brain structure and functional connectivity patterns have been found in ASD [51] and to a lesser degree in ADHD [52] or OCD [53]. Previous evidence has suggested a protective effect for females, or required increased biological ‘hit’ related to resilience to developing NDDs [54, 55]. However, interpretation of any differences found amongst males versus females in the current study is limited due to the small number of represented females with NDDs.

We took a series of approaches to determine the stability of data-driven grouping and potential meaningfulness. We found that top contributing features could reliably identify our groups (mean sensitivity 42–86%), which may improve with larger sample sizes. All groups were identifiable using the full spectrum of features included in our SNF analysis; however, sensitivity for prediction of Groups 1 and 2 remained stable when based on a more constrained set of top contributing features (i.e., cortical thickness and inattention/hyperactivity-impulsivity symptom scores), suggesting that these groups may be identifiable in another sample using more constrained behavioral and biological information. Although Group 1 was the most stable across resampling, participants in this group were not classified with the highest sensitivity, perhaps due to their intermediate values on biological measures. Sensitivity for prediction of Groups 3 and 4 improved when the top two features from all four data types were included in the model. In particular, inclusion of the top two FA measures in the classification model improved sensitivity for Group 4. This finding suggests that, in contrast to Groups 1 and 2, a fuller spectrum of data may be needed to identify children with more complex presentations (and perhaps more overlapping behavioral and functional impairment profiles) as in Groups 3 and 4.

LIMITATIONS

Results should be interpreted in the context of study limitations. Sample sizes were unequal across diagnostic groups, and a larger sample could have provided a number of statistical advantages and likely improved clustering robustness. Adaptive functioning was moderately correlated to one of the top behavioral features that contributed to data-driven groups (i.e., SWAN inattention, see Fig. S10a). Future work could extend the dimensionality of input features to include cognitive, genetic, environmental and other neuroimaging features, further refine selection of input features, and reserve additional meaningful and independent measures for out-of-model testing. Repetition of analyses in new samples is needed to test agreement with the current study results. We were unable to take into account potential confounding variables such as socioeconomic status due to a lack of available data. Visualizing similarities between participants showed that although participants within the four data-driven groups identified featured more similar and separable brain-behavior profiles than found using conventional DSM-5 diagnostic categories, some participants did not cluster ‘cleanly’ into a specific data-driven group. Although studying individuals on a spectrum may be more informative for characterizing the continuum of brain-behavioral relationships present across the population (and shown to be relevant for clustering children with different NDDs [22]), others argue that biotypes (i.e., new subgroups) may be needed to parse multi-dimensional brain-behavior profiles into groupings that can be useful for clinical translation [10].

CONCLUSION

The current study adds to recent work suggesting biological and behavioral convergences across NDDs, as well as divergences within them [16, 22, 56]. We identified new groups cutting across NDDs characterized by multi-level neuroimaging and behavioral data. Children within these newly identified groups featured more similar profiles on brain and behavioral measures than found among conventional diagnostic (NDD) groupings. The current study provides an early demonstration of the feasibility and utility of using multi-level data integration to redefine clinical boundaries of NDDs. Future work is needed to build on this early attempt through prospective replication of our findings in independent samples (a high priority for the POND Network) and testing longitudinally for prognostic value.

FUNDING AND DISCLOSURES

This research was supported by the grant IDS-I I-02 from the Ontario Brain Institute. The Ontario Brain Institute is an independent non-profit corporation funded partially by the Ontario government. The opinions, results and conclusions are those of the authors and no endorsement by the Ontario Brain Institute is intended or should be inferred. GRJ currently receives funding from an Ontario Graduate Scholarship and the Ontario Student Opportunity Trust Fund. ANV currently receives funding from the National Institute of Mental Health (1/3R01MH102324 & 1/5R01MH114970), Canadian Institutes of Health Research, Canada Foundation for Innovation, CAMH Foundation, and University of Toronto. NJF receives funding from the Centre for Addiction and Mental Health Discovery Fund Postdoctoral Grant. M-CL receives funding from the Ontario Brain Institute via the POND Network, Canadian Institutes of Health Research, the Academic Scholars Award from the Department of Psychiatry, University of Toronto, and CAMH Foundation. PS has received royalties from Guilford Press. RS has consulted to Highland Therapeutics, Eli Lilly and Co., and Purdue Pharma. He has commercial interest in a cognitive rehabilitation software company, “eHave”. PDA receives funding from the Alberta Innovates Translational Health Chair in Child and Youth Mental Health and holds a patent for ‘SLCIAI Marker for Anxiety Disorder’ granted May 6, 2008. SHA currently receives funding from the National Institute of Mental Health (R01MH114879), Canadian Institutes of Health Research, the Academic Scholars Award from the Department of Psychiatry, University of Toronto, Cundill Centre for Child and Youth Depression and CAMH Foundation. Other authors report no financial interests or potential conflicts of interest.

ACKNOWLEDGEMENTS

We thank the following individuals for research support and data collection: Tara Goodale, M.Sc., Reva Schachter, M.Sc., Mithula Sriskandarajah, B.Sc., Marlina Colasanto, M.Sc., Jennifer Gomez, M.A., and Laura Park, M.Sc. from The Hospital for Sick Children; Susan Day Fragiadakis, M.A., Naomi Peleg, M.Sc., and Leanne Ristic, B.A., from Holland Bloorview Kids Rehabilitation Hospital; Richa Mehta, B.A., Christina Sommerdyk, M.Sc., from the Lawson Health Research Institute; Carolyn Russell, B.Sc., Alessia Greco, M.A., Mike Chalupka, B.A., B.Sc., Christina Chrysler, B.A., Irene O’Connor, M.Ed. Psych., from McMaster Children’s Hospital. We thank Margot J. Taylor for her review of a draft version of the manuscript. We thank Jason P. Lerch and Evdokia Anagnostou for their input into the analyses undertaken and manuscript preparation. We thank Hajer Nakua for her work on imaging quality control.

AUTHOR CONTRIBUTIONS

Conceptualization: SHA, ANV, GRJ; Methodology: GRJ, EWD, AG, LE, CH, NF, LS; Formal Analysis: GRJ, CH, NF; Resources: SHA, ANV; Drafting the manuscript: GRJ, SHA, ANV; Manuscript revisions and finalization: GRJ, ANV, CH, LS, NJF, EWD, M-CL, PS, RS, JC, PDA, AG, LE, SHA; Visualizations: GRJ; Supervision: SHA, ANV.

ADDITIONAL INFORMATION

Supplementary Information accompanies this paper at (<https://doi.org/10.1038/s41386-020-00902-6>).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Szatmari P, Georgiades S, Duku E, Bennett TA, Bryson S, Fombonne E, et al. Developmental trajectories of symptom severity and adaptive functioning in an inception cohort of preschool children with autism spectrum disorder. *JAMA Psychiatry*. 2015;72:276–83.
2. Karalunas SL, Fair D, Musser ED, Aykes K, Iyer SP, Nigg JT. Subtyping attention-deficit/hyperactivity disorder using temperament dimensions: toward biologically based nosologic criteria. *JAMA Psychiatry*. 2014;71:1015–24.
3. Di Rezze B, Duku E, Szatmari P, Volden J, Georgiades S, Zwaigenbaum L, et al. Examining trajectories of daily living skills over the preschool years for children with autism spectrum disorder. *J Autism Dev Disord*. 2019;49:4390–9.
4. Schwartzman CM, Boisseau CL, Sibrava NJ, Mancebo MC, Eisen JL, Rasmussen SA. Symptom subtype and quality of life in obsessive-compulsive disorder. *Psychiatry Res*. 2017;249:307–10.
5. Anholt GE, Cath DC, Van Oppen P, Eikelenboom M, Smit JH, Van Megen H, et al. Autism and adhd symptoms in patients with ocd: are they associated with specific oc symptom dimensions or oc symptom severity. *J Autism Dev Disord*. 2010;40:580–9.
6. Ashwood KL, Tye C, Azadi B, Cartwright S, Asherson P, Bolton P. Brief report: adaptive functioning in children with ASD, ADHD and ASD+ ADHD. *J Autism Dev Disord*. 2015;45:2235–42.
7. Van Der Meer JMJ, Oerlemans AM, Van Steijn DJ, Lappenschaar MGA, De Sonnevile LMJ, Buitelaar JK, et al. Are autism spectrum disorder and attention-deficit/hyperactivity disorder different manifestations of one overarching disorder? Cognitive and symptom evidence from a clinical and population-based sample. *J Am Acad Child Adolesc Psychiatry*. 2012. <https://doi.org/10.1016/j.jaac.2012.08.024>.
8. Lionel AC, Crosbie J, Barbosa N, Goodale T, Thiruvahindrapuram B, Rickaby J, et al. Rare Copy Number Variation Discovery and Cross-Disorder Comparisons Identify Risk Genes for ADHD. *Sci Transl Med*. 2010;40:580–9.
9. Constantino JN, Charman T. Diagnosis of autism spectrum disorder: reconciling the syndrome, its diverse origins, and variation in expression. *Lancet Neurol*. 2016;15:279–91.
10. Feczko E, Miranda-Dominguez O, Marr M, Graham AM, Nigg JT, Fair DA. The heterogeneity problem: approaches to identify psychiatric subtypes. *Trends Cogn Sci*. 2019;23:584–601.
11. Lombardo MV, Lai M-C, Baron-Cohen S. Big data approaches to decomposing heterogeneity across the autism spectrum. *Mol Psychiatry*. 2019;24:1435–50.
12. Ameis SH, Lerch JP, Taylor MJ, Lee W, Viviano JD, Pipitone J, et al. A diffusion tensor imaging study in children with ADHD, Autism Spectrum Disorder, OCD, and Matched Controls: distinct and non-distinct white matter disruption and dimensional brain-behavior relationships. *Am J Psychiatry*. 2016;173:1213–22.
13. Dajani DR, Burrows CA, Odriozola P, Baez A, Beth M, Mostofsky SH, et al. NeuroImage: clinical Investigating functional brain network integrity using a traditional and novel categorical scheme for neurodevelopmental disorders. *NeuroImage Clin*. 2019;21:101678.
14. Di Martino A, Zuo X-N, Kelly C, Grzadzinski R, Mennes M, Schvarcz A, et al. Shared and distinct intrinsic functional network centrality in autism and attention-deficit/hyperactivity disorder. *Biol Psychiatry*. 2013;74:623–32.
15. Chantiluke K, Christakou A, Murphy CM, Giampietro V, Daly EM, Ecker C, et al. Disorder-specific functional abnormalities during temporal discounting in youth with Attention Deficit Hyperactivity Disorder (ADHD), Autism and comorbid ADHD and Autism. *Psychiatry Res*. 2014;223:113–20.
16. Boedhoe PSW, van Rooij D, Hoogman M, Twisk JWR, Schmaal L, Abe Y, et al. Subcortical brain volume, regional cortical thickness and cortical surface area across attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), and obsessive-compulsive disorder (OCD). *bioRxiv*. 2019. p. 673012. <https://doi.org/10.1101/673012>.
17. Baribeau DA, Dupuis A, Paton TA, Hammill C, Scherer SW, Schachar RJ, et al. Structural neuroimaging correlates of social deficits are similar in autism spectrum disorder and attention-deficit/hyperactivity disorder: analysis from the POND Network. *Transl Psychiatry*. 2019;9:72.
18. Aoki Y, Yoncheva YN, Chen B, Nath T, Sharp D, Lazar M, et al. Association of white matter structure with autism spectrum disorder and attention-deficit/hyperactivity disorder. *JAMA Psychiatry*. 2017;74:1120–8.
19. Feczko E, Balba NM, Miranda-Dominguez O, Cordova M, Karalunas SL, Irwin L, et al. Subtyping cognitive profiles in autism spectrum disorder using a functional random forest algorithm. *Neuroimage*. 2018;172:674–88.
20. Stefanik L, Erdman L, Ameis SH, Fossias G, Mulsant BH, Behdinan T, et al. Brain-behavior participant similarity networks among youth and emerging adults with schizophrenia spectrum, autism spectrum, or bipolar disorder and matched controls. *Neuropsychopharmacology*. 2017;43:1180.
21. Hawco C, Buchanan RW, Calarco N, Mulsant BH, Viviano JD, Dickie EW, et al. Separable and replicable neural strategies during social brain function in people with and without severe mental illness. *Am J Psychiatry*. 2019;176:521–30.
22. Kushki A, Anagnostou E, Hammill C, Duez P, Brian J, Iaboni A, et al. Examining overlap and homogeneity in ASD, ADHD, and OCD: a data-driven, diagnosis-agnostic approach. *Transl Psychiatry*. 2019;9:318.
23. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, et al. Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods*. 2014;11:333.
24. Lord C, Rutter M, Le Couteur A. Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord*. 1994;24:659–85.
25. Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC, et al. The autism diagnostic observation schedule—generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord*. 2000;30:205–23.
26. Ickowicz A, Schachar RJ, Sugarman R, Chen SX, Millette C, Cook L. The parent interview for child symptoms: a situation-specific clinical research interview for attention-deficit hyperactivity and related disorders. *Can J Psychiatry*. 2006;51:325–8.
27. Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, et al. Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry*. 1997;36:980–8.
28. Scahill L, Riddle MA, McSwiggin-Hardin M, Ort SI, King RA, Goodman WK, et al. Children's Yale-Brown Obsessive Compulsive Scale: reliability and validity. *J Am Acad Child Adolesc Psychiatry*. 1997;36:844–52.
29. Achenbach TM, Edelbrock CS, Others. Manual for the child behavior checklist and revised child behavior profile. 1983. <https://pdfs.semanticscholar.org/e06f/18f950ee20811acd25b9671c14a80b681e3c.pdf>.
30. Park LS, Burton CL, Dupuis A, Shan J, Storch EA, Crosbie J, et al. The Toronto obsessive-compulsive scale: psychometrics of a dimensional measure of obsessive-compulsive traits. *J Am Acad Child Adolesc Psychiatry*. 2016;55:310. e4
31. Berument SK, Rutter M, Lord C, Pickles A, Bailey A. Autism screening questionnaire: diagnostic validity. *Br J Psychiatry*. 1999;175:444–51.
32. Bodfish JW, Symons FJ, Parker DE, Lewis MH. Varieties of repetitive behavior in autism: comparisons to mental retardation. *J Autism Dev Disord*. 2000;30:237–43.
33. Swanson J, Schuck S, Mann M, Carlson C, Hartman K, Sergeant J, et al. Categorical and dimensional definitions and evaluations of symptoms of ADHD: the SNAP and SWAN rating scales. Irvine: University of California; 2006.
34. Harrison PL, Oakland T. (eds) Adaptive Behavior Assessment System-II. In Practical Resources for the Mental Health Professional. Academic Press, 2008. <https://doi.org/10.1016/B978-012373586-7.00031-X>.
35. Baribeau DA, Doyle-Thomas KAR, Dupuis A, Iaboni A, Crosbie J, McGinn H, et al. Examining and comparing social perception abilities across childhood-onset neurodevelopmental disorders. *J Am Acad Child Adolesc Psychiatry*. 2015;54:479–86. e1
36. Fischl B. FreeSurfer. *Neuroimage* 2012;62:774–81.
37. Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, et al. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage* 2006;31:1487–505.
38. Johns Hopkins University. School of Medicine, Jean S. The Johns Hopkins Atlas of Human Functional Anatomy. Johns Hopkins University Press; 1997.
39. Steinley D, Brusco MJ, Hubert L. The variance of the adjusted Rand index. *Psychol Methods*. 2016;21:261–72.
40. Sussman D, Leung RC, Chakravarty MM, Lerch JP, Taylor MJ. Developing human brain: age-related changes in cortical, subcortical, and cerebellar anatomy. *Brain Behav*. 2016;6:e00457.
41. Oyefiade AA, Ameis S, Lerch JP, Rockel C, Szulc KU, Scantlebury N, et al. Development of short-range white matter in healthy children and adolescents. *Hum Brain Mapp*. 2018;39:204–17.
42. Shaw P, Greenstein D, Sharp W, Clasen L, Giedd J, Rapoport J, et al. Longitudinal mapping of cortical thickness and clinical outcome in children and adolescents with attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry*. 2006;63:540–9.
43. Shaw P, Eckstrand K, Sharp W, Blumenthal J, Lerch JP, Greenstein D, et al. Attention-deficit/hyperactivity disorder is characterized by a delay in cortical maturation. *Proc Natl Acad Sci*. 2007;104:19649–54.

44. Van Rooij D, Anagnostou E, Arango C, Auzias G, Behrmann M, Busatto GF, et al. Cortical and subcortical brain morphometry differences between patients with autism spectrum disorder and healthy individuals across the lifespan: Results from the ENIGMA ASD working group. *Am J Psychiatry*. 2018;175:359–69.
45. Zhu Y, Jiang X, Ji W. The mechanism of cortico-striato-thalamo-cortical neuro-circuitry in response inhibition and emotional responding in attention deficit hyperactivity disorder with comorbid disruptive behavior disorder. *Neurosci Bull*. 2018;34:566–72.
46. Kuo H-Y, Liu F-C. Synaptic wiring of corticostriatal circuits in basal ganglia: insights into the pathogenesis of neuropsychiatric disorders. *eNeuro*. 2019;6. <https://doi.org/10.1523/ENEURO.0076-19.2019>.
47. Fettes P, Schulze L, Downar J. Cortico-Striatal-Thalamic loop circuits of the orbitofrontal cortex: promising therapeutic targets in psychiatric illness. *Front Syst Neurosci*. 2017;11:25.
48. Hoogman M, Bralten J, Hibar DP, Mennes M, Zwiers MP, Schwaren LSJ, 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.
49. Boedhoe PSW, Schmaal L, Abe Y, Ameis SH, Arnold PD, Batistuzzo MC, et al. Distinct subcortical volume alterations in pediatric and adult OCD: a worldwide meta- and mega-analysis. *Am J Psychiatry*. 2017;174:60–9.
50. Bethlehem RAI, Romero-Garcia R, Mak E, Bullmore ET, Baron-Cohen S, Bethlehem RAI. Structural covariance networks in children with autism or ADHD. *Cereb Cortex*. 2017;27:4267–76.
51. Lai M-C, Lerch JP, Floris DL, Ruigrok ANV, Pohl A, Lombardo MV, et al. Imaging sex/gender and autism in the brain: etiological implications. *J Neurosci Res*. 2017;95:380–97.
52. Scofield JE, Johnson JD, Wood PK, Geary DC. Latent resting-state network dynamics in boys and girls with attention-deficit/hyperactivity disorder. *PLoS One*. 2019;14:e0218891.
53. Hawco C, Voineskos AN, Radhu N, Rotenberg D, Ameis S, Backhouse FA, et al. Age and gender interactions in white matter of schizophrenia and obsessive compulsive disorder compared to non-psychiatric controls: commonalities across disorders. *Brain Imaging Behav*. 2017;11:1836–48.
54. Werling DM. The role of sex-differential biology in risk for autism spectrum disorder. *Biol Sex Differ*. 2016;7:58.
55. Taylor MJ, Lichtenstein P, Larsson H, Anckarsäter H, Greven CU, Ronald A. Is there a female protective effect against attention-deficit/hyperactivity disorder? Evidence from two representative twin samples. *J Am Acad Child Adolesc Psychiatry*. 2016;55:504–12. e2.
56. Brem S, Grünblatt E, Drechsler R, Riederer P, Walitza S. The neurobiological link between OCD and ADHD. *Atten Defic Hyperact Disord*. 2014;6:175–202.