Radiology

ORIGINAL RESEARCH • PEDIATRIC IMAGING

Brain Growth Charts for Quantitative Analysis of Pediatric Clinical Brain MRI Scans with Limited Imaging Pathology

Jenna M. Schabdach, PhD • J. Eric Schmitt, PhD, MD • Susan Sotardi, MD, MS • Arastoo Vossough, PhD, MD • Savvas Andronikou, MD, PhD • Timothy P. Roberts, PhD • Hao Huang, PhD • Viveknarayanan Padmanabhan, MS • Alfredo Ortiz-Rosa, BS • Margaret Gardner, BS • Sydney Covitz, BA • Saashi A. Bedford, MSc • Ayan S. Mandal, PhD • Barbara H. Chaiyachati, MD, PhD • Simon R. White, PhD • Edward Bullmore, MB, PhD • Richard A. I. Bethlehem, PhD • Russell T. Shinohara, PhD • Benjamin Billot, PhD • J. Eugenio Iglesias, PhD • Satrajit Ghosh, PhD • Raquel E. Gur, MD, PhD • Theodore D. Satterthwaite, MD • David Roalf, PhD • Jakob Seidlitz, PhD • Aaron Alexander-Bloch, MD, PhD • for the Lifespan Brain Chart Consortium

From the Lifespan Brain Institute (LiBI) of the Children's Hospital of Philadelphia (CHOP) and Penn Medicine, Philadelphia, Pa (J.M.S., A.O.R., M.G., A.S.M., B.H.C., R.E.G., T.D.S., J.S., A.A.B.); Department of Child and Adolescent Psychiatry and Behavioral Science (J.M.S., J.S., A.A.B.), Department of Radiology (S.S., A.V., S.A., T.P.R., H.H.), PolicyLab and Clinical Futures, CHOP Research Institute (B.H.C.), and Department of Biomedical and Health Informatics (J.E.S., S.S., V.P.), Children's Hospital of Philadelphia, Philadelphia, Pa; Department of Psychiatry (J.E.S., R.E.G., T.D.S., D.R., J.S., A.A.B.), Department of Radiology (J.E.S., S.S., A.V., S.A., T.P.R., H.H.), Lifespan Informatics and Neuroimaging Center (PennLINC), Department of Psychiatry (S.C., T.D.S.), and Department of Pediatrics (B.H.C.), Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pa; Departments of Psychiatry (S.A.B., S.R.W., E.B., R.A.I.B.) and Psychology (R.A.I.B.), University of Cambridge, Cambridge, United Kingdom; Center for Biomedical Image Computation and Analytics (R.T.S.), Penn Statistics in Imaging and Visualization Center, Department of Biostatistics, Epidemiology and Informatics (R.T.S.), and Lifespan Brain Chart Consortium (S.R.W., E.B., R.A.I.B., R.T.S., T.D.S., J.S., A.A.B.), University of Pennsylvania, Philadelphia, Pa; Centre for Medical Image Computing, Department of Medical Physics and Biomedical Engineering, University College London, London, United Kingdom (B.B., J.E.I.); Martinos Center for Biomedical Imaging and Department of Radiology (J.E.I.) and Department of Otolaryngology–Head and Neck Surgery (S.G.), Massachusetts General Hospital and Harvard Medical School, Boston, Mass; and McGovern Institute for Brain Research (S.G.) and Computer Science & Artificial Intelligence Laboratory (B.B., J.E.I.), Massachusetts Institute of Technology, Cambridge, Mass. Received January 27, 2023; revision requested March 13; final revision received August 21; accepted September 12. **Address correspondence to** A.A.B., Richards Medical Research Laboratories, 3700 Hamilton Walk, Philadelphia, PA 19104 (email: *aaron.alexander-bloch@pennmedicine.upenn.edu*).

J.M.S., J.S., and A.A.B. were supported in part by K08MH120564 and R01MH133843 from the National Institute of Mental Health. R.T.S. was supported by R01MH123550, R01NS112274, and R01MH112847 from the National Institutes of Health (NIH). H.H. was supported by R01MH092535, R01EB031284, and R01MH125333 from the NIH. S.R.W. was supported by the National Institute for Health and Care Research (NIHR) Cambridge Biomedical Research Centre (BRC-1215-20014) and United Kingdom Research and Innovation Medical Research Council (MC_UU_00002/2). B.B. and J.E.I. were supported in part by the European Research Council (ERC Starting Grant 677697) and the Engineering and Physical Sciences Research Council–funded University College London Centre for Doctoral Training in Medical Imaging (EP/L016478/1). T.D.S. was supported in part by R01MH120482 from the NIH. All research at the Department of Psychiatry in the University of Cambridge is supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014) and NIHR Applied Research Centre. Additional monetary support was provided by the Children's Hospital of Philadelphia Penn Lifespan Brain Institute. The views expressed are those of the author(s) and not necessarily those of the National Institute for Health and Care Research or the Department of Health and Social Care.

Conflicts of interest are listed at the end of this article.

See also the editorial by Ertl-Wagner and Pai in this issue.

Radiology 2023; 309(1):e230096 • https://doi.org/10.1148/radiol.230096 • Content codes: NR PD MR

Background: Clinically acquired brain MRI scans represent a valuable but underused resource for investigating neurodevelopment due to their technical heterogeneity and lack of appropriate controls. These barriers have curtailed retrospective studies of clinical brain MRI scans compared with more costly prospectively acquired research-quality brain MRI scans.

Purpose: To provide a benchmark for neuroanatomic variability in clinically acquired brain MRI scans with limited imaging pathology (SLIPs) and to evaluate if growth charts from curated clinical MRI scans differed from research-quality MRI scans or were influenced by clinical indication for the scan.

Materials and Methods: In this secondary analysis of preexisting data, clinical brain MRI SLIPs from an urban pediatric health care system (individuals aged ≤22 years) were scanned across nine 3.0-T MRI scanners. The curation process included manual review of signed radiology reports and automated and manual quality review of images without gross pathology. Global and regional volumetric imaging phenotypes were measured using two image segmentation pipelines, and clinical brain growth charts were quantitatively compared with charts derived from a large set of research controls in the same age range by means of Pearson correlation and age at peak volume.

Results: The curated clinical data set included 532 patients (277 male; median age, 10 years [IQR, 5–14 years]; age range, 28 days after birth to 22 years) scanned between 2005 and 2020. Clinical brain growth charts were highly correlated with growth charts derived from research data sets (22 studies, 8346 individuals [4947 male]; age range, 152 days after birth to 22 years) in terms of normative developmental trajectories predicted by the models (median $r = 0.979$).

Conclusion: The clinical indication of the scans did not significantly bias the output of clinical brain charts. Brain growth charts derived from clinical controls with limited imaging pathology were highly correlated with brain charts from research controls, suggesting the potential of curated clinical MRI scans to supplement research data sets.

Supplemental material is available for this article.

Abbreviations

GAMLSS = generalized additive models for location, scale, and shape; LBCC = Lifespan Brain Chart Consortium; SLIP = scan with limited imaging pathology

Summary

Brain growth charts derived from pediatric clinical brain MRI scans with limited imaging pathology were highly correlated with charts derived from scans in research control participants.

Key Results

- Clinical brain MRI scans with limited imaging pathology (n = 532; median patient age, 10 years [IQR, 5–14 years]; age range, 0.07–22 years) were identified using signed radiology reports and processed using two segmentation pipelines.
- Clinical brain growth charts of automated measures of brain volume were highly correlated with those of research controls (n = 8346; median phenotype trajectory correlation, r = 0.979).
- There was no evidence of differences in imaging measures related to the four most prevalent clinical indications for scans (P = .502–.877).

A principal challenge in brain MRI research is recruiting the large numbers of participants necessary to support valid scientific inference and generalizability. Retrospective studies of clinically acquired brain MRI scans may supplement costly prospective neuroimaging studies by harnessing existing health care data. For example, individuals with specific diseases may lack the support or resources to participate in research studies, but their clinical data could be anonymized for secondary use in research applications. From this perspective, the millions of brain MRI scans acquired yearly in clinical settings represent a valuable and vastly underused resource.

A major obstacle to using clinical MRI scans is the lack of appropriate controls necessary to rigorously test hypotheses in patient groups of interest. In research settings, it is customary to recruit healthy, typically developing participants explicitly for this purpose. Clinical MRI studies might have a control group composed of demographically matched patients who underwent brain MRI to rule out serious neuropathologic abnormalities and were found to have unremarkable MRI findings (1). A critical unresolved question is whether the difference in ascertainment process between research and clinical controls biases inferences about patient groups of interest.

Other challenges in using clinical brain MRI data include their technical heterogeneity and variable quality. Statistical approaches have proven successful in harmonizing scanner and sequence differences in MRI data (2–4), and deep learning–based image segmentation tools may provide robustness to scan quality (5,6). Recently, research-control brain growth charts were developed to quantitatively benchmark brain MRI phenotypes against population norms while controlling for differences between sites in an aggregated neuroimaging data set of 123 984 MRI scans from 100 studies (Lifespan Brain Chart Consortium [LBCC]) (7).

The aim of this study was to benchmark neuroanatomic differences in curated clinically acquired brain MRI scans with limited imaging pathology (SLIPs) and evaluate if growth charts developed from these clinical MRI scans demonstrate brain development differences when quantitatively compared with research-quality brain MRI scans. With use of a large set of clinically acquired brain MRI SLIPs from an urban pediatric health care system, clinical growth charts were quantitatively compared with the LBCC research growth charts. Similarities between the SLIP and LBCC charts were evaluated in terms of growth trajectories, age at peak cortical region volumes, and the generalizability of the clinical growth trajectories to out-of-sample clinical data.

Materials and Methods

This study was reviewed by the Children's Hospital of Philadelphia institutional review board and determined to be exempt from institutional review board oversight because it consisted of secondary analyses of preexisting clinical data. More detailed methods are available in Appendix S1.

Data Set Curation

The initial request for MRI SLIPs from the Children's Hospital of Philadelphia Department of Radiology consisted of 2013 randomly sampled scan sessions determined to lack clinically significant pathology by means of a randomly ordered review of their signed radiology reports (Appendix S1, Table S1). Scans whose reports detailed large artifacts due to motion or dental hardware, mentions of brain surgery, or disorders associated with tumor growth were excluded. They were limited to nine 3.0-T MRI scanners that used a harmonized magnetization-prepared rapid acquisition gradient-echo, or MPRAGE, T1-weighted sequence deployed in 2008 for routine brain MRI examinations (Appendix S1, Table S2). Scans were organized into Brain Imaging Data Structure, or BIDS (8), format using heudiconv (9) and divided into primary (n = 731) and out-of-sample (n = 739) data sets.

To generate growth chart models, primary SLIPs were filtered using CuBIDS (10) to isolate noncontrast MPRAGE scans with high spatial resolution (1 × 1 × 1 mm) and manually graded by two independent raters (J.M.S. and M.G., with 6 and 2 years of experience, respectively) to remove low-quality scans (Appendix S1, Fig 1) (11,12). Out-of-sample SLIP data were reserved to test the generalizability of growth chart models.

Clinical MRI SLIPs: Processing and Analysis

Two parallel processing pipelines were used for quantitative analyses of MRI SLIPs. The first pipeline reoriented and aligned scans to the MNI152 atlas (13), removed facial features, and performed segmentation using either FreeSurfer 6.0.0 (patient age, >3 years) (14,15) or Infant FreeSurfer 6.0.0 (age, ≤3 years) (16). The second pipeline performed segmentation using SynthSeg+ with pretrained models (5).

Both pipelines produced quantitative volumetric measurements, subsequently referred to as imaging phenotypes. Global imaging phenotypes quantified by each pipeline included total cortical gray matter volume, white matter volume, subcortical gray matter volume, ventricular cerebrospinal fluid volume, and total cerebrum volume. Additionally, 34 regional cortical volumes were quantified by the FreeSurfer pipeline with use of sulcal-based Desikan-Killiany parcellation (17).

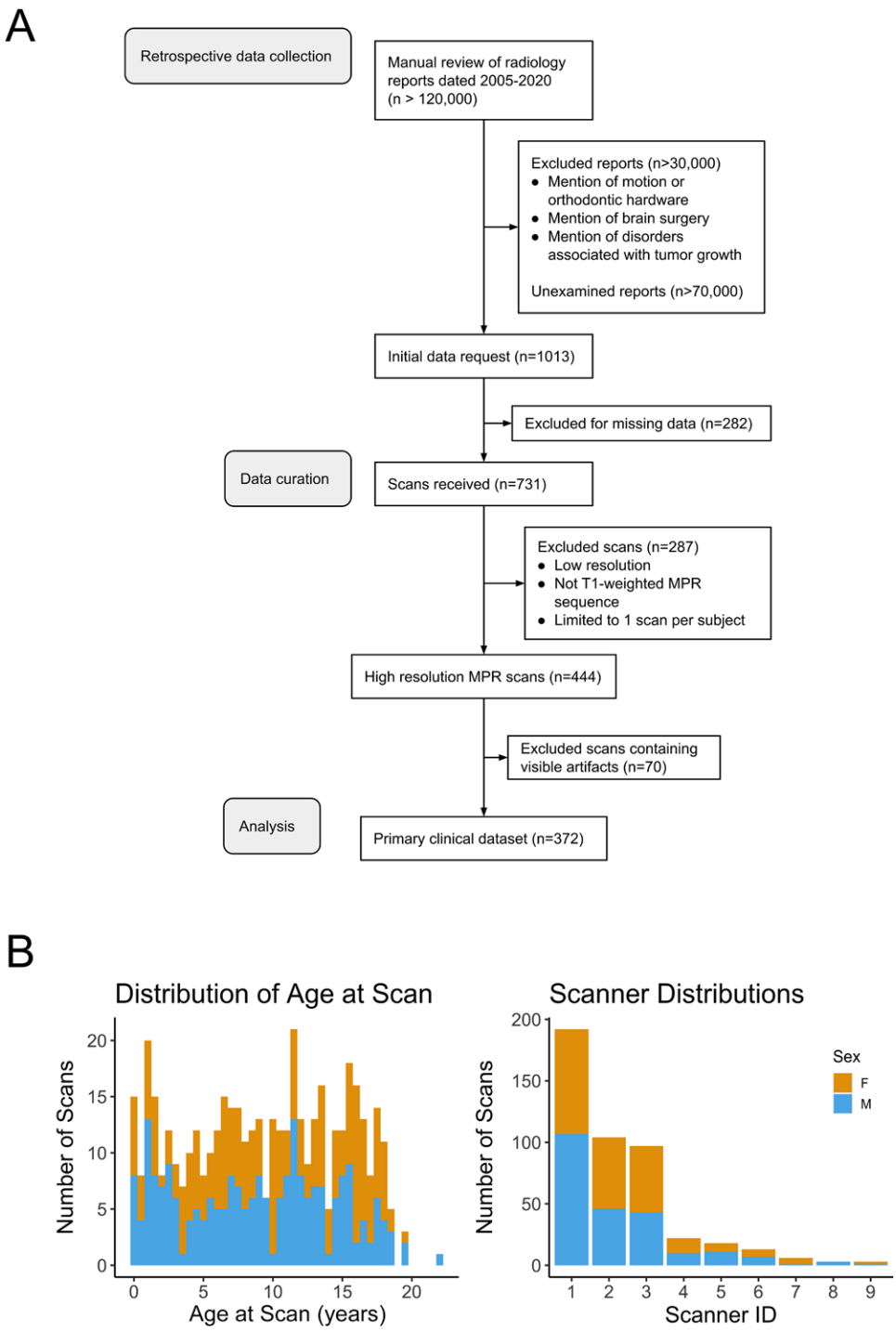

Figure 1: **(A)** Flowchart shows overview of the data curation and processing pipeline. The initial request for sessions whose signed radiology reports contained no gross pathology was submitted to the honest broker. The honest broker returned a set of anonymized MRI scans, which were then filtered to identify only high-spatial-resolution T1-weighted scans from 3.0-T scanners. Next, the high-resolution scans were manually reviewed by independent raters to remove visually low-quality images. This finalized set of curated clinical brain MRI scans with limited reported imaging pathology was processed using two neuroimaging processing pipelines (FreeSurfer or Infant FreeSurfer 6.0.0 and SynthSeg+) to produce two sets of imaging phenotypes for high-quality clinically acquired scans. MPR = magnetization-prepared rapid acquisition gradient echo. **(B)** Bar graphs show the distribution of age at scan (left) and the distribution of scans obtained across all scanners (right) labeled by sex. ID = identification number.

Table 1: Demographic Characteristics of the Curated Data Set of Clinical Brain MRI Scans with Limited Imaging Pathology

	Primary Data Set			Out-of-Sample Data Set		
Characteristic	Total	Male	Female	Total	Male	Female
No. of individuals scanned	372	187	185	160	90	70
Age (y)*	10 ± 5	9 ± 5	10 ± 5	9 ± 6	9 ± 5	9 ± 6
Age group						
0–2 years	41	25	16	22	12	10
2–5 years	41	22	19	28	18	10
5–10 years	102	51	51	33	17	16
10–13 years	63	37	26	34	23	11
13–18 years	113	44	69	41	19	22
18 years or older	12	8	4	2	1	1
Primary reason for scan†						
Developmental disorder	22	13	9			
Clinical eye or vision finding	31	18	13			
Headache	156	67	89			
Suspected seizure	33	18	15			
Other	130	71	59			
Year of scan						
2005	24	13	11	22	11	11
2006	58	26	32	17	12	5
2007	63	34	29	26	15	11
2008	66	34	32	21	14	7
2009	61	33	28	35	20	15
2010	78	38	40	28	13	15
2011 or later	22	9	13	11	5	6
Race						
American Indian or Alaska Native	1	1	0	0	0	0
Asian	7	4	3	4	3	1
Black or African American	63	33	30	40	27	13
Multiple races	2	1	1	4	1	3
Other	32	19	13	17	8	9
White	267	129	138	95	51	44

Note.—Unless otherwise specified, data are numbers of individuals scanned. Age of individual and year of scan were obtained from the electronic health record. For each age group, the lower limit is inclusive, and the upper limit is exclusive. The primary reason for scan was assessed based on manual review of radiology reports. The four most prevalent scan reasons were reported, and the remaining scan reasons were placed in an "other" group. Information about race was obtained from the electronic medical record, where self-reported data are recorded as part of routine clinical care in the following categories: American Indian or Alaska Native, Asian, Black or African American, multiple races, White, or other (unspecified).

* Data are means ± SDs.

† The primary reason for scan was not recorded for the out-of-sample data set.

LBCC Data

Clinical brain growth charts from primary SLIP data were compared with a subset of the previously published LBCC reference data (*http://www.brainchart.io* [7]). The full publicly available LBCC growth charts were used for comparison of cortical region phenotypes. The subset was limited to individuals with scans processed using FreeSurfer 6.0.0 or Infant FreeSurfer in the same age range as the SLIP data set.

Statistical Analysis

After data curation, quality control metrics were compared across demographic categories with use of analysis of variance and *t* tests. Imaging phenotypes underwent batch correction using ComBat, an approach adapted from statistical genomics, to control for the effect of MRI scanners while preserving the effect of other model covariates (2–4). Growth charts of primary SLIP and age-limited LBCC data were fit separately for each phenotype with use of generalized additive models for location, scale, and shape (GAMLSS), a distributional regression approach that models the mean, variance, and higher-order statistical moments in terms of flexible nonlinear associations using covariates of interest (18). Generalized gamma distributions linked each imaging phenotype to predictor variables of age and sex (plus Euler number for FreeSurfer models, which has been validated as a robust and automated measure of image quality) (19). In keeping with

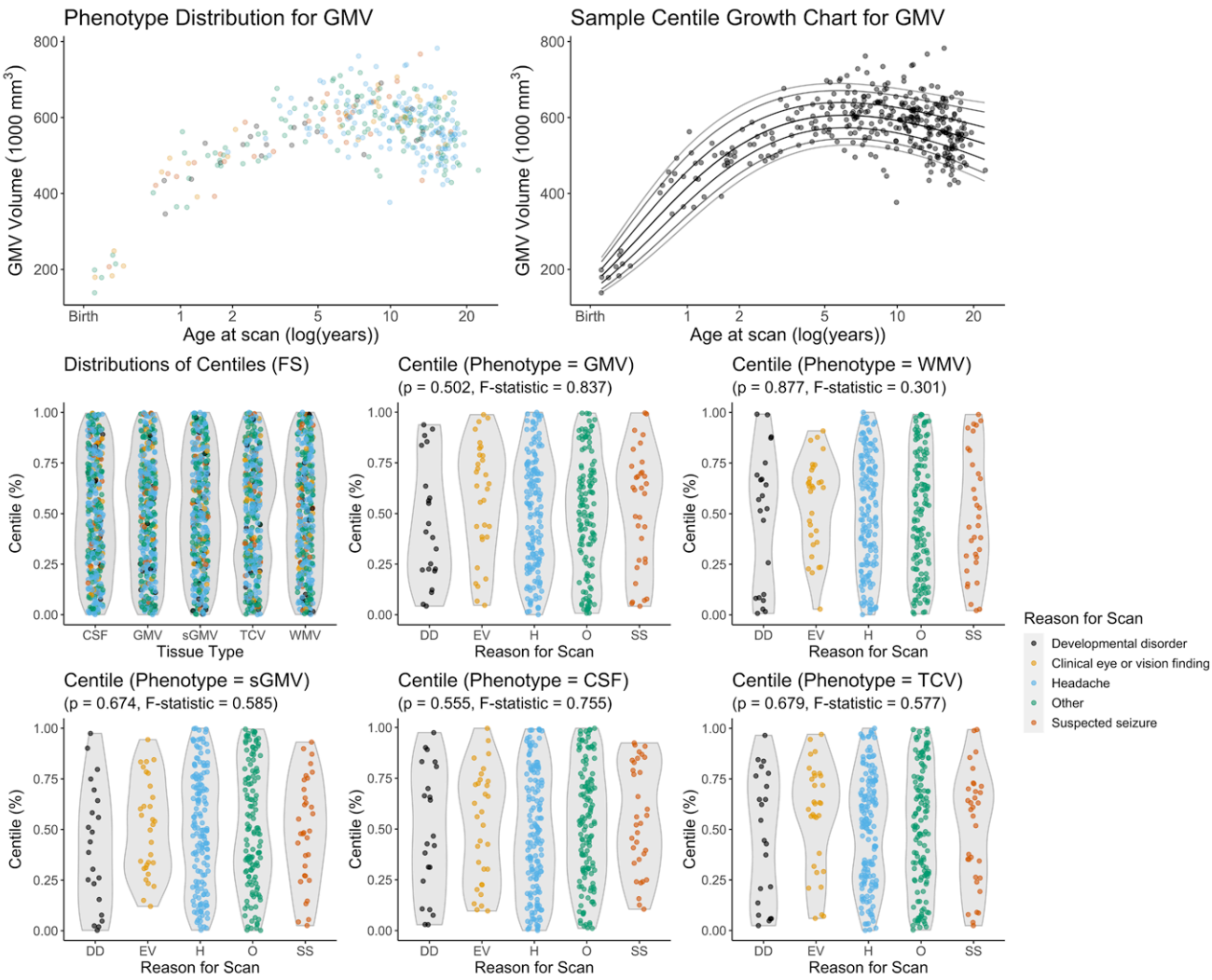

Figure 2: Scatterplots (top row) and violin plots show the distribution of phenotype centiles calculated using generalized additive models for location, scale, and shape, or GAMLSS, for the global FreeSurfer (FS) imaging phenotypes. There was no evidence of a statistical difference in the centile distributions for each phenotype based on the clinical indication for the scan (developmental disorder [DD], clinical eye or vision finding [EV], headache [H], suspected seizure [SS], and other [O]). For all five scan reason categories, there was no evidence of differences between phenotype distributions detected (gray matter volume [GMV]: $P = .502$, F statistic = 0.837; white matter volume [WMV]: $P = .877$, F statistic = 0.301; subcortical gray matter volume [sGMV]: $P = .674$, F statistic = 0.585; ventricular cerebrospinal fluid volume [CSF]: $P = .555$, F statistic = 0.755; and total cerebrum volume [TCV]: $P = .679$, F statistic = 0.577).

LBCC analyses (7), the mean and variance of each GAMLSS were modeled by a nonlinear age effect using third-order fractional polynomials, while an intercept term was used for model skewness, allowing for shared skewness across the age range *(https://github.com/BGDlab/clinical_brain_analysis)*.

Centile scores produced by GAMLSS were evaluated for potential bias related to year of scan and reason for scan with use of analysis of variance. Sensitivity analyses were conducted with and without implementation of ComBat harmonization to assess for the potential effect of MRI scanner. Centile scores produced for primary SLIP data were compared with those produced for the out-of-sample data to measure the generalizability of clinical growth charts to new data using the Kolmogorov-Smirnov test. To compare clinical and research charts, corresponding centile lines were plotted, and Pearson correlation coefficients between 50th centiles estimated for each chart across age were calculated. In addition, for cortical brain regions in the Desikan-Killiany atlas, age at peak volume for clinical and research controls was compared (Appendix S1). To test the anatomic correspondence in age at peak volume for clinical and research controls, "spin test" spatial null models based on random rotations of spherical projections of the cortical surface were used as described in detail in previous publications *(https://github.com/spin-test/spin-test)* (20,21). To supplement this approach, the intraclass correlation coefficient and Deming regression coefficient between age at peak volume for clinical and research controls were also calculated. To evaluate the generalizability of the clinical growth charts, SynthSeg+ growth chart models derived from the primary clinical data set were used to estimate centiles of SynthSeg+ phenotypes derived for the out-of-sample clinical data set. Bonferroni correction for multiple

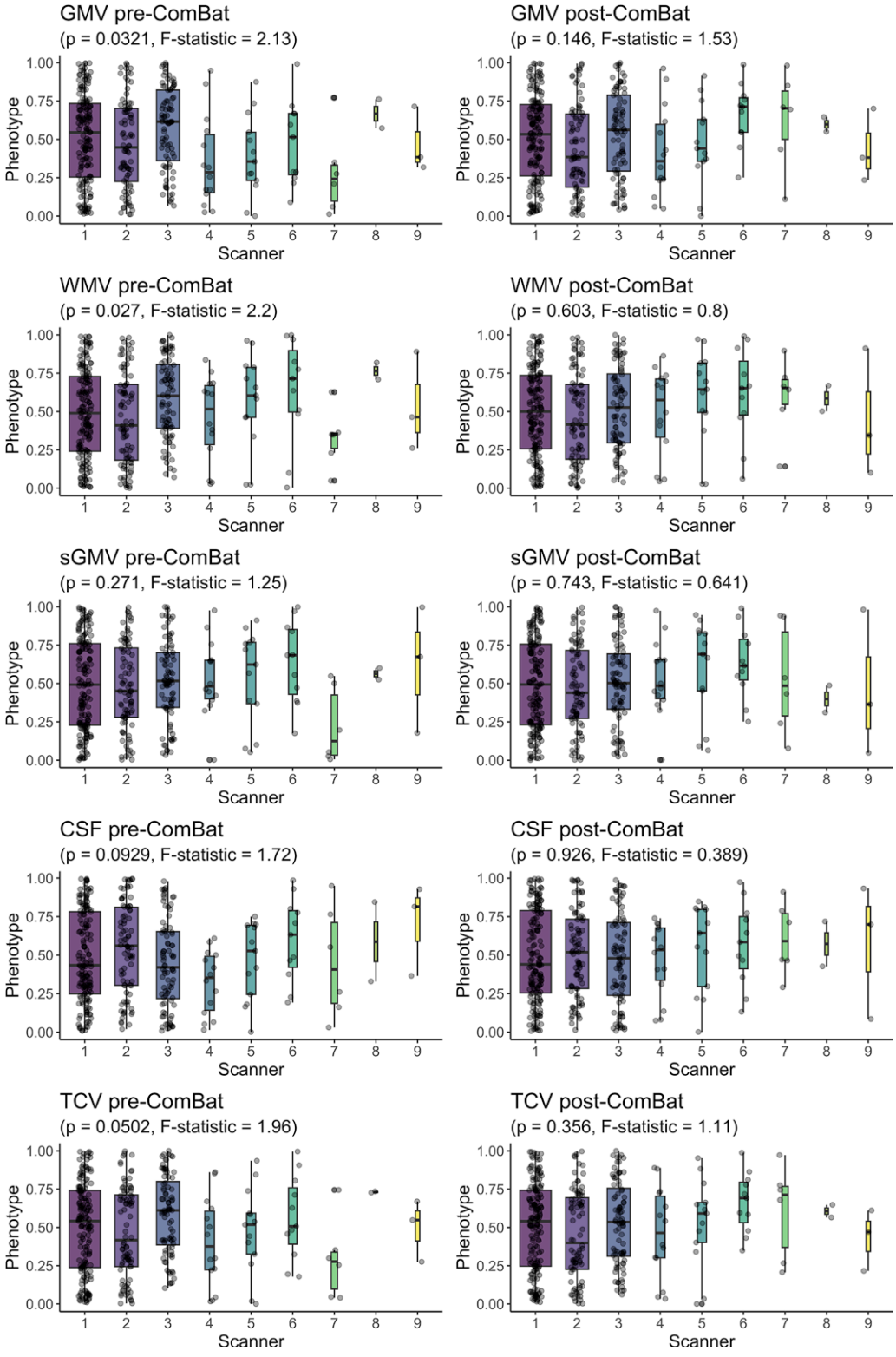

Figure 3: Centiles of global phenotypes by MRI scanner before and after harmonization with ComBat across scanners. Boxplots of centiles from each scanner are represented individually, and the scanners are presented left to right in decreasing order of the number of scans. The left column shows the centiles of FreeSurfer imaging phenotypes before ComBat harmonization (pre-ComBat), while the right column shows the centiles of FreeSurfer after ComBat harmonization (post-ComBat). The effect of scanner was tested with analysis of variance, showing no significant difference after harmonization (gray matter volume [GMV]: $P = .146$, F statistic = 1.53; white matter volume [WMV]: $P = .603$, F statistic = 0.8; subcortical gray matter volume [sGMV]: $P = .743$, F statistic = 0.641; ventricular cerebrospinal fluid volume [CSF]: $P = .926$, F statistic = 0.389; and total cerebrum volume [TCV]: $P = .356$, F statistic = 1.11). In each boxplot, the horizontal line in the box represents the median, the two hinges represent the 25th and 75th centiles, and the whiskers extend out from the hinges to the furthest values from the median up to 1.5 times the IQR.

comparisons was used, with a $P < .05$ threshold for a statistically significant difference.

Results

Data Set Demographic Characteristics and Quality Control

After curation and quality control (Fig 1, Appendix S1), the SLIP data set consisted of 532 individuals (277 male; median age, 10 years [IQR, 5–14 years]; age range, 28 days after birth to 22 years) scanned between 2005 and 2020 (Table 1). SLIP data were acquired across nine 3.0-T MRI scanners (Table S2) and split into primary (n = 372; 187 male) and out-of-sample (n = 160; 90 male) groups. There was no evidence of age or sex biases relating to scanners or manual or automated measures of image quality (Fig S1). Evidence of biases in centile scores related to year of the scan (Fig S2) or the four most prevalent clinical indications (developmental disorder, clinical eye or vision finding, headache, and suspected seizure) was not found (analysis of variance P = .502–.877) (Fig 2). There was evidence of subtle but significant scanner effects on imaging phenotypes, which were mitigated using ComBat harmonization (Fig 3). The age-limited LBCC data used in the GAMLSS consisted of 22 studies with 8346 individuals (4947 male; median age, 14 years [IQR, 10–17 years]; age range, 152 days after birth to 22 years) (Table 2).

Hospital Growth Charts versus Research Growth Charts

The characteristics of SLIP and age-limited LBCC growth charts were compared to assess the validity of clinical brain growth charts and the impact of the image processing pipeline (Fig 4). The charts displayed similar key milestones and developmental trajectory shapes, demonstrating robustness to data source and processing pipeline. In general, the charts were highly correlated between LBCC and SLIP data (Table 3) and between SLIP segmentation using FreeSurfer and SynthSeg+ (median phenotype trajectory correlation, r = 0.979) (Table 3). Arguably, differences due to processing pipeline outweighed differences attributable to data set, most notably for ventricular cerebrospinal fluid volume (correlation between LBCC and SLIP FreeSurfer, r = 0.338; correlation between LBCC and SLIP SynthSeg+, r = 0.882), although cerebrospinal fluid volumes segmented by FreeSurfer and SynthSeg+ in the same individuals remained highly correlated (r = 0.975). Additionally, a clear age-related discontinuity between FreeSurfer and Infant FreeSurfer was observed for subcortical gray matter volume (Fig 4), which has been reported previously (7).

For the cortical regions on both SLIP and LBCC brain charts, a clear maturational gradient in age at peak volume was observed from early maturation in the sensorimotor cortex to late maturation in the association cortex (Fig 5A, 5B). Interregional differences in age at peak volume were strongly correlated between SLIP and LBCC (Pearson r = 0.726 [95% CI: 0.496, 0.861]; intraclass correlation coefficient, 0.841; spin-test $P < .001$) (Fig 5C) (20,21). Generally, smaller cortical regions showed greater deviation between SLIP and LBCC brain charts (Fig 5).

Table 2: Demographic Characteristics of the Lifespan Brain Chart Consortium Subset (Research Controls) of the Same Age Range as Those with Brain MRI Scans with Limited Imaging Pathology (Clinical Controls) Used to Generate the Median Centiles in the Global Imaging Phenotype Trajectories

Study	Age (y)*	Total	Female	Male
Total	0.4–22 (14)	8346	3399	4947
ABCD (22)	9–11 (10)	376	155	221
ABIDE1 (23)	7–22 (14)	859	136	723
ABIDE2 (24)	5–22 (12)	647	139	508
ADHD200 (25)	7–22 (11)	809	294	515
AOBA (7)	11–21 (19)	132	39	93
AOMIC 1000 (26)	19–22 (21)	362	187	175
ASRB (27)	18–22 (21)	37	21	16
BHRCS (28)	6–21 (10)	717	314	403
BSNIP (29)	9–22 (18)	147	71	76
CAMFT (30)	13–16 (15)	68	35	33
cVEDA (31)	5–22 (15)	1162	482	680
EMBARC (32)	17–22 (19)	41	30	11
Female ASD (33)	6–17 (12)	358	180	178
IBIS (34)	0.4–2 (0.6)	394	148	246
ICBM (35)	17–21 (19)	149	82	67
IMAGEN (36)	12–22 (14)	1792	912	880
LA5c (37)	21–22 (21)	29	18	11
MCIC (38)	17–21 (19)	47	14	33
Narratives (39)	2–22 (20)	135	75	60
SALD (40)	19–22 (21)	40	26	14
STRIVE (41)	18–22 (21)	35	35	0
WAYNE (42)	18–22 (20)	10	6	4

Note.—Unless otherwise specified, data are numbers of participants. Participants older than 22 years in these studies were excluded from the comparison with the clinical controls. ABCD = Adolescent Brain Cognitive Development, ABIDE = Autism Brain Imaging Data Exchange, ADHD = Neuro Bureau Attention Deficit Hyperactivity Disorder 200 preprocessed data set, AOBA = Aoba Brain Imaging Research Center, AOMIC = Amsterdam Open MRI Collection, ASD = autism spectrum disorder, ASRB = Australian Schizophrenia Research Bank, BHRCS = Brazilian High Risk Cohort Study for Childhood Psychiatric Disorders, BSNIP = Bipolar-Schizophrenia Network on Intermediate Phenotypes, CAMFT = Cambridge Fetal Testosterone Study, cVEDA = Consortium on Vulnerability to Externalizing Disorders and Addictions, EMBARC = Establishing Moderators and Biosignatures of Antidepressant Response in Clinical Care, IBIS = Infant Brain Imaging Study, ICBM = International Consortium for Brain Mapping, IMAGEN = IMAGEN Consortium, LA5c = University of California Los Angeles Consortium for Neuropsychiatric Phenomics LA5c Study, MCIC = Mental Illness and Neuroscience Discovery Institute Clinical Imaging Consortium, SALD = Southwest University Adult Lifespan Dataset, STRIVE = Stress in Eating Study, WAYNE = Wayne State longitudinal data set of the Brain Aging in Detroit longitudinal study.

* Data are ranges, with medians in parentheses.

Growth Charts from Clinically Acquired Data Are Generalizable to Out-of-Sample Clinical Data

SynthSeg+ GAMLSS built using primary SLIP data were used to derive centile scores for each out-of-sample scan. The

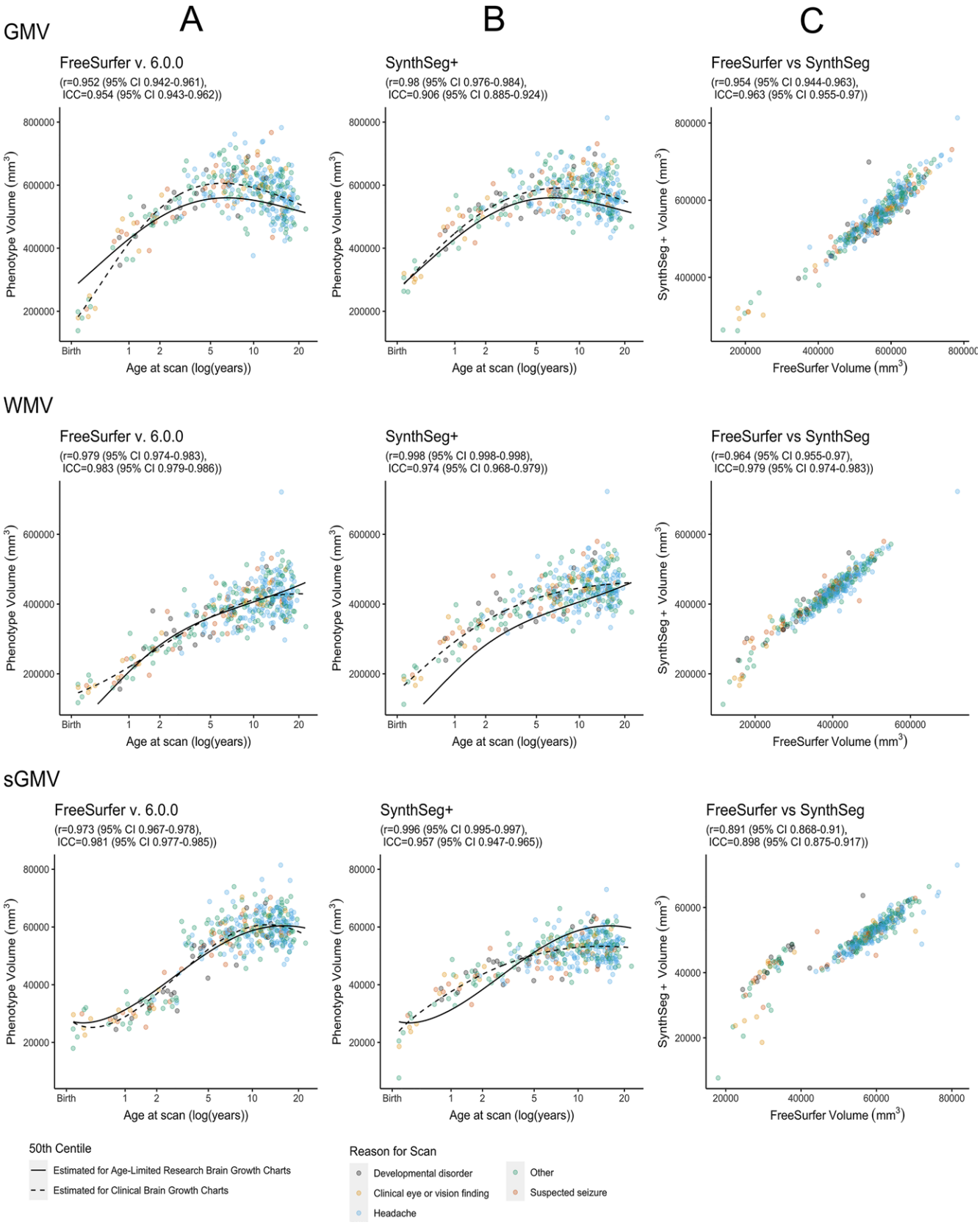

Figure 4: Growth trajectories of the five global imaging phenotypes compared across data sets and image processing pipelines. The clinical brain MRI scans with limited imaging pathology were processed with FreeSurfer and SynthSeg+ and compared with the models generated from the Lifespan Brain Chart Consortium data with the same age range as the clinical data set. **(A, B)** Scatterplots show the clinical brain MRI global imaging phenotypes, with the 50th centile line estimated with generalized additive models for location, scale, and shape, or GAMLSS, for the clinical brain growth charts of that phenotype (solid line) and the 50th centile line estimated from the research brain growth charts (dashed line) for the FreeSurfer 6.0.0 and SynthSeg+ processing pipelines *(Fig 4 continues)*.

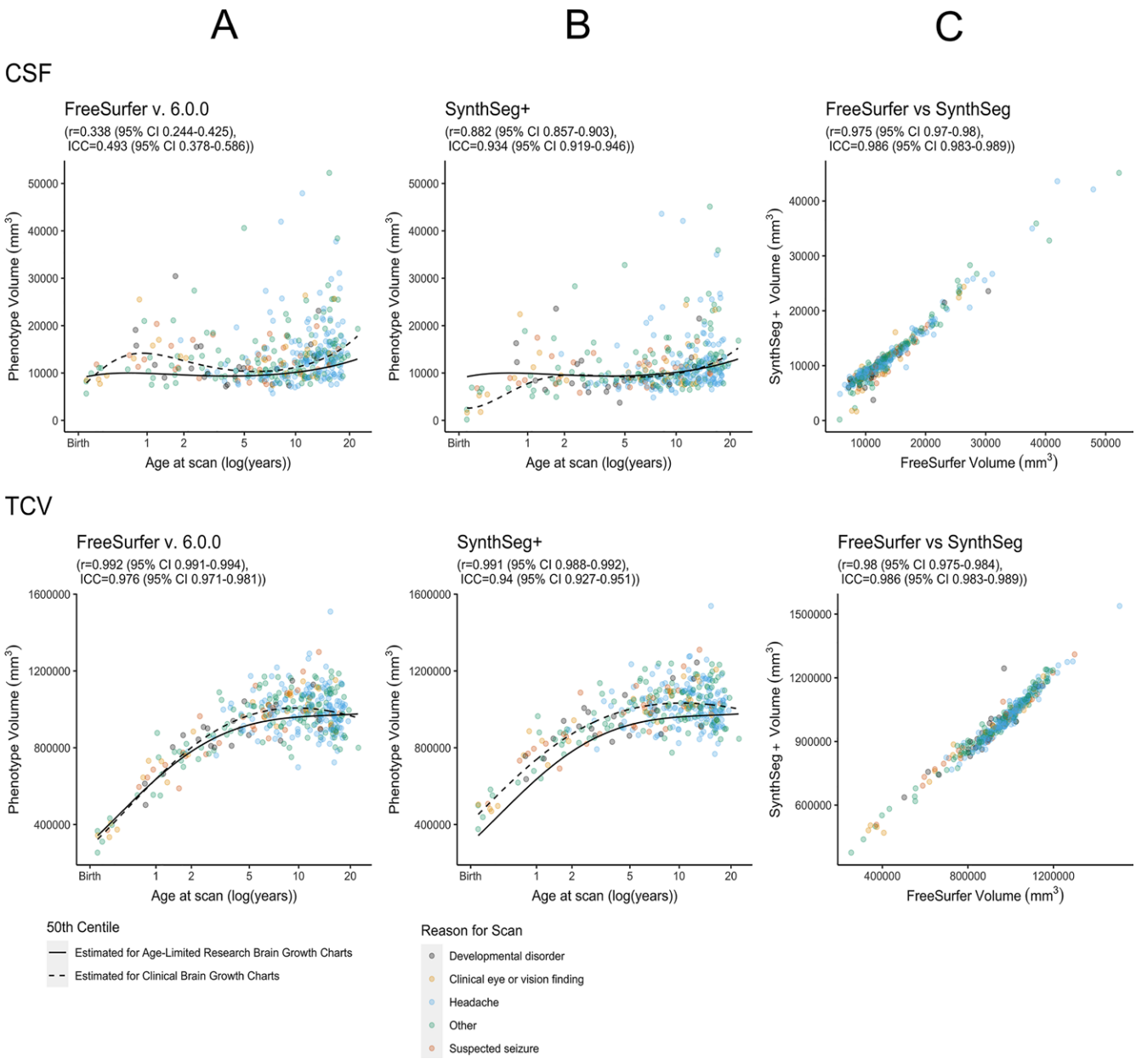

Figure 4 *(continued)*: (C) Scatterplots show direct comparison of the clinical imaging phenotypes derived from FreeSurfer and SynthSeg+. The Pearson correlation coefficient, intraclass correlation coefficient (ICC), and 95% CIs for each comparison can be seen in Table 3. CSF = ventricular cerebrospinal fluid volume, GMV = gray matter volume, sGMV = subcortical gray matter volume, TCV = total cerebrum volume, WMV = white matter volume.

distributions of centiles for both SLIP data sets showed no evidence of differences for any phenotype according to the Kolmogorov-Smirnov test after Bonferroni correction (all $P > .10$) (Fig 6).

Discussion

We generated clinical brain growth charts to address the challenge of using existing clinically acquired data for quantitative brain MRI research. We developed a curation process to identify clinical brain MRI scans with limited imaging pathology, then performed rigorous quality control and processed them with neuroimaging research pipelines. We used a principled distributional regression technique recommended by the World Health Organization to model growth charts for global and regional imaging phenotypes (19). We compared this clinical data set with one of the largest aggregated brain MRI research data sets and found a high degree of convergence between growth charts from each data set (median Pearson $r = 0.979$). Moreover, the clinical indication for the scans did not significantly bias the output of clinical brain charts (analysis of variance $P = .502$–$.877$).

Our study suggests multiple areas for future investigation. Tens of thousands of additional health care system scans meet inclusion criteria to be included in future growth charts. Curating more heterogeneous scans will increase the generalizability of these charts. Metrics of individual-level deviation from brain charts can be combined with other information available from the electronic health record, such as clinical genetics to investigate altered patterns of brain development in subgroups of patients.

Table 3: Brain Developmental Milestones Estimated for the Clinical Brain MRI SLIPs Using Generalized Additive Models for Location, Shape, and Scale

Comparison and Imaging Phenotype	Pearson Correlation Coefficient *(r)*	ICC
SLIP FreeSurfer versus LBCC		
Gray matter volume	0.952 (0.942, 0.961)	0.954 (0.943, 0.962)
White matter volume	0.979 (0.974, 0.983)	0.983 (0.979, 0.986)
Subcortical gray matter volume	0.973 (0.967, 0.978)	0.981 (0.977, 0.985)
Ventricular cerebrospinal fluid volume	0.338 (0.244, 0.425)	0.493 (0.378, 0.586)
Total cerebrum volume	0.992 (0.991, 0.994)	0.976 (0.971, 0.981)
SLIP SynthSeg+ versus LBCC		
Gray matter volume	0.980 (0.976, 0.984)	0.906 (0.885, 0.924)
White matter volume	0.998 (0.998, 0.998)	0.974 (0.968, 0.979)
Subcortical gray matter volume	0.996 (0.995, 0.997)	0.957 (0.947, 0.965)
Ventricular cerebrospinal fluid volume	0.882 (0.857, 0.903)	0.934 (0.919, 0.946)
Total cerebrum volume	0.991 (0.988, 0.992)	0.940 (0.927, 0.951)
SLIP FreeSurfer versus SLIP SynthSeg+		
Gray matter volume	0.954 (0.944, 0.963)	0.963 (0.955, 0.97)
White matter volume	0.964 (0.955, 0.97)	0.979 (0.974, 0.983)
Subcortical gray matter volume	0.891 (0.868, 0.91)	0.898 (0.875, 0.917)
Ventricular cerebrospinal fluid volume	0.975 (0.97, 0.98)	0.986 (0.983, 0.989)
Total cerebrum volume	0.980 (0.975, 0.984)	0.986 (0.983, 0.989)

Note.—Data in parentheses are 95% CIs. Models based on imaging phenotypes derived from FreeSurfer and SynthSeg+ image processing pipelines were compared with aggregated research data from the Lifespan Brain Chart Consortium (LBCC). The Pearson correlation coefficient, intraclass correlation coefficient (ICC), and their 95% CIs were calculated for the five global volumetric phenotypes based on the 50th centile predicted for primary MRI scans with limited imaging pathology (SLIPs) (n = 372) according to SLIP and LBCC brain chart models.

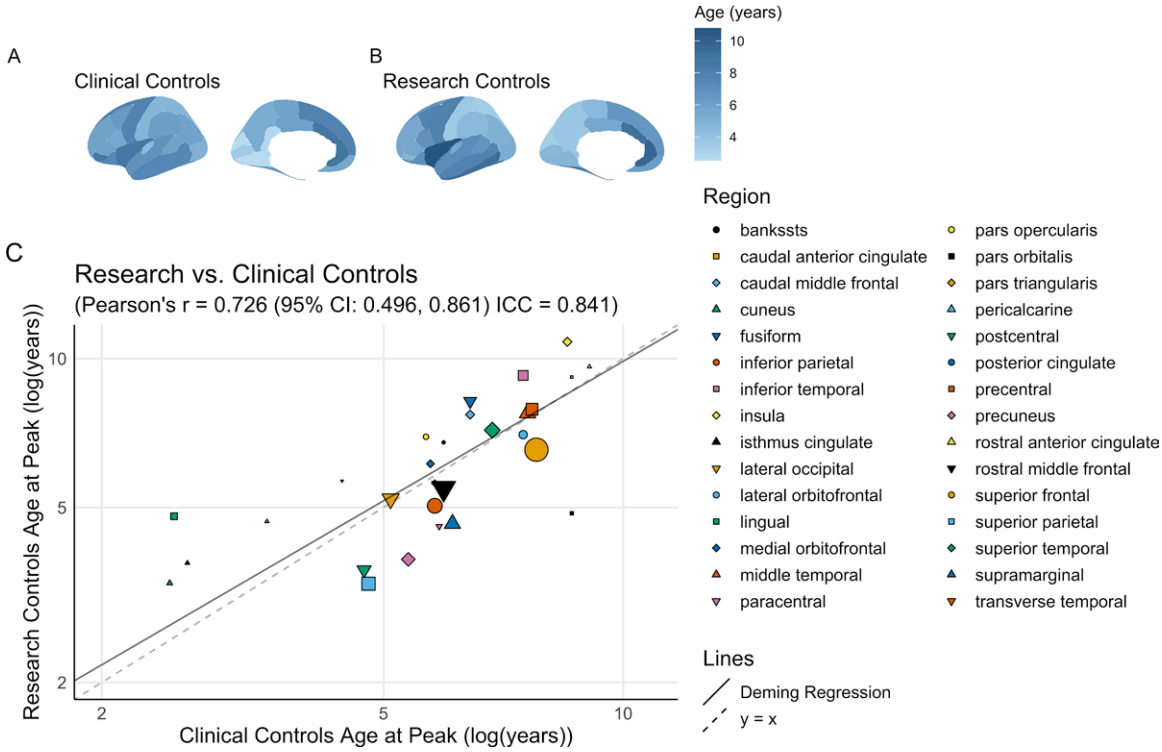

Figure 5: Comparison of regional brain development modeled on scans with limited imaging pathology (SLIPs) and Lifespan Brain Chart Consortium (LBCC) scans. Desikan-Killiany atlases for **(A)** SLIPs and **(B)** LBCC scans show the age at which the volume of each cortical region peaked with values averaged across left and right hemispheres. The correspondence of the pair of age at peak volume maps was statistically significant according to spatially informed null models (see Materials and Methods section; spin-test $P < .001$). **(C)** Scatterplot shows the age at peak regional volume in clinical and research controls (Pearson $r = 0.726$ [95% CI: 0.496, 0.861]; intraclass correlation coefficient [ICC] = 0.841). The two lines in **C** are for the line y = x and the Deming regression (slope = 0.938 [95% CI: 0.533, 1.34]; intercept = 0.220 [95% CI: −1.17, 1.60]). In the scatterplot, the size of each point is proportional to the average size of the brain region it represents. bankssts = banks of superior temporal sulcus.

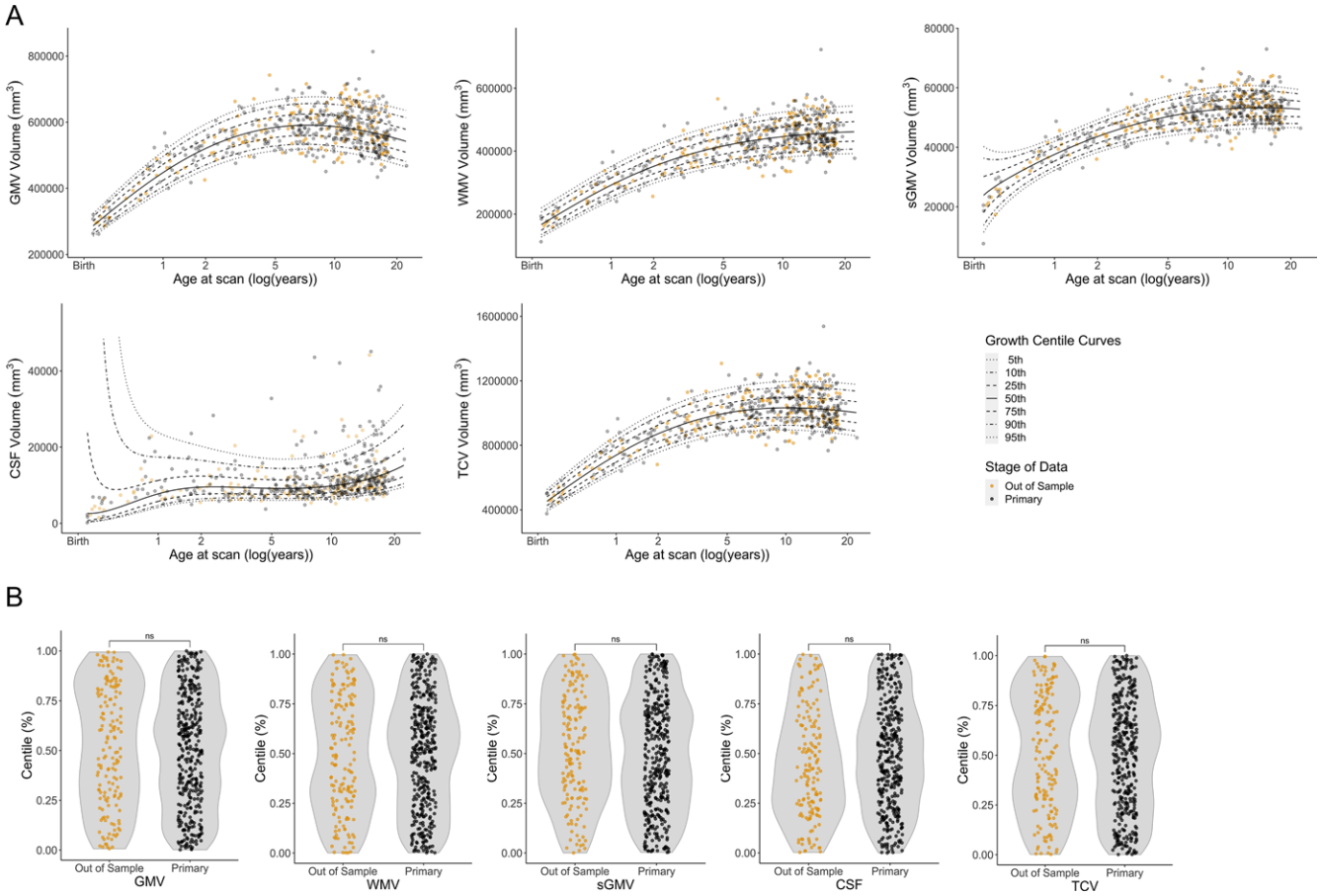

Figure 6: **(A)** Scatterplots show growth trajectories of the five global imaging phenotypes as produced by SynthSeg+ overlaid with out-of-sample data processed by SynthSeg+. Each growth trajectory chart shows how the out-of-sample clinical data fit on the growth charts generated using the primary clinical data. The curves on each chart are the 5th, 10th, 25th, 50th, 75th, 90th, and 95th centiles, respectively. **(B)** Violin plots show the comparison between the centiles of the primary data and the out-of-sample data as estimated using generalized additive models for location, scale, and shape trained on the primary data. The Kolmogorov-Smirnov test comparing the distributions of centiles from the primary data and the out-of-sample data showed no statistically significant differences between the distributions of centiles for all phenotypes after Bonferroni correction (five classes of phenotypes; $P = .430$, .338, .854, .118, and .515 for gray matter volume [GMV], white matter volume [WMV], subcortical gray matter volume [sGMV], ventricular cerebrospinal fluid volume [CSF], and total cerebrum volume [TCV], respectively). ns = not significant.

Our study had limitations. First, we compared clinical and research controls in a clinical center using high-quality isotropic T1-weighted MRI sequences for routine MRI examinations. Thus, growth charts presented herein may generalize less well to centers using lower spatial resolution or nonisotropic scanning protocols or serving clinically dissimilar populations. However, SynthSeg+ is expected to generalize well to lower-quality scans and diverse clinical scanning protocols (5). Second, biases due to the year of scan could pose challenges in comparisons with more recent data. Finally, short-term utility of growth chart benchmarks of quantitative imaging features is limited to the research domain, although it is conceivable that contemporaneous benchmarks could prove clinically useful.

In conclusion, brain growth charts derived from clinical controls with limited imaging pathology were highly correlated with brain charts from research controls, and their output was robust to the clinical indication for scans. Harnessing clinically acquired brain MRI scans vastly increases the amount of data available to brain imaging researchers, especially for clinical subpopulations where prospective research studies are currently lacking. Our results address a longstanding need in brain MRI research by suggesting the utility of clinical brain growth charts as suitable benchmarks for developmental norms.

Acknowledgments: The authors thank Jisoo Kang, MS; Nadia Ngom; Sreya Subramanian, BS; and Caleb Schmitt for contributing to the review of signed radiology reports. The live list of authors included in the Lifespan Brain Chart Consortium can be found at *https://github.com/brainchart/Lifespan*.

Author contributions: Guarantors of integrity of entire study, **J.M.S., A.A.B.**; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, **J.M.S., S.S., H.H., A.O.R., D.R., J.S., A.A.B.**; clinical studies, **A.V., S.A., V.P., A.S.M.**; experimental studies, **J.M.S., S.S., T.P.R., H.H., M.G., R.E.G., D.R.**; statistical analysis, **J.M.S., J.E.S., A.O.R., S.C., S.R.W., R.A.I.B., R.T.S., J.E.I., T.D.S., J.S., A.A.B.**; and manuscript editing, **J.M.S., J.E.S., S.S., A.V., S.A., T.P.R., H.H., M.G., S.C., S.A.B., A.S.M., B.H.C., S.R.W., E.B., R.A.I.B., R.T.S., B.B., J.E.I., S.G., R.E.G., T.D.S., D.R., J.S., A.A.B.**

Disclosures of conflicts of interest: **J.M.S.** No relevant relationships. **J.E.S.** No relevant relationships. **S.S.** No relevant relationships. **A.V.** No relevant relationships. **S.A.** No relevant relationships. **T.P.R.** Stock or stock options in Proteus Neurodynamics, FieldLine, and Prism Clinical Imaging. **H.H.** No relevant relationships. **V.P.** No relevant relationships. **A.O.R.** No relevant relationships. **M.G.** No relevant relationships. **S.C.** No relevant relationships. **S.A.B.** No relevant relationships. **A.S.M.** No relevant

relationships. **B.H.C.** No relevant relationships. **S.R.W.** Equity in Centile Bio. **E.B.** Royalties from Hachette and Elsevier; consulting fees from Boehringer Ingelheim, Sosei Heptares, SR One, and GlaxoSmithKline; chair of the Advisory Board on Mental Health and Adolescent Brain Development for the Medical Research Council UK; honorary treasurer and member of the Academy of Medical Sciences, director of Academy of Medical Sciences Trading, and director and co-founder of Centile Bio. **R.A.I.B.** Board member for Centile Bio; equity in Centile Bio. **R.T.S.** Grants to institution from the National Institutes of Health and the Multiple Sclerosis Society; consulting fees from Octave Bioscience; compensation for reviewing manuscripts from the American Medical Association. **B.B.** No relevant relationships. **J.E.I.** No relevant relationships. **S.G.** Grant from the National Institutes of Health. **R.E.G.** No relevant relationships. **T.D.S.** No relevant relationships. **D.R.** Grants from the National Institutes of Health. **J.S.** Grants from the National Institutes of Health; board member for Centile Bio; equity in Centile Bio. **A.A.B.** Consulting fees from Octave Bioscience; payment for lecture from the University of Oregon and for grand rounds from Vanderbilt University; board member for the Hyacinth Fellowship and Centile Bio; stock in Centile Bio.

References

1. Kerr WT, Tatekawa H, Lee JK, et al. Clinical MRI morphological analysis of functional seizures compared to seizure-naïve and psychiatric controls. Epilepsy Behav 2022;134:108858.
2. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2007;8(1):118–127.
3. Fortin JP, Cullen N, Sheline YI, et al. Harmonization of cortical thickness measurements across scanners and sites. Neuroimage 2018;167:104–120.
4. Pomponio R, Erus G, Habes M, et al. Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. Neuroimage 2020;208:116450.
5. Billot B, Magdamo C, Cheng Y, Arnold SE, Das S, Iglesias JE. Robust machine learning segmentation for large-scale analysis of heterogeneous clinical brain MRI datasets. Proc Natl Acad Sci USA 2023;120(9):e2216399120.
6. McClure P, Rho N, Lee JA, et al. Knowing what you know in brain segmentation using Bayesian deep neural networks. Front Neuroinform 2019;13:67.
7. Bethlehem RAI, Seidlitz J, White SR, et al. Brain charts for the human lifespan. Nature 2022;604(7906):525–533. [Published correction appears in Nature 2022;610(7931):E6.]
8. Gorgolewski KJ, Auer T, Calhoun VD, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci Data 2016;3(1):160044.
9. Halchenko Y, Goncalves M, Visconti di Oleggio C, Ghosh S, Salo T. A Flexible DICOM Converter for Organizing Brain Imaging Data Into Structured Directory Layouts. Zenodo. https://doi.org/10.5281/zenodo.5557588. Published October 8, 2021. Accessed October 2021.
10. Covitz S, Tapera TM, Adebimpe A, et al. Curation of BIDS (CuBIDS): a workflow and software package for streamlining reproducible curation of large BIDS datasets. Neuroimage 2022;263:119609.
11. Rosen AFG, Roalf DR, Ruparel K, et al. Quantitative assessment of structural image quality. Neuroimage 2018;169:407–418.
12. Bedford SA, Ortiz-Rosa A, Schabdach JM, et al. The impact of quality control on cortical morphometry comparisons in autism. medRxiv 2022.12.05.22283091 [preprint] https://doi.org/10.1101/2022.12.05.22283091. Posted December 6, 2022. Accessed December 2022.
13. Fonov V, Evans AC, Botteron K, et al. Unbiased average age-appropriate atlases for pediatric studies. Neuroimage 2011;54(1):313–327.
14. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. FSL. Neuroimage 2012;62(2):782–790.
15. Fischl B, van der Kouwe A, Destrieux C, et al. Automatically parcellating the human cerebral cortex. Cereb Cortex 2004;14(1):11–22.
16. Zöllei L, Iglesias JE, Ou Y, Grant PE, Fischl B. Infant FreeSurfer: an automated segmentation and surface extraction pipeline for T1-weighted neuroimaging data of infants 0-2 years. Neuroimage 2020;218:116946.
17. Desikan RS, Ségonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 2006;31(3):968–980.
18. Stasinopoulos MD, Rigby RA, Heller GZ, Voudouris V, De Bastiani F. Flexible regression and smoothing: using GAMLSS in R 21. CRC Press, 2017.
19. Borghi E, de Onis M, Garza C, et al. Construction of the World Health Organization child growth standards: selection of methods for attained growth curves. Stat Med 2006;25(2):247–265. [Published correction appears in Stat Med 2008;27(16):3216.]
20. Alexander-Bloch AF, Shou H, Liu S, et al. On testing for spatial correspondence between maps of human brain structure and function. Neuroimage 2018;178:540–551.
21. Markello RD, Misic B. Comparing spatial null models for brain maps. Neuroimage 2021;236:118052.
22. Casey BJ, Cannonier T, Conley MI, et al. The Adolescent Brain Cognitive Development (ABCD) study: imaging acquisition across 21 sites. Dev Cogn Neurosci 2018;32:43–54.
23. Di Martino A, Yan CG, Li Q, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol Psychiatry 2014;19(6):659–667.
24. Di Martino A, O'Connor D, Chen B, et al. Enhancing studies of the connectome in autism using the autism brain imaging data exchange II. Sci Data 2017;4:170010.
25. Bellec P, Chu C, Chouinard-Decorte F, Benhajali Y, Margulies DS, Craddock RC. The Neuro Bureau ADHD-200 preprocessed repository. Neuroimage 2017;144(Pt B):275–286.
26. Snoek L, van der Miesen MM, Beemsterboer T, van der Leij A, Eigenhuis A, Steven Scholte H. The Amsterdam Open MRI Collection, a set of multimodal MRI datasets for individual difference analyses. Sci Data 2021;8(1):85.
27. Loughland C, Draganic D, McCabe K, et al. Australian Schizophrenia Research Bank: a database of comprehensive clinical, endophenotypic and genetic data for aetiological studies of schizophrenia. Aust N Z J Psychiatry 2010;44(11):1029–1035.
28. Salum GA, Gadelha A, Pan PM, et al. High risk cohort study for psychiatric disorders in childhood: rationale, design, methods and preliminary results. Int J Methods Psychiatr Res 2015;24(1):58–73.
29. Tamminga CA, Pearlson G, Keshavan M, Sweeney J, Clementz B, Thaker G. Bipolar and schizophrenia network for intermediate phenotypes: outcomes across the psychosis continuum. Schizophr Bull 2014;40 Suppl 2(Suppl 2):S131–S137.
30. Lombardo MV, Ashwin E, Auyeung B, et al. Fetal testosterone influences sexually dimorphic gray matter in the human brain. J Neurosci 2012;32(2):674–680.
31. Sharma E, Vaidya N, Iyengar U, et al. Consortium on Vulnerability to Externalizing Disorders and Addictions (cVEDA): a developmental cohort study protocol. BMC Psychiatry 2020;20(1):2.
32. Trivedi MH, McGrath PJ, Fava M, et al. Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): rationale and design. J Psychiatr Res 2016;78:11–23.
33. Pelphrey K. NIMH Data Archive. National Institute of Mental Health. https://nda.nih.gov/edit_collection.html?id=2021. Posted November 23, 2012. Accessed October 22, 2021.
34. Hazlett HC, Gu H, McKinstry RC, et al. Brain volume findings in 6-month-old infants at high familial risk for autism. Am J Psychiatry 2012;169(6):601–608.
35. Mazziotta J, Toga A, Evans A, et al. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philos Trans R Soc Lond B Biol Sci 2001;356(1412):1293–1322.
36. Schumann G, Loth E, Banaschewski T, et al. The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. Mol Psychiatry 2010;15(12):1128–1139.
37. Bilder R, Poldrack R, Cannon T, et al. UCLA Consortium for Neuropsychiatric Phenomics LA5c Study. OpenNeuro ds000030. https://openneuro.org/datasets/ds000030/versions/00016. Updated July 18, 2018. Accessed December 5, 2020.
38. Gollub RL, Shoemaker JM, King MD, et al. The MCIC collection: a shared repository of multi-modal, multi-site brain image data from a clinical investigation of schizophrenia. Neuroinformatics 2013;11(3):367–388.
39. Nastase SA, Liu YF, Hillman H, et al. The "Narratives" fMRI dataset for evaluating models of naturalistic language comprehension. Sci Data 2021;8(1):250.
40. Wei D, Zhuang K, Ai L, et al. Structural and functional brain scans from the cross-sectional Southwest University adult lifespan dataset. Sci Data 2018;5:180134.
41. Westwater ML, Mancini F, Gorka AX, et al. Prefrontal responses during proactive and reactive inhibition are differentially impacted by stress in anorexia and bulimia nervosa. J Neurosci 2021;41(20):4487–4499.
42. Raz N. Brain Aging in Detroit Longitudinal Studies. NeuroImaging Tools and Resource Collaboratory. https://www.nitrc.org/doi/landing_page.php?table=groups&id=1542. Posted November 4, 2021. Accessed July 5, 2020. .