Dataset factors associated with age-related changes in brain structure and function in neurodevelopmental conditions

Marlee M. Vandewouw[1,2*], Yifan (Julia) Ye[1,3], Jennifer Crosbie[4,5], Russell J. Schachar[4,5^], Alana Iaboni[1], Stelios Georgiades[6], Robert Nicolson[7], Elizabeth Kelley[8,9,10], Muhammad Ayub[10,11], Jessica Jones[8,9,10], Paul D. Arnold[12], Margot J. Taylor[13,14,15,16], Jason P. Lerch[14,17,18], Evdokia Anagnostou[1,14,19], Azadeh Kushki[1,2]

1 Autism Research Centre, Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, ON, Canada
2 Institute of Biomedical Engineering, University of Toronto, Toronto, Canada
3 Division of Engineering Science, University of Toronto, Toronto, Canada
4 Department of Psychiatry, University of Toronto, Toronto, Canada
5 Department of Psychiatry, The Hospital for Sick Children, Toronto, Ontario, Canada
6 Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Canada
7 Department of Psychiatry, Western University, London, Canada
8 Department of Psychology, Queen's University, Kingston, Canada
9 Centre for Neuroscience Studies, Queen's University, Kingston, Canada
10 Department of Psychiatry, Queen's University, Kingston, Canada
11 Division of Psychiatry, University of College London, London, United Kingdom
12 The Mathison Centre for Mental Health Research & Education, Cumming School of Medicine, University of Calgary, Calgary, Canada
13 Department of Diagnostic and Interventional Radiology, The Hospital for Sick Children, Toronto, Canada
14 Program in Neurosciences & Mental Health, The Hospital for Sick Children, Toronto, Canada
15 Department of Psychology, University of Toronto, Toronto, Canada
16 Department of Medical Imaging, University of Toronto, Toronto, Canada
17 Wellcome Centre for Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom,
18 Department of Medical Biophysics, University of Toronto, Toronto, Canada
19 Institute of Medical Science, University of Toronto, Toronto, Canada

* Corresponding author:
Azadeh Kushki
Autism Research Centre, Holland Bloorview Kids Rehabilitation Hospital
150 Kilgour Road, Toronto, ON M4G 1R8
Phone: 416-425-6220, extension 3563
Email: akushki@hollandbloorview.ca

^Emeritus

Data availability statement: Data is available from the Province of Ontario Neurodevelopmental Disorders (POND) network (exported April 2023; now available via a controlled data release through Ontario Brain Institute's Brain-CODE: https://www.braincode.ca/) and the Healthy Brain Network (Release 9; http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/) datasets.

Funding statement: Funding was provided by the Ontario Brain Institute, Canadian Institutes of Health Research, and New Frontiers in Research Fund.

Conflict of interest disclosure: R. Nicolson has received grants from Brain Canada, Hoffman La Roche, Otsuka Pharmaceuticals, and Maplight Therapeutics. E. Anagnostou has received grants from Roche and Anavex, served as a consultant to Roche, Quadrant Therapeutics, Ono, and Impel Pharmaceuticals, has received in-kind support from AMO Pharma and CRA-Simons Foundation, holds a patent for the device, “Anxiety Meter”, has received royalties from APPI and Springer, and has received an editorial honorarium from Wiley. A. Kushki received grants from the National Science and Engineering Research Council and holds a patent for Anxiety Meter with royalties paid from Awake Labs. R. Schachar has served as a consultant to Ehave, Lilly Corporation, and Highland Therapeutics. The remaining authors have reported no biomedical financial interests or potential conflicts of interest.

Ethics approval statement: The current study was approved by the institution’s research ethics board, and both consortia studies were approved by the appropriate boards.

Abstract

With brain structure and function undergoing complex changes throughout childhood and adolescence, age is a critical consideration in neuroimaging studies, particularly for those of individuals with neurodevelopmental conditions. However, despite the increasing use of large, consortium-based datasets to examine brain structure and function in neurotypical and neurodivergent populations, it is unclear whether age-related changes are consistent between datasets, and whether inconsistencies related to differences in sample characteristics, such as demographics and phenotypic features, exist. To address this, we built models of age-related changes of brain structure (regional cortical thickness and regional surface area; *N*=1,218) and function (resting-state functional connectivity strength; *N*=1,254) in two neurodiverse datasets: the Province of Ontario Neurodevelopmental network (POND) and the Healthy Brain Network (HBN). We examined whether deviations from these models differed between the datasets, and explored whether these deviations were associated with demographic and clinical variables. We found significant differences between the two datasets for measures of cortical surface area and functional connectivity strength throughout the brain. For regional measures of cortical surface area, the patterns of differences were associated with race/ethnicity, while for functional connectivity strength, positive associations were observed with head motion. Our findings highlight that patterns of age-related changes in the brain may be influenced by demographic and phenotypic characteristics, and thus future studies should consider these when examining or controlling for age effects in analyses.

Introduction

The brain's structure and function undergo complex and protracted changes throughout childhood and adolescence due to neurobiological processes such as synaptogenesis, myelination, and synaptic pruning (Grayson & Fair, 2017; Mills & Tamnes, 2020; Norbom et al., 2021). Neurodevelopmental differences, such as those associated with autism spectrum disorder (autism), attention-deficit/hyperactivity disorder (ADHD), and obsessive-compulsive disorder (OCD), can further impact the presentation of age-related changes, particularly given the heterogeneity in etiology, biology, and behaviour (Bruin et al., 2020; Lombardo et al., 2019; Wolfers et al., 2019). As such, age is a critical consideration in neuroimaging studies of neurotypical and neurodivergent populations.

While many studies directly examine age-related effects, age is often treated as a nuisance covariate (Hyatt et al., 2020). With increasing awareness of the importance of replicability in neuroimaging studies (Poldrack et al., 2017), it is becoming more common to leverage multiple consortium-based datasets to examine brain structure and function in neurotypical and neurodivergent populations (Abrol et al., 2023; Grotzinger et al., 2023; Marek et al., 2022; Nicolaisen-Sobesky et al., 2022; Romer et al., 2019; Vandewouw et al., 2023). However, these datasets differ in composition across factors which may influence age-related effects, such as participants' demographics (e.g., age, sex and gender, race and ethnicity, socioeconomic status), diagnosis, co-occurring conditions, and other phenotypic variables. As a result, it is possible that the patterns of age-related changes in brain structure and function differ between datasets. Since it is common to apply the same assumptions about the development of neurobiology across datasets (e.g., linearity), the presence of dataset-specific developmental effects would be problematic for downstream analyses, challenging replicability.

Studies have examined the consistency of age-related changes in brain structure and function across neurotypical cohorts (Cao et al., 2017; Herting et al., 2018; Madan & Kensinger, 2017; Mills et al., 2016; Morandini et al., 2021; Tamnes et al., 2017; Wang et al., 2021). For example, yearly percentage changes in cortical thickness and surface area across childhood and adulthood have been shown to differ between four samples

of typically developing individuals, with the precise patterns varying across lobes of the brain (Tamnes et al., 2017). White matter volume has also been shown to peak at different age ranges in different cohorts, with those from the United States showing decreased volume across age compared to European samples (Mills et al., 2016). These findings have important implications for those studies which are not only directly examining age-related effects, but also those which treat age as a covariate, as controlling for age in the same manner does not consider these differences. Yet, no studies have examined whether these changes are consistent between cohorts of neurodivergent children, and if not, what factors may be contributing to the inconsistencies. This is particularly critical to investigate given the large variability in demographic and phenotypic characteristics of samples across different cohorts, even when the same diagnostic labels are studied.

While there is a wealth of evidence showing that demographic and clinical factors can impact brain development (e.g., neurodivergence), most of this work has taken a case-control approach, where differences from neurotypical trajectories are established at the “average” level. On the other hand, there is a considerable amount of developmental variation at the individual level (Foulkes & Blakemore, 2018), with emerging evidence that clinical features of autism and ADHD are associated with individual deviations from neurotypical maturational trajectories (Bethlehem et al., 2020; Floris et al., 2021; Kessler et al., 2016; Looden et al., 2022; Marquand et al., 2016; Oliveira-Saraiva & Ferreira, 2023; Parkes et al., 2021; Tunç et al., 2019; Wolfers et al., 2019; Zabihi et al., 2019). Thus, to understand why datasets may exhibit distinct age-related changes of neurobiology, it is important to consider how demographic and clinical factors impact individual variations.

To this end, we examined age-related changes of brain structure (regional cortical thickness and surface area) and function (functional connectivity strength) in two transdiagnostic datasets of neurodiverse children and adolescents, namely the Province of Ontario Neurodevelopmental network (POND) and the Healthy Brain Network (HBN). We first built dataset-specific developmental models and evaluated whether deviations from these models differed between the POND and HBN individuals, separately for

males and females. Next, we explored whether the patterns of between-dataset differences were associated with demographic and phenotypic variables. We have previously reported that POND and HBN differ in their phenotypic profiles, specifically full-scale intelligence quotient (FSIQ), social communication difficulties, and ADHD symptoms (Vandewouw et al., 2023). Thus, we hypothesize that deviations from age-related models of brain structure and function will differ between these datasets, and the deviations will be associated with these phenotypic measures.

Methods

Participants

Neuroimaging and phenotype data were obtained from two independently collected datasets: the Province of Ontario Neurodevelopmental network (POND; exported April 2023), and the Healthy Brain Network (HBN; Release 9). For both datasets, participants who were between 5-19 years of age and who were either neurotypical or had diagnoses of autism, attention-deficit/hyperactivity disorder (ADHD), or obsessive-compulsive disorder (OCD) were selected. Details on the diagnostic assessments are provided elsewhere (Vandewouw et al., 2023). Structural magnetic resonance imaging (sMRI) data were available for 868 POND (255 ADHD, 377 autism, 75 OCD, 161 TD) and 1,505 HBN (1,100 ADHD, 162 autism, 36 OCD, 207 TD) participants and resting-state functional magnetic imaging (rs-fMRI) data were available for 864 POND (211 ADHD, 385 autism, 92 OCD, 185 TD) and 1,459 HBN (1,068 ADHD, 160 autism, 29 OCD, 202 TD) participants. Informed consent and/or assent was obtained from the participants and/or caregivers. The current study was approved by the institution’s research ethics board, and both consortia studies were approved by the appropriate boards.

Data acquisition

For both datasets, T1-weighted magnetization-prepared rapid gradient echo (MPRAGE) sequences (sMRI data) and five minutes of resting-state data with an echo planar imaging (EPI) sequence (rs-fMRI data) were acquired. In POND, the rs-fMRI data were obtained as participants watched either a movie of their choosing or the

naturalistic movie paradigm akin to a screensaver, *Inscapes* (Vanderwal et al., 2015), while HBN participants viewed a fixation cross. While HBN collected two five-minute resting-state acquisitions, only one was selected to be consistent with POND. Acquisition parameters for the sMRI and rs-fMRI protocols for each POND and HBN site are presented in **Supplemental Table 1**.

Biological sex was collected as part of both studies and used in the current analysis; insufficient data on gender was available at the time of export. Phenotypic measures of core and co-occurring differences for the diagnostic categories (see **Supplemental Table 2**), socioeconomic status, and race and ethnicity data were also available for both datasets and were used to characterize each sample. Socioeconomic status was determined using the caregiver's highest level of education and annual household income. Race and ethnicity data were collected for both datasets according to each country's census guidelines and were categorized into five racial/ethnic categories: Hispanic/Latino, non-Hispanic Asian, non-Hispanic Black, non-Hispanic White, and Multi-Racial/Other (Cardenas-Iniguez & Gonzalez, 2024).

Image processing

The sMRI data were processed using the CIVET pipeline (version 2.1.0; (Ad-Dab'bagh et al., 2006)) on CBRAIN (Sherif et al., 2014). CIVET first registers the T1-weighted image to stereotaxic space (Collins et al., 1994) using the Montreal Neurological Institute template (MNI ICBM152; (Fonov et al., 2009)) and corrected for non-uniformities (Sled et al., 1998). A brain mask was extracted (Smith, 2002), and tissues were classified into white matter, gray matter, and cerebrospinal fluid using a discrete tag point classification and a correction for partial volumes (Tohka et al., 2004; Zijdenbos et al., 1998). The white matter and pial surfaces (40,962 vertices per hemisphere) were extracted using the marching-cubes algorithm (Kabani et al., 2001; Kim et al., 2005; MacDonald et al., 2000). Vertex-wise cortical thickness was calculated as the Euclidean distance between reference vertices of the white and pial surfaces (Ad-Dab'bagh et al., 2005; Lerch & Evans, 2005). Vertex-wise cortical surface area and volumes were calculated as local variations of contraction and expansion relative to the distribution of vertices (Lyttelton et al., 2009). Surface-based diffusion smoothing

kernels were applied (thickness: 30mm full-width at half-maximum (FWHM); area and volume: 40mm FWHM) (Boucher et al., 2009). Regional measures of cortical thickness and surface area were extracted for 76 brain regions defined using the Automated Anatomical Labelling (AAL) atlas (Tzourio-Mazoyer et al., 2002) by averaging across the vertices per region. The AAL atlas was chosen to be consistent with most studies using the CIVET pipeline, including work from our group (Kushki et al., 2019, 2021; Sadat-Nejad et al., 2023; Sanjeevan et al., 2020; Sussman et al., 2015). The CIVET automatic quality control pipeline was used, details of which are described elsewhere (Hammill et al., 2021; Kushki et al., 2019, 2021). Of the remaining participants that passed quality control, propensity score matching (Rosenbaum & Rubin, 1985) was used to identify a smaller subset of participants to ensure the datasets showed no significant difference in median age or male-to-female ratio (see (Vandewouw et al., 2023), for further details). Across both datasets, ComBat harmonization (Fortin et al., 2018; Johnson et al., 2007) was used to adjust all measures for site effects, followed by the regression of mean cortical thickness and total surface area for the regional measures of thickness and area, respectively. For both steps, an iterative sampling procedure was performed to increase robustness (10,000 iterations, each selecting 63.2% of the sample (Da-ano et al., 2020)).

The rs-fMRI data preprocessing was performed using fMRIPrep (Esteban et al., 2019, 2020), and is described in detail elsewhere (Vandewouw et al., 2023). As part of the pipeline, measures of framewise displacement (FD; (Power et al., 2014)) and the standardized derivative of root mean square variance over voxels (DVARS; (Power et al., 2014)) were obtained. Participants were required to have at least 2/3 of their data within the recommended motion thresholds (FD: 0.5mm; DVARS: 1.5) to pass quality control. Of the remaining participants, propensity score matching was used to identify a smaller subset of individuals for which there was not a statistically significant between-dataset differences in median age, male-to-female ratio, and median head motion (Vandewouw et al., 2023). The Schaefer cortical (Schaefer et al., 2018) and Melbourne subcortical (Tian et al., 2020) atlases were used to define 232 brain regions, which have been shown to produce functional network topology that is more similar across network construction pipelines compared to the AAL and other alternative functional

151 parcellations (Luppi & Stamatakis, 2021). Pairwise Pearson correlations were computed
152 between region-averaged time series to construct rs-fMRI connectomes. ComBat
153 harmonization (Da-ano et al., 2020; Fortin et al., 2018; Johnson et al., 2007) was used
154 to adjust the connectivity values for site effects across both datasets (10,000 iterations,
155 each selecting 63.2% of the sample). The resulting connectomes were thresholded with
156 orthogonal minimum spanning trees to remove spurious connections (Dimitriadis et al.,
157 2017), and the functional connectivity strength of each brain region was extracted.

158 ***Modeling age-related changes***

159 Models of age-related changes in the regional sMRI (area and thickness) and rs-
160 fMRI (connectivity strength) measures were built using Bayesian linear regression (BLR;
161 (Huertas et al., 2017)). Unlike traditional regression, BLR estimates the posterior
162 probability of the regression coefficient, capturing the expected variation or uncertainty
163 around the predicted age-related change; this enables the computation of a statistical *z*-
164 score, or deviation score, that is similar to a residual but also considers the uncertainty
165 around the prediction (Huertas et al., 2017; Rutherford et al., 2022). BLR is comparable
166 but less computationally expensive than similar techniques such as Gaussian process
167 regression (Fraza et al., 2021; Huertas et al., 2017). We used the BLR implementation
168 from the Predictive Clinical Neuroscience toolkit (version 0.25;
169 https://pcntoolkit.readthedocs.io/en/latest) in Python (version 3.9), applying a B-spline
170 basis expansion of age with three evenly spaced knots to model non-linearities, using
171 warping to model normal and non-normal distributions in the brain measures, and using
172 the Powell method for optimization of the Bayesian hyper-parameters.

173 Dataset (POND and HBN; *N*=2) and sex (males and females; *N*=2) specific
174 models were built for each measure (regional cortical thickness, regional cortical
175 surface area, and functional connectivity strength; *N*=3), resulting in 12 models. To
176 promote generalizability, 80% of participants of a given dataset were used to build a
177 model, with 20% withheld for testing. This procedure was repeated 10,000 times, each
178 time with a different randomly selected subset of participants. To describe the
179 performance of the models, the model fit was evaluated by examining the mean
180 standardized log-loss (MSLL; (Rasmussen & Williams, 2006)), which quantifies the error

between the predicted and true mean of the data while also accounting for the predictive variances, and has an upper bound of 0, with negative values indicating a better model fit (Rutherford et al., 2022). Deviation scores (*z*-scores) were calculated from a built model under two test scenarios: the model built on 80% of a given dataset was tested on 1) the remaining 20% of participants from the same dataset, and 2) the participants from the other dataset. **Figure 1** depicts an example of this process where the model is built using 80% of the participants in POND and tested on the remaining 20% of participants from POND, and 100% of the HBN participants. For robustness, the analysis was also performed with a 70%/30% train/test split. Median model fit metrics and deviation scores were computed across all iterations. This procedure was performed separately for males and females, with both the full samples and smaller, matched samples.

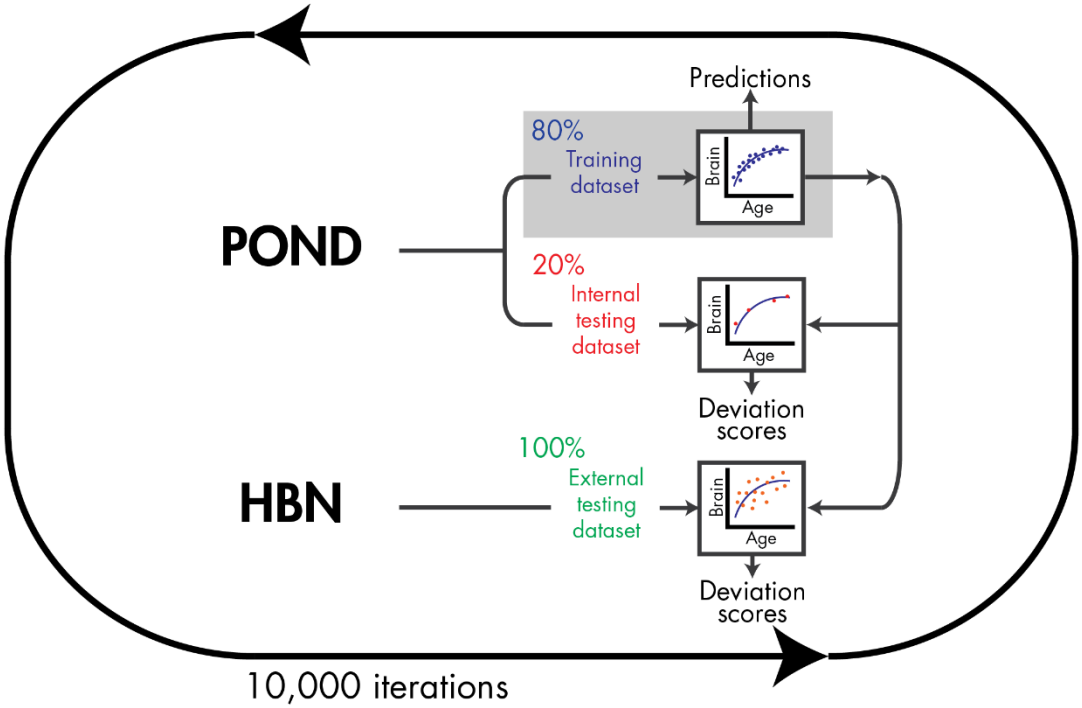


Figure 1: An example of the approach for constructing comparing the dataset- and sex-specific BLR models.

Statistics

First, differences between the two datasets in their demographic and clinical characteristics were examined, separately for males and females. Chi-squared tests,

Kruskal-Wallis tests (given non-normality according to Shapiro-Wilks tests), and ordinal regressions were used for categorical, continuous, and ordinal variables, respectively, reporting Cramer's *v*, eta-squared, and odds ratio effect sizes; in all cases, significance was held at $p<0.05$.

For each of the brain measures, we compared the deviations between the datasets; this was performed with both POND and HBN used as the training dataset, and separately for males and females. One-way ANOVAs or Kruskal-Wallis tests were used where appropriate, resulting *p*-values were adjusted using the false discovery rate (FDR) method, holding significance at $q<0.05$ and reporting eta-squared effect sizes.

Given differences in demographic (socioeconomic status and race/ethnicity) and clinical (diagnosis and phenotypic measures) variables between POND and HBN, we explored what factors may be contributing to any identified between-dataset differences. To do so, for each brain measure, we extracted the deviations for each participant from their dataset- and sex-specific model and used generalized linear mixed-effects models to examine their associations with these variables; dataset and sex were modeled as random factors. Resulting *p*-values were FDR-corrected across the brain regions, and significance was held at $q < 0.05$.

Results

Participants

For the sMRI dataset, 754 HBN (214 female, 540 male) and 825 POND (270 female, 555 male) children and adolescents were included in the analysis. Given the difference in age between the two datasets (females: $H(1)=104.35$, $p<.001$, $\eta^2=.04$; males: $H(1)=203.75$, $p<.001$, $\eta^2=.03$), a smaller, matched subset of participants was identified for which no significant difference remained, with each dataset consisting of 609 (178 females, 431 males) individuals. Participant demographics for the matched sample are presented in **Table 1**, with phenotypic measures, socioeconomic status, and race/ethnicity for each dataset are presented in **Table 2**. Co-occurring conditions are summarized in **Supplemental Table 3**.

Table 1: Participant demographics for the sMRI sample.

			HBN	POND	Test-Statistic[1]	*p*-value	Effect size
Females	***N***		178	178	–	–	–
	Dx (*N*)	**ADHD**	124	51	76.33	<.001	.46
		Autism	7	56			
		OCD	9	24			
		TD	38	47			
	Age[2] (years)		10.1 [8.5, 13.3]	11.1 [9.0, 13.3]	2.61	.107	<.01
Males	***N***		431	431	–	–	–
	Dx (*N*)	**ADHD**	304	145	141.73	<.001	.41
		Autism	55	190			
		OCD	11	33			
		TD	61	63			
	Age[2] (years)		10.6 [8.6, 13.0]	10.9 [9.0, 13.0]	2.07	.150	<.01

HBN: Healthy Brain Network; POND: Province of Ontario Neurodevelopmental network; ADHD: attention-deficit/hyperactivity disorder; OCD: obsessive-compulsive disorder; TD: typically developing; [1]Kruskal-Wallis *H*-statistic for age with eta-squared effect size, and chi-squared statistic for diagnosis with Cramer's *v* effect size; [2]Age is reported as median [IQR].

Table 2: Socioeconomic status, race/ethnicity, and phenotypic measures for the sMRI sample.

		HBN	POND	Test-Statistic[1]	*p*-value	Effect size
Females	**Race/ethnicity[2]**					
	Hispanic/Latino	30	4			
	Non-Hispanic Asian	7	11			
	Non-Hispanic Black	15	1	40.39	<.001	0.39
	Non-Hispanic White	67	90			
	Other/Multi-Racial	12	24			
	Caregiver income[3]					
	<$50k/year	31	18			
	$50k-$150k	65	57	0.03	.974	0.99
	>$150k	51	34			
	Caregiver education[4]					
	< High school	6	1			
	High school	19	15	3.88	<.001	2.44
	College	58	72			
	Graduate	94	27			
	SCQ	6.0 [4.0, 8.0]	5.0 [2.0, 14.2]	0.18	.675	.02
	RBS-R	11.0 [2.0, 28.0]	11.5 [2.0, 29.0]	0.22	.639	.01
	SWAN-I	2.0 [0.0, 5.0]	2.0 [0.0, 7.0]	0.04	.843	<.01
	SWAN-HI	0.0 [0.0, 2.0]	1.0 [0.0, 4.0]	7.16	.007	.02
	TOCS	–	-6.0 [-40.0, 6.0]	–	–	–
	CBCL-OCS	2.0 [1.0, 4.8]	–	–	–	–
	FSIQ	102.0 [87.0, 113.5]	106.0 [94.0, 115.0]	1.90	.168	.01
	CBCL-I	56.0 [48.0, 65.0]	60.0 [52.0, 69.0]	6.91	.009	.02
	CBCL-E	57.0 [47.0, 64.8]	55.0 [44.0, 64.0]	0.80	.372	<.01
Males	**Race/ethnicity[2]**					
	Hispanic/Latino	91	25			
	Non-Hispanic Asian	8	25			
	Non-Hispanic Black	41	7	88.90	<.001	0.38
	Non-Hispanic White	148	158			
	Other/Multi-Racial	30	75			
	Caregiver income[3]					
	<$50k/year	67	47			
	$50k-$150k	133	123	3.09	.002	1.64
	>$150k	146	58			
	Caregiver education[4]					
	< High school	6	4			
	High school	79	30	3.39	.001	1.66
	College	143	149			
	Graduate	196	57			
	SCQ	8.0 [5.0, 11.0]	10.0 [4.0, 20.0]	19.42	<.001	.06
	RBS-R	14.5 [4.0, 44.0]	16.0 [5.0, 31.0]	0.77	.381	<.01
	SWAN-I	3.0 [0.0, 6.0]	4.0 [1.0, 7.0]	11.37	<.001	.01
	SWAN-HI	1.0 [0.0, 4.0]	2.0 [0.0, 6.0]	22.60	<.001	.03
	TOCS	–	-15.0 [-45.0, 2.0]	–	–	–
	CBCL-OCS	2.0 [1.0, 5.0]	–	–	–	–
	FSIQ	100.5 [87.0, 111.0]	101.0 [88.0, 111.0]	0.07	.794	<.01
	CBCL-I	58.0 [50.0, 67.0]	62.0 [52.0, 68.0]	10.87	.001	.01
	CBCL-E	56.0 [48.0, 65.0]	56.0 [48.0, 66.0]	0.06	.814	<.01

sMRI: structural magnetic resonance imaging; HBN: Healthy Brain Network; POND: Province of Ontario Neurodevelopmental network; SCQ: social communication questionnaire; RBS-R: repetitive Behaviours Scale – Revised; SWAN-I: Strengths and Weaknesses of ADHD-symptoms and Normal Behaviour inattention subscale; SWAN-HI: Strengths and Weaknesses of ADHD-symptoms and Normal Behaviour hyperactivity/impulsivity subscale; TOCS: Toronto Obsessive-Compulsive Scale; CBCL-OCS: Child Behaviour Checklist obsessive-compulsive subscale; FSIQ: full-scale intelligence quotient; CBCL-I: Child Behaviour Checklist internalizing problems subscale; CBCL-E: Child Behaviour Checklist externalizing problems subscale; [1]Kruskal-Wallis *H*-statistic for the non-normally distributed continuous variables with eta-squared effect sizes, chi-squared statistic for race/ethnicity with Cramer's *v* effect size, and ordinal regression with odds ratios for the ordinal variables; [2]Collected according to the respective country's census guidelines; [3]In the respective country's currency; [4]College includes associate and bachelor's degrees

For the fMRI dataset, 1072 HBN (349 females, 723 males) and 720 POND (207 females, 513 males) participants were included in the analysis. Significant between-dataset differences in age (females: $H(1)=21.34$, $p<.001$, $\eta^2=.04$; males: $H(1)=40.82$, $p<.001$, $\eta^2=.03$) and motion (females: $H(1)=36.50$, $p<.001$, $\eta^2=.03$; males: $H(1)=29.50$, $p<.001$, $\eta^2=.01$) were observed. Thus, a smaller, matched subset of children and adolescents with no such differences was identified, with each dataset consisting of 627 (189 female, 438 male) individuals (**Tables 3** and **4**). Co-occurring conditions are summarized in **Supplemental Table 4**. In this matched sample, there were no significant between-diagnosis differences in head motion in the POND ($H(3)=0.51$, $p=.676$, $\eta^2=0.01$) nor HBN ($H(3)=1.07$, $p=.362$, $\eta^2=0.02$) females. In the males, diagnosis was marginally significant in POND ($H(3)=2.68$, $p=.046$, $\eta^2=0.02$) but not significant in HBN ($H(3)=1.79$, $p=.149$, $\eta^2=0.01$).

Table 3: Participant demographics for the fMRI sample.

			HBN	POND	Test-Statistic[1]	*p*-value	Effect size
Females	***N***		189	189	–	–	–
	Dx (*N*)	**ADHD**	130	38	107.87	<.001	.53
		Autism	7	65			
		OCD	12	31			
		TD	40	55			
	Age[2] (years)		10.7 [8.7, 13.4]	11.7 [9.4, 14.0]	2.77	.096	.01
	Motion[2] (mm)		0.1 [0.1, 0.2]	0.1 [0.1, 0.2]	2.09	.148	<.01
Males	***N***		438	438	–	–	–
	Dx (*N*)	**ADHD**	321	121	202.91	<.001	.48
		Autism	53	192			
		OCD	7	48			
		TD	57	77			
	Age[2] (years)		10.9 [9.0, 13.1]	11.4 [9.4, 13.6]	2.75	.097	<.01
	Motion[2] (mm)		0.2 [0.1, 0.2]	0.2 [0.1, 0.3]	0.53	.466	<.01

HBN: Healthy Brain Network; POND: Province of Ontario Neurodevelopmental network; ADHD: attention-deficit/hyperactivity disorder; OCD: obsessive-compulsive disorder; TD: typically developing; [1]Kruskal-Wallis *H*-statistic for age with eta-squared effect size, and chi-squared statistic for diagnosis with Cramer's *v* effect size; [2]Age and motion is reported as median [IQR].

Table 4: Socioeconomic status, race/ethnicity, and phenotypic measures for the fMRI sample.

		HBN	POND	Test-statistic[1]	*p*-value	Effect size
Females	**Race/ethnicity[2]**					
	Hispanic/Latino	35	7			
	Non-Hispanic Asian	7	13			
	Non-Hispanic Black	18	2	37.68	<.001	0.35
	Non-Hispanic White	88	100			
	Other/Multi-Racial	14	27			
	Caregiver income[3]					
	<$50k/year	32	20			
	$50k-$150k	68	62	0.00	.999	1.00
	>$150k	71	52			
	Caregiver education[4]					
	< High school	2	1			
	High school	16	16	4.04	<.001	2.45
	College	69	77			
	Graduate	102	38			
	SCQ	5.0 [3.0, 8.5]	5.0 [2.0, 14.2]	0.01	.917	.03
	RBS-R	10.0 [2.0, 27.2]	11.0 [2.0, 29.0]	0.11	.741	.01
	SWAN-I	2.0 [0.0, 5.8]	2.0 [0.0, 6.0]	0.00	.987	<.01
	SWAN-HI	0.0 [0.0, 2.0]	1.0 [0.0, 3.0]	6.81	.009	.03
	TOCS	–	-5.0 [-39.8, 7.8]	–	–	–
	CBCL-OCS	2.0 [1.0, 5.0]	–	–	–	–
	FSIQ	104.0 [91.0, 116.0]	106.0 [94.0, 115.0]	0.37	.544	<.01
	CBCL-I	56.0 [48.0, 65.0]	62.0 [52.0, 70.0]	10.73	.001	.03
	CBCL-E	54.0 [47.0, 64.2]	54.0 [44.0, 64.0]	0.94	.333	<.01
Males	**Race/ethnicity[2]**					
	Hispanic/Latino	100	28			
	Non-Hispanic Asian	11	29			
	Non-Hispanic Black	48	5	99.32	<.001	0.38
	Non-Hispanic White	189	183			
	Other/Multi-Racial	33	80			
	Caregiver income[3]					
	<$50k/year	66	46			
	$50k-$150k	158	124	1.37	.171	1.23
	>$150k	169	97			
	Caregiver education[4]					
	< High school	5	4			
	High school	77	33	3.24	.001	1.60
	College	144	155			
	Graduate	205	73			
	SCQ	7.0 [4.0, 10.0]	9.0 [3.0, 18.5]	14.08	<.001	.06
	RBS-R	12.0 [2.0, 40.8]	15.0 [5.0, 27.0]	0.51	.474	<.01
	SWAN-I	3.0 [0.0, 6.0]	4.0 [1.0, 7.0]	5.85	.016	.01
	SWAN-HI	0.0 [0.0, 3.0]	2.0 [0.0, 6.0]	38.64	<.001	.05
	TOCS	–	-15.0 [-44.0, 2.0]	–	–	–
	CBCL-OCS	2.0 [1.0, 4.0]	–	–	–	–
	FSIQ	99.0 [86.0, 110.0]	101.0 [89.5, 111.5]	1.32	.250	<.01
	CBCL-I	58.0 [49.0, 67.0]	61.0 [52.0, 68.0]	8.84	.003	.01
	CBCL-E	54.0 [46.0, 63.0]	56.0 [46.0, 66.2]	1.97	.161	<.01

fMRI: functional magnetic resonance imaging; HBN: Healthy Brain Network; POND: Province of Ontario Neurodevelopmental network; SCQ: social communication questionnaire; RBS-R: repetitive Behaviours Scale – Revised; SWAN-I: Strengths and Weaknesses of ADHD-symptoms and Normal Behaviour inattention subscale; SWAN-HI: Strengths and Weaknesses of ADHD-symptoms and Normal Behaviour hyperactivity/impulsivity subscale; TOCS: Toronto Obsessive-Compulsive Scale; CBCL-OCS: Child Behaviour Checklist obsessive-compulsive subscale; FSIQ: full-scale intelligence quotient; CBCL-I: Child Behaviour Checklist internalizing problems subscale; CBCL-E: Child Behaviour Checklist externalizing problems subscale; [1]Kruskal-Wallis *H*-statistic for the non-normally distributed continuous variables with eta-squared effect sizes, chi-squared statistic for race/ethnicity with Cramer's *v* effect size, and ordinal regression with odds ratios for the ordinal variables; [2]Collected according to the respective country's census guidelines; [3]In the respective country's currency; [4]College includes associate and bachelor's degrees

236 ***Dataset differences in deviations***

237 Models of age-related changes in cortical thickness, surface area, and functional
238 connectivity strength were fit separately for each dataset and sex. To evaluate the
239 model performance, the distributions of MSLL across all brain regions are presented in
240 **Supplemental Figure 1**, and distributions of deviation scores are presented in
241 **Supplemental Figure 2**. The deviation scores for the age models were compared
242 between datasets, separately for the males and females. Significant differences were
243 observed for cortical surface area and functional connectivity, but not cortical thickness,
244 with significant differences presented in **Figure 2**. Findings were robust to the size of
245 the training and test sets (**Supplemental Figure 3**). Significant between-dataset
246 differences within each diagnostic label are presented in **Supplemental Figure 4**. To
247 explore what factors may be contributing to the observed differences, associations
248 between deviations and the demographic variables (age, sex, race/ethnicity, caregiver
249 highest level of education, and annual household income) and phenotypic variables
250 (social communication difficulties, repetitive behaviours, inattention,
251 hyperactivity/impulsivity, obsessive-compulsive traits, intelligence, internalizing
252 problems, and externalizing problems) were investigated.

253 For regional measures of cortical surface area (**Figure 2A**), deviation scores
254 differed between the datasets throughout the brain. When examining the median
255 deviation scores for the significant regions, deviations for the POND participants were
256 negative when the models were built using the HBN participants, indicating that the
257 HBN models overestimated surface area for the POND individuals; on the other hand,
258 POND models underestimated surface area in the HBN participants (positive deviation
259 scores). Significant associations between deviations and race/ethnicity were observed
260 in the left orbital part of the middle frontal gyrus, left medial superior frontal gyrus, right
261 supplementary motor area, right angular gyrus, right middle and inferior occipital gyri,
262 right fusiform gyrus, right parahippocampal gyrus, and right median cingulate and
263 paracingulate gyrus (**Supplemental Figure 5A**). Post-hoc tests indicated that while
264 deviations in White participants were closer to zero, deviations in minoritized
265 participants were increased in magnitude. All other associations were insignificant. For

266 regional measures of cortical thickness, no regions showed significant between-dataset
267 differences in the deviations.

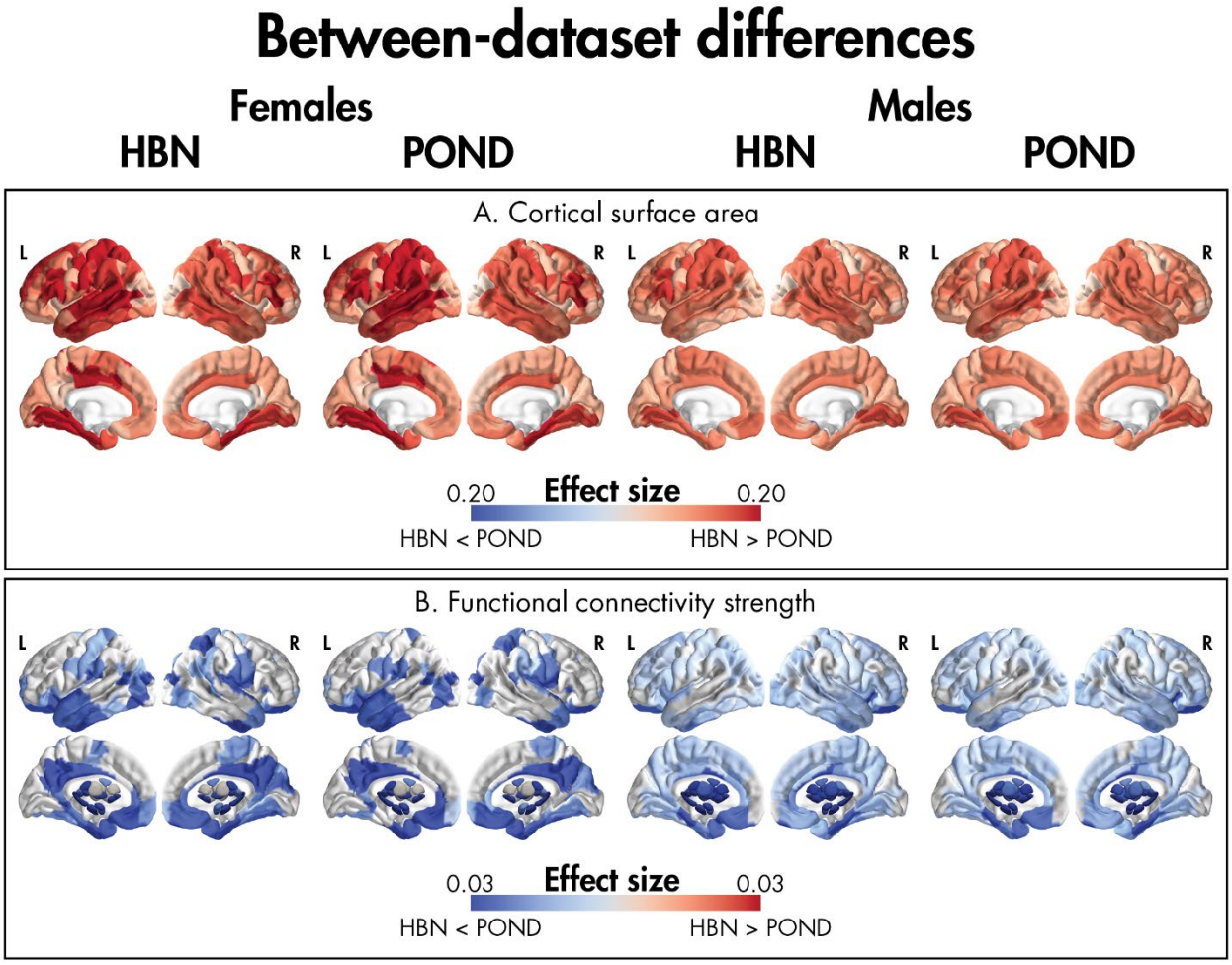


Figure 2: Effect sizes (eta-squared) for significant between-dataset differences in deviations from age-related models of cortical surface area (A) and functional connectivity strength (B) for the females (left) and males (right) when the models were built using each dataset (HBN: Healthy Brain Network, POND: Province of Ontario Neurodevelopmental network). No differences in cortical thickness were observed.

268 For functional connectivity strength (**Figure 2B**), significant differences in the
269 deviation scores between POND and HBN were observed throughout the brain, with
270 more widespread effects observed in the males (HBN: 202 regions, POND: 199
271 regions), albeit with reduced overall effect sizes, compared to females (HBN: 135
272 regions, POND: 122 regions). When models were built using the HBN individuals,
273 median deviation scores indicated that they underestimated functional connectivity
274 strength in POND; conversely, POND models overestimated connectivity strength in the

HBN participants. Deviation scores were significantly positively correlated with head motion (**Supplemental Figure 5B**).

Discussion

In this study, using data from two consortia, namely POND and HBN, we built dataset-specific models that describe age-related changes in measures of brain structure and function. We determined that deviations from the models were significantly different between the two datasets for measures of regional cortical surface area and functional connectivity strength throughout the brain, but not regional cortical thickness; these differences were observed within both males and females. For cortical surface area, the patterns of differences were associated with race/ethnicity, suggesting that it may be contributing to the between-dataset findings. For functional connectivity strength, positive associations were observed with head motion.

Overall, our findings highlight that patterns of age-related changes in the brain may not be consistent between cohorts of neurodiverse individuals, and these inconsistencies may be related to differences in the demographic and phenotypic composition of the sample. While previous work has identified factors such as methodology and medication (Bednarz & Kana, 2018) as contributors to the lack of convergence in the literature on the nature and directionality of brain differences in neurodevelopmental conditions, we hypothesize that differences in age-related changes between cohorts, potentially driven by differences in measures like race/ethnicity and head motion, are also contributing to the lack of consistent findings. Even though age may not be the focus of a particular study, adjusting for age in the same manner in multiple cohorts despite the presence of cohort-specific age-related effects can produce downstream effects in subsequent analyses.

For the regional measures of cortical surface area, differences in deviations from age models differed between POND and HBN in every brain region, irrespective of which group was used to construct the model. The magnitude of this effect differed throughout the brain, for example, larger effects were observed in parietal regions. The development of cortical surface area shows the largest changes across childhood, adolescents, and early adulthood in the parietal lobe (Tamnes et al., 2017). Thus, we

hypothesize that while dataset-specific age-related changes of cortical surface area are pervasive, brain regions which show more drastic developmental differences may show more exaggerated effects.

Deviations from age models of cortical surface area were also larger in magnitude for minoritized compared to White individuals, including frontal, occipital, and cingulate regions. Given that deviations from the models were higher in magnitude in the minoritized compared to white participants, we hypothesize that this may reflect the effect of underrepresentation of minoritized individuals in the two datasets. Interestingly, despite race/ethnicity being significantly associated with measures of socioeconomic status (household income, highest education level), we did not observe an association between socioeconomic status and the deviations alongside race/ethnicity. Potential mechanisms contributing to racial/ethnic differences in brain structure have been suggested to include, for example, childhood adversity (Dumornay et al., 2023), although we did not explicitly test this in our study.

While between-dataset differences in deviations from age models were observed for regional cortical surface area, this finding did not extend to regional cortical thickness. The development of cortical thickness and surface area are driven by different genetic and biological processes and show distinct maturational trajectories (Geschwind & Rakic, 2013; Panizzon et al., 2009; Tamnes et al., 2017; Vijayakumar et al., 2016; Wierenga et al., 2014). Surface area has also been shown to uniquely relate to general psychopathology, while this association is not observed with cortical thickness in several well-powered studies (Jalbrzikowski et al., 2019; Mewton et al., 2022; Romer et al., 2023; Schmaal et al., 2016), although we acknowledge that smaller studies have reported a link between psychopathology and cortical thickness (e.g., (Luking et al., 2022)). Thus, given the differences between POND and HBN in the core and co-occurring domains of neurodevelopmental conditions, we hypothesize that the unique relationship between psychopathology with surface area is why differences in deviations are restricted to this structural measure rather than cortical thickness.

We also observed differences between POND and HBN in deviations from models describing changes in functional connectivity strength with age, with the largest

effects observed in limbic, default mode, and the subcortical regions. Deviations from the models of age-related effects were also shown to be positively associated with head motion, with, again, the largest effect sizes in limbic, default, mode, and subcortical regions. Head motion is known to increase functional connectivity (Satterthwaite et al., 2013), and head motion is higher in HBN compared to POND, possibly due to the use of a fixation-cross resting-state (HBN) compared to movie paradigms (POND) (Vanderwal et al., 2015). Thus, together with the noticeable overlap in the spatial pattern of effects, we suggest that head motion may be driving deviations in age-related changes of functional connectivity. This conclusion is supported by the largest effect sizes being localized to the subcortical structures, which are located distally from other brain regions, and thus may be more impacted by head motion due to its effect on long-range connectivity (van Dijk et al., 2012). Note that while differences in head motion are often reported between children with and without neurodevelopmental conditions (Vandewouw et al., 2023), in our matched samples, this was not the case – only marginal differences were observed in the POND males, with no other groups showing this pattern. Coupled with the lack of associations between diagnosis and model deviations, it does not appear that the correlations between deviations and head motion were being driven by diagnosis.

A strength of this study is the stratification of the analyses by biological sex. Sex and gender differences in the prevalence rates of neurodevelopmental conditions have long contributed to biases in existing research samples and neuroimaging literature (Mo et al., 2021). Specifically, the male preponderance of many of these conditions has resulted in a poor understanding of females with these conditions (Mo et al., 2021). This has resulted in calls for stratifying by sex and gender in neurodevelopmental research (Bölte et al., 2023), which motivated the decision to fit sex-specific age-related models. However, we recognize that the binarized measure of biological sex used in this study is an oversimplification of sex and gender; future work should account for more nuanced identities.

The results of this study should be interpreted considering several limitations. This study is cross-sectional, and future studies should leverage longitudinal data to

validate our developmental findings. The resting-state acquisitions were also different between the two datasets: while POND participants viewed a movie, HBN participants were instructed to fixate on a cross. Thus, the harmonization that was applied across both datasets may be inadvertently removing effects specific to each type of acquisition. We also only considered regional measures of cortical thickness and surface area, which greatly reduced the computational expenses of our analysis; future work could take a less limited spatial resolution and examine between-dataset differences in vertex-wise measures of brain structure. Finally, given the sample size required to establish reproducible brain-behaviour associations (Marek et al., 2022), the contribution of the identified factors to deviations from age models should be validated in larger samples.

The replicability crisis has posed a significant challenge to the neuroscience field and to the broader scientific community. In this context, we have questioned the extent to which age-related changes of brain structure and function are consistent between two independent cohorts of individuals with neurodevelopmental conditions. While deriving the neurobiological changes that occur during development is often a specific research goal, the presence or absence of replicable developmental effects is also highly relevant for all investigations which include age at any step in the methodology. Given our findings that age-related differences are not consistent across samples, this urges future consortium-based studies to carefully examine age-related changes within each dataset prior to applying the same assumptions (e.g., linearity) across datasets. This is particularly pertinent in the context of between-dataset differences in race/ethnicity and head motion, which may be important covariates given their observed associations with deviations from age-related models. While previous work has established cohort-specific effects in neurotypical samples, the heterogeneity associated with neurodivergence makes this finding specifically relevant to neurodivergent populations, where it is less likely that two samples have comparable characteristics. While inconsistencies were observed in both males and females, effect sizes were, on average, larger in the females compared to the males, suggesting that stratification approaches may be more suitable for addressing dataset-specific age-related changes in neurobiology.

Acknowledgements: We would like to thank all participants and their families for their involvement in this study. We would also like to thank all individuals who work for POND for their invaluable assistance in data collection.

This manuscript was prepared using a limited access dataset obtained from the Child Mind Institute Biobank, the Healthy Brain Network (HBN). This manuscript reflects the views of the authors and does not necessarily reflect the opinions or views of the Child Mind Institute. This research was conducted with the support of the Ontario Brain Institute (POND, PIs: Anagnostou/Lerch), 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.

References

Abrol, A., Fu, Z., Du, Y., Wilson, T. W., Wang, Y. P., Stephen, J. M., & Calhoun, V. D. (2023). Developmental and aging resting functional magnetic resonance imaging brain state adaptations in adolescents and adults: A large N (>47K) study. *Human Brain Mapping*, *1*, 0. https://doi.org/10.1002/HBM.26200

Ad-Dab'bagh, Y., Einarson, D., Lyttelton, O., Muehlboeck, J.-S., Mok, K., Ivanov, O., Vincent, R. D., Lepage, C., Lerch, J., Fombonne, E., & Evans, A. C. (2006). The CIVET image-processing environment: A fully automated comprehensive pipeline for anatomical neuroimaging research. *Proceedings of the 12th Annual Meeting of the Organization for Human Brain Mapping*.

Ad-Dab'bagh, Y., Singh, V. K., Robbins, S., Lerch, J. A., Lyttelton, O. C., Fombonne, E. J., & Evans, A. (2005). Native space cortical thickness measurement and the absence of correlation to cerebral volume. *Proceedings of the 11th Annual Meeting of the Organization for Human Brain Mapping*.

Bednarz, H. M., & Kana, R. K. (2018). Advances, challenges, and promises in pediatric neuroimaging of neurodevelopmental disorders. *Neuroscience & Biobehavioral Reviews*, *90*, 50–69. https://doi.org/10.1016/J.NEUBIOREV.2018.03.025

Bethlehem, R. A. I., Seidlitz, J., Romero-Garcia, R., Trakoshis, S., Dumas, G., & Lombardo, M. V. (2020). A normative modelling approach reveals age-atypical cortical thickness in a subgroup of males with autism spectrum disorder. *Communications Biology*, *3*(1). https://doi.org/10.1038/s42003-020-01212-9

Bölte, S., Neufeld, J., Marschik, P. B., Williams, Z. J., Gallagher, L., & Lai, M. C. (2023). Sex and gender in neurodevelopmental conditions. *Nature Reviews Neurology*, *19*(3), 136–159. https://doi.org/10.1038/s41582-023-00774-6

Boucher, M., Whitesides, S., & Evans, A. (2009). Depth potential function for folding pattern representation, registration and analysis. *Medical Image Analysis*, *13*(2), 203–214. https://doi.org/10.1016/j.media.2008.09.001

Bruin, W. B., Taylor, L., Thomas, R. M., Shock, J. P., Zhutovsky, P., Abe, Y., Alonso, P., Ameis, S. H., Anticevic, A., Arnold, P. D., Assogna, F., Benedetti, F., Beucke, J. C., Boedhoe, P. S. W., Bollettini, I., Bose, A., Brem, S., Brennan, B. P., Buitelaar, J. K., … van Wingen, G. A. (2020). Structural neuroimaging biomarkers for obsessive-compulsive disorder in the ENIGMA-OCD consortium: Medication matters. *Translational Psychiatry 2020 10:1*, *10*(1), 1–12. https://doi.org/10.1038/s41398-020-01013-y

Cao, B., Mwangi, B., Passos, I. C., Wu, M. J., Keser, Z., Zunta-Soares, G. B., Xu, Di., Hasan, K. M., & Soares, J. C. (2017). Lifespan gyrification trajectories of human brain in healthy individuals and patients with major psychiatric disorders. *Scientific Reports*, *7*(1), 1–8. https://doi.org/10.1038/s41598-017-00582-1

Cardenas-Iniguez, C., & Gonzalez, M. R. (2024). Recommendations for the responsible use and communication of race and ethnicity in neuroimaging research. *Nature Neuroscience*, *27*(4), 615–628. https://doi.org/10.1038/s41593-024-01608-4

Collins, D. L., Neelin, P., Peters, T. M., & Evans, A. C. (1994). Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *Journal of Computer Assisted Tomography*, *18*(2), 192–205. https://doi.org/10.1097/00004728-199403000-00005

Da-ano, R., Masson, I., Lucia, F., Doré, M., Robin, P., Alfieri, J., Rousseau, C., Mervoyer, A., Reinhold, C., Castelli, J., De Crevoisier, R., Rameé, J. F., Pradier, O., Schick, U., Visvikis, D., & Hatt, M. (2020). Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies. *Scientific Reports*, *10*(1), 1–12. https://doi.org/10.1038/s41598-020-66110-w

Dimitriadis, S. I., Antonakakis, M., Simos, P., Fletcher, J. M., & Papanicolaou, A. C. (2017). Data-driven topological filtering based on orthogonal minimal spanning trees: Application to multigroup magnetoencephalography resting-state connectivity. *Brain Connectivity*, *7*(10), 661–670. https://doi.org/10.1089/brain.2017.0512

Dumornay, N. M., Lebois, L. A. M., Ressler, K. J., & Harnett, N. G. (2023). Racial disparities in adversity during childhood and the false appearance of race-related differences in brain structure. *American Journal of Psychiatry*, *180*(2), 127–138. https://doi.org/10.1176/APPI.AJP.21090961/ASSET/IMAGES/LARGE/APPI.AJP.21090961F3.JPEG

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Esteban, O., Markiewicz, C. J., Goncalves, M., & DuPre, Elizabeth Kent, James D. Salo, Taylor Ciric, Rastko Pinsard, Basile Blair, Ross W. Poldrack, Russell A. Gorgolewski, K. J. (2020). *fMRIPrep: A robust preprocessing pipeline for functional MRI* (20.2.0). Zenodo. https://doi.org/10.5281/zenodo.4252786

Floris, D. L., Wolfers, T., Zabihi, M., Holz, N. E., Zwiers, M. P., Charman, T., Tillmann, J., Ecker, C., Dell'Acqua, F., Banaschewski, T., Moessnang, C., Baron-Cohen, S., Holt, R., Durston, S., Loth, E., Murphy, D. G. M., Marquand, A., Buitelaar, J. K., Beckmann, C. F., … Wooldridge, C. (2021). Atypical brain asymmetry in autism - A candidate for clinically meaningful stratification. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *6*(8), 802–812. https://doi.org/10.1016/J.BPSC.2020.08.008

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R., & Collins, D. L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, *47*, S102. https://doi.org/10.1016/s1053-8119(09)70884-5

Fortin, J. P., Cullen, N., Sheline, Y. I., Taylor, W. D., Aselcioglu, I., Cook, P. A., Adams, P., Cooper, C., Fava, M., McGrath, P. J., McInnis, M., Phillips, M. L., Trivedi, M. H., Weissman, M. M., & Shinohara, R. T. (2018). Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage*, *167*, 104–120. https://doi.org/10.1016/J.NEUROIMAGE.2017.11.024

Foulkes, L., & Blakemore, S. J. (2018). Studying individual differences in human adolescent brain development. *Nature Neuroscience*, *21*(3), 315–323. https://doi.org/10.1038/s41593-018-0078-4

Fraza, C. J., Dinga, R., Beckmann, C. F., & Marquand, A. F. (2021). Warped Bayesian linear regression for normative modelling of big data. *NeuroImage*, *245*, 118715. https://doi.org/10.1016/J.NEUROIMAGE.2021.118715

Geschwind, D. H., & Rakic, P. (2013). Cortical evolution: Judge the brain by its cover. *Neuron*, *80*(3), 633–647. https://doi.org/10.1016/j.neuron.2013.10.045

Grayson, D. S., & Fair, D. A. (2017). Development of large-scale functional networks from birth to adulthood: A guide to the neuroimaging literature. *NeuroImage*, *160*, 15–31. https://doi.org/10.1016/j.neuroimage.2017.01.079

Grotzinger, A. D., Mallard, T. T., Liu, Z., Seidlitz, J., Ge, T., & Smoller, J. W. (2023). Multivariate genomic architecture of cortical thickness and surface area at multiple levels of analysis. *Nature Communications*, *14*(1), 1–13. https://doi.org/10.1038/s41467-023-36605-x

Hammill, C., Lerch, J. P., Taylor, M. J., Ameis, S. H., Chakravarty, M. M., Szatmari, P., Anagnostou, E., & Lai, M. C. (2021). Quantitative and qualitative sex modulations in the brain anatomy of autism. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, *6*(9), 898–909. https://doi.org/10.1016/J.BPSC.2021.03.001

Herting, M. M., Johnson, C., Mills, K. L., Vijayakumar, N., Dennison, M., Liu, C., Goddings, A. L., Dahl, R. E., Sowell, E. R., Whittle, S., Allen, N. B., & Tamnes, C. K. (2018). Development of subcortical volumes across adolescence in males and females: A multisample study of longitudinal changes. *NeuroImage*, *172*, 194. https://doi.org/10.1016/J.NEUROIMAGE.2018.01.020

Huertas, I., Oldehinkel, M., van Oort, E. S. B., Garcia-Solis, D., Mir, P., Beckmann, C. F., & Marquand, A. F. (2017). A Bayesian spatial model for neuroimaging data based on biologically informed basis functions. *NeuroImage*, *161*, 134–148. https://doi.org/10.1016/J.NEUROIMAGE.2017.08.009

Hyatt, C. S., Owens, M. M., Crowe, M. L., Carter, N. T., Lynam, D. R., & Miller, J. D. (2020). The quandary of covarying: A brief review and empirical examination of

covariate use in structural neuroimaging studies on psychological variables. *NeuroImage*, *205*, 116225. https://doi.org/10.1016/J.NEUROIMAGE.2019.116225

Jalbrzikowski, M., Freedman, D., Hegarty, C. E., Mennigen, E., Karlsgodt, K. H., Olde Loohuis, L. M., Ophoff, R. A., Gur, R. E., & Bearden, C. E. (2019). Structural brain alterations in youth with psychosis and bipolar spectrum symptoms. *Journal of the American Academy of Child & Adolescent Psychiatry*, *58*(11), 1079–1091. https://doi.org/10.1016/J.JAAC.2018.11.012

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, *8*(1), 118–127. https://doi.org/10.1093/biostatistics/kxj037

Kabani, N., Le Goualher, G., MacDonald, D., & Evans, A. C. (2001). Measurement of cortical thickness using an automated 3-D algorithm: A validation study. *NeuroImage*, *13*(2), 375–380. https://doi.org/10.1006/NIMG.2000.0652

Kessler, D., Angstadt, M., & Sripada, C. (2016). Growth Charting of Brain Connectivity Networks and the Identification of Attention Impairment in Youth. *JAMA Psychiatry*, *73*(5), 481–489. https://doi.org/10.1001/JAMAPSYCHIATRY.2016.0088

Kim, S. J., Singh, V., Lee, J. K., Lerch, J., Ad-Dab'bagh, Y., MacDonald, D., Jong, M. L., Kim, S. I., & Evans, A. C. (2005). Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification. *NeuroImage*, *27*, 210–221. https://doi.org/10.1016/j.neuroimage.2005.03.036

Kushki, A., Anagnostou, E., Hammill, C., Duez, P., Brian, J., Iaboni, A., Schachar, R., Crosbie, J., Arnold, P., & Lerch, J. P. (2019). Examining overlap and homogeneity in ASD, ADHD, and OCD: A data-driven, diagnosis-agnostic approach. *Translational Psychiatry*, *9*(1). https://doi.org/10.1038/s41398-019-0631-2

Kushki, A., Cardy, R. E., Panahandeh, S., Malihi, M., Hammill, C., Brian, J., Iaboni, A., Taylor, M. J., Schachar, R., Crosbie, J., Arnold, P., Kelley, E., Ayub, M., Nicolson, R., Georgiades, S., Lerch, J. P., & Anagnostou, E. (2021). Cross-diagnosis structural correlates of autistic-like social communication differences. *Cerebral Cortex*, *31*(11), 5067–5076. https://doi.org/10.1093/cercor/bhab142

Lerch, J. P., & Evans, A. C. (2005). Cortical thickness analysis examined through power analysis and a population simulation. *NeuroImage*, *24*, 163–173. https://doi.org/10.1016/j.neuroimage.2004.07.045

Lombardo, M. V., Lai, M. C., & Baron-Cohen, S. (2019). Big data approaches to decomposing heterogeneity across the autism spectrum. *Molecular Psychiatry*, *24*(10), 1435–1450. https://doi.org/10.1038/s41380-018-0321-0

Looden, T., Floris, D. L., Llera, A., Chauvin, R. J., Charman, T., Banaschewski, T., Murphy, D. G. M., Marquand, A. F., Buitelaar, J. K., Beckmann, C. F., Ahmad, J.,

Ambrosino, S., Auyeung, B., Baron-Cohen, S., Baumeister, S., Bölte, S., Bourgeron, T., Bours, C., Brammer, M., … Zwiers, M. P. (2022). Patterns of connectome variability in autism across five functional activation tasks: Findings from the LEAP project. *Molecular Autism*, *13*(1), 1–11. https://doi.org/10.1186/S13229-022-00529-Y/FIGURES/5

Luking, K. R., Jirsaraie, R. J., Tillman, R., Luby, J. L., Barch, D. M., & Sotiras, A. (2022). Timing and type of early psychopathology symptoms predict longitudinal change in cortical thickness from middle childhood Into early adolescence. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *7*(4), 397–405. https://doi.org/10.1016/J.BPSC.2021.06.013

Luppi, A. I., & Stamatakis, E. A. (2021). Combining network topology and information theory to construct representative brain networks. *Network Neuroscience*, *5*(1), 96–124. https://doi.org/10.1162/netn_a_00170

Lyttelton, O. C., Karama, S., Ad-Dab'bagh, Y., Zatorre, R. J., Carbonell, F., Worsley, K., & Evans, A. C. (2009). Positional and surface area asymmetry of the human cerebral cortex. *NeuroImage*, *46*(4), 895–903. https://doi.org/10.1016/J.NEUROIMAGE.2009.03.063

MacDonald, D., Kabani, N., Avis, D., & Evans, A. C. (2000). Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. *NeuroImage*, *12*, 340–356. https://doi.org/10.1006/nimg.1999.0534

Madan, C. R., & Kensinger, E. A. (2017). Age-related differences in the structural complexity of subcortical and ventricular structures. *Neurobiology of Aging*, *50*, 87–95. https://doi.org/10.1016/J.NEUROBIOLAGING.2016.10.023

Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., Donohue, M. R., Foran, W., Miller, R. L., Hendrickson, T. J., Malone, S. M., Kandala, S., Feczko, E., Miranda-Dominguez, O., Graham, A. M., Earl, E. A., Perrone, A. J., Cordova, M., Doyle, O., … Dosenbach, N. U. F. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature*, *603*(7902), 654. https://doi.org/10.1038/S41586-022-04492-9

Marquand, A. F., Rezek, I., Buitelaar, J., & Beckmann, C. F. (2016). Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies. *Biological Psychiatry*, *80*(7), 552. https://doi.org/10.1016/J.BIOPSYCH.2015.12.023

Mewton, L., Lees, B., Squeglia, L. M., Forbes, M. K., Sunderland, M., Krueger, R., Koch, F. C., Baillie, A., Slade, T., Hoy, N., & Teesson, M. (2022). The relationship between brain structure and general psychopathology in preadolescents. *Journal of Child Psychology and Psychiatry*, *63*(7), 734–744. https://doi.org/10.1111/JCPP.13513

Mills, K. L., Goddings, A. L., Herting, M. M., Meuwese, R., Blakemore, S. J., Crone, E. A., Dahl, R. E., Güroğlu, B., Raznahan, A., Sowell, E. R., & Tamnes, C. K. (2016). Structural brain development between childhood and adulthood: Convergence across four longitudinal samples. *NeuroImage*, *141*, 273–281. https://doi.org/10.1016/J.NEUROIMAGE.2016.07.044

Mills, K. L., & Tamnes, C. K. (2020). Longitudinal structural and functional brain development in childhood and adolescence. In *The Oxford Handbook of Developmental Cognitive Neuroscience*. Oxford University Press. https://academic.oup.com/edited-volume/34742/chapter/296561676

Mo, K., Sadoway, T., Bonato, S., Ameis, S. H., Anagnostou, E., Lerch, J. P., Taylor, M. J., & Lai, M. C. (2021). Sex/gender differences in the human autistic brains: A systematic review of 20 years of neuroimaging research. *NeuroImage: Clinical*, *32*, 102811. https://doi.org/10.1016/J.NICL.2021.102811

Morandini, H. A. E., Rao, P., Hood, S. D., Zepf, F. D., Silk, T. J., & Griffiths, K. R. (2021). Age-related resting-state functional connectivity of the Vigilant Attention network in children and adolescents. *Brain and Cognition*, *154*, 105791. https://doi.org/10.1016/J.BANDC.2021.105791

Nicolaisen-Sobesky, E., Mihalik, A., Kharabian-Masouleh, S., Ferreira, F. S., Hoffstaedter, F., Schwender, H., Maleki Balajoo, S., Valk, S. L., Eickhoff, S. B., Yeo, B. T. T., Mourao-Miranda, J., & Genon, S. (2022). A cross-cohort replicable and heritable latent dimension linking behaviour to multi-featured brain structure. *Communications Biology*, *5*(1), 1–11. https://doi.org/10.1038/s42003-022-04244-5

Norbom, L. B., Ferschmann, L., Parker, N., Agartz, I., Andreassen, O. A., Paus, T., Westlye, L. T., & Tamnes, C. K. (2021). New insights into the dynamic development of the cerebral cortex in childhood and adolescence: Integrating macro- and microstructural MRI findings. *Progress in Neurobiology*, *204*, 102109. https://doi.org/10.1016/J.PNEUROBIO.2021.102109

Oliveira-Saraiva, D., & Ferreira, H. A. (2023). Normative model detects abnormal functional connectivity in psychiatric disorders. *Frontiers in Psychiatry*, *14*, 1068397. https://doi.org/10.3389/FPSYT.2023.1068397/BIBTEX

Panizzon, M. S., Fennema-Notestine, C., Eyler, L. T., Jernigan, T. L., Prom-Wormley, E., Neale, M., Jacobson, K., Lyons, M. J., Grant, M. D., Franz, C. E., Xian, H., Tsuang, M., Fischl, B., Seidman, L., Dale, A., & Kremen, W. S. (2009). Distinct genetic influences on cortical surface area and cortical thickness. *Cerebral Cortex*, *19*(11), 2728–2735. https://doi.org/10.1093/CERCOR/BHP026

Parkes, L., Moore, T. M., Calkins, M. E., Cook, P. A., Cieslak, M., Roalf, D. R., Wolf, D. H., Gur, R. C., Gur, R. E., Satterthwaite, T. D., & Bassett, D. S. (2021). Transdiagnostic dimensions of psychopathology explain individuals' unique

deviations from normative neurodevelopment in brain structure. *Translational Psychiatry*, *11*(1), 1–13. https://doi.org/10.1038/s41398-021-01342-6

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., Nichols, T. E., Poline, J. B., Vul, E., & Yarkoni, T. (2017). Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, *18*(2), 115–126. https://doi.org/10.1038/nrn.2016.167

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*.

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. The MIT Press. https://direct-mit-edu.myaccess.library.utoronto.ca/books/book/2320/Gaussian-Processes-for-Machine-Learning

Romer, A. L., Knodt, A. R., Sison, M. L., Ireland, D., Houts, R., Ramrakha, S., Poulton, R., Keenan, R., Melzer, T. R., Moffitt, T. E., Caspi, A., & Hariri, A. R. (2019). Replicability of structural brain alterations associated with general psychopathology: Evidence from a population-representative birth cohort. *Molecular Psychiatry*, *26*(8), 3839–3846. https://doi.org/10.1038/s41380-019-0621-z

Romer, A. L., Ren, B., & Pizzagalli, D. A. (2023). Brain Structure Relations With Psychopathology Trajectories in the Adolescent Brain Cognitive Development Study. *Journal of the American Academy of Child & Adolescent Psychiatry*. https://doi.org/10.1016/J.JAAC.2023.02.002

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician*, *39*(1), 33–38. https://doi.org/10.1080/00031305.1985.10479383

Rutherford, S., Mostafa Kia, S., Wolfers, T., Fraza, C., Zabihi, M., Dinga, R., Berthet, P., Worker, A., Verdi, S., Ruhe, H. G., Beckmann, C. F., & Marquand, A. F. (2022). The normative modeling framework for computational psychiatry. *Nature Methods*, *17*, 1711–1734. https://doi.org/10.1038/s41596-022-00696-5

Sadat-Nejad, Y., Vandewouw, M. M., Cardy, R., Lerch, J., Taylor, M. J., Iaboni, A., Hammill, C., Syed, B., Brian, J. A., Kelley, E., Ayub, M., Crosbie, J., Schachar, R., Georgiades, S., Nicolson, R., Anagnostou, E., & Kushki, A. (2023). Investigating heterogeneity across autism, ADHD, and typical development using measures of cortical thickness, surface area, cortical/subcortical volume, and structural covariance. *Frontiers in Child and Adolescent Psychiatry*, *2*, 1171337. https://doi.org/10.3389/FRCHA.2023.1171337

Sanjeevan, T., Hammill, C., Brian, J., Crosbie, J., Schachar, R., Kelley, E., Liu, X., Nicolson, R., Iaboni, A., Day Fragiadakis, S., Ristic, L., Lerch, J. P., & Anagnostou, E. (2020). Exploring the Neural Structures Underlying the Procedural Memory Network as Predictors of Language Ability in Children and Adolescents With Autism Spectrum Disorder and Attention Deficit Hyperactivity Disorder. *Frontiers in Human Neuroscience*, *14*, 587019. https://doi.org/10.3389/FNHUM.2020.587019/BIBTEX

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. E. R. C., Gur, R. E. R. C., & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*.

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X. N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral Cortex*, *28*(9), 3095–3114. https://doi.org/10.1093/cercor/bhx17

Schmaal, L., Hibar, D. P., Sämann, P. G., Hall, G. B., Baune, B. T., Jahanshad, N., Cheung, J. W., Van Erp, T. G. M., Bos, D., Ikram, M. A., Vernooij, M. W., Niessen, W. J., Tiemeier, H., Hofman, A., Wittfeld, K., Grabe, H. J., Janowitz, D., Bülow, R., Selonke, M., … Veltman, D. J. (2016). 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. *Molecular Psychiatry*, *22*(6), 900–909. https://doi.org/10.1038/mp.2016.60

Sherif, T., Rioux, P., Rousseau, M.-E., Kassis, N., Beck, N., Adalat, R., Das, S., Glatard, T., & Evans, A. C. (2014). CBRAIN: A web-based, distributed computing platform for collaborative neuroimaging research. *Frontiers in Neuroinformatics*, *8*, 54. https://doi.org/10.3389/fninf.2014.00054

Sled, J. G., Zijdenbos, A. P., & Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging*, *17*(1), 87–97. https://doi.org/10.1109/42.668698

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 143–155. https://doi.org/10.1002/hbm.10062

Sussman, D., Leung, R. C., Vogan, V. M., Lee, W., Trelle, S., Lin, S., Cassel, D. B., Chakravarty, M. M., Lerch, J. P., Anagnostou, E., & Taylor, M. J. (2015). The autism puzzle: Diffuse but not pervasive neuroanatomical abnormalities in children with ASD. *NeuroImage: Clinical*, *8*, 170–179. https://doi.org/10.1016/J.NICL.2015.04.008

Tamnes, C. K., Herting, M. M., Goddings, A. L., Meuwese, R., Blakemore, S. J., Dahl, R. E., Güroğlu, B., Raznahan, A., Sowell, E. R., Crone, E. A., & Mills, K. L. (2017). Development of the cerebral cortex across adolescence: A multisample study of

inter-related longitudinal changes in cortical volume, surface area, and thickness. *Journal of Neuroscience*, *37*(12), 3402–3412. https://doi.org/10.1523/JNEUROSCI.3302-16.2017

Tian, Y., Margulies, D. S., Breakspear, M., & Zalesky, A. (2020). Topographic organization of the human subcortex unveiled with functional connectivity gradients. *Nature Neuroscience*, *23*(11), 1421–1432. https://doi.org/10.1038/s41593-020-00711-6

Tohka, J., Zijdenbos, A., & Evans, A. (2004). Fast and robust parameter estimation for statistical partial volume models in brain MRI. *NeuroImage*, *23*(1), 84–97. https://doi.org/10.1016/j.neuroimage.2004.05.007

Tunç, B., Yankowitz, L. D., Parker, D., Alappatt, J. A., Pandey, J., Schultz, R. T., & Verma, R. (2019). Deviation from normative brain development is associated with symptom severity in autism spectrum disorder. *Molecular Autism*, *10*(1), 1–14. https://doi.org/10.1186/S13229-019-0301-5

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*, 273–289. https://doi.org/10.1006/nimg.2001.0978

van Dijk, K. R. A., Sabuncu, M. R., & Buckner, R. L. (2012). The influence of head motion on intrinsic functional connectivity MRI. *NeuroImage*, *59*(1), 431–438. https://doi.org/10.1016/j.neuroimage.2011.07.044

Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L. C., & Castellanos, F. X. (2015). Inscapes: A movie paradigm to improve compliance in functional magnetic resonance imaging. *NeuroImage*, *122*, 222–232. https://doi.org/10.1016/j.neuroimage.2015.07.069

Vandewouw, M. M., Brian, J., Crosbie, J., Schachar, R. J., Iaboni, A., Georgiades, S., Nicolson, R., Kelley, E., Ayub, M., Jones, J., Taylor, M. J., Lerch, J. P., Anagnostou, E., & Kushki, A. (2023). Identifying replicable subgroups in neurodevelopmental conditions using resting-state functional magnetic resonance imaging data. *JAMA Network Open*, *6*(3), e232066–e232066. https://doi.org/10.1001/JAMANETWORKOPEN.2023.2066

Vijayakumar, N., Allen, N. B., Youssef, G., Dennison, M., Yücel, M., Simmons, J. G., & Whittle, S. (2016). Brain development during adolescence: A mixed-longitudinal investigation of cortical thickness, surface area, and volume. *Human Brain Mapping*, *37*(6), 2027–2038. https://doi.org/10.1002/HBM.23154

Wang, Q., Hu, K., Wang, M., Zhao, Y., Liu, Y., Fan, L., & Liu, B. (2021). Predicting brain age during typical and atypical development based on structural and functional

neuroimaging. *Human Brain Mapping*, *42*(18), 5943–5955. https://doi.org/10.1002/HBM.25660

Wierenga, L. M., Langen, M., Oranje, B., & Durston, S. (2014). Unique developmental trajectories of cortical thickness and surface area. *NeuroImage*, *87*, 120–126. https://doi.org/10.1016/J.NEUROIMAGE.2013.11.010

Wolfers, T., Beckmann, C. F., Hoogman, M., Buitelaar, J. K., Franke, B., & Marquand, A. F. (2019). Individual differences v. the average patient: Mapping the heterogeneity in ADHD using normative models. *Psychological Medicine*, *50*(2), 314–323. https://doi.org/10.1017/S0033291719000084

Zabihi, M., Oldehinkel, M., Wolfers, T., Frouin, V., Goyard, D., Loth, E., Charman, T., Tillmann, J., Banaschewski, T., Dumas, G., Holt, R., Baron-Cohen, S., Durston, S., Bölte, S., Murphy, D., Ecker, C., Buitelaar, J. K., Beckmann, C. F., & Marquand, A. F. (2019). Dissecting the Heterogeneous Cortical Anatomy of Autism Spectrum Disorder Using Normative Models. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *4*(6), 567–578. https://doi.org/10.1016/J.BPSC.2018.11.013

Zijdenbos, A., Forghani, R., & Evans, A. (1998). Automatic quantification of MS lesions in 3D MRI brain data sets: Validation of INSECT. *Medical Image Computing and Computer-Assisted Intervention*, 439–448.

Supplemental information

Supplemental Table 1: Acquisition parameters for the sMRI and rs-fMRI protocols for each site in the POND and HBN datasets.

Dataset	Scanner	Scan type	Sequence	TR (ms)	TE (ms)	FA (°)	FOV (mm)	Voxel size (mm)	Scan time (min)
sMRI protocol									
	SK: 3T TimTrio	–	MPRAGE	2300	2.96	9	192×240×256	1	5.0
	QU: 3T TimTrio	–	MPRAGE	2300	3.14	9	192×240×256	0.8	6.2
	SK: 3T Prisma^{fit}	–	MPRAGE	1870	3.14	9	192×240×256	0.8	5.0
	QU: 3T Prisma^{fit}	–	MPRAGE	1870	3.10	9	192×240×256	0.8	5.0
POND	HB: 3T Prisma^{fit}	–	MPRAGE	1870	3.10	9	192×240×256	0.8	5.0
	CBIC: 3T Prisma^{fit}	–	MPRAGE	2500	3.15	8	179×256×256	0.8	7.00
	CUNY: 3T Prisma^{fit}	–	MPRAGE	2500	3.15	8	179×256×256	0.8	7.00
	RU: 3T TimTrio	–	MPRAGE	2500	3.15	8	179×256×256	0.8	7.00
HBN	SI: 1.5T Avanto	–	MPRAGE	2730	1.64	7	176×256×256	1	6.53
fMRI protocol									
	SK: 3T TimTrio	Movies	EPI	2340	30	70	224×224×140	3.5	5.1
	QU: 3T TimTrio	Inscapes	EPI	2340	30	70	224×224×140	3.5	5.1
	SK: 3T Prisma^{fit}	Inscapes	EPI	1500	30	70	222×222×150	3.0	5.1
	QU: 3T Prisma^{fit}	Inscapes	EPI	1500	30	70	222×222×150	3.0	5.1
POND	HB: 3T Prisma^{fit}	Inscapes	EPI	1500	30	70	222×222×150	3.0	5.1
	CBIC: 3T Prisma^{fit}	Fixation	EPI	800	30	31	202×202×144	2.4	2×5.1
	CUNY: 3T Prisma^{fit}	Fixation	EPI	800	30	31	202×202×144	2.4	2×5.1
	RU: 3T TimTrio	Fixation	EPI	800	30	31	202×202×144	2.4	2×5.1
	SI: 1.5T Avanto	Fixation	EPI	1450	40	55	195×195×135	2.5	10.3
	SK: 3T TimTrio	Movies	EPI	2340	30	70	224×224×140	3.5	5.0
HBN	QU: 3T TimTrio	Inscapes	EPI	2340	30	70	224×224×140	3.5	5.0

sMRI: structural magnetic resonance imaging; rs-fMRI: resting-state functional magnetic resonance imaging; POND: Province of Ontario Neurodevelopmental network; HBN: Healthy Brain Network; SK: Hospital for Sick Children; QU: Queen's University; HB: Holland Bloorview Kids Rehabilitation Hospital; CBIC: CitiGroup Cornell Brain Imaging Center; CUNY: the City University of New York Advanced Science Research Center; RU: Rutgers University; SI: Staten Island; TR: repetition time; TE: echo time; FA: flip angle; FOV: field of view

Supplemental Table 2: Summary of the phenotypic measures of core and co-occurring differences for the diagnostic categories.

Domain	Assessment	Score	Range	Interpretation[1]	Acronym
Social communication difficulties	Social Communication Questionnaire (Rutter et al., 2003)	Total score	0-39	More severe difficulties with social communication	SCQ
Repetitive behaviours	Repetitive Behaviours Scale – Revised (Bodfish et al., 2000)	Total score	0-129	More severe repetitive behaviours	RBS-R
Inattention	Strengths and Weaknesses of ADHD-symptoms and Normal Behaviour (Swanson et al., 2012)	Inattention subscale	0-9	More severe difficulties with inattention	SWAN-I
Hyperactivity	Strengths and Weaknesses of ADHD-symptoms and Normal Behaviour (Swanson et al., 2012)	Hyperactivity subscale	0-9	More severe difficulties with hyperactivity/impulsivity	SWAN-HI
Obsessive-compulsive traits	Toronto Obsessive-Compulsive Scale (Nelson et al., 2001)	Total score	-63-63	More severe obsessive-compulsive traits	TOCS
	Child Behaviour Checklist (Nelson et al., 2001)	Obsessive-compulsive subscale	0-16	More severe obsessive-compulsive traits	CBCL-OCS
Intelligence	Age-appropriate Wechsler or Stanford-Binet scales	Full-scale intelligence quotient	40-160	Higher intelligence	FSIQ
Internalizing problems	Child Behaviour Checklist (Achenbach & Ruffle, 2000)	Internalizing problems subscale	40-75	Worse internalizing problems	CBCL-I
Externalizing problems	Child Behaviour Checklist (Achenbach & Ruffle, 2000)	Externalizing problems subscale	40-75	Worse externalizing problems	CBCL-E

[1]Interpretation: relative to higher scores

Supplemental Table 3: Co-occurring diagnosis in the sMRI sample.

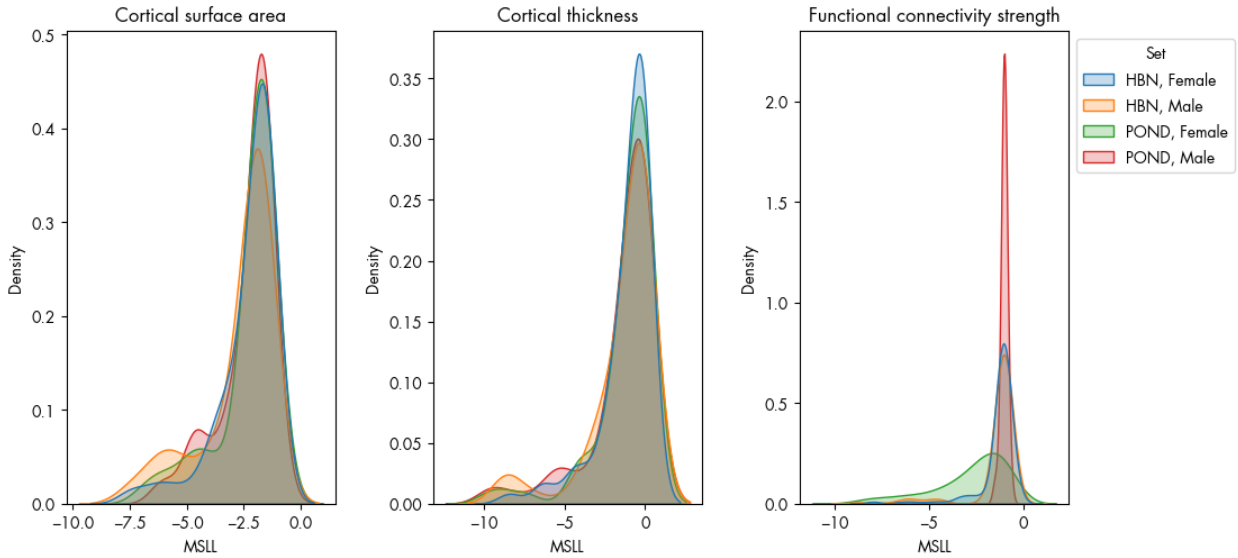
	Primary diagnosis	Data available	Co-occurring diagnosis Autism	ADHD	OCD
POND females	Autism	26	–	4	0
	ADHD	50	0	–	1
	OCD	23	0	3	–
POND males	Autism	104	–	34	3
	ADHD	143	2	–	3
	OCD	33	0	9	–
HBN females	Autism	7	–	2	2
	ADHD	124	9	–	7
	OCD	9	0	2	–
HBN males	Autism	55	–	39	5
	ADHD	304	49	–	10
	OCD	11	2	5	–

HBN: Healthy Brain Network; POND: Province of Ontario Neurodevelopmental network; ADHD: attention-deficit/hyperactivity disorder; OCD: obsessive-compulsive disorder

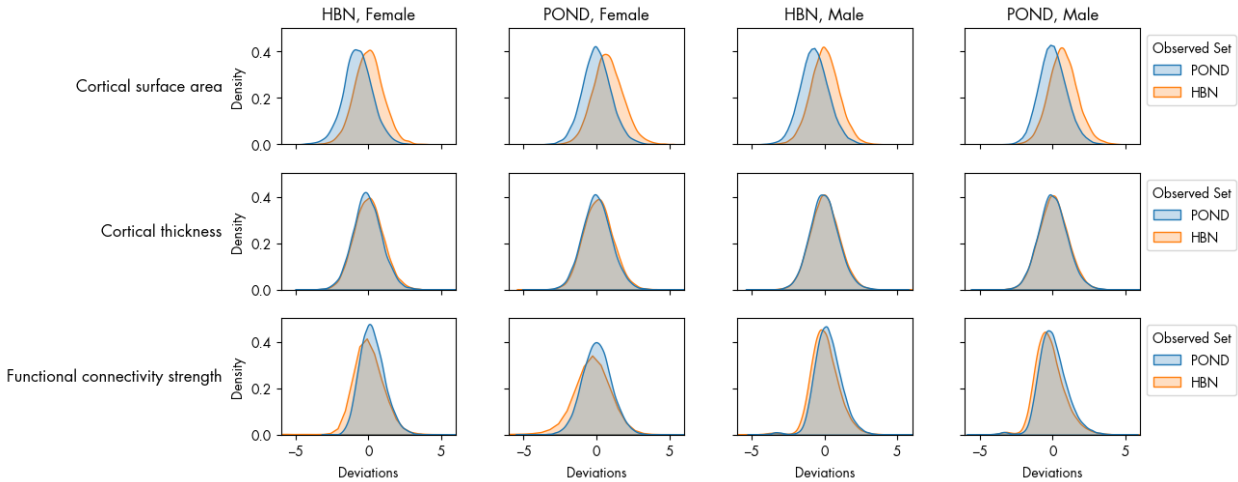
Supplemental Table 4: Co-occurring diagnosis in the fMRI sample.

	Primary diagnosis	Data available	Co-occurring diagnosis Autism	ADHD	OCD
POND females	Autism	36	–	0	0
	ADHD	37	0	–	1
	OCD	30	0	8	–
POND males	Autism	114	–	49	4
	ADHD	119	2	–	2
	OCD	45	0	17	–
HBN females	Autism	7	–	4	2
	ADHD	130	8	–	6
	OCD	12	0	1	–
HBN males	Autism	53	–	35	5
	ADHD	321	49	–	10
	OCD	7	2	4	–

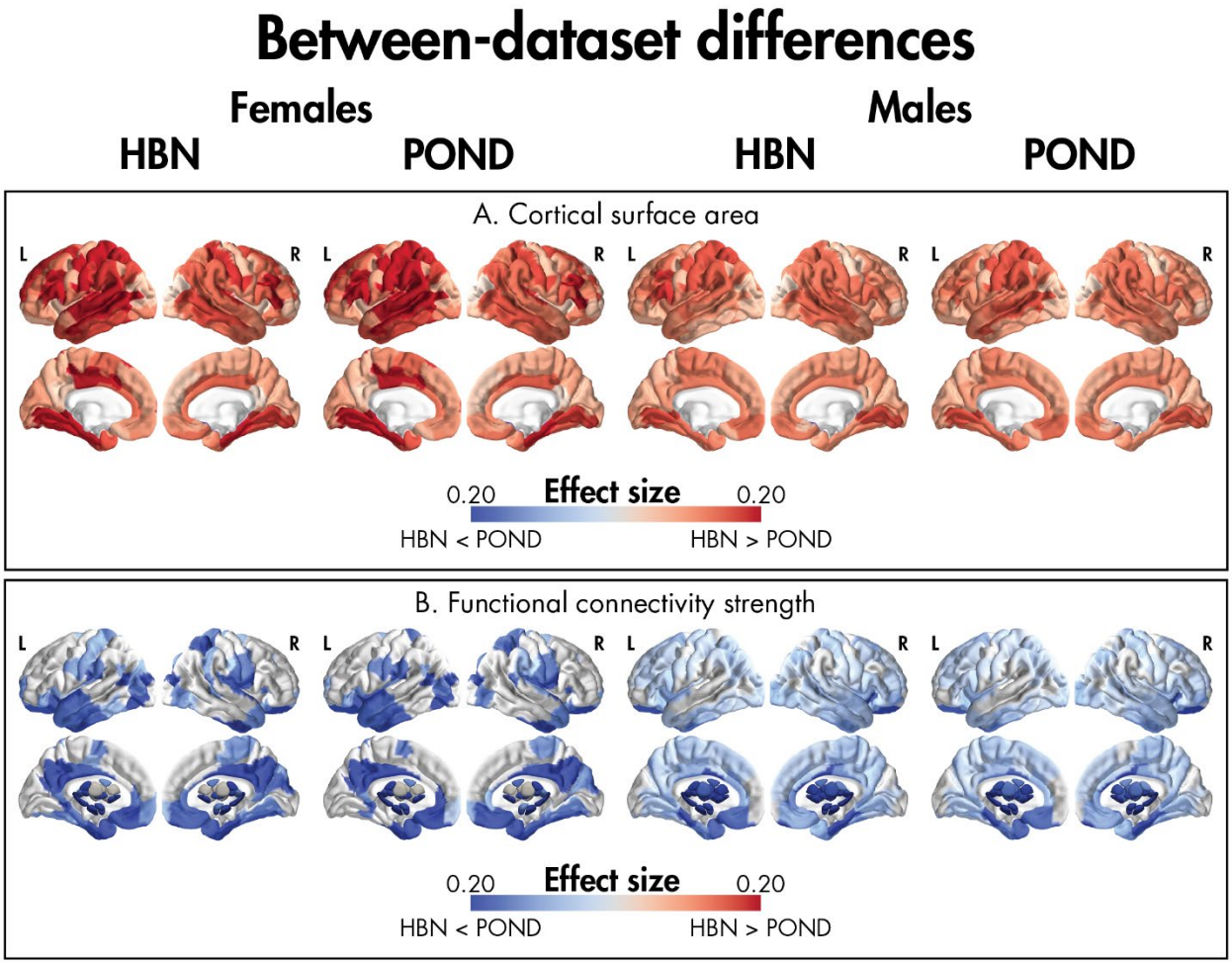
HBN: Healthy Brain Network; POND: Province of Ontario Neurodevelopmental network; ADHD: attention-deficit/hyperactivity disorder; OCD: obsessive-compulsive disorder



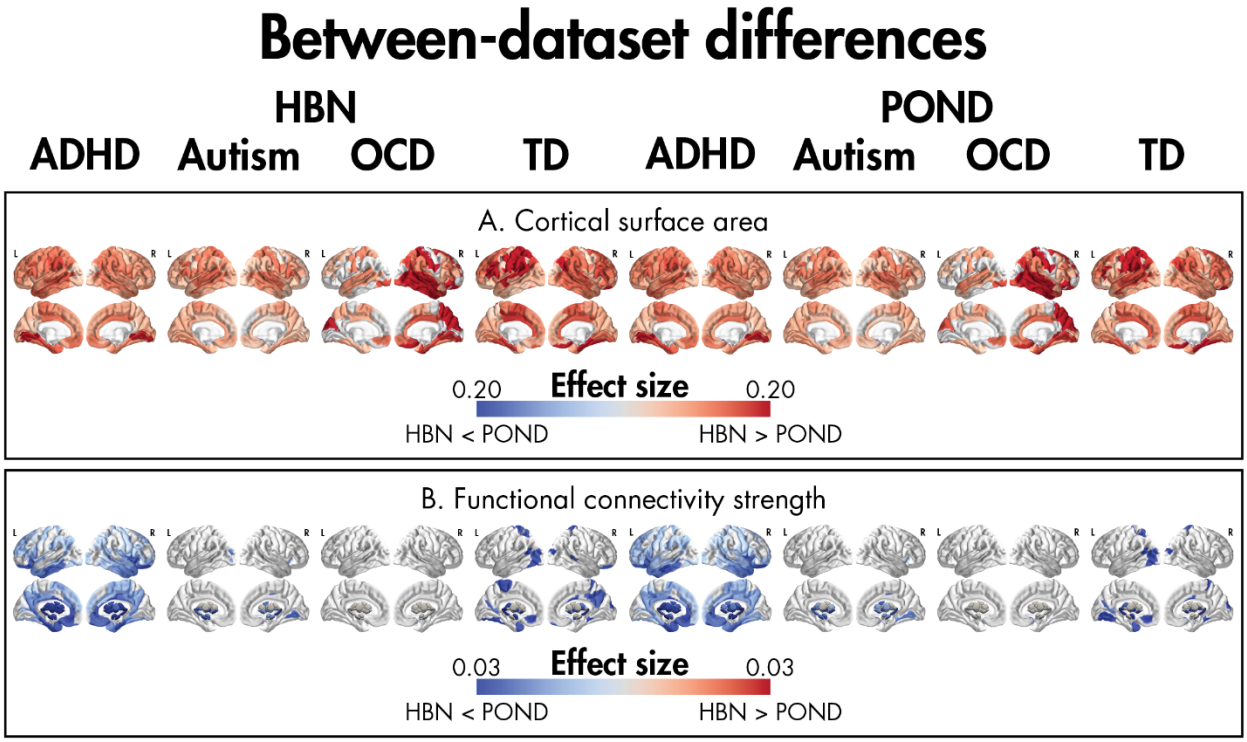
Supplemental Figure 1: The distribution of mean squared log loss (MSLL) across all brain regions for the BLR models for cortical surface area, cortical thickness, and functional connectivity strength. MSLL was computed using data from the inner test set (HBN females, HBN males, POND females, and POND males).



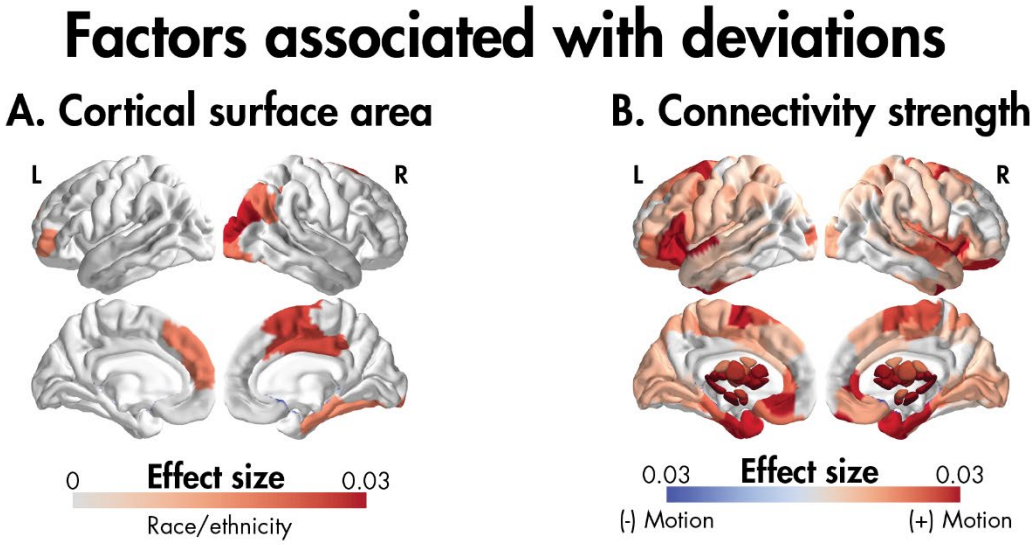
Supplemental Figure 2: The distribution of deviation scores across all brain regions for the BLR models for cortical surface area, cortical thickness, and functional connectivity strength. Deviations were computed for the inner test set (HBN females, HBN males, POND females, and POND males) and its corresponding external test set.



Supplemental Figure 3: Effect sizes (eta-squared) for significant between-dataset differences in deviations from age-related models of cortical surface area (A) and functional connectivity strength (B) using a 70%/30% train/test split for the females (left) and males (right) when the models were built using each dataset (HBN: Healthy Brain Network, POND: Province of Ontario Neurodevelopmental network). No differences in cortical thickness were observed.



Supplemental Figure 4: Effect sizes (eta-squared) for significant between-dataset differences in deviations from age-related models of cortical surface area (A) and functional connectivity strength (B) within each diagnosis and across sex when the models were built using each dataset (HBN: Healthy Brain Network, POND: Province of Ontario Neurodevelopmental network).



Supplemental Figure 5: Significant association between (A) the deviations from age models of cortical surface area and race/ethnicity and (B) the deviations from connectivity strength models and head motion.

References

Achenbach, T. M., & Ruffle, T. M. (2000). The Child Behavior Checklist and Related Forms for Assessing Behavioral/Emotional Problems and Competencies. *Pediatrics In Review*, *21*(8), 265–271. https://doi.org/10.1542/PIR.21-8-265

Bodfish, J. W., Symons, F. J., Parker, D. E., & Lewis, M. H. (2000). Varieties of repetitive behavior in autism: Comparisons to mental retardation. *Journal of Autism and Developmental Disorders*, *30*(3), 237–243. https://doi.org/10.1023/A:1005596502855

Nelson, E. C., Hanna, G. L., Hudziak, J. J., Botteron, K. N., Heath, A. C., & Todd, R. D. (2001). Obsessive-compulsive scale of the child behavior checklist: specificity, sensitivity, and predictive power. *Pediatrics*, *108*(1). https://doi.org/10.1542/PEDS.108.1.E14

Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire*. Western Psychological Services.

Swanson, J. M., Schuck, S., Porter, M. M., Carlson, C., Hartman, C. A., Sergeant, J. A., Clevenger, W., Wasdell, M., McCleary, R., Lakes, K., & Wigal, T. (2012). Categorical and Dimensional Definitions and Evaluations of Symptoms of ADHD: History of the SNAP and the SWAN Rating Scales. *The International Journal of Educational and Psychological Assessment*, *10*(1), 51–70. http://www.ncbi.nlm.nih.gov/pubmed/26504617%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4618695