

Original Research Articles

Large-scale open science simulated data to study the development of brain and behavior relationships: Epidemiologic, clinical, and neuroimaging characteristics

Tonya White, MD, PhD¹^a, Neda Sadeghi, PhD¹, Luke Norman, PhD¹, Vince Calhoun, PhD², Philip Shaw, MD, PhD³

¹ Section on Social and Cognitive Developmental Neuroscience, National Institute of Mental Health, Bethesda, MD, USA, ² Tri-institutional Center for Translational Research in Neuroimaging and Data Science, Georgia State University, Atlanta, GA, USA, ³ Department of Child and Adolescent Psychiatry, King's College London, London, UK

Keywords: neurodevelopment, simulation, longitudinal studies, brain behavior relationships, epidemiology, missing data, sex effects

<https://doi.org/10.52294/001c.165375>

Aperture Neuro

Vol. 6, Issue SI 1, 2026

Both brain and behavior undergo considerable changes during adolescence, a time synonymous with the emergence several psychiatric disorders, including mood and psychotic disorders. Multiple studies have explored the interaction between brain and behavior during childhood and adolescence, with both mixed and convergent findings between the studies. One challenge is that different longitudinal models can be applied to both the same or similar dataset and yet yield different results. Without knowing the ground truth, the best modeling approach to understanding the brain/behavior interplay during development remains unclear. Within this context, simulated datasets can provide a mechanism to test many different neurodevelopmental features, not only regarding the interplay between brain and behavior, but also how specific epidemiologic factors can influence study findings. Simulated datasets have the ‘ground truth’ relationships embedded within the models used to create these datasets. Thus, the goal of this study was to assess the epidemiologic and clinical characteristics of 15 large simulated longitudinal neuroimaging datasets, independently created from five research groups from three different continents. We found that while most of the neuroimaging and behavioral trajectories in the simulated datasets reflected growth patterns identified in the literature, there was also considerable variability between how the groups modeled the epidemiologic, clinical, and neurodevelopmental growth patterns. The conclusion is that while simulation studies are an underused strategy to test assumptions related to neurodevelopmental disorders, simulation studies require iterative approaches, testing simulated against real-world data and challenging the different assumptions associated with the neurodevelopmental models, until there is convergence between actual and simulated datasets.

INTRODUCTION

Adolescence, a period involving considerable neurodevelopment changes, is a stage of life when a number of major psychiatric disorders can emerge.¹ Disorders such as depression, generalized anxiety, psychotic, eating, and substance use disorders are not uncommonly present during adolescence.¹⁻⁵ While numerous neuroimaging studies have captured the neurobiological characteristics of these psychiatric disorders, there remains a considerable gap in

our understanding of the pre-illness characteristics of the brain. Most studies recruit patients who are either prodromal, or those who have already developed their illness. Studies assessing the pre-illness characteristics of the brain are rare and yet are critical to understand early differences in neurodevelopmental trajectories and the interplay between brain and behavior over time.

One approach to assess the neurodevelopmental trajectories of emerging psychopathology is using large, longitudinal studies that recruit participants prior to the expected

^a Corresponding Author:

Tonya White, MD, PhD (tonya.white@nih.gov)

Section on Social and Cognitive Developmental Neuroscience, National Institute of Mental Health
Bethesda, MD, USA

Table 1. Large Datasets of Child and Adolescent Brain Development: Clinical, Typically Developing, and Population-based.

Study Name	Study Type	Baseline Neuroimaging Sample Size	Neuroimaging waves of Data Collection	Open Science Rating	References
ABCD	Population-based	11,878	3 (partial 4 th) (ongoing study)		10
ABIDE	Clinical (Autism)				11,12
Generation R	Population-based	3996	4 (ongoing study)		6,13
GUSTO	Population-based	1,247	4 (ongoing study)		14-16
HBCD	Mixed	~ 1200			17
HBN	Clinical	664 (Goal 10,000)	1 (ongoing study)		18
IMAGEN		2,223	4 (ongoing study)		19-21
NIMH	Typically Developing	> 2,000	4 + (ongoing study)		22
NKI Rockland	Clinical	300 (pediatric)	3 (ongoing study)		23
PNC	Clinical	1,445	1		7,24
PING	Typically Developing*	1,493	1		25
SYS	Population-based	1,029	2		26

- ABCD = Adolescent Brain and Cognitive Development Study; GUSTO = Growing Up in Singapore Towards Healthy Outcomes; HBCD = Healthy Brain and Child Development Study; HBN = Healthy Brain Network; NIMH = National Institute of Mental Health Study of Typical Development; PGC = Philadelphia Neurodevelopmental Cohort; PING = Pediatric Imaging, Neurocognition, and Genetics (* Children with attention-deficit hyperactivity disorder (ADHD) and learning disabilities were not excluded); SYS = Saguenay Youth Study
- Open Science Rating (Green = Data access is free and can be accessed and downloaded via a data transfer agreement; Yellow = Some challenges in accessing data, but can be downloaded and shared; Red = Significant challenges in accessing data and cannot be downloaded).

age-of-onset of the disorders.⁶ Examples of large pediatric datasets are provided in Table 1, including those datasets that are accessible for data sharing. Several of these studies now have up to four waves of neuroimaging data collection.⁷ The goal of these large studies is to capture a large, population-based sample of children and to study these children in multiple waves of data collection into adolescence and adulthood. The goals of these studies is to not only map typical neurodevelopment, but to assess deviations from typical neurodevelopment as a result of adverse exposures or psychiatric illnesses. Studying these individuals well past the median age of onset for major psychiatric disorders offers the opportunity to map the temporal characteristics of emerging psychiatric disorders. These studies are critical for enhancing our understanding of whether the neurobiological features are causal or predictive of the psychopathology, or alternatively, are downstream features of having the disorder.^{8,9}

These existing and ongoing studies can be used to apply a wide array of longitudinal models to test brain/behavior relationships during development.²⁷ Such models offer the opportunity to test both linear and nonlinear deviations of developmental trajectories, delays in development, and time-variant relationships between brain, behavior, and other variables.²⁷ In fact, since there are multiple statistical

models that can be applied to assess brain/behavior relationships over time, one of the challenges is knowing whether the selected model has internal validity to detect specific brain/behavior relationships. As an example, we now have several studies that utilize cross-lagged panel models to assess the bidirectional relationship between brain and behavior.^{8,9,28,29} Interestingly, rather than demonstrating that brain predicts downstream behavior, which was our original hypothesis, we found that behavior predicts downstream brain changes. However, had we used linear mixed models with the hypothesis that brain predicts behavior, we would have missed the alternative hypothesis. Simulated datasets, where the true direction of the brain-behavior relationship is known, will allow researchers to test whether cross-lagged panel models or linear mixed models more accurately recover the ground truth under different conditions.

Neuroimaging studies have at times been challenging due to small sample sizes, small effect sizes, and the considerable noise, variability, and heterogeneity of the study results. These factors contribute to the mixed findings in the literature.^{30,31} However, even with the availability of large samples for neuroimaging studies, we often lack the 'ground truth' involving the longitudinal interplay between brain and behavior in neurodevelopmental disorders. While

pre-clinical and replication studies can help clarify the internal validity of specific neurodevelopmental models, factors such as the extension of animal models to humans, the limited number of large longitudinal datasets of child development, the differences in variables that have been collected, and questions surrounding generalizability between study populations can limit the interpretation. Within this framework, one additional approach to assess brain/behavior relationships is through the creation of simulated datasets that offers the opportunity to test data generated under different assumptions and to model which approach best reflects real-world data.

Simulated datasets have, by definition, a known ‘ground truth’ and can be modeled to include both causal and predictive relationships. Even though simulated datasets have *a priori* assumptions that may or may not accurately reflect the actual brain/behavior relationships, they are an under-used approach to assess whether longitudinal models can accurately measure different longitudinal trajectories of the interplay between brain and behavior. Furthermore, when actual longitudinal datasets with multiple waves become available, the different variables that are entered into the models, including the characteristics of missing data, demographic features, and clinical measures can be tweaked until they reflect the actual data, and can then be tested to predict outcomes in independent datasets.

To enhance our knowledge of longitudinal modelling of large studies of child development, we recently created a consortium of five sites across three continents to independently simulate and share 15 large longitudinal datasets.³² Each of the five sites created three datasets, each containing 10,000 individuals. The age range of each dataset spanned the ages of 7 to 20 years with two-year intervals and a total of 7 waves per dataset. The instructions provided to each site was that they were to create the simulated datasets based on how they believe brain development, behavior, and the interplay between brain and behavior occurs over time.³² The brain measures consisted of five regions that are typically derived from structural MRI (total gray matter volume, total white matter volume, intracranial volume, and volumes of the hippocampus and amygdala). In addition, we also included frontal lobe cortical thickness. Measures of behavior consisted of three subscales from the Child Behavior Checklist (CBCL) (internalizing, externalizing, and attention problems scale) and a discrete measure of a diagnosis of autism spectrum disorder (ASD) (see Table 2). With the exception of the ASD diagnosis, brain and behavior measures were time varying, offering the opportunity for earlier datapoints to influence later datapoints, both within and between the brain and behavior domains.

The simulated datasets were independently created, without discussion between the groups regarding the assumptions built into their simulated datasets. Many of the studies modelled specific baseline characteristics from existing longitudinal studies, including the Adolescent Brain and Cognitive Development (ABCD) Study,¹⁰ the “Hjernens Udvikling hos Børn og Unge” (Brain Maturation in Children and Adolescents, HUBU) study,³³ or using brain growth

data combined from many existing studies.^{34,35} The data and code to generate the datasets are freely available at <https://github.com/SoCoDeN/Simulation>. An advantage of having multiple groups independently simulate large datasets is to limit inherent biases based on underlying assumptions within the different groups. If a single group both simulates datasets and subsequently uses these datasets to test different models, there is the possibility for circularity in the design and analyses. This is limited by multiple independent groups being blinded to how other groups create the simulated datasets. However, there is a balance to the *a priori* creation of boundary conditions for the simulated variables, versus allowing the groups to have flexibility in their modeling approach.

The primary goal of this paper is to assess the epidemiological, demographic, and brain and behavior trajectories of these 15 simulated datasets and to assess how well each dataset meshes with actual human-subjects data. In addition, we determined how each dataset dealt with missing variables. Missing data can be clustered into categories of missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR).³⁶ An additional goal was to generate and compare developmental trajectories of internalizing, externalizing, attention problems, and each of the six brain measures (Table 2) from 7 to 20 years-of-age. The neuroimaging datasets were compared to actual neuroimaging data published by the ENIGMA Lifespan data project^{34,35} to assess how well each simulated neuroimaging dataset reflected actual longitudinal trajectories. Finally, an additional goal was to assess sex-based differences in neurodevelopmental trajectories.

METHODS

SIMULATED SUBJECTS

The data for this study encompasses simulated demographic, behavioral, and neuroimaging variables from a multi-site study of brain development.³² In this study, five different groups across three continents independently simulated datasets based on their hypotheses of the longitudinal interplay between brain and behavior. Each group created three longitudinal datasets, with each dataset containing 10,000 individuals and seven waves of data beginning at age 7 years up to 20 years-of-age in two-year increments. The simulated datasets contained up to 20% missing data. In total, the dataset includes 150,000 individuals and seven waves of data and is freely available through the Open Science Foundation (OSF) via the following link: <https://osf.io/yjt9p/>. We analyzed each of the 15 datasets individually.

VARIABLES

All fifteen datasets contain the same select group of variables as described in Sadeghi et al.³² The demographic variables include age, sex and parental education, with the latter having four categories (Table 2). Cognitive and behavioral measures include IQ, a binary classification of

Table 2. Description of the Variables

Demographic Variables		Units	Range	Time Varying
Age	Continuous	months	7-20	Yes
Sex	Dichotomous	Male / Female		No
Parental Education	Categorical	< 12 th grade High School (HS) Bachelors (BS/BA) Master's Degree +		No
Behavioral Measures				
CBCL – Internalizing	Continuous	Raw Scores	0 - 64	Yes
CBCL – Externalizing	Continuous	Raw Scores	0 - 66	Yes
CBCL – Attention Problems	Dichotomous	Raw Scores	0 - 20	Yes
Autism Spectrum Disorder	Dichotomous	Present / Absent		No
Cognitive Measure			Distribution	
IQ	Continuous		Normal $\mu=100$ $\sigma=15$	No
Structural MRI Measures				
Total Gray Matter Volume	Continuous	mm ³	Normal	Yes
Total White Matter Volume	Continuous	mm ³	Normal	Yes
Hippocampal Volume	Continuous	mm ³	Normal	Yes
Amygdala Volume	Continuous	mm ³	Normal	Yes
Intracranial Volume	Continuous	mm ³	Normal	Yes
Cortical Thickness	Continuous	mm	Normal	Yes

autism spectrum disorder (ASD), and three measures from the child behavior checklist (CBCL),^{37,38} namely raw scores for the internalizing, externalizing, and attention problems subscales. While the CBCL has been designed for ages 4 to 18 years of age,³⁷ The CBCL was also used for the seventh wave (19 to 20 years-of-age) to enhance continuity. Extending the age limits slightly around transition zones has been used in other studies to enhance continuity.³⁹ Six neuroimaging measures were defined, including the global measures of intracranial volume (ICV), total gray matter volume (GM_{tot}), total white matter volume (WM_{tot}) and subcortical regions, including the hippocampus and amygdala. Finally, longitudinal measures of cortical thickness of the frontal lobe was provided for each individual. A description of the demographic variables, percent missing data, and the prevalence of ASD for each dataset are shown in [Table 3](#).

STATISTICAL ANALYSES

Missingness: Although some of the groups provided information regarding how missingness was modeled within each of the datasets, we utilized a consistent approach for all datasets. This involved first assessing the total number and pattern of missing data; using Little's MCAR test⁴⁰ to assess whether the datasets fitted a pattern of missing completely at random (MCAR); and for those datasets which did not fit the pattern of MCAR, further assessments were performed to see if the data was missing at random (MAR) versus missing not at random (MNAR). These further as-

sessments were performed by using multiple linear regression to assess whether missingness was associated with sex, parental education, CBCL score, using the highest CBCL externalizing score for each individual across all seven waves. In addition, assessment of attrition or dropouts were evaluated. In the Supplementary Figures 1 through 9, four of the subfigures (B,C,E, and F) were created using the python package 'missingno' (<https://360digitmg.com/blog/missingno>).

Behavioral and Neurodevelopmental Trajectories: Longitudinal trajectories for the three continuous behavioral measures (CBCL internalizing, externalizing, and attention problems scores) and the six brain measures were created using R version 4.3.1 ('Beagle Scouts'). Linear mixed models were performed using the python package 'mixedlm' with subject ID as the random variable and sex, educational status, and IQ as fixed variables. A sex by age interaction model was also included. Finally, to illustrate the trajectories of sex differences, a sliding window analyses was performed across the age spectrum assessing effect size measures of the behavior and brain measures between males and females. The sliding window analyses had one month incremental steps and averaged over a window of ± 12 months.

To assess how the datasets meshed with actual human-subjects MRI data, each of the datasets were compared with the ENIGMA Lifespan dataset.^{34,35} Several steps were followed in order to perform this comparison. First, the simulated datasets were restructured to a format compatible with the BrainChart App, which is a web based shiny app

Table 3. Overview of the Demographic and Clinical Characteristics of the Sample. Measures included are the sex distribution, educational attainment, percent of participants who refused to respond or had missing data, intelligence quotient, and the percentage of those with a diagnosis of autism spectrum disorder.

Dataset	% Male		Education Percentile				IQ		% Dx ASD (m / f)
		<12 th	HS	BS/BA	M+	Re	%miss	(mean/SD)	
dpn_data1	52.6	5.5	35.4	24.8	34.3	0	0	100.1 (15.1)	0.4 / 1.6
dpn_data2	50.7	5.2	35.6	25.3	33.9	0	0	100.1 (14.9)	0.4 / 1.6
dpn_data3	50.6	5.3	35.8	24.8	34.1	0	0	100.0 (15.0)	0.4 / 1.6
leer_data1	49.7	4.7	30.2	23.0	30.9	1.9	9.2	99.9 (15.2)	1.3 / 4.9
leer_data2	49.9	4.8	30.9	23.5	29.4	1.9	9.5	99.8 (14.8)	1.5 / 4.8
leer_data3	50.4	5.1	29.8	23.0	30.1	2.0	10.0	100.1 (14.9)	1.3 / 5.0
OSA_data1	50.5	4.2	28.0	20.2	27.1	7.9	19.4	102.6 (12.8)	3.7 / 0.1
OSA_data2	50.5	4.2	28.0	20.2	27.3	7.9	19.5	103.9 (12.8)	4.0 / 0.2
OSA_data3	50.5	4.2	28.0	20.2	27.3	7.9	19.5	105.2 (12.9)	4.4 / 0.1
paint_data1	52.4	4.9	35.7	25.2	34.2	0	0	100.0 (15.6)	3.1 / 2.7
paint_data2	52.2	4.8	35.4	25.5	34.2	0	0	100.0 (15.6)	2.9 / 3.3
paint_data3	52.2	4.9	35.6	25.3	34.2	0	0	100.0 (15.6)	2.8 / 2.7
site4802_data1	50.2	5.3	35.3	23.8	34.8	0	0.86	100.1 (15.0)	50.1 / 49.7
site4802_data2	50.2	5.4	35.6	24.0	35.0	0	0	100.1 (15.0)	50.1 / 49.7
site4802_data3	50.2	5.4	35.6	24.0	35.0	0	0	100.1 (15.0)	50.1 / 49.7

The names of the datasets are the same as those that have been uploaded and are freely available via the Open Science Foundation Website [link: <https://osf.io/yjt9p/>]. A description for each of the **abbreviations** are as follows: '<12th' refers to education up to but not including a high school degree, or equivalent in other countries; 'HS' refers to a high school (or equivalent) diploma as the highest degree attained; 'BS/BA' refers to bachelors degree as the highest degree attained; 'M+' refers to a Masters Degree or higher. 'Re' refers to refused to answer; '%miss' refers to the percent of missing variables from each of the datasets; 'IQ' refers to intelligence quotient; and '% Dx ASD' is the percent of children with a diagnosis of autism spectrum disorder (ASD) split in the percentage within males (m) and the percentage within females (f).

(<https://brainchart.shinyapps.io/brainchart/>). Second, the simulation data for each site was uploaded to the BrainChart App. Since BrainChart has data from both clinical and population-based studies, we compared our data in the BrainChart App with typically developing or population-based studies or participants. The App calculates an age-normalized percentage for each individual. Third, the age-normalized percentages were downloaded for each individual from the BrainChart site. Fourth, we performed a z-transform on each individuals percentile score and calculated the median value and converted this median into a z-score. We then calculated the absolute value of the z-score. This score, one for each dataset, reflects the median distance away from the 50th percentile line from the ENIGMA lifespan trajectory. A score of zero would reflect identical medians across all ages for the simulated and the ENIGMA dataset. A score of 1.0 would reflect that over half of the age-related datapoints have a standard deviation of 1.0 away from the ENIGMA dataset. Note, this does not re-

flect the direction of the deviation away from the ENIGMA lifespan dataset.

There were a total number of 270 statistical tests performed (18 per dataset), including main effects for sex and interaction effects. Multiple testing correction was performed for all 270 tests using both the conservative Bonferroni approach as well as the false discovery rate (FDR) using the Benjamini-Hochberg⁴¹ method. All analyses were performed using either python version 3.11.4 or R version 4.3.1 ('Beagle Scouts') and the code to perform all analyses are freely available on GitHub, <https://github.com/SoCo-DeN/Simulation>.

RESULTS

Based on the primary aims of the study, we evaluated demographic and clinical characteristics, missing data, the distributions of the CBCL domains, and how well the neu-

roimaging trajectories meshed with actual neuroimaging data.

DEMOGRAPHIC AND CLINICAL CHARACTERISTICS

The characteristics for sex, parental education, and IQ are shown in [Table 3](#). In most clinical populations, diagnoses of ASD are three to four times higher in boys compared to girls.⁴² Two of the sites reflected this ratio; the male to female ratio from the three datasets at the dpn site ranged from 3.3:1 to 4.2:1. The male to female ratio from the three datasets at the OSA site was 4.6:1. Two sites, paint and site4802 had male to female ratios of ASD that ranged between 1:1 and 1.2:1. The leer site was oversampled for females with autism, having ratios ranging from 1:30 to 1:50 males to females. Finally, site4802 had three datasets with nearly equal numbers of males and females with ASD ([Table 3](#)). Thus, all three datasets from site4802 are clinic datasets rather than population-based datasets, which can have a dramatic influence on the trajectories.^{6,13}

MISSINGNESS

The frequency and type of missingness for each dataset is presented in Supplemental Table 1. In addition, the characteristics of missing for each dataset are provided in Supplementary Figures 1-9. Except for one dataset (paint_data3), the amount of missing data for both the CBCL and the neuroimaging data remained relatively constant over the nine data collection waves (Supplementary Figure 10). Six of the datasets had no missing data, three datasets had data that were missing completely at random; six of the datasets failed Little's MCAR test and were thus either MAR or MNAR. Linear regression analyses comparing individuals with missing data based on sex and parental education resulted in no significant associations (Supplemental Table 1). Two datasets (OSA_data2 and OSA_data3) had data that were considered MAR, since higher rates of missingness were significantly associated with greater externalizing ($\beta = -0.032$, $t = -17.8$, $p < 0.0001$), internalizing ($\beta = -0.033$, $t = -17.2$, $p < 0.0001$), and attention problem scores ($\beta = -0.09$, $t = -17.5$, $p < 0.0001$). The other three datasets were labeled MNAR (Supplemental Table 1). There were few complete dropouts until the 5th wave, when there was increasing numbers of attrition in the datasets containing missing data (Supplementary Table 2).

Supplementary Figures 1 through 9 provide information regarding missing data and dropouts for each of the nine datasets with missing data. Each of these 9 Figures contain five subfigures that provide an overview of the characteristics of missing data for that specific dataset. The first subfigure (A) in each figure is a histogram providing the number of missing data per individual. The second subfigure (B) demonstrates the number of missing data per variable. Subfigure C shows a correlational heatmap indicating the relationship between missing variables; Subfigure D provides a matrix with participants and waves along the y-axis and the variables along the x-axis, which demonstrates patterns of missingness. Finally, subfigure E is a dendrogram

demonstrating the hierarchy of the missing values for each dataset.

BEHAVIORAL TRAJECTORIES AND SEX DIFFERENCES

The CBCL measures for internalizing, externalizing, and attention problems demonstrated different developmental trajectories and distributions ([Figure 1](#)). The internalizing scores either remained relatively constant over the age range from 7 to 20 years-of-age or consisted of inverted U-shaped trajectories. The CBCL externalizing and attention problems scores showed a different pattern in contrast to the internalizing domain, being either relatively constant or showing on-average decreasing problems over time.

Results of the linear mixed models for main and interaction effects of behavioral measures with sex for each dataset are shown in Tables 4 and 5, respectively. Most datasets modeled females as having significantly greater internalizing symptoms (dpn, leer, and paint datasets 2 & 3), whereas the other showed either no difference (OSA datasets 2 & 3 and Paint dataset 1) or greater internalizing symptoms in males (site4802 and OSA dataset 1). The datasets from two sites, paint and site4802 modeled sex by age interactions with internalizing symptoms, with all but paint_data1 showing females having greater internalizing symptoms with age (Table 5).

Externalizing symptoms were significantly greater in males for the dpn and site4802 datasets, as well as paint_data2. Leer had greater externalizing symptoms in females, and the remaining (OSA datasets 2 & 3 and paint datasets 1 & 3) showed no sex differences (Table 4). The interaction effects were more subtle, with greater symptoms over time in males in the leer datasets, and less in the dpn datasets, and in the paint_data3 and site4802_data1 datasets. Finally, main effects related to the attention problems score tended to parallel the externalizing symptoms scores (Table 5). The interaction effects of sex with age show no effects at the dpn site and OSA data 1 & 2, increasing attention problems in males at with the leer and paint_data1 datasets, and decreasing over time in paint_data2 and data3 and site4802 1 & 3 (Table 5).

BRAIN GROWTH TRAJECTORIES AND SEX DIFFERENCES

Different model assumptions were used to generate the neurodevelopmental trajectories for behavior over time ([Figure 2](#)). The most consistent pattern across all datasets was seen in total white matter volume, which demonstrated non-linear increases with several datasets reaching a plateau before the age of 20 years ([Figure 2-B](#)). Total gray matter volume ([Figure 2-A](#)) and frontal lobe cortical thickness ([Figure 2-E](#)) decreased with age in most datasets, although gray matter volume also showed inverted-U and non-linear increases. In most datasets, volumes of the two subcortical structures, the hippocampus and amygdala increased to approximately 17 years, whereafter they show differing levels of decreasing volume ([Figure 2-C & D](#)).

The fit metrics for gray matter, white matter, and cortical thickness for each of the datasets are presented in Table 6 and demonstrated visually per year in [Figure 3](#). Three sites

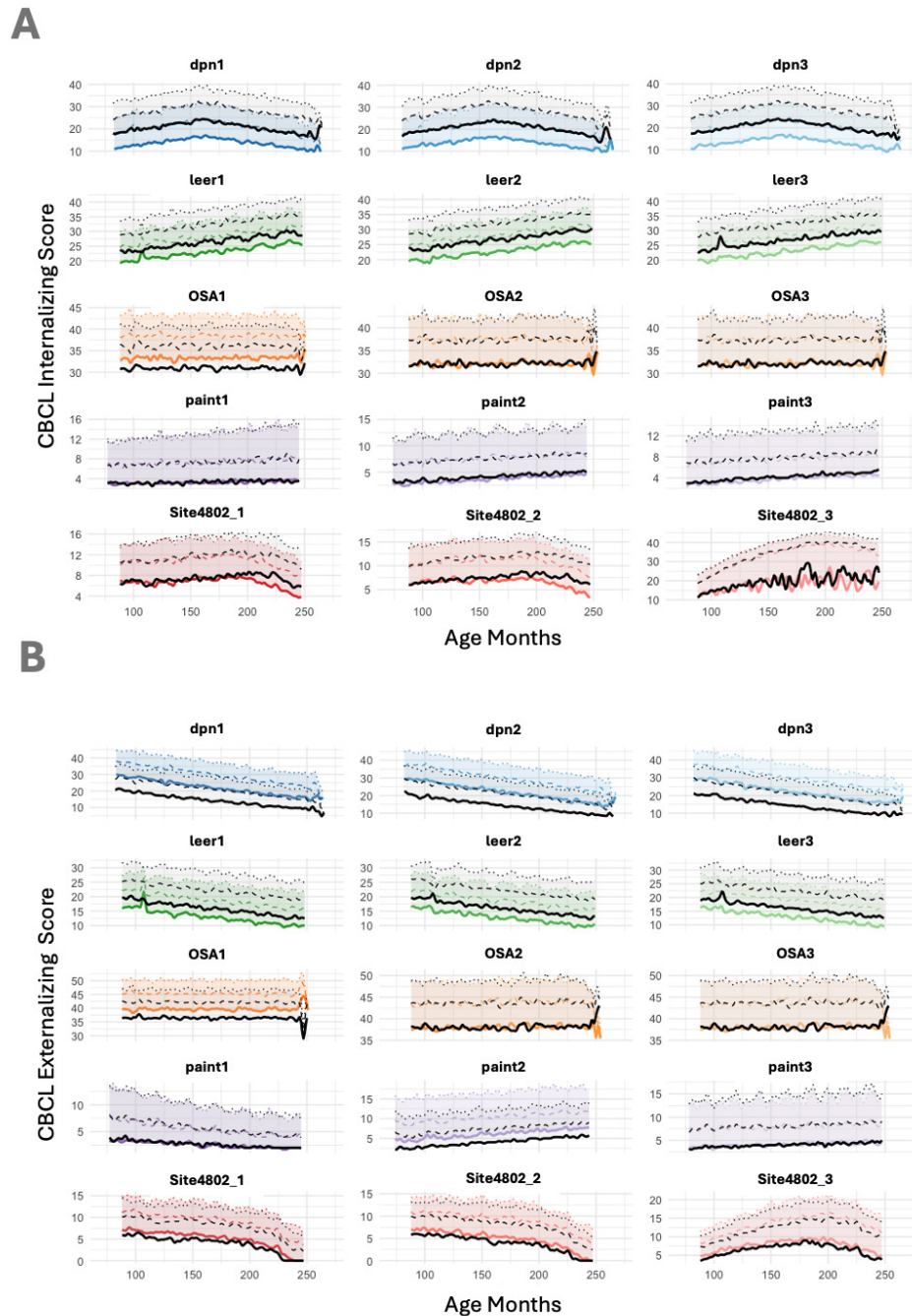


Figure 1. Trajectories of the mean CBCL raw scores for each of the 15 datasets, with (A-1) median internalizing symptoms by age; (A-2) Distribution of CBCL internalizing scores within each dataset; (B-1) median externalizing symptoms by age; (B-2) Distribution of CBCL externalizing scores within each dataset; (C-1) median attention problem symptom by age; (B-2) Distribution of CBCL attention problem scores within each dataset. Males have colored line with light-colored ribbons. Females have black lines with gray ribbons.

(dpn, paint, and site4802) showed excellent correspondence to the lifespan growth curves. The other two sites, while having reasonably good fit metrics (Table 6), demonstrated non-linear fits at the lower and upper ages.

Nearly all datasets had highly significant main and interaction effects with sex (Table 4). Regions that were significantly larger in males at all sites included the ICV, hippocampus, and the amygdala (see Supplemental Table 3 for main effects and Supplemental Table 4 for age/sex interac-

tion effects). Gray and white matter volumes were larger in males in all sites except the OSA site (Supplemental Table 3 & Figure 3-A). One approach to assess longitudinal trajectories of sex differences is comparing effect size differences of brain regions across age.⁴³ This approach was performed for all imaging features in each of the datasets and are shown in Supplemental Figure 11. Within the OSA site, females had significantly larger volumes in datasets 1 & 3 and no difference in dataset 2 (Supplemental Figure

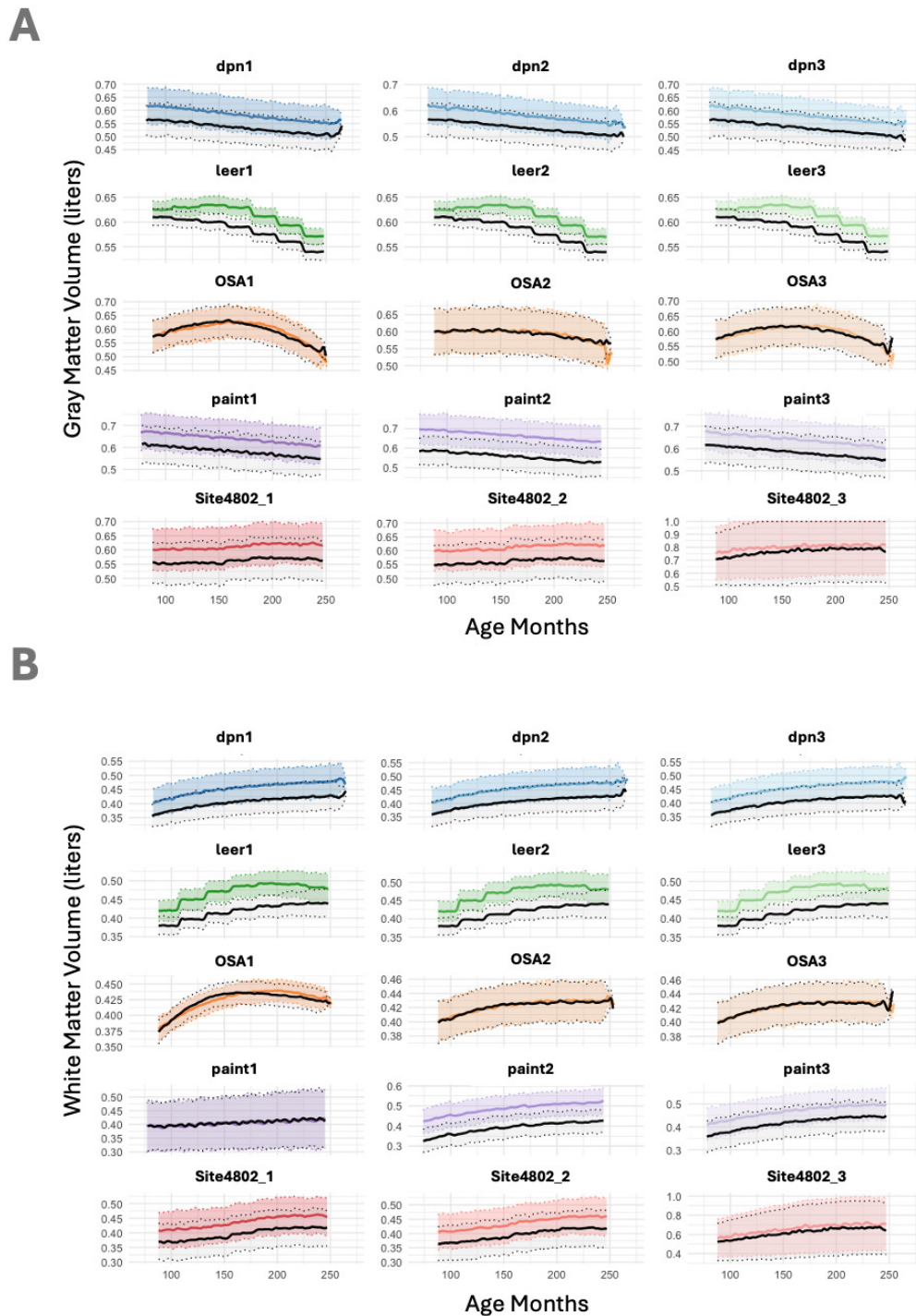


Figure 2. Trajectories of the mean volume or cortical thickness measures for each of the 15 datasets. The measures include the mean volume in mm^3 for each wave of the (A) total gray matter (B) total white matter; (C) hippocampus; (D) Amygdala; and (E) intracranial volume. (F) reflects the frontal lobe cortical thickness (mm). Males have colored line with light-colored ribbons. Females have black lines with gray ribbons.

11). The greatest main effects for sex were seen in frontal lobe cortical thickness measures, with no sex differences in either the dpn site and paint datasets 1 and 3; smaller cortical thickness in females in the OSA datasets, and larger cortical thickness in females in site4802 and paint 2 datasets (Supplementary Figure 11 & Supplemental Table 3).

DISCUSSION

Considering that the simulation data contained only six structural neuroimaging measures, four behavioral measures, age, sex, parental education, and IQ, these 14 variables are orders of magnitude smaller than the vast number of variables collected in typical longitudinal population-

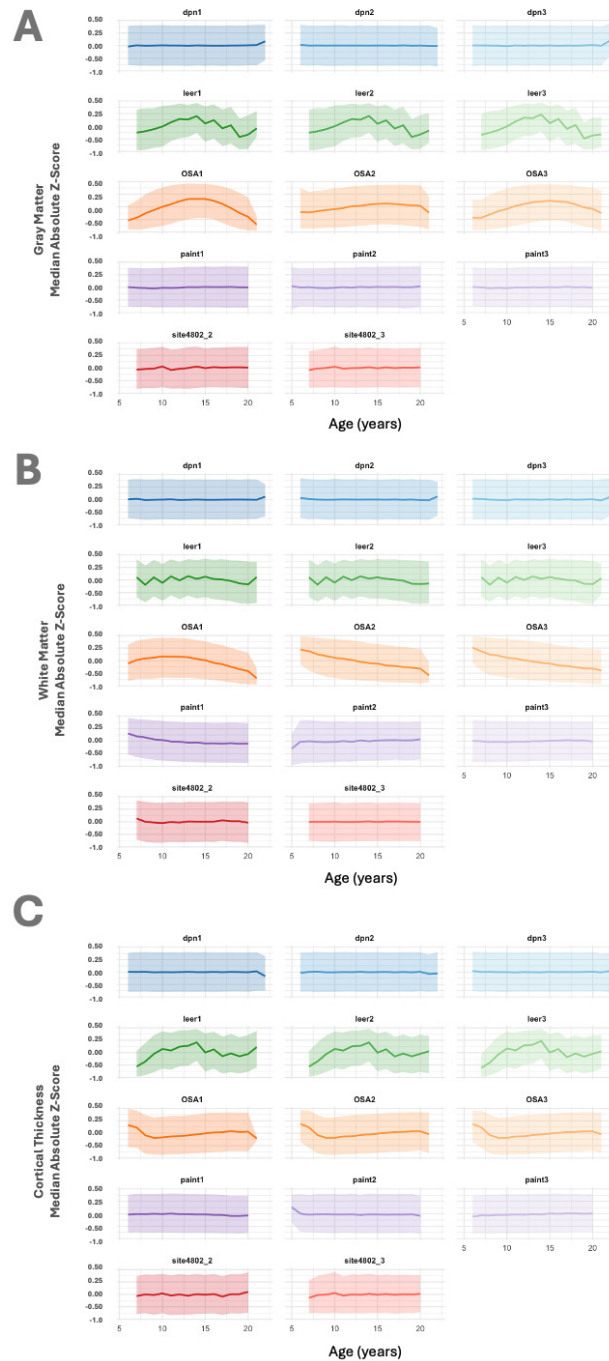


Figure 3. Comparison of Calculated Centiles for each Dataset with the ENIGMA Lifespan Data. Centiles were calculated for each dataset and all datapoints using BrainChart, a web based shiny app (<https://brainchart.shinyapps.io/brainchart/>). (A-1) Age-related centiles for gray matter volume for each dataset; (A-2) Depiction of the age-related best-fit for gray matter volume for each of the datasets. (B-1) Age-related centiles for white matter volume for each dataset; (B-2) Depiction of the age-related best-fit for gray matter volume for each of the datasets. (C-1) Age-related centiles for cortical thickness for each dataset; (C-2) Depiction of the age-related best-fit for cortical thickness for each of the datasets. **Note:** one dataset (site4802 dataset 1) yielded no results using the BrainChart App and thus results are not presented.

based studies.^{10,13} Yet even with this small number of variables, we found both similarities and differences in how groups modeled the brain and behavior trajectories. These differences translated into mixed results between studies, such as the differences for main and interaction effects of sex with brain and behavior. In addition, each group ap-

plied different models for missing data and utilized different distributions for the behavioral measures. Understanding how the different groups modeled the data can help determine the underlying assumptions related to neurodevelopmental processes, especially since these assumptions likely will influence how groups analyze real-world data.

NEUROIMAGING MEASURES

While there was also some variability in the trajectories of brain morphology, many of the curves did approximate those found in the literature.^{34,44,45} Cerebral gray matter volumes typically demonstrate an inverted 'u' shape curve, with a gradual increase in volume up to early adolescence, followed by a gradual decrease.^{34,46} Two of the groups (leer & OSA) modeled trajectories with a u-shaped curve, whereas the other three tended to have a linear decline (Figure 2-A). Studies have shown that during childhood and adolescence, total white matter volume tends to demonstrate a gradual increase and hippocampal and amygdala volumes tend to increase into adolescence and reach a plateau.^{22,34} These patterns were overall reflected in the simulated datasets (Figure 2-B,C,D). Frontal lobe cortical thickness demonstrated a gradual decrease with age in all datasets (Figure 2-F), which is consistent with the literature.^{45,47} Finally, many datasets showed trajectories that were very consistent with the ENIGMA Lifespan trajectories, even at sites that did not model their data using the lifespan data. The dataset that most diverged from the ENIGMA lifespan study was the case/control dataset with 50% of the population having ASD. Not surprisingly, datasets that modeled their curves using data that overlapped with the ENIGMA consortium datasets tended to show the best fit. One approach to create greater harmonization in simulated datasets would be for the different sites to agree a priori not only on the actual datasets used for the simulations, but also to have consensus regarding the study population (i.e., population-based versus case/control designs), and the prevalence of clinical diagnoses.

The sex-based differences in volumes did show considerable variability between the different datasets (Supplementary Figure 11). Longitudinal measures of effect sizes between males and females between 7 and 20 years-of-age were relatively constant in nine of the datasets, with males showing greater volumes and effect sizes varying from small to very large between the studies. While both constant and age-dependent differences in brain volumes between males and females have been reported,^{35,48-50} a recent meta-analysis found that global effect size differences to be on the order of 2.1.⁵⁰ Six datasets showed non-linear changes of effect sizes, with the leer datasets, showing inverted-u shaped curves for effect sizes for both total gray and total white matter volumes. These measures showed an increase in volumes in boys up to mid-adolescence, followed by a decline. The hippocampus showed an increase up to 180 months after which there was a plateau (Supplementary Figure 11-C). The amygdala volume was considerably larger in boys up to 140 months, where it decreased thereafter. Frontal lobe cortical thickness was not different across the age range in 12 of the 15 datasets, with the leer datasets showing females have a greater than one standard deviation larger cortical thickness (Supplementary Figure 11-F).

While numerous neuroimaging studies have captured the changes and differences in those who develop psychiatric disorders, there remains a gap related to the pre-

illness characteristics of the brain. Behavior is inherently driven by the brain, but using our existing neuroimaging approaches does not immediately translate into our ability to detect the underlying neural processes and the location of these processes using our existing neuroimaging tools. For example, evidence from several studies has shown that whereas cross-sectional baseline brain/behavior relationships were not present, it was the presence of the behavior at baseline that predicted downstream brain changes, rather than the opposite.^{8,9,28,51} Thus, the presence or persistence of certain behaviors at an earlier timepoint, even those that are problematic, could result in downstream brain changes. These changes could be due to Hebbian processes or other factors inherent in neuroplasticity. It's not that the brain-based differences at baseline were not present, but rather, they were not able to be detected using the current neuroimaging technology.

BEHAVIORAL MEASURES

Trajectories of the mean levels of internalizing, externalizing, and attention problem behavior and the distribution of scores for each of the datasets are shown in Figure 1. These curves demonstrate the age-related median trajectories for each dataset. Studies have shown that internalizing symptom scores tend to be low to moderate leading into mid-adolescence, followed by an increase during mid-to-late adolescence, especially in girls, with a decline or plateau in late adolescence.^{37,52-55} We found this pattern to be reflected in several datasets (Figure 1-A₁). Externalizing symptoms can be high during childhood in boys, followed by a peak in mid-adolescence for a subset of children, with an overall decline following childhood.^{37,54} Most of the datasets in our study reflected a gradual increase in internalizing symptoms with age and a gradual decline in externalizing and attention problems symptoms. Studies have found that attention problems tend to show an overall slow decline into late-adolescence in most, but not all children.⁵⁶ Nine of the datasets showed patterns consistent with this trajectory (Figure 1-C₁). While several datasets matched the overall mean pattern of behavioral traits, very few datasets modelled what would be considered expected sex differences in behavioral trajectories (Supplemental Figure 11 & Table 4). Finally, the CBCL tends to be highly skewed in population-based studies, however, only two groups (paint & site4802) modeled the distributions to fit a general pediatric population-based sample.

MISSING DATA

Missing data in longitudinal studies is common³⁶ and in fact, a complete dataset is extremely rare, and nearly impossible in large longitudinal studies. Types of missingness have been classified into three different categories: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR).^{36,57} One advantage of simulating large datasets is the ability to titrate the type and amount of missing data to test different underlying assumptions. In our study, six of the simulated datasets had no missing data, and thus these datasets could be used

to explore brain/behavior relationships without potential bias due to missing data. Little's test⁴⁰ categorized three datasets to be MCAR. Of the remaining datasets, two were defined to be MAR and four were MNAR (Supplemental Table 1). When comparing our findings to the description of how missingness was modeled in these datasets from Sadeghi et al.,³² the site4802_data1 dataset accurately matched the description of being generated completely at random (Supplemental Figure 9). However, three of the leer datasets were described as MAR, but were found using standard approaches to be MNAR. While Little's test is commonly used to assess for MCAR, the test has been described as being both underpowered to detect differences for small numbers of variables and when applied to variables have a lower covariance structure,^{58,59} which may have mislabeled these datasets. Since approaches to account for missing data, i.e., imputation, do not greatly influence the primary predictors of interest,⁵⁸ we accounted for missing data to assess the differences in trajectories between sex using and behavior and brain trajectories listwise deletion.

Supplementary Figures 1 through 9 provide a graphic overview for the nine datasets with missing data and could be used to assess longitudinal population-based studies. When a complete wave for a subject is missing, for example when a participant does not show up for the visit, then all behavior and imaging data will be missing for that wave. This would result in high correlations within and between all imaging and behavioral variables in the correlation heatmap of missing, horizontal lines in the matrix pattern of missingness, and a collapsed dendrogram (Supplementary Figures 1-3). When there is a combination of both missing waves and participants deciding not to take part in the neuroimaging component, there is a drop in the correlation between the brain and behavior measures, less horizontal structure in the matrix pattern of missingness, and clusters within behavior and imaging in the dendrogram (Supplementary Figures 4-6). Finally, data missing completely at random results in zero correlation in the heatmap of missingness, random structure of the matrix pattern of missingness, and little clustering within the dendrogram of missing values (Supplementary Figure 9).

STRENGTHS AND LIMITATIONS

While the considerable variability on some measures between the groups is a weakness of the study, this could also be considered a strength. One of the challenges of simulation is circularity, meaning the same site that creates a simulated dataset uses a specific statistical framework and subsequently the same model to test the predefined assumptions. Having five different sites generating data independently from the other sites offers the opportunity to test different assumptions related to the data models used and reduces the chance for circularity. However, it also increases variability between the datasets, which is what we found. With 15 simulated datasets, there are a vast array of relationships that could be assessed, which is one of the challenges. Approaches using multiverse-model interactions to simulate data, or data driven approaches such as

deep learning models that try to learn the underlying structure of the data (e.g., Neural ODEs) could provide evidence for the number of false positives and the power necessary to achieve robust findings across different models.

An additional weakness of our study is that we assessed mean behavioral problems and did not parse clinical or sub-clinical scales within the internalizing, externalizing, and attention problem scores. Longitudinal analyses of behavior changes are typically much more extensive. For example, clustering algorithms, such as latent profile analyses or latent transition analyses can parse individuals into different groups and to assess changes within these groups, respectively.⁶⁰ Using these approaches, most children are clustered into a 'no problems' group, whereas smaller groups of children fall into a problem behavior group, which is often also associated with cognitive deficits.^{61,62} However, as an example of the challenge, single publications have been written about sex effects of just the internalizing scale in one study with much smaller sample sizes and fewer waves.

Real-life studies have multiple variables that were not included with the simulation data. First, in real-life studies, harmonizing the different demographic, neuroimaging, cognitive, and behavioral data is not always straightforward. Many of these variables found in other studies were not included as additional variables in our simulated dataset. Second, many of these datasets have limited waves of data collection, different sampling techniques (i.e., longer periods between follow-up periods), and different study designs. Third, there can be varying levels of data quality and, especially in children, neuroimaging data is inherently noisy. Fourth, the simulation dataset did not specify a specific source for generating the data (i.e., FreeSurfer, FSL, SPM, etc.), which could infuse variability between the datasets. Fifth, whereas four of the sites assumed a population-based study, one site simulated a case/control study, making comparisons with this study more difficult. This may be the reason why this site, site4802 was least similar to the ENIGMA lifespan trajectories. Finally, whereas some large longitudinal neuroimaging studies have data sharing policies that easily allow researchers access to the data,^{10, 18,23} whereas other datasets are more restrictive.⁶³ Approaches to overcome these challenges is important, since high-quality replication studies are one key element to foster convergence on 'ground truth' relationships between brain and behavior. In parallel to acquiring and harmonizing longitudinal data, simulated datasets can be very valuable to lay the foundation for the selection of the best models to determine specific relationships are present between different variables.

FUTURE DIRECTIONS & RECOMMENDATIONS

There are a number of possibilities for future directions that can be taken to utilize simulated datasets to augment our understanding of neurodevelopmental trajectories, especially related to the interaction between brain and behavior over time. First, simulations performed iteratively, with each subsequent version tested as new waves of data

become available from ongoing studies, such as the ABCD Study, would offer the opportunity to adapt the simulated models such that they can better reflect the actual data. This also offers the opportunity for researchers to evaluate their biases and assumptions, which are embedded into the simulated models which may or may not mesh with existing data. Having the independent groups meet to discuss their assumptions as a part of the iterative feedback loop could be very valuable. Meeting would be especially important to help determine the underlying assumptions within and between the groups, which can be both implicit and explicit assumptions.

While site-related differences in behavioral and neuroimaging trajectories support differences in the underlying assumptions, it does not necessarily extract exactly what those assumptions are, especially implicit assumptions. Thus, discussions among the group members regarding theories in modeling neurodevelopmental data can be beneficial to better understand the differing assumptions.⁶⁴ To best understand the different assumptions, an iterative approach in developing simulation models from independent sites is recommended. Such an iterative approach would involve the creation of initial models, which is then followed by meetings and discussions among team members at the different sites to discuss the underlying assumptions and to then test those models that most represent the actual data. Additional variables could also be added to assess how the addition of these variables alters the predictive qualities of the models. One example is that the 3rd and part of the 4th wave of the ABCD Study has recently been released and an interesting question is whether simulated models, including implementing machine and deep learning algorithms, could be used predict subsequent waves.

One important future direction is to test different longitudinal statistical models using the simulated datasets to study longitudinal brain/behavior relationships of neurodevelopment.²⁷ Simulated datasets offer the opportunity to modify or eliminate variables to assess changes to effect size estimates based on these changes. Since each of the longitudinal models have specific assumptions related to the underlying structure of the data, simulated datasets that are based on the neuroscience literature provide a controlled structure to help define and test different assumptions, such as the number of recruitment waves necessary to demonstrate changes using different longitudinal modeling approaches or if different longitudinal models should be combined in order to assess different elements of the brain/behavior interplay. The site differences identified in our study reflect different assumptions related to neurodevelopmental trajectories. Whereas our study focused on the realism of independently generated neurodevelopmental growth trajectories, we did not perform formal statistical analyses to test causal or predictive models of brain/behavior relationships. It is optimal to first demonstrate realistic simulations of neurodevelopmental trajectories prior to testing causal models. Further, while we compared the growth curves to large-scale ENIGMA lifespan trajectories, direct comparisons with other studies (i.e., the ABCD¹⁰ or Generation R Study¹³) may have slightly different results.

Finally, simulation studies can help quantify the number of false positives and negatives, and to assess the influence of missing data.

Improvements in algorithms that assess and predict missingness, including the relationship between missing data and demographic and clinical variables would also be beneficial. Six datasets in our study had no missing data and thus missing data could be added to these datasets to assess how they alter the outcome. While having no missing data is beneficial to test models and is easily done in simulated datasets, it does not mesh well with actual longitudinal studies of child development. The limit of 20% missing data was likely an optimistic value in relation to actual longitudinal studies.⁵⁸ That said some studies have very low attrition rates in some countries whereas other studies tend to have larger rates of attrition and it would be helpful to compare different studies on the rates of missingness and further to test how the missingness can interfere with outcome measures.

Finally testing and adapting to best approximate real data will provide better insight into the interplay between brain and behavior over time. The ultimate 'ground truth' test would be to alter some aspects of the temporal relationship within specific variables within the longitudinal simulated model while subsequently performing an interventional study in humans to show that the model can predict the outcome in actual human subjects studies.

CONCLUSIONS

The simulation of longitudinal datasets that embed assumptions regarding the interplay between brain and behavior is a valuable but underused approach to better understand neurodevelopment. It can help parse the incredible complexity inherent within neurodevelopment, although even with limited variables, it is humbling to consider the level of complexity that must be navigated. Consider additional variables in developmental studies that were not included in this study, such as adverse life experiences, parenting characteristics, school and peer relationships, bullying, and a myriad of additional behavioral and emotional problems and genetic susceptibilities. Simulation has many advantages, even for datasets with limited variables. Not only do simulated datasets provide the ground truth relationships between brain and behavior, where the causal relationship is known, they also offer the opportunity to test specific statistical models to determine best suited to identify specific relationships. We found that it is challenging for independent groups to create simulated datasets that have considerable overlap, even with a small number of variables. An optimal approach would first involve groups independently creating datasets, as was done in this study, but followed by the groups meeting to discuss the assumptions they used in modeling the data. This would be followed by an iterative approach with closed loop feedback, until the simulated data approximates real-world data. Finally, once simulated data approximates real-world data, variables in the simulated datasets could be altered to assess how downstream brain and behavior is changed. Us-

ing this information, it may then be possible to predict and test interventions.

Submitted: December 22, 2025 CDT. Revised: March 09, 2026 CDT. Accepted: May 12, 2026 CDT. Published: August 27, 2026 CDT.

.....

ACKNOWLEDGEMENTS

This project was supported by the Intramural Research Program of the National Institute of Mental Health (ZIA MH002986).



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

REFERENCES

1. Casey BJ, Getz S, Galvan A. The adolescent brain. *Dev Rev.* 2008;28(1):62-77. doi:[10.1016/j.dr.2007.08.003](https://doi.org/10.1016/j.dr.2007.08.003). PMID:18688292
2. Solmi M et al. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. *Mol Psychiatry.* 2022;27(1):281-295. doi:[10.1038/s41380-021-01161-7](https://doi.org/10.1038/s41380-021-01161-7). PMID:34079068
3. Kessler RC et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry.* 2007;6(3):168-176.
4. Casey BJ, Jones RM. Neurobiology of the adolescent brain and behavior: implications for substance use disorders. *J Am Acad Child Adolesc Psychiatry.* 2010;49(12):1189-1201. doi:[10.1016/j.jaac.2010.08.017](https://doi.org/10.1016/j.jaac.2010.08.017). PMID:21093769
5. Malter Cohen M, Tottenham N, Casey BJ. Translational developmental studies of stress on brain and behavior: implications for adolescent mental health and illness? *Neuroscience.* 2013;249:53-62.
6. White T et al. Pediatric population-based neuroimaging and the Generation R Study: the intersection of developmental neuroscience and epidemiology. *Eur J Epidemiol.* 2013;28(1):99-111. doi:[10.1007/s10654-013-9768-0](https://doi.org/10.1007/s10654-013-9768-0)
7. Satterthwaite TD et al. The Philadelphia Neurodevelopmental Cohort: A publicly available resource for the study of normal and abnormal brain development in youth. *Neuroimage.* 2016;124(Pt B):1115-1119. doi:[10.1016/j.neuroimage.2015.03.056](https://doi.org/10.1016/j.neuroimage.2015.03.056). PMID:25840117
8. Blok E et al. Stage 2 Registered Report: The Bidirectional Relationship Between Brain Features and the Dysregulation Profile: A Longitudinal, Multimodal Approach. *J Am Acad Child Adolesc Psychiatry.* 2023;62(12):1363-1375. doi:[10.1016/j.jaac.2023.03.024](https://doi.org/10.1016/j.jaac.2023.03.024). PMID:37339753
9. Muetzel RL et al. Tracking Brain Development and Dimensional Psychiatric Symptoms in Children: A Longitudinal Population-Based Neuroimaging Study. *Am J Psychiatry.* 2018;175(1):54-62. doi:[10.1176/appi.ajp.2017.16070813](https://doi.org/10.1176/appi.ajp.2017.16070813)
10. Casey BJ et al. The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Dev Cogn Neurosci.* 2018;32:43-54. doi:[10.1016/j.dcn.2018.03.001](https://doi.org/10.1016/j.dcn.2018.03.001). PMID:29567376
11. Di Martino A et al. Enhancing studies of the connectome in autism using the autism brain imaging data exchange II. *Sci Data.* 2017;4:170010.
12. Di Martino A 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. doi:[10.1038/mp.2013.78](https://doi.org/10.1038/mp.2013.78). PMID:23774715
13. White T et al. Paediatric population neuroimaging and the Generation R Study: the second wave. *Eur J Epidemiol.* 2018;33(1):99-125. doi:[10.1007/s10654-017-0319-y](https://doi.org/10.1007/s10654-017-0319-y). PMID:29064008
14. Broekman BF et al. Gestational age and neonatal brain microstructure in term born infants: a birth cohort study. *PLoS One.* 2014;9(12):e115229. doi:[10.1371/journal.pone.0115229](https://doi.org/10.1371/journal.pone.0115229). PMID:25535959
15. Goh SKY et al. Infant night sleep trajectory from age 3-24 months: evidence from the Singapore GUSTO study. *Sleep Med.* 2017;33:82-84. doi:[10.1016/j.sleep.2017.01.013](https://doi.org/10.1016/j.sleep.2017.01.013)
16. White T et al. Automated quality assessment of structural magnetic resonance images in children: Comparison with visual inspection and surface-based reconstruction. *Hum Brain Mapp.* 2018;39(3):1218-1231. doi:[10.1002/hbm.23911](https://doi.org/10.1002/hbm.23911). PMID:29206318
17. Dean DC 3rd, et al. Quantifying brain development in the HEALthy Brain and Child Development (HBCD) Study: The magnetic resonance imaging and spectroscopy protocol. *Dev Cogn Neurosci.* 2024;70:101452.
18. Alexander LM et al. An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Sci Data.* 2017;4:170181.
19. Mascarell Maricic L et al. The IMAGEN study: a decade of imaging genetics in adolescents. *Mol Psychiatry.* 2020;25(11):2648-2671. doi:[10.1038/s41380-020-0822-5](https://doi.org/10.1038/s41380-020-0822-5). PMID:32601453
20. Schumann G et al. The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Mol Psychiatry.* 2010;15(12):1128-1139. doi:[10.1038/mp.2010.4](https://doi.org/10.1038/mp.2010.4)
21. Backhausen LL et al. Adolescent to young adult longitudinal development of subcortical volumes in two European sites with four waves. *Hum Brain Mapp.* 2024;45(3):e26574. doi:[10.1002/hbm.26574](https://doi.org/10.1002/hbm.26574). PMID:38401132

22. Giedd JN et al. Child psychiatry branch of the National Institute of Mental Health longitudinal structural magnetic resonance imaging study of human brain development. *Neuropsychopharmacology*. 2015;40(1):43-49. doi:[10.1038/npp.2014.236](https://doi.org/10.1038/npp.2014.236). PMID:25195638
23. Nooner KB et al. The NKI-Rockland Sample: A Model for Accelerating the Pace of Discovery Science in Psychiatry. *Front Neurosci*. 2012;6:152.
24. Satterthwaite TD et al. Neuroimaging of the Philadelphia neurodevelopmental cohort. *Neuroimage*. 2014;86:544-553. doi:[10.1016/j.neuroimage.2013.07.064](https://doi.org/10.1016/j.neuroimage.2013.07.064). PMID:23921101
25. Jernigan TL et al. The Pediatric Imaging, Neurocognition, and Genetics (PING) Data Repository. *Neuroimage*. 2016;124(Pt B):1149-1154. doi:[10.1016/j.neuroimage.2015.04.057](https://doi.org/10.1016/j.neuroimage.2015.04.057). PMID:25937488
26. Paus T et al. Saguenay Youth Study: a multi-generational approach to studying virtual trajectories of the brain and cardio-metabolic health. *Dev Cogn Neurosci*. 2015;11:129-144. doi:[10.1016/j.dcn.2014.10.003](https://doi.org/10.1016/j.dcn.2014.10.003). PMID:25454417
27. Hawes SW et al. Longitudinal analysis of the ABCD(R) study. *Dev Cogn Neurosci*. 2025;72:101518.
28. Blok E et al. The Bidirectional Relationship Between Brain Features and the Dysregulation Profile: A Longitudinal, Multimodal Approach. *J Am Acad Child Adolesc Psychiatry*. 2022;61(6):830-831. doi:[10.1016/j.jaac.2022.03.008](https://doi.org/10.1016/j.jaac.2022.03.008)
29. Durkut M et al. The longitudinal bidirectional relationship between autistic traits and brain morphology from childhood to adolescence: a population-based cohort study. *Mol Autism*. 2022;13(1):31. doi:[10.1186/s13229-022-00504-7](https://doi.org/10.1186/s13229-022-00504-7). PMID:35790991
30. White T. Behavioral phenotypes, stochastic processes, entropy, evolution, and individual variability: Toward a unified field theory for neurodevelopment and psychopathology. *Aperture Neuro*. 2022;2:1-3.
31. Marek S et al. Reproducible brain-wide association studies require thousands of individuals. *Nature*. 2022;603(7902):654-660. doi:[10.1038/s41586-022-04492-9](https://doi.org/10.1038/s41586-022-04492-9). PMID:35296861
32. Sadeghi N et al. The interplay between brain and behavior during development: A multisite effort to generate and share simulated datasets. *Sci Data*. 2025;12(1):473. doi:[10.1038/s41597-025-04740-3](https://doi.org/10.1038/s41597-025-04740-3). PMID:40118942
33. Fuhrmann D et al. The midpoint of cortical thinning between late childhood and early adulthood differs between individuals and brain regions: Evidence from longitudinal modelling in a 12-wave neuroimaging sample. *Neuroimage*. 2022;261:119507.
34. Bethlehem RAI et al. Publisher Correction: Brain charts for the human lifespan. *Nature*. 2022;610(7931):E6.
35. Bethlehem RAI et al. Brain charts for the human lifespan. *Nature*. 2022;604(7906):525-533. doi:[10.1038/s41586-022-04554-y](https://doi.org/10.1038/s41586-022-04554-y). PMID:35388223
36. Ibrahim JG, Molenberghs G. Missing data methods in longitudinal studies: a review. *Test (Madr)*. 2009;18(1):1-43. doi:[10.1007/s11749-009-0138-x](https://doi.org/10.1007/s11749-009-0138-x). PMID:21218187
37. Achenbach TM. *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Dept. of Psychiatry, University of Vermont; 1991.
38. Ivanova MY et al. Syndromes of Self-Reported Psychopathology for Ages 18-59 in 29 Societies. *J Psychopathol Behav Assess*. 2015;37(2):171-183. doi:[10.1007/s10862-014-9448-8](https://doi.org/10.1007/s10862-014-9448-8). PMID:29805197
39. Mous SE et al. Cortical thickness and inattention/hyperactivity symptoms in young children: a population-based study. *Psychol Med*. 2014;44(15):3203-3213. doi:[10.1017/S0033291714000877](https://doi.org/10.1017/S0033291714000877)
40. Little RJA. A test of missing completely at random for multivariate data with missing values. *J Am Statist Assoc*. 1988;83(404):1198-1202. doi:[10.1080/01621459.1988.10478722](https://doi.org/10.1080/01621459.1988.10478722)
41. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med*. 1990;9(7):811-818. doi:[10.1002/sim.4780090710](https://doi.org/10.1002/sim.4780090710)
42. Loomes R, Hull L, Mandy WPL. What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. *J Am Acad Child Adolesc Psychiatry*. 2017;56(6):466-474. doi:[10.1016/j.jaac.2017.03.013](https://doi.org/10.1016/j.jaac.2017.03.013)
43. Vucic SN et al. Sex differences in global metrics of brain size across the lifespan. *Frontiers in Neuroscience*. 2026;20:1646144.
44. Giedd JN et al. Trajectories of anatomic brain development as a phenotype. *Novartis Found Symp*. 2008;289:101-112. doi:[10.1002/9780470751251.ch9](https://doi.org/10.1002/9780470751251.ch9). PMID:18497098
45. Thompson PM et al. Structural MRI and brain development. *Int Rev Neurobiol*. 2005;67:285-323. doi:[10.1016/S0074-7742\(05\)67009-2](https://doi.org/10.1016/S0074-7742(05)67009-2)

46. Shaw P, Rabin C. New insights into attention-deficit/hyperactivity disorder using structural neuroimaging. *Curr Psychiatry Rep.* 2009;11(5):393-398. doi:[10.1007/s11920-009-0059-0](https://doi.org/10.1007/s11920-009-0059-0)
47. Shaw P et al. Intellectual ability and cortical development in children and adolescents. *Nature.* 2006;440(7084):676-679. doi:[10.1038/nature04513](https://doi.org/10.1038/nature04513)
48. Ritchie SJ et al. Sex Differences in the Adult Human Brain: Evidence from 5216 UK Biobank Participants. *Cereb Cortex.* 2018;28(8):2959-2975. doi:[10.1093/cercor/bhy109](https://doi.org/10.1093/cercor/bhy109). PMID:29771288
49. Gennatas ED et al. Age-Related Effects and Sex Differences in Gray Matter Density, Volume, Mass, and Cortical Thickness from Childhood to Young Adulthood. *J Neurosci.* 2017;37(20):5065-5073. doi:[10.1523/JNEUROSCI.3550-16.2017](https://doi.org/10.1523/JNEUROSCI.3550-16.2017). PMID:28432144
50. Ruigrok AN et al. A meta-analysis of sex differences in human brain structure. *Neurosci Biobehav Rev.* 2014;39(100):34-50. doi:[10.1016/j.neubiorev.2013.12.004](https://doi.org/10.1016/j.neubiorev.2013.12.004). PMID:24374381
51. Afzali MH et al. Cross-lagged Relationships Between Depressive Symptoms and Altered Default Mode Network Connectivity Over the Course of Adolescence. *Biol Psychiatry Cogn Neurosci Neuroimaging.* 2022;7(8):774-781. doi:[10.1016/j.bpsc.2021.10.018](https://doi.org/10.1016/j.bpsc.2021.10.018)
52. Nivard MG et al. Joint developmental trajectories of internalizing and externalizing disorders between childhood and adolescence. *Dev Psychopathol.* 2017;29(3):919-928. doi:[10.1017/S0954579416000572](https://doi.org/10.1017/S0954579416000572)
53. Bista S et al. Joint developmental trajectories of internalizing and externalizing problems from mid-childhood to late adolescence and childhood risk factors: Findings from a prospective pre-birth cohort. *Dev Psychopathol.* 2025;37(1):176-191. doi:[10.1017/S0954579423001505](https://doi.org/10.1017/S0954579423001505)
54. Achenbach TM et al. Internalizing/Externalizing Problems: Review and Recommendations for Clinical and Research Applications. *J Am Acad Child Adolesc Psychiatry.* 2016;55(8):647-656. doi:[10.1016/j.jaac.2016.05.012](https://doi.org/10.1016/j.jaac.2016.05.012)
55. Basten M et al. The Stability of Problem Behavior Across the Preschool Years: An Empirical Approach in the General Population. *J Abnorm Child Psychol.* 2016;44(2):393-404. doi:[10.1007/s10802-015-9993-y](https://doi.org/10.1007/s10802-015-9993-y). PMID:25832625
56. Robbers SC et al. Trajectories of CBCL attention problems in childhood. *Eur Child Adolesc Psychiatry.* 2011;20(8):419-427. doi:[10.1007/s00787-011-0194-0](https://doi.org/10.1007/s00787-011-0194-0). PMID:21713506
57. Rubin DB. Inference and missing data. *Biometrika.* 1976;63(3):581-592. doi:[10.1093/biomet/63.3.581](https://doi.org/10.1093/biomet/63.3.581)
58. Li L et al. Missing data approaches for longitudinal neuroimaging research: Examples from the Adolescent Brain and Cognitive Development (ABCD) Study(R). *Dev Cogn Neurosci.* 2025;74:101563.
59. Enders CK. Missing Data: An Update on the State of the Art. *Psychological Methods.* 2025;30(2):322-339. doi:[10.1037/met0000563](https://doi.org/10.1037/met0000563)
60. Blok E et al. Stability and Change of Psychopathology Symptoms Throughout Childhood and Adolescence. *Child Psychiatry Hum Dev.* 2022;53(6):1330-1339. doi:[10.1007/s10578-021-01212-8](https://doi.org/10.1007/s10578-021-01212-8). PMID:34184159
61. Basten M et al. Nonverbal intelligence in young children with dysregulation: the Generation R Study. *Eur Child Adolesc Psychiatry.* 2014;23(11):1061-1070. doi:[10.1007/s00787-014-0551-x](https://doi.org/10.1007/s00787-014-0551-x)
62. Blanken LM et al. Cognitive functioning in children with internalising, externalising and dysregulation problems: a population-based study. *Eur Child Adolesc Psychiatry.* 2017;26(4):445-456. doi:[10.1007/s00787-016-0903-9](https://doi.org/10.1007/s00787-016-0903-9). PMID:27645702
63. White T, Blok E, Calhoun VD. Data sharing and privacy issues in neuroimaging research: Opportunities, obstacles, challenges, and monsters under the bed. *Hum Brain Mapp.* 2022;43(1):278-291. doi:[10.1002/hbm.25120](https://doi.org/10.1002/hbm.25120). PMID:32621651
64. Kahneman D, Klein G. Conditions for intuitive expertise: a failure to disagree. *Am Psychol.* 2009;64(6):515-526. doi:[10.1037/a0016755](https://doi.org/10.1037/a0016755)

FIGURE LEGENDS

Supplementary Figures 1-9 – An overview of missingness for each of the 15 datasets. Each dataset has six separate sub-figures that highlight different characteristics of the missingness for that specific dataset. These entail (A) the frequency of missing variables per individual in the dataset; (B) the number of missing data per variable; (C) a correlation heatmap of missing data where a value approaching zero reflects that the missingness between the two variables is uncorrelated, whereas a value of one reflects that when one variable is missing, the other variable is also missing; (D) the matrix pattern of missingness, which demonstrates the nature of the structure (or randomness) of missing data; and (E) a dendrogram of missing values, which groups variables that are jointly missing. A dendro-

gram uses hierarchical clustering to show the similarities of missingness between variables.

Supplementary Figure 10 – The number of missing data per wave, with (A) reflecting missing CBCL data and (B) reflecting missing MRI datasets.

Supplementary Figure 11 – Longitudinal trajectories of male/female differences as reflected in effect sizes (positive values reflect males > females) for (A) total gray matter volume (B) total white matter volume; (C) hippocampal volume; (D) Amygdala volume; (E) intracranial volume, and (F) cortical thickness.

SUPPLEMENTARY MATERIALS

Supplementary Figures

Download: <https://apertureneuro.org/article/165375-large-scale-open-science-simulated-data-to-study-the-development-of-brain-and-behavior-relationships-epidemiologic-clinical-and-neuroimaging-charac/attachment/354965.pdf>

Supplemental Tables

Download: <https://apertureneuro.org/article/165375-large-scale-open-science-simulated-data-to-study-the-development-of-brain-and-behavior-relationships-epidemiologic-clinical-and-neuroimaging-charac/attachment/354966.pdf>

Tables 3 to 6

Download: <https://apertureneuro.org/article/165375-large-scale-open-science-simulated-data-to-study-the-development-of-brain-and-behavior-relationships-epidemiologic-clinical-and-neuroimaging-charac/attachment/354967.docx>
