JMI Journal of Medical Imaging

RESEARCH PAPER

Charting confidence in white matter brain charts: enabling study planning through stability validation

Michael E. Kim,[a,*] Gaurav Rudravaram,[b] Adam M. Saunders,[b] Chenyu Gao,[b] Karthik Ramadass,[b] Nancy R. Newlin,[c] Praitayini Kanakaraj,[d] Sam Bogdanov,[e] Angela L. Jefferson,[f] Victoria L. Morgan,[g] Alexandra Roche,[h] Dario J. Englot,[i] Susan M. Resnick,[j] Lori L. Beason Held,[j] Murat Bilgel,[j] Laurie E. Cutting,[k] Laura A. Barquero,[k] Micah A. D'Archangel,[k] Tin Q. Nguyen,[k] Kathryn L. Humphreys,[l] Yanbin Niu,[l] Sophia Vinci-Booher,[l] Carissa J. Cascio,[m] Kimberly R. Pechman,[n] Niranjana Shashikumar,[n] The HABS-HD Study Team,[†] Alzheimer's Disease Neuroimaging Initiative,[‡] The BIOCARD Study Team,[§] Zhiyuan Li,[o] Panpan Zhang,[p] John C. Gore,[g] Yihao Liu,[b] Lianrui Zuo,[b] Yuankai Huo,[a] Derek B. Archer,[n] Timothy J. Hohman,[n] L. Taylor Davis,[g] Kurt G. Schilling,[g] Daniel C. Moyer,[a] and Bennett A. Landman[a,b,g,h,n]

[a]Vanderbilt University, Department of Computer Science, Nashville, Tennessee, United States
[b]Vanderbilt University, Department of Electrical and Computer Engineering, Nashville, Tennessee, United States
[c]Digital Informatics & Technology Solutions, Memorial Sloan Kettering Cancer Center, New York, United States
[d]Bioscope.ai, Boston, Massachusetts, United States
[e]Vanderbilt University, Medical Scientist Training Program, Nashville, Tennessee, United States
[f]Vanderbilt University Medical Center, Department of Neurology, Nashville, Tennessee, United States
[g]Vanderbilt University Medical Center, Department of Radiology and Radiological Sciences, Nashville, Tennessee, United States
[h]Vanderbilt University Institute of Imaging Science, Nashville, Tennessee, United States
[i]Vanderbilt University Medical Center, Department of Neurological Surgery, Nashville, Tennessee, United States
[j]National Institute on Aging, National Institutes of Health, Laboratory of Behavioral Neuroscience, Baltimore, Maryland, United States
[k]Peabody College of Education and Human Development, Department of Special Education, Nashville, Tennessee, United States
[l]Vanderbilt University, Department of Psychology and Human Development, Nashville, Tennessee, United States
[m]University of Kansas, Life Span Institute and Department of Psychology, Lawrence, Kansas, United States

*Address all correspondence to Michael E. Kim, michael.kim@vanderbilt.edu

[§]Data used in preparation of this article were derived from BIOCARD study, supported by grant U19 – AG033655 from the National Institute on Aging. The BIOCARD study team did not participate in the analysis or writing of this report, however, they contributed to the design and implementation of the study. A listing of BIOCARD investigators can be found on the BIOCARD website (on the 'BIOCARD Data Access Procedures' page, 'Acknowledgement Agreement' document).

[‡]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

[†]HABS-HD MPIs: Sid E O'Bryant, Kristine Yaffe, Arthur Toga, Robert Rissman, & Leigh Johnson; and the HABS-HD Investigators: Meredith Braskie, Kevin King, James R Hall, Melissa Petersen, Raymond Palmer, Robert Barber, Yonggang Shi, Fan Zhang, Rajesh Nandy, Roderick McColl, David Mason, Bradley Christian, Nicole Phillips, Stephanie Large, Joe Lee, Badri Vardarajan, Monica Rivera Mindt, Amrita Cheema, Lisa Barnes, Mark Mapstone, Annie Cohen, Amy Kind, Ozioma Okonkwo, Raul Vintimilla, Zhengyang Zhou, Michael Donohue, Rema Raman, Matthew Borzage, Michelle Mielke, Beau Ances, Ganesh Babulal, Jorge Llibre-Guerra, Carl Hill, and Rocky Vig.

[n]Vanderbilt University Medical Center, Vanderbilt Memory and Alzheimer's Center, Nashville, Tennessee, United States
[o]Park University, Department of Computer Science, Parkville, Missouri, United States
[p]Vanderbilt University Medical Center, Department of Biostatistics, Nashville, Tennessee, United States

ABSTRACT. **Purpose:** Normative modeling of quantitative brain measurements is being widely adopted as a promising method for charting population-level developmental trajectories and identifying abnormalities in groups and individuals. Generalized additive models for location, scale, and shape (GAMLSS) are a statistical framework that has been identified by the World Health Organization as a robust method for large-scale population modeling. Validating the stability of normative models is essential for appropriate benchmarking and detecting anomalies, especially for patients with neurodegenerative or developmental disorders. As these models continue to expand in the number of features and amount of data, it becomes increasingly important to assess the stability of the models and their predictions.

Approach: An analytic approach is extended from previous work to assess variability of the GAMLSS framework for normative modeling of white matter measurements across the lifespan. The analytic variance and bias are compared with empirical bootstrapping estimates and those of an analogous model that is a nonparametric, data-driven analog.

Results: Across all models, the analytic approach showed low model variability (<0.5% Coefficient of Variation) for lifespan trajectories that increased with age. Empirical variability of trajectories was slightly elevated by comparison, especially at the beginning of the lifespan. The trajectory variability for the data-driven analog was even larger, showing a ten-fold increase from the empirical GAMLSS variability. For individual centile prediction, the majority of data points for the analytic assessment had a mean absolute difference (MAD) in centile score prediction as z-scores that was less than 0.003 with the standard deviation of these errors below 0.06. The empirical assessment showed an elevated standard deviation (0.15) and MAD of z-score centile predictions (0.06) for the majority of datapoints. By contrast, when compared with the empirical assessment, the nonparametric, data-driven analog showed over a two-fold increase in the MAD of centile z-score predictions (0.14) and a three-fold increase in standard deviation (0.45) for the majority of data points. The divergences in variability across assessments suggest that individual centile predictions are more sensitive to random fluctuations in the data, which the nonparametric data-driven approach exacerbates.

Conclusions: Results suggest that the GAMLSS brain charts are a stable method of defining and charting normative white matter development as well as planning future clinical or research studies. Overall, this study provides an assessment of GAMLSS-derived brain chart variance in a large-scale population.

© The Authors. Published by SPIE under the terms of the Creative Commons Attribution NonCommercial-NoDerivatives (CC-BY-NC-ND) 4.0 International License. Use, distribution, or reproduction of this work in whole or in part in any medium is permitted for noncommercial purposes only, provided the work is not modified and the original work is properly cited, including the DOI. [DOI: 10.1117/1.JMI.13.5.054001]

Keywords: normative modeling; magnetic resonance imaging; white matter; diffusion mRI

Paper 260214 received Apr. 23, 2026; revised Jun. 30, 2026; accepted Aug. 13, 2026; published Sep. 21, 2026.

1 Introduction

Brain charts, which were inspired by pediatric growth charts,[1] are valuable tools that allow researchers to leverage large populations to investigate lifespan brain development and benchmark individuals or groups. Specifically, these charts are computational models that describe how quantitative measures derived from medical imaging data are expected to change across the human lifespan, providing age-specific reference distributions against which an individual's measurements can be compared. Existing models have already allowed for identification of

important developmental and aging milestones in the human brain using magnetic resonance imaging (MRI) data.[2] These brain charts also have promising clinical applications for detecting anomalies. Group-level analyses have shown the utility of these charts for detecting quantifiable changes in structural brain measurements across a range of clinical groups.[2–4] Furthermore, normative modeling of quantitative brain measurements has promise for detecting abnormalities at the individual level.[5]

There are several methods that are used to normatively model MRI-derived measurements, including Gaussian process regression, polynomial regression, hierarchical Bayesian regression, and others.[6,7] One notable candidate is generalized additive models for location, scale, and shape (GAMLSS).[8] In addition to being recognized by the World Health Organization (WHO) as a method for constructing growth curves,[9] the GAMLSS framework provides flexibility for modeling many types of distributions. This flexibility includes the ability to model higher-order distributional moments, such as skewness and kurtosis, as well as the more standard mean and variance. Due to its utility, GAMLSS has been used widely in normative modeling for brain charts.[2,7,10]

Many previous works have focused on whole brain or gray matter measurements derived from structural MRI,[2,3,7,11,12] with some models beginning to expand into functional MRI data.[13–15] Brain charts for white matter (WM) derived from diffusion MRI (dMRI) data are emerging as well,[10,16,17] which are important for charting lifespan changes in microstructural, or micron-level biological architecture, and macrostructural, or tract-level morphology, organization of white matter.[10] Although they provide complementary information to structural MRI brain charts, WM brain charts similarly show promise in providing population references and identifying group-level abnormalities and normative trajectories. Furthermore, the study of developing microstructure allows for studying neurobiological hypotheses regarding regional myelination in the brain.[10]

Although these WM brain charts can provide several benefits to researchers and clinicians, validations of the stability of these models are paramount to make sure individual benchmarking and normative trajectories are robust. High variability in model fits yields low precision centile scores, reducing the utility of the brain charts (Fig. 1). In addition to model stability, there are numerous quantitative features to be extracted from dMRI, both microstructural[18–20] and macrostructural.[21] Coupled with the large number of potential regions of interest, collective normative WM brain charts could yield thousands of different models that require stability assessments. This high-dimensional brain phenotyping at high spatial resolution makes empirical variability assessments computationally burdensome. Furthermore, although GAMLSS has proven valuable for constructing WM brain charts, the stability and robustness of these models, especially for individual-level benchmarking, require further characterization.

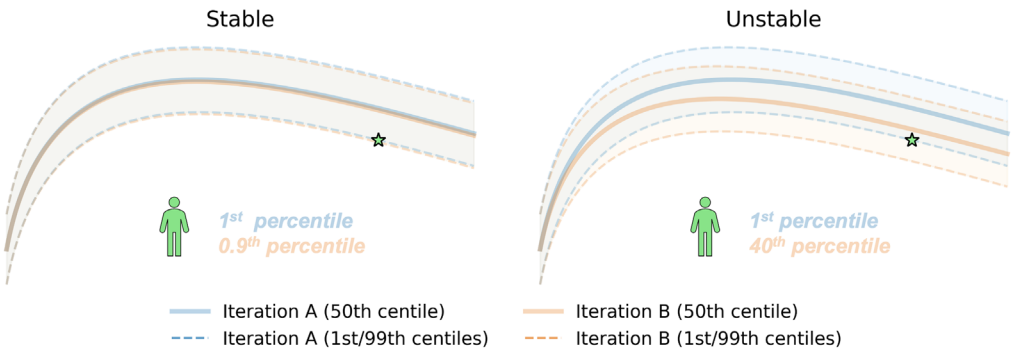


Fig. 1 Normative models are important tools for clinical use and research to benchmark individuals against a greater population. Stability of these normative models is paramount for downstream anomaly detection, as unstable models can misclassify individuals or produce variable quantitative results during inference. (Left) In a stable normative model, confidence in individual benchmarking is high, where variability assessments yield near-identical trajectories. (Right) For unstable normative models, drastic shifts can occur due to small variations in the data, making these models much less reliable for assessing abnormalities and development.

To address these concerns, we conduct a characterization of the stability of these normative models for white matter created using GAMLSS. Specifically, we extend[10] the creation of these WM brain charts for microstructure and macrostructure, to variability and stability assessments of centile scoring and normative trajectories from GAMLSS WM brain charts in two ways. First, using a standard empirical approach based on bootstrapping, and second, using an analytic approach extended from our prior work[22] that relies on the assumptions of the GAMLSS model framework. In addition to these stability assessments, we compare the variability of our brain charts to a deep learning analog, which serves as a high-variance, nonparametric data-driven benchmark for model stability.

2 Methods

2.1 Data

We use data from our previous work[10] consisting of $N = 35,120$ participants with no known neurological disorders from 50 population datasets (see Table S1 in the Supplementary Material). Each participant has demographic information available for age and sex and has at least one dMRI scan; for datasets where participants received multiple scans, only one was chosen for analysis. Further details regarding data preprocessing can be found in our previous papers.[10,23,24] Briefly, dMRI data were preprocessed for motion and distortion artifacts using the PreQual pipeline.[25] With the MRtrix3 library,[26] diffusion tensors were fit to the data to obtain microstructural values of fractional anisotropy (FA) and mean, radial, and axial diffusivities (MD, RD, and AxD). The TractSeg[27] method was then used to perform subject-specific tractography for the 72 different WM tracts defined in the published algorithm. Finally, the *scilpy* Python library[28] was used to calculate macrostructural measurements (volume, surface area, and average length) and average microstructural measurements for each tract. A manual, high-throughput quality control pipeline was used to reject failed outputs at each processing step.[23]

2.2 GAMLSS Normative Model Fitting

We fit 504 models in total using the GAMLSS framework,[8] one for each combination of the 72 different WM tracts and seven microstructural/macrostructural features. The GAMLSS framework estimates the parameters μ, σ, ν, τ that define a distributional family $D(\mu, \sigma, \nu, \tau)$, where the parameters are functions of the independent variables X (age, sex, and dataset) and the dependent variable Y (WM tract measurement) is distributed according to $D(\mu, \sigma, \nu, \tau)$. This flexibility is important for capturing non-Gaussian distributions that have been shown to be present for quantitative WM metrics.[29] For increased stability in model fitting, we normalize input age to be log-scaled, post-conception days, using the following equation:

$$\text{Age}_{\text{normalized}} = \ln([\text{Age}_{\text{years}} * 365.25] + 280). \tag{1}$$

Following our previous work, D follows the generalized gamma distribution, with the probability density function $GG(x; \mu, \sigma, \nu)$ that is implemented in the GAMLSS framework.[8] Notably, although GAMLSS currently supports distributional families of up to four parameters, the generalized gamma distribution is parameterized by three terms. The parameters, μ, σ, and ν, roughly corresponding to the location, scale, and skewness of the distribution (see Supplementary Material for further details), are modeled as follows:

$$\log(\mu) = X\beta_\mu + Zb_\mu + U_\mu f_\mu + \epsilon_\mu, \tag{2}$$

$$\log(\sigma) = X\beta_\sigma + Zb_\sigma + U_\sigma f_\sigma + \epsilon_\sigma, \tag{3}$$

$$\nu = \alpha + \epsilon_\nu, \tag{4}$$

where $\beta_{\{\mu,\sigma\}}$ represents a vector for intercept and fixed effect for sex; X is the design matrix for $\beta_{\{\mu,\sigma\}}$; $b_{\{\mu,\sigma\}}$ represents a random intercept vector for dataset; Z is a design matrix for $b_{\{\mu,\sigma\}}$; $f_{\{\mu,\sigma\}}$ is a vector for fractional polynomial coefficients describing nonlinear age effects; $U_{\{\mu,\sigma\}}$ is a design matrix for each $f_{\{\mu,\sigma\}}$; α is the intercept term for v; and $\epsilon_{\{\mu,\sigma,\nu\}}$ corresponds to the error for a given distributional parameter. Model fitting was performed using the code in the Docker for WM brain charts.[30]

2.3 Analytic Assessment

In GAMLSS, each block of parameters in the set $\hat{\theta} = [\hat{\beta}_\mu \quad \hat{\beta}_\sigma \quad \hat{b}_\mu \quad \hat{b}_\sigma \quad \hat{f}_\mu \quad \hat{f}_\sigma \quad \hat{\alpha}]$ are estimated independently based on a backfitting process.[8] Directly calculating V, the joint variance-covariance matrix of all fit parameters in the GAMLSS model, becomes prohibitive because each parameter batch has its own weight matrix. However, we can still estimate V as the inverse of the Hessian matrix H for the negative log likelihood l of the GAMLSS model, where

$$(H_l)_{i,j} = -\frac{\partial^2 l}{\partial\theta_i \partial\theta_j} \tag{5}$$

represents each entry in the Hessian matrix (see Supplementary Material section "Analytic Model – Variance-Covariance Matrix Calculation" for more details). Approximating V as $\hat{V} \approx (H_l)^{-1}$, each entry $\hat{V}_{i,j}$ represents the variance of the i^{th} parameter when $i = j$ and the covariance between the i^{th} and the j^{th} parameter when $i \neq j$. For numerical stability, we calculate the Moore–Penrose pseudoinverse[31] of H_l. We then model the GAMLSS parameters together as the vector $\hat{\theta}$, a k-dimensional random variable that is normally distributed, with k being the number of parameters in θ. The final, converged model estimates of each parameter θ_i serve as the mean of each $\hat{\theta}_i$, with $\hat{V}$ as the variance structure for the k-dimensional distribution.

Finally, to estimate the analytic variability of each model, we sample 1000 times from the k-dimensional Gaussians to obtain new models that are used in variability assessments. We choose to perform this parametric sampling procedure because Eqs. (2) and (3) employ a log-link function, meaning that the respective parameters in θ undergo a nonlinear transformation to compute the final μ and σ. This transformation will break assumptions that direct variance assessments, such as the Delta method,[32] make regarding normality. As we have assumed θ to be normally distributed, we choose this sampling approach to fully capture the resulting nonnormal distributions of μ and σ.

2.4 Empirical Assessment

To obtain estimates of empirical variance for GAMLSS models, we perform bootstrap resampling for each of the 504 models. For each model, we generate 100 bootstrapped samples by sampling 95% of the data pool without replacement and refitting model parameters on each data subset. Our selection of 95% sampling follows the variability assessment of the GAMLSS models from Bethlehem et al.[2] Each bootstrapped model is then used for downstream variability assessments.

2.5 Learning Analog

To provide a comparative benchmark for the variability observed in the GAMLSS models, we implement a deep learning architecture that performs the same normative modeling task while replacing the predefined parametric functional forms of GAMLSS with a flexible neural network. Comparing the variability of the two approaches helps place the stability of the GAMLSS models in context: If a more flexible neural network implementation of the same normative modeling task exhibits substantially greater variability when trained on the same data, this suggests that the low variability of the GAMLSS models is consistent with the regularizing effect of their statistical structure, rather than being solely a consequence of the underlying dataset. The architecture of the model is split into three parts: 1) a biological network, 2) a site network, and 3) the centile-trajectory equation (Fig. 2). For the biological network, a vector of age and sex is input to a multilayer perceptron (MLP) consisting of two hidden layers. ELU activation and 20% dropout followed each hidden layer. Age in years is normalized via a mean shift and scaling to more easily allow the network to learn age effects, whereas sex is encoded as 0 for female and 1 for male. The outputs of the biological network, δ and c, are then used for the centile trajectory equation, where δ alters the width of the learned distribution as biological covariates change and c alters the position. To make the δ and c outputs strictly positive, both are sent through a *softplus* layer.

The dataset variable is encoded as a one-hot vector that is sent as input through the site network, which is a single linear layer that outputs Δ and γ, which correspond to dataset-specific

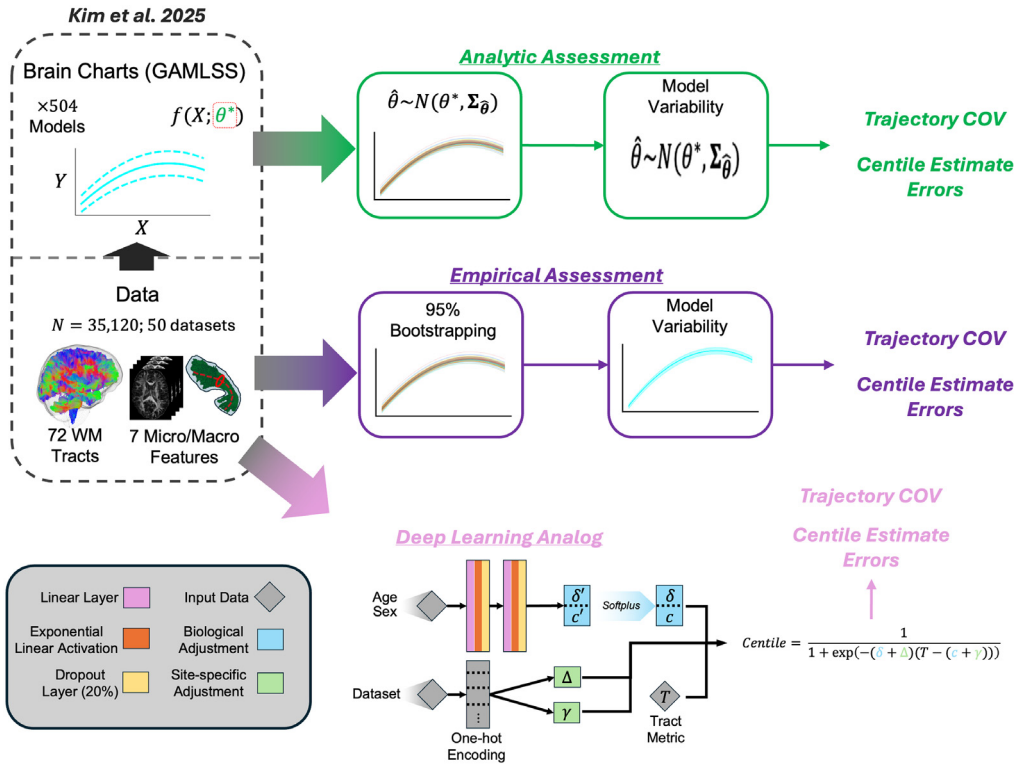


Fig. 2 We use dMRI data from 35,120 participants to extract seven micro/macrostructural features of 72 different WM tracts that are used to create GAMLSS-based brain charts, where data pre- and post-processing have been described in our previous work. From the fit GAMLSS models (with best model fit parameters θ^*), we conduct the analytic assessment of variability, whereas for the empirical assessment, we perform bootstrapping of models through 95% sampling without replacement. We also train deep learning analog to the GAMLSS models as nonparametric, data-driven comparisons for model variability. From each of these assessments, we assess the stability of the trajectories and the centile scoring methods by calculating the coefficient of variation (COV).

changes in δ and c, respectively. Finally, outputs of the biological and site networks, as well as the micro- or macrostructural quantitative tract metric T, are sent as inputs to the centile-trajectory equation:

$$F_T(t) = \frac{1}{1+e^{(-[(\delta+\Delta)(t-(c+\gamma))])}}, \tag{6}$$

where t is the tract metric T normalized to the range $(0, 10]$ and F_T is a centile value in the range (0,1). We choose F_T to be a logistic function, as it satisfies the criteria of increasing monotonicity and a (0,1) range for a centile. Thus, the deep learning analog can be interpreted as learning a conditional logistic distribution for tract metrics, where the network predicts the distribution shape conditioned on age, sex, and site. To extract the kth centile trajectory, Eq. (6) can be rearranged as follows:

$$t = (c+\gamma) + \left(\frac{-1}{(\delta+\Delta)}\right)\left(\ln\left(\frac{1}{F_T}-1\right)\right). \tag{7}$$

With this invertible architecture, the deep learning analog can be compared against the GAMLSS variability assessments for both trajectory and centile-scoring stability.

The loss function used for training is based on maximizing the likelihood of the centile-trajectory network outputs. Letting $A = \delta + \Delta$ and $C = c + \gamma$, we can rewrite Eq. (6) as

$$F_T(t) = \frac{1}{1+e^{(-A(t-\mathrm{C}))}}. \tag{8}$$

Because Eq. (8) outputs a centile score, defined as the probability that a value is less than or equal to a given value, it corresponds to a cumulative distribution function (CDF) over the tract metric t. Taking the derivative of the CDF with respect to t yields the corresponding probability density function (PDF) f_T:

$$f_T(t) = \frac{A * e^{(-A(t-C))}}{(1 + e^{(-A(t-C))})^2}, \tag{9}$$

which can be simplified using

$$1 - F_T(t) = \frac{e^{(-A(t-C))}}{1 + e^{(-A(t-C))}}, \tag{10}$$

and Eq. (8) to yield

$$f_T(t) = A * (F_T)(1 - F_T). \tag{11}$$

Models are optimized by minimizing the negative of Eq. (11). Each deep learning analog is trained using the Adam optimizer with a learning rate of $1e-3$ and evaluated via five-fold cross validation. During training, the data were divided into three equal batches per epoch. Training was stopped once the validation loss failed to improve for 20 consecutive epochs, and the model from that point was used for validation.

2.6 Assessing Model Variability

For each model of a given tract metric, we calculate the variability as the coefficient of variation (COV) (Fig. 2),

$$\mathrm{COV}_{\mathrm{tract_metric}} = \frac{\sigma_{\mathrm{tract_metric}}}{\mu_{\mathrm{tract_metric}}}, \tag{12}$$

where $\sigma_{\mathrm{tract_metric}}$ is the standard deviation of the model for the given tract metric we are assessing and $\mu_{\mathrm{tract_metric}}$ is the mean. There are two ways in which we assess variability of the brain charts. First is to assess the stability of the normative population model trajectories, which describe the ways in which the quantitative white matter measurements change throughout the lifespan. Second is to assess the stability of the centile scoring of the normative models, as stability in the centile scoring process is essential for proper anomaly detection.

We calculate the COV at different ages across the lifespan for the median trajectory to assess whether stability of the normative trajectories depends on the age being evaluated. For centile scoring, we calculate the mean absolute difference (MAD) of the centile and standard deviation of those errors. For the analytic assessment of variability, we evaluate the model at each of the 1000 iterations from our parametric sampling procedure and calculate the COV across these parametric samples. The empirical assessment is similar, instead using the 95% resampling bootstrapped model iterations. For the deep learning analog, we evaluate models 100 times using 5% node dropout in the biological network during evaluation.

2.7 Performing Power Analyses

The WM brain charts can be used to calculate power analyses for planning sample sizes of future microstructural or macrostructural analyses at any age in the lifespan, and we demonstrate how calculation of the model variability is important when estimating the necessary sample size to determine a population effect. To appropriately estimate the required number of participants, it is important to account for the model variability in centile scoring:

$$\sigma_{\mathrm{total}} = \sqrt{\sigma^2_{\mathrm{model}} + \sigma^2_{\mathrm{population}}}, \tag{13}$$

where σ_{total} is the total standard deviation used for a power analysis, σ_{model} is the variability (as standard deviation) introduced by the model that comes from the stability estimates, and $\sigma_{\mathrm{population}}$ is the population standard deviation coming from the distribution modeling of the WM brain charts (see Supplementary Material for both $\sigma_{\mathrm{population}}$ calculation and σ_{total} derivation). For use in a Z-test power analysis, the required sample size N to detect an effect d is given by

$$N = \frac{(Z_\alpha + Z_\beta)^2 \left(\frac{1}{q_0} + \frac{1}{q_1}\right)}{(d/\sigma_{\text{total}})^2}, \tag{14}$$

where α is the desired false positive rate, β is the desired false negative rate, Z_α is the Z-score corresponding to a two-tailed test at the α significance level, Z_β is the Z-score corresponding to the desired statistical power β, and q_0, q_1 are the proportions of the total sample size allocated to each group, with $q_0 + q_1 = 1$.

3 Results

3.1 Stability of Model Trajectories

GAMLSS normative model trajectories all appear stable across the majority of the lifespan (Fig. 3). The COV for the analytic assessment is similar to the empirical assessment (<0.6%) at ages greater than 60 years but tends to yield lower variability estimates from infancy to around 40 years of age (below 0.2% for analytic and 0.8% for empirical). For the deep learning analog, while trajectories are stable across all five folds for all age ranges except 0 to 2 years (Fig. S1 in the Supplementary Material), the COV is notably increased when compared with the analytic and empirical assessments of variability. Furthermore, the COV increases more at the earliest (<10%) and oldest (<4%) ages in the lifespan when compared with the GAMLSS normative models.

3.2 Stability of Model Centiles

For centile scoring, GAMLSS models showed high stability. Centile scores from the model can be converted to Z-scores; we report the stability results for both. Mean absolute differences (MAD) in centile Z-scores were consistently below 0.003 (0.1% as centile) for the analytic assessment and 0.06 (1% as centile) for the empirical assessment (Fig. 4; Fig. S2 in the Supplementary Material). The variabilities of these differences also showed high stability, with most datapoints having a standard deviation of less than 0.06 and 0.15 for analytic and empirical, respectively, (2% as centile for both). In contrast, the deep learning analog was much less stable for centile scoring, with most datapoints having a MAD less than 0.14 (6% as centile) and standard deviation of 0.45 (15% as centile).

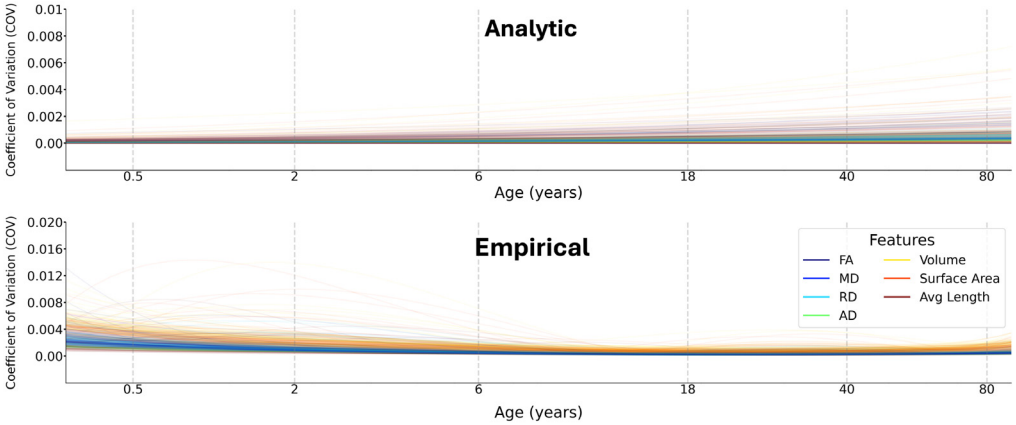


Fig. 3 GAMLSS model trajectories are both empirically and analytically stable across the lifespan, with the COV for median trajectories sitting below 0.6% for analytic and 1.6% for empirical variability estimates. However, at most timepoints the variability of models is much lower, with the largest COV occurring at the edges of the lifespan range. This "edge effect" is a known artifact in normative modeling. We observe the edge effect to be more pronounced at older ages for the analytic analyses, whereas it is more apparent at younger ages for the empirical analysis. Shown in the plot are the COVs of all models (every tract), colored by the corresponding microstructural or macrostructural feature.

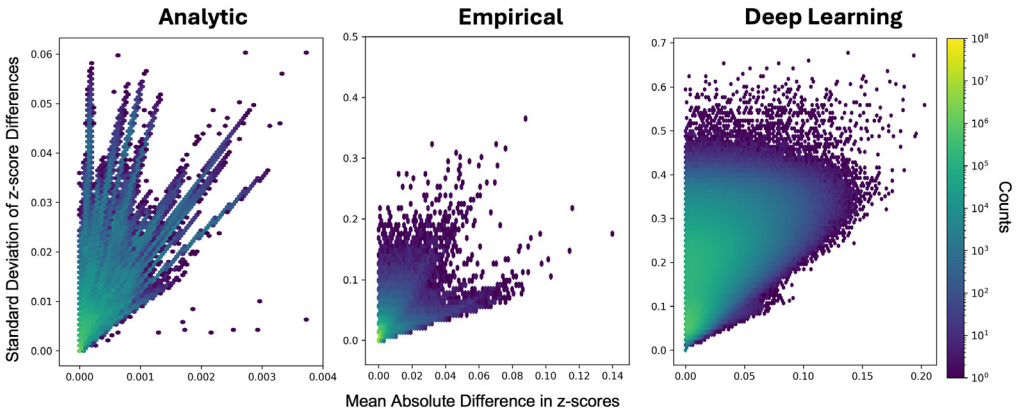


Fig. 4 GAMLSS models demonstrate stable centile scoring across participants and data models, with mean absolute differences consistently below 0.003 and 0.08 *z*-scored units for analytic and empirical assessments, respectively. The variabilities in these differences are low as well, with the standard deviation for most datapoints sitting below 0.06 for analytic variability and 0.15 for empirical variability. Deep learning analogs show less stability, with mean differences mostly below 0.14 and standard deviation of differences below 0.45. However, as the deep learning analog architecture is more flexible than the GAMLSS framework, this increased variability is expected.

4 Discussion

We demonstrate that the recently available GAMLSS WM brain charts[30] are highly stable models across the entire human lifespan. The notably low COV of trajectories (<1%, mostly <0.4%) for both analytic and empirical assessments indicates high confidence in the previously identified developmental and aging milestones.[10] Furthermore, although the centile score variability for analytic and empirical assessments differed, both were highly stable in MAD and standard deviation of errors. This stability in centile scoring suggests that researchers and clinicians can be confident using the brain charts as a normative reference for benchmarking individuals. Although the edge effects we observed are common to normative modeling,[33] the increased variability could also, in part, be due to more rapid changes in development[34] and the beginning emergence of a variety of pathologies in older ages.[35]

As mentioned previously, analytic variability assessments offer a more rapid way to examine model stability, which is important when considering normative charts that model individual quantitative measurements for several separate regions of interest. Our results indicate that analytic variability assessments tend to yield less conservative, or smaller, estimates of model variability when compared to empirical assessments. Although prior work suggests that this divergence may be for only the most stable models,[22] the results from this report suggest that the analytic method underpredicts variability of both normative trajectories and individual centile estimations (Figs. 3 and 4). Although we observe this dissonance between the analytic and empirical variability estimates, we note that these results are assessing only the highly stable GAMLSS WM brain charts.[7]

The increased stability of the GAMLSS brain charts compared with the deep learning analogs is expected due to the greatly increased flexibility and number of degrees of freedom that a neural network possesses, especially given the increased sensitivity that a neural network has to non-linear effects such as age. Our results suggest that the deep learning analogs, especially for individual centile prediction, are more sensitive to random fluctuations in the data due to the increased variability. In addition, the analogs were trained using 20% less data than the GAMLSS models due to the five-fold cross-validation that was used in our methodology. We also observed that the edge effects normative models suffer from Ref. 33 appeared to be exacerbated in the deep learning analogs (Fig. S1 in the Supplementary Material), which aligns with our expectations, given that the GAMLSS framework adheres to preset distributional assumptions that are not present in the analogs. These findings suggest that the modeling framework, and not solely the underlying data, contributes to the observed stability of the GAMLSS brain charts. However, as the two approaches differ in multiple respects beyond their functional parameterization, our comparison to the deep learning analogs should be interpreted as

supporting evidence rather than a definitive attribution of the source of the variability. We note that we are not disparaging the use of deep learning for normative modeling; there are examples of promising works in the literature that use deep learning frameworks for normative brain modeling.[5,36]

Beyond identifying normative trajectories and performing centile scoring of new data, these WM brain charts facilitate sample size planning for future microstructural or macrostructural analyses at any age in the lifespan. Employing Eq. (13) to account for model variance, investigators can perform robust power analyses to guide study design. We provide Table 1 as an example of this, where we demonstrate power analysis results at several ages and effect sizes for the fractional anisotropy of the left arcuate fasciculus. Specifically, we use the higher empirical variability level of $\sigma_{z,\text{model}} = 0.15$ as a conservative estimate for the power analysis (see Supplementary Material for translating $\sigma_{z,\text{model}}$ to σ_{model}), with a false positive rate of $\alpha = 0.05$, false negative rate of $\beta = 0.2$, and assume that the control cohort size is equal to the experimental cohort size. We facilitate the ability to perform these power analyses by releasing the code alongside our available WM brain charts through Zenodo[10,30] (see Supplementary Material for how to run these power analyses).

As mentioned at the beginning of this paper, normative modeling of quantitative brain measures extends beyond the cerebral white matter measures in this paper, encompassing asymmetry indices of white matter pathways,[37] whole-brain volumetric measurements,[2] gray matter morphology,[7,38] and even functional and structural connectivity measures.[15,39] These charts have been extending beyond the brain as well, with recent works encompassing morphological measures from abdominal organs such as the pancreas, liver, and spleen.[40,41] Because these separate quantitative medical imaging normative charts have been derived from differing populations,

Table 1 Example power analysis conducted using the GAMLSS brain charts, while conservatively accounting for the model variability. The results are for the brain chart model of the mean FA of the left arcuate fasciculus (AF left).

Age	Effect Size ([FA])	Sample Size Required
8	0.020	51
8	0.010	202
8	0.005	808
15	0.020	45
15	0.010	180
15	0.005	718
25	0.020	44
25	0.010	173
25	0.005	691
50	0.020	50
50	0.010	199
50	0.005	794
65	0.020	58
65	0.010	231
65	0.005	922
80	0.020	69
80	0.010	275
80	0.005	1098

modalities, datasets, frameworks, sample sizes, and organ systems, evaluating stability of centile scores derived from these other charts could provide valuable insight into the reliability and uncertainty of their downstream applications. More broadly, the analytic and empirical stability assessment framework presented here is not specific to white matter brain charts and could be applied to other normative modeling applications in medical imaging. Furthermore, incorporating the stability estimates of these different modality- and organ-specific models into power analyses may facilitate uncertainty-aware study planning for multimodal and multiorgan imaging studies.

There are a few considerations related to this work. Regarding the differences between the analytic and empirical variability assessments, our experiments were conducted only on the GAMLSS framework for WM normative models. Future work will need to assess if analytic methods are more appropriate for less stable models or different statistical frameworks that have been used with normative modeling, such as those detailed by Ge et al..[7] In addition, although the GAMLSS brain charts were more stable than the deep learning variants, a more complex architecture, training method, or loss function might provide increased stability compared with the models presented in this paper, which were designed to be direct analogs to the GAMLSS brain charts. Investigating the long-term stability of this deep learning architecture, or alternative ones, represents a compelling avenue for future research.

5 Conclusion

We have demonstrated that previously released GAMLSS WM brain charts have high stability in both identifying normative trajectories and centile scoring of individuals using two different methods, analytic and empirical. As we found, the empirical assessment offers a more conservative estimate of model variability; researchers who wish to avoid overstating the precision of models should favor this method of assessment. Furthermore, we use a deep learning analog as a comparative baseline to highlight model stability. Our results suggest that the analytic method underestimates variability when compared to the empirical method, and GAMLSS WM brain chart stability is much greater than that of a deep learning analog. This high degree of stability suggests that sensitivity and specificity of the WM brain charts are high for individual centile predictions. Finally, when designing studies, it is important to account for age-dependent model uncertainty. As the normative WM brain charts capture this age-dependent population variability in WM microstructure and macrostructure, they have the capability to facilitate study planning. By publicly providing our method for performing power analyses with the WM brain charts while accounting for model uncertainty, we enable researchers to generate age-dependent sample size estimates unique to their own target study populations.

Disclosures

The authors declare no conflicts of interest. Bennett A. Landman is Editor-in-Chief of JMI. To maintain transparency and uphold the highest standards of editorial integrity, the peer-review process for this paper was managed independently by an editor who oversaw the review and editorial decision-making process.

Code and Data Availability

Code for obtaining centile curves and fitting the GAMLSS models is available in a containerized Docker image that is downloadable at https://zenodo.org/records/15367425. Instructions for running the Docker image can also be found at that link and in the Supplementary Material. The post-processing pipeline to obtain the microstructural and macrostructural measurements from dMRI data is also available as a Docker image downloadable at: https://hub.docker.com/r/kimm58/wm_lifespan_processing. Instructions for running the Docker image are available at https://zenodo.org/records/17144461. Code for the deep learning analog models can be found here: https://github.com/MASILab/MAGCE_WMBrainChart.

Data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) are available upon request from https://adni.loni.usc.edu/. Data from the AOMIC-PIOP1 dataset are freely available for download on OpenNEURO: https://openneuro.org/datasets/ds002785/versions/2.0.0. Data from

the AOMIC-PIOP2 dataset are freely available for download on OpenNEURO: https://openneuro.org/datasets/ds002790/versions/2.0.0. Data from the AOMIC-ID1000 dataset are freely available for download on OpenNEURO: https://openneuro.org/datasets/ds003097/versions/1.2.1. Data from the Boston Adolescent Neuroimaging of Depression and Anxiety (BANDA) dataset are available upon request from https://www.humanconnectome.org/study/connectomes-related-anxiety-depression. Data from BIOCARD are available upon request after filling out a data use application: https://www.gaaindata.org/partner/BIOCARD. Data from the Baltimore Longitudinal Study of Aging (BLSA) are available upon request from https://www.blsa.nih.gov/. Data from the Calgary Preschool MRI Dataset are freely available for download at: https://doi.org/10.17605/OSF.IO/AXZ5R. Data from the Centre for Attention Learning and Memory (CALM) dataset are available upon request from https://calm.mrc-cbu.cam.ac.uk/researchers/. Data from the Cambridge Center for Ageing Neuroscience (CAMCAN) dataset are available upon request from https://camcan-archive.mrc-cbu.cam.ac.uk/dataaccess/. Data from the Dallas Lifespan Brain Study (DLBS) dataset are freely available for download on OpenNEURO: https://openneuro.org/datasets/ds004856/versions/1.2.0. Data from the Health & Aging Brain Study - Health Disparities (HABS-HD) dataset are available upon request from https://apps.unthsc.edu/itr/reports. Imaging data and basic demographic information for the Healthy Brain Network (HBN) dataset are freely available to download from https://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/. Phenotypic data are available upon request by filling out a data use agreement: https://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/Phenotypic.html. Data from the Human Connectome Project – Aging (HCPA) dataset are available upon request from https://www.humanconnectome.org/study/hcp-lifespan-aging. Data from the Lifespan Baby Connectome Project (HCPBaby) dataset are available upon request https://www.humanconnectome.org/study/lifespan-baby-connectome-project. Data from the Human Connectome Project – Development (HCPD) dataset are available upon request from https://www.humanconnectome.org/study/hcp-lifespan-development. Data from the Human Connectome Project – Young Adult (HCP) dataset are freely available for download from https://www.humanconnectome.org/study/hcp-young-adult. Data from the Infant Brain Imaging Study (IBIS) are available for download from the National Institutes of Mental Health data archive upon request from https://nda.nih.gov/edit_collection.html?id=19. Data from the International Consortium for Brain Mapping (ICBM) dataset are available upon request from www.loni.usc.edu/ICBM. Data from the Longitudinal Brain Correlates of Multisensory Lexical Processing in Children study (shortened to Lexical in this paper) are freely available for download on OpenNEURO: https://openneuro.org/datasets/ds001894/versions/1.4.2. Data from the Memory and Aging Project (MAP), Religious Orders Study (ROS), and the Minority Aging Research Study (MARS) datasets are available upon request from https://www.radc.rush.edu/. Data from the Multisite, Multiscanner, and Multisubject Acquisitions for Studying Variability in Diffusion Weighted Magnetic Resonance Imaging (MASiVar) dataset are freely available for download on OpenNEURO: https://openneuro.org/datasets/ds003416/versions/2.0.2. Data from the National Alzheimer's Coordinating Center (NACC) and the Standardized Centralized Alzheimer's & Related Dementias Neuroimaging (SCAN) are available upon request from https://naccdata.org/requesting-data/data-request-process. Imaging data and basic demographic information for the Nathan Kline Institute – Rockland Sample (NKI) dataset are freely available to download from https://fcon_1000.projects.nitrc.org/indi/enhanced/sharing_neuro.html. Phenotypic data are available upon request by filling out a data use agreement: https://fcon_1000.projects.nitrc.org/indi/enhanced/sharing_phenotypic.html. Data from the Pediatric Imaging, Neurocognition, and Genetics dataset (PING) are available for download from the National Institutes of Mental Health data archive upon request from https://nda.nih.gov/edit_collection.html?id=2607. Data from the Queensland Twin Adolescent Brain (QTAB) dataset are freely available for download on OpenNEURO: https://openneuro.org/datasets/ds004146/versions/1.0.4. Data from the UCLA Consortium for Neuropsychiatric Phenomics LA5c Study (UCLA) dataset are freely available for download on OpenNEURO: https://openneuro.org/datasets/ds000030/versions/1.0.0. Data from the Southwest University (SWU) Longitudinal Imaging Multimodal dataset are freely available for download here: https://fcon_1000.projects.nitrc.org/indi/retro/southwestuni_qiu_index.html. Data from the Social Reward and Nonsocial Reward Processing Across the Adult Lifespan: An Interim Multi-echo fMRI and Diffusion Dataset (referred to as TempleSocial in this paper) are freely available for download on OpenNEURO: https://openneuro.org/datasets/ds005123/versions/1.1.3. Data from UK Biobank (UKBB) are available upon request from https://www.ukbiobank.ac.uk/. Data from the UPennRisk dataset are freely available for download on OpenNEURO: https://openneuro.org/datasets/ds002843/versions/1.0.1. Data from the dataset titled "A longitudinal neuroimaging

dataset on language processing in children ages 5, 7, and 9 years old" (referred to as UTAustin579 in this paoer) are freely available for download on OpenNEURO: https://openneuro.org/datasets/ds003604/versions/1.0.7. Data from the Vanderbilt Memory and Aging Project (VMAP_JEFFERSON, VMAP_2.0, TN Aging Project) are available for download upon request from https://vmacdata.org/vmap/data-requests. Data from the Wisconsin Registry for Alzheimer's Prevention (WRAP) are available upon request from https://wrap.wisc.edu/data-requests-2/. Data from the HEALthy Brain and Child Development (HBCD) Study are available for download upon request from (https://hbcdstudy.org/data-sharing/). Data from the Adolescent Brain Cognitive Development (ABCD) Study are available for download upon request from (https://abcdstudy.org/scientists/data-sharing/). Data used in this paper from the Ageility Project (Phase 1 only) are available for download from (https://www.nitrc.org/projects/age-ility). Data from the Early Brain Development in Twins (EBDT) dataset are available for download from the National Institutes of Mental Health data archive upon request from https://nda.nih.gov/edit_collection.html?id=2384. Data from the Bipolar & Schizophrenia Consortium for Parsing Intermediate Phenotypes dataset (BSNIP1) and its renewal (BSNIP2) are available for download from the National Institutes of Mental Health data archive upon request from https://nda.nih.gov/edit_collection.html?id=2274 and https://nda.nih.gov/edit_collection.html?id=2165. Data from the Developing Human Connectome Project (dHCP) are available for download from the National Institutes of Mental Health data archive upon request from https://nda.nih.gov/edit_collection.html?id=3955. Vanderbilt University data (MORGAN, BABIES-ABC, CUTTING, VUMC-ASD) subject to third party restrictions. Please contact corresponding authors for data requests.

Author Contribution Statement

M.E.K. performed the bulk of the data processing and organization, performed the quality control, conducted analyses, created figures, created the software for power analyses and the associated Docker image, and wrote and edited the paper and Supplementary Material. G.R., A.M.S., D.C.M., and B.A.L. contributed to discussions about the methodology for the paper. C.G., K.R., P.K., N.R.N., G.R., and Z.L. helped with the processing and preprocessing of the data. Y.L. and L.Z. helped with the creation of the figures. Y.H., D.B.A., T.J.H., L.T.D., K.G.S., D.C.M., and B.A.L. provided a supervisory role for this project. P.Z. provided direction on the design of the statistical analyses. All other authors, including those mentioned above, made substantial contributions to the conception or design of the work or data acquisition. All authors provided substantial revisions to the paper.

Acknowledgments

This work was supported in part by the National Institutes of Health (NIH) through NIH awards K01-EB032898 (Schilling) and K01-AG073584 (Archer), Grant Nos. 1R01EB017230-01A1 (Landman), K24-AG046373 (Jefferson), and ViSE/VICTR VR3029, UL1-TR000445, and UL1-TR002243. This work was supported by the National Institute on Aging (NIA), through Grant Nos. R01-AG034962 (Jefferson), and R01-AG056534 (Jefferson), R01-AG062826; by the Alzheimer's Association IIRG-08-88733 (Jefferson); and by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), through Grant No. R01-HD114489 (Vinci-Booher). This work was supported by the Alzheimer's Disease Sequencing Project Phenotype Harmonization Consortium (ADSP-PHC) that is funded by NIA Grant Nos. U24 AG074855, U01 AG068057, and R01 AG059716. This work was supported by the Department of Defense through Grant No. HT94252410563. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education (ACCRE) at Vanderbilt University, Nashville, Tennessee. We appreciate the NIH S10 Shared Instrumentation grant 1S10OD020154-01, and grant 1S10OD023680-01 (Vanderbilt's High-Performance Computer Cluster for Biomedical Research). This work was also supported in part by Intramural Research Program of the National Institute on Aging, NIH. BABIES and ABC datasets were supported by the Jacobs Foundation Early Career Research Fellowship (2017-1261-05) (Humphreys); National Science Foundation CAREER Award (2042285) (Humphreys); Brain and Behavior Research Foundation John and Polly Sparks Foundation Investigator Award (29593) (Humphreys); Vanderbilt Institute for Clinical and Translational Research Grant (VR53419) (Humphreys); Vanderbilt Strong Grant; Vanderbilt Kennedy Center Grant (Humphreys); National Institute of Mental Health (R01MH129634) (Humphreys). This work was supported by the NIH National Institute of Neurological Disorders and Stroke (NINDS) through award numbers R01 NS108445 and R01 NS110130 (Morgan), and R01 NS134625 and R01 NS112252 (Englot). Data collected from the MORGAN dataset were acquired at Vanderbilt University, with the aim

of studying cognitive patterns in patients with epilepsy both before and after clinical treatment. This work was supported by the NIH NICHD through Grant Nos. F31 HD104385 (Nguyen), P50HD103537 (PI: Neul), R37 HD095519 MERIT Award, R01 HD089474, R01 HD109151, R01 HD067254, and R01 HD044073 (Cutting). This work was also supported by the NIH NINDS through Grant No. R01 NS049096 (Cutting). This research was supported in part by the Intramural Research Program of the NIH. The contributions of the NIH author(s) were made as part of their official duties as NIH federal employees, are in compliance with agency policy requirements, and are considered works of the United States Government. However, the findings and conclusions presented in this paper are those of the author(s) and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services. Research reported in this publication was supported by the National Institute of General Medical Sciences (NIGMS) of the NIH, under award numbers T32GM007347 and T32GM152284. Data and/or research tools used in the preparation of this paper were obtained from the National Institute of Mental Health (NIMH) Data Archive (NDA). NDA is a collaborative informatics system created by the National Institutes of Health to provide a national resource to support and accelerate research in mental health. This paper reflects the views of the authors and may not reflect the opinions or views of the NIH or of the Submitters submitting original data to NDA. The Pediatric Imaging, Neurocognition, and Genetics (PING) dataset was collected and released openly to contribute to the assessment of typical brain development in a pediatric sample (RC2DA029475-01), (https://www.sciencedirect.com/science/article/pii/S1053811915003572). The data used in this study come from the Human Connectome Project, which aims to map the structural connections and circuits of the brain and their relationships to behavior by acquiring high-quality magnetic resonance images. We used diffusion MRI data from the Baby Connectome Project (HCPBaby), the Human Connectome Project Development (HCPD) study, the Human Connectome Project Young Adult (HCP) study, and the Human Connectome Project Aging (HCPA) study. Research data from the Infant Brain Imaging Study (IBIS) dataset come from the IBIS Autism project, a collaborative effort by investigators to conduct a longitudinal MRI/DTI and behavioral study of infants at high risk for autism (R01HD055741-01). The Enhanced NKI-RS is a large cross-sectional sample of brain development, maturation and aging, that is currently funded by the NIMH (BRAINS R01MH094639-01; PI Milham) and Child Mind Institute (PI Milham) to characterize 1000 community-ascertained participants using state-of-the-art multiplex imaging-based resting state fMRI (R-fMRI) and diffusion tensor imaging (DTI), genetics, and a broad neurobehavioral phenotypic characterization protocol. Data were acquired from the DSI studio website: (https://brain.labsolver.org/nki_rockland.html). The Healthy Brain Network (HBN) is an ongoing initiative focused on building a biobank of data from 10,000 children and adolescents (ages 5-21) in the New York City area (https://www.nature.com/articles/sdata2017181). Data were acquired from the DSI studio website: (https://brain.labsolver.org/hbn.html). Data collection and sharing for this project was provided by the Cambridge Centre for Ageing and Neuroscience (CamCAN). CamCAN funding was provided by the UK Biotechnology and Biological Sciences Research Council (Grant No. BB/H008217/1), together with support from the UK Medical Research Council and University of Cambridge, UK. Data used in the preparation of this work were obtained from the CamCAN repository (available at http://www.mrc-cbu.cam.ac.uk/datasets/camcan/). The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) is a large-scale collaborative research project at the University of Cambridge. VMAP data were collected by Vanderbilt Memory and Alzheimer's Center Investigators at Vanderbilt University Medical Center. VMAP began in 2012 with the goal of investigating vascular health and brain aging. BLSA is a prospective cohort study with continuous enrollment that began in 1958. Comprehensive data from BLSA are available upon request by a proposal submission through the cohort website (www.blsa.nih.gov). The BLSA is supported by the Intramural Research Program of the National Institute on Aging, NIH. The BIOCARD study is designed to identify biomarkers associated with progression from normal cognitive status to cognitive impairment or dementia, with a particular focus on Alzheimer's Disease. Data collection and sharing for ADNI were supported by National Institutes of Health Grant U01-AG024904 and Department of Defense (award number W81XWH-12-2-0012). ADNI is also funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.,; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research;

Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public private partnership, led by Principal Investigator Michael W. Weiner, MD. The original goal of ADNI was to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). The current goals include validating biomarkers for clinical trials, improving the generalizability of ADNI data by increasing diversity in the participant cohort, and to provide data concerning the diagnosis and progression of Alzheimer's disease to the scientific community. For up-to-date information, see adni.loni.usc.edu. Research reported in this publication as a part of the Health and Aging Brain Study: Health Disparities (HABS-HD) was supported by the National Institute on Aging of the National Institutes of Health under Award Numbers R01AG054073 and R01AG058533, R01AG070862, P41EB015922, and U19AG078109. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The NACC database is funded by NIA/NIH Grant U24 AG072122. NACC data are contributed by the NIA-funded ADRCs: P30 AG062429 (PI James Brewer, MD, PhD), P30 AG066468 (PI Oscar Lopez, MD), P30 AG062421 (PI Bradley Hyman, MD, PhD), P30 AG066509 (PI Thomas Grabowski, MD), P30 AG066514 (PI Mary Sano, PhD), P30 AG066530 (PI Helena Chui, MD), P30 AG066507 (PI Marilyn Albert, PhD), P30 AG066444 (PI John Morris, MD), P30 AG066518 (PI Jeffrey Kaye, MD), P30 AG066512 (PI Thomas Wisniewski, MD), P30 AG066462 (PI Scott Small, MD), P30 AG072979 (PI David Wolk, MD), P30 AG072972 (PI Charles DeCarli, MD), P30 AG072976 (PI Andrew Saykin, PsyD), P30 AG072975 (PI David Bennett, MD), P30 AG072978 (PI Ann McKee, MD), P30 AG072977 (PI Robert Vassar, PhD), P30 AG066519 (PI Frank LaFerla, PhD), P30 AG062677 (PI Ronald Petersen, MD, PhD), P30 AG079280 (PI Eric Reiman, MD), P30 AG062422 (PI Gil Rabinovici, MD), P30 AG066511 (PI Allan Levey, MD, PhD), P30 AG072946 (PI Linda Van Eldik, PhD), P30 AG062715 (PI Sanjay Asthana, MD, FRCP), P30 AG072973 (PI Russell Swerdlow, MD), P30 AG066506 (PI Todd Golde, MD, PhD), P30 AG066508 (PI Stephen Strittmatter, MD, PhD), P30 AG066515 (PI Victor Henderson, MD, MS), P30 AG072947 (PI Suzanne Craft, PhD), P30 AG072931 (PI Henry Paulson, MD, PhD), P30 AG066546 (PI Sudha Seshadri, MD), P20 AG068024 (PI Erik Roberson, MD, PhD), P20 AG068053 (PI Justin Miller, PhD), P20 AG068077 (PI Gary Rosenberg, MD), P20 AG068082 (PI Angela Jefferson, PhD), P30 AG072958 (PI Heather Whitson, MD), P30 AG072959 (PI James Leverenz, MD). This research has been conducted using the UK Biobank resource, application 16315. Data contributed from MAP/ROS/MARS was supported by NIA R01AG017917, P30AG10161, P30AG072975, R01AG022018, R01AG056405, UH2NS100599, UH3NS100599, R01AG064233, R01AG15819 and R01AG067482, and the Illinois Department of Public Health (Alzheimer's Disease Research Fund). Data can be accessed at www.radc.rush.edu. More information about participant demographics and study information can be found here: https://www.rushu.rush.edu/research-rush-university/departmental-research/rush-alzheimers-disease-center/rush-alzheimers-disease-center-research/epidemiologic-research. The data contributed from the Wisconsin Registry for Alzheimer's Prevention was supportedby NIA AG021155, AG0271761, AG037639, and AG054047. Data collection and sharing for this project was provided by the Centre for Attention, Learning and Memory (CALM). CALM funding was provided by the UK Medical Research Council and University of Cambridge, UK. Data used in the preparation of this work were obtained from CALM resource – https://calm.mrc-cbu.cam.ac.uk/. The study protocol is reported in Holmes et al. (2019). Data used in the preparation of this work were obtained from the International Consortium for Brain Mapping (ICBM) database (www.loni.usc.edu/ICBM). The ICBM project (Principal Investigator John Mazziotta, MD, PhD, University of California, Los Angeles) is supported by the National Institute of Biomedical Imaging and BioEngineering. ICBM is the result of efforts of co-investigators from UCLA, Montreal Neurologic Institute, University of Texas at San Antonio, and the Institute of Medicine, Juelich/Heinrich Heine University - Germany. Data collection and sharing for this project was provided by the International Consortium for Brain Mapping (ICBM; Principal Investigator: John

Mazziotta, MD, PhD). ICBM funding was provided by the National Institute of Biomedical Imaging and BioEngineering. ICBM data are disseminated by the Laboratory of Neuro Imaging at the University of Southern California. The dataset that we refer to as UTAustin in this paper is a longitudinal neuroimaging dataset on language processing in children ages 5, 7, and 9 years old collected at The University of Texas at Austin. The dataset is openly available from OpenNEURO here: https://openneuro.org/datasets/ds003604/versions/1.0.7. We use version 1.0.7. The Queensland Twin Adolescent Brain (QTAB) Project was established with the purpose of promoting the conduct of health-related research in adolescence. The QTAB dataset comprises multimodal neuroimaging, as well as cognitive and mental health data collected in adolescent twins over two sessions. The MRI protocol consisted of T1-weighted (MP2RAGE), T2-weighted, FLAIR, high-resolution TSE, SWI, resting-state fMRI, DWI, and ASL scans. The QTAB project resource was produced as a result of i) the goodwill and contribution of 422 twin/triplet participants and their parents, ii) funding from the National Health and Medical Research Council, Australia (APP1078756) and the Queensland Brain Institute, University of Queensland, iii) access to several key resources, including the Centre for Advanced Imaging, the Human Studies Unit, Institute of Molecular Bioscience, and the Queensland Cyber Infrastructure Foundation, at the University of Queensland, local and national twin registries at the QIMR Berghofer Medical Research Institute and Twin Research Australia, as well as the many assessments made available by researchers worldwide, and iv) was established with the purpose of promoting the conduct of health-related research in adolescence. The imaging data and basic demographics are openly accessibly on OpenNEURO here: https://openneuro.org/datasets/ds004146/versions/1.0.4. We use version 1.0.4. The Social Reward and Nonsocial Reward Processing Across the Adult Lifespan: An Interim Multi-echo fMRI and Diffusion Dataset (referred to as TempleSocial in this paper) comes from a study that aims to investigate whether older adults have a blunted response to some features of social reward. The dataset is openly accessibly on OpenNEURO here: https://openneuro.org/datasets/ds005123/versions/1.1.3. In particular, we use version 1.1.3. The Longitudinal Brain Correlates of Multisensory Lexical Processing in Children study (shortened to Lexical in this paper) aims to explore developmentally dependent changes in lexical processing for adolescents. The dataset is openly accessibly on OpenNEURO here: https://openneuro.org/datasets/ds001894/versions/1.4.2. We use version 1.4.2. The UCLA Consortium for Neuropsychiatric Phenomics LA5c Study (UCLA) is focused on understanding the dimensional structure of memory and cognitive control (response inhibition) functions in both healthy individuals and individuals with neuropsychiatric disorders including schizophrenia, bipolar disorder, and attention deficit/hyperactivity disorder. Neuroimaging data were downloaded from OpenNEURO here: https://openneuro.org/datasets/ds000030/versions/1.0.0. We use version 1.0.0. The dataset we refer to as UPennRisk comes from a study at University of Pennsylvania that investigated whether training executive cognitive function could influence choice behavior and brain responses. Neuroimaging data were downloaded from OpenNEURO here: https://openneuro.org/datasets/ds002843/versions/1.0.1. We use version 1.0.1. The Dallas Lifespan Brain Study (DLBS) is a longitudinal multi-modal neuroimaging study of the aging mind, which was initiated in 2008 (referred to as Wave 1). Participants returned for two additional waves of data collection with an approximate interval of 4-5 years between waves. The DLBS protocol encompasses various imaging modalities, including structural MRI, diffusion MRI, and functional MRI, as well as comprehensive cognitive and psychosocial assessments. DLBS data can be downloaded from OpenNEURO here: https://openneuro.org/datasets/ds004856. Specifically, we use version 1.2.0. The Multisite, Multiscanner, and Multisubject Acquisitions for Studying Variability in Diffusion Weighted Magnetic Resonance Imaging (MASiVar) dataset consists of 319 diffusion scans acquired at 3T from b = 1000 to 3000 s/mm2 across 14 healthy adults, 83 healthy children (5 to 8 years), three sites, and four scanners curated to promote investigation of diffusion MRI variability. In particular, we used the data coming only from healthy children (Cohort IV) for version 2.0.2 of the dataset. Data are available to download from OpenNEURO here: https://openneuro.org/datasets/ds003416/versions/2.0.2. Data coming from the Southwestern University (SWU) dataset, part of the Consortium for Reliability and Reproducibility (CoRR), were downloaded via NITRC-IR from the 1000 Functional Connectomes Project. Specifically, the data come from the Emotion and Creativity One Year Retest Dataset subset, comprised of 235 subjects, all of whom were college students. Each subject underwent two sessions of anatomical, resting state fMRI, and DTI scans, spaced one year apart. In order to access the CoRR datasets through NITRC, users must be logged into NITRC at the time of download and registered with the 1000 Functional Connectomes Project / INDI website. More information about this subset can be found here: https://fcon_1000.projects.nitrc.org/indi/CoRR/html/swu_4.html. The Preschool MRI study in The Developmental Neuroimaging Lab at the University of Calgary (https://www.developmentalneuroimaginglab.ca) uses different magnetic resonance imaging (MRI) techniques to study brain structure and function in early childhood. The study

aims to characterize brain development in early childhood, and to offer baseline data that can be used to understand cognitive and behavioural development, as well as to identify deviations from normal development in children with various diseases, disorders, or brain injuries. The MRI techniques used include diffusion tensor imaging (DTI), anatomical imaging, arterial spin labeling (ASL), and resting state functional MRI (rsfMRI). Data can be downloaded from here: https://osf.io/axz5r/. The Amsterdam Open MRI Collection (AOMIC) is a collection of three datasets with multimodal (3T) MRI data including structural (T1-weighted), diffusion-weighted, and (resting-state and task-based) functional BOLD MRI data, as well as detailed demographics and psychometric variables from a large set of healthy participants. All raw data is publicly available from the OpenNEURO data sharing platform: ID1000: https://openneuro.org/datasets/ds003097, PIOP1: https://openneuro.org/datasets/ds002785, PIOP2: https://openneuro.org/datasets/ds002790. We use version 1.2.1 for ID1000 and 2.0.0 for PIOP1 and PIOP2. The Boston Adolescent Neuroimaging of Depression and Anxiety (BANDA) is a study of 215 adolescents ages 14-17, 152 of whom had a current diagnosis of a DSM-5 (APA, 2013) anxious and/or depressive disorder. The BANDA study collected a rich dataset of brain, clinical, and cognitive/neuropsychological measures from these adolescent subjects. The dataset is available to download upon request on the NDA. We thank Knight ADRC for providing neuroimaging data to us (WASHU dataset). The data contributed through WASHU (Knight ADRC) was supported by Grant Nos. P30 AG066444, P01 AG03991, and P01 AG026276. For WASHU, Clinical Dementia Ratings (CDRs) are obtained from assessments by experienced clinicians trained in the use of the CDR. The NACC database is funded by NIA/NIH Grant No. U24 AG072122. SCAN is a multiinstitutional project that was funded as a U24 grant (AG067418) by the National Institute on Aging in May 2020. Data collected by SCAN and shared by NACC are contributed by the NIA-funded ADRCs as follows: Arizona Alzheimer's Center - P30 AG072980 (PI: Eric Reiman, MD); R01 AG069453 (PI: Eric Reiman (contact), MD); P30 AG019610 (PI: Eric Reiman, MD); and the State of Arizona which provided additional funding supporting our center; Boston University - P30 AG013846 (PI Neil Kowall MD); Cleveland ADRC - P30 AG062428 (James Leverenz, MD); Cleveland Clinic, Las Vegas – P20AG068053; Columbia - P50 AG008702 (PI Scott Small MD); Duke/UNC ADRC – P30 AG072958; Emory University - P30AG066511 (PI Levey Allan, MD, PhD); Indiana University - R01 AG19771 (PI Andrew Saykin, PsyD); P30 AG10133 (PI Andrew Saykin, PsyD); P30 AG072976 (PI Andrew Saykin, PsyD); R01 AG061788 (PI Shannon Risacher, PhD); R01 AG053993 (PI Yu-Chien Wu, MD, PhD); U01 AG057195 (PI Liana Apostolova, MD); U19 AG063911 (PI Bradley Boeve, MD); and the Indiana University Department of Radiology and Imaging Sciences; Johns Hopkins - P30 AG066507 (PI Marilyn Albert, Phd.); Mayo Clinic - P50 AG016574 (PI Ronald Petersen MD PhD); Mount Sinai - P30 AG066514 (PI Mary Sano, PhD); R01 AG054110 (PI Trey Hedden, PhD); R01 AG053509 (PI Trey Hedden, PhD); New York University - P30AG066512-01S2 (PI Thomas Wisniewski, MD); R01AG056031 (PI Ricardo Osorio, MD); R01AG056531 (PIs Ricardo Osorio, MD; Girardin Jean-Louis, PhD); Northwestern University - P30 AG013854 (PI Robert Vassar PhD); R01 AG045571 (PI Emily Rogalski, PhD); R56 AG045571, (PI Emily Rogalski, PhD); R01 AG067781, (PI Emily Rogalski, PhD); U19 AG073153, (PI Emily Rogalski, PhD); R01 DC008552, (M.-Marsel Mesulam, MD); R01 AG077444, (PIs M.-Marsel Mesulam, MD, Emily Rogalski, PhD); R01 NS075075 (PI Emily Rogalski, PhD); R01 AG056258 (PI Emily Rogalski, PhD); Oregon Health and Science University - P30 AG008017 (PI Jeffrey Kaye MD); R56 AG074321 (PI Jeffrey Kaye, MD); Rush University - P30 AG010161 (PI David Bennett MD); Stanford – P30AG066515; P50 AG047366 (PI Victor Henderson MD MS); University of Alabama, Birmingham – P20; University of California, Davis - P30 AG10129 (PI Charles DeCarli, MD); P30 AG072972 (PI Charles DeCarli, MD); University of California, Irvine - P50 AG016573 (PI Frank LaFerla PhD); University of California, San Diego - P30AG062429 (PI James Brewer, MD, PhD); University of California, San Francisco - P30 AG062422 (Rabinovici, Gil D., MD); University of Kansas - P30 AG035982 (Russell Swerdlow, MD); University of Kentucky - P30 AG028283-15S1 (PIs Linda Van Eldik, PhD and Brian Gold, PhD); University of Michigan ADRC - P30AG053760 (PI Henry Paulson, MD, PhD) P30AG072931 (PI Henry Paulson, MD, PhD) Cure Alzheimer's Fund 200775 - (PI Henry Paulson, MD, PhD) U19 NS120384 (PI Charles DeCarli, MD, University of Michigan Site PI Henry Paulson, MD, PhD) R01 AG068338 (MPI Bruno Giordani, PhD, Carol Persad, PhD, Yi Murphey, PhD) S10OD026738-01 (PI Douglas Noll, PhD) R01 AG058724 (PI Benjamin Hampstead, PhD) R35 AG072262 (PI Benjamin Hampstead, PhD) W81XWH2110743 (PI Benjamin Hampstead, PhD) R01 AG073235 (PI Nancy Chiaravalloti, University of Michigan Site PI Benjamin Hampstead, PhD) 1I01RX001534 (PI Benjamin Hampstead, PhD) IRX001381 (PI Benjamin Hampstead, PhD); University of New Mexico - P20 AG068077 (Gary Rosenberg, MD); University of Pennsylvania - State of PA project 2019NF4100087335 (PI David Wolk, MD);

Rooney Family Research Fund (PI David Wolk, MD); R01 AG055005 (PI David Wolk, MD); University of Pittsburgh - P50 AG005133 (PI Oscar Lopez MD); University of Southern California - P50 AG005142 (PI Helena Chui MD); University of Washington - P50 AG005136 (PI Thomas Grabowski MD); University of Wisconsin - P50 AG033514 (PI Sanjay Asthana MD FRCP); Vanderbilt University – P20 AG068082; Wake Forest - P30AG072947 (PI Suzanne Craft, PhD); Washington University, St. Louis - P01 AG03991 (PI John Morris MD); P01 AG026276 (PI John Morris MD); P20 MH071616 (PI Dan Marcus); P30 AG066444 (PI John Morris MD); P30 NS098577 (PI Dan Marcus); R01 AG021910 (PI Randy Buckner); R01 AG043434 (PI Catherine Roe); R01 EB009352 (PI Dan Marcus); UL1 TR000448 (PI Brad Evanoff); U24 RR021382 (PI Bruce Rosen); Avid Radiopharmaceuticals / Eli Lilly; Yale - P50 AG047270 (PI Stephen Strittmatter MD PhD); R01AG052560 (MPI: Christopher van Dyck, MD; Richard Carson, PhD); R01AG062276 (PI: Christopher van Dyck, MD); 1Florida - P30AG066506-03 (PI Glenn Smith, PhD); P50 AG047266 (PI Todd Golde MD PhD). Data used in the preparation of this article were obtained from the HEALthy Brain and Child Development (HBCD) Study (https://hbcdstudy.org/), held in the NIH Brain Development Cohorts Data Sharing Platform. This is a multisite, longitudinal study designed to recruit approximately 7000 families and follow them from pregnancy to early childhood. The HBCD Study is supported by the NIH and additional federal partners under award numbers U01DA055352, U01DA055353, U01DA055366, U01DA055365, U01DA055362, U01DA055342, U01DA055360, U01DA055350, U01DA055338, U01DA055355, U01DA055363, U01DA055349, U01DA055361, U01DA055316, U01DA055344, U01DA055322, U01DA055369, U01DA055358, U01DA055371, U01DA055359, U01DA055354, U01DA055370, U01DA055347, U01DA055357, U01DA055367, U24DA055325, and U24DA055330. A full list of supporters is available at https://hbcdstudy.org/federal-partners/. A full list of participating sites is available at: https://hbcdstudy.org/recruitment-sites/. HBCD Study Consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This paper reflects the views of the authors and may not reflect the opinions or views of the NIH or the HBCD Study Consortium investigators. Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive DevelopmentSM (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9-10 and follow them over 10 years into early adulthood. The ABCD Study® is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This paper reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The Bipolar & Schizophrenia Consortium for Parsing Intermediate Phenotypes (BSNIP1; 10.15154/tnzs-a323) dataset (R01MH078113-01, R01MH077852-01, R01MH077851-01, R01MH077945-01, R01MH077862-01) and its renewal (BSNIP2; 10.15154/8v3w-et72) (R01MH103368-01, R01MH103366-01) were collected with the aim to improve diagnosis and clinical management of psychosis by defining biologically based biotypes that better predict symptoms, course, and treatment response across major psychoses. The Early Brain Development in Twins (EBDT) dataset was collected to understand the role of genetic and environmental contributions to brain structure and function in the crucial period of development that is early childhood (U01MH070890-11). (10.15154/1822-fs72). The Age-ility Project (Phase 1) dataset was collected with the aim of developing novel analysis approaches that integrate information across multiple brain imaging modalities (e.g., dMRI, fMRI, electrophysiology) to build a more complete picture of how brain networks are structurally and functionally organized to support cognition. The Developing Human Connectome Project (dHCP) is an open science study, funded by the European Research Council to obtain and disseminate Magnetic Resonance Imaging (MRI) data which map the brain's structural and functional development across the period from 20 weeks gestational age to full term. By coupling advances in imaging with bespoke solutions developed for the fetal and neonatal population, principally but not exclusively solving the problems of subject motion, the dHCP captures the development of brain anatomy and connectivity at a systems level (10.15154/92vw-g837). Data from the VUMC-ASD dataset are collected seeking to elucidate the neural mechanisms underlying atypical sensory processing in autism spectrum disorders (ASD), focusing on tactile and interoceptive sensitivity and their relationship to social and behavioral symptoms. By integrating behavioral,

neurophysiological, and neuroimaging approaches, the studies aim to identify how alterations in thalamocortical and salience networks contribute to sensory hypo- and hyper-responsiveness in ASD, ultimately informing early identification and novel intervention strategies. Collection of these data were supported by K01MH090232, R21MH101321, and R01MH102272. We use generative AI to create code segments based on task descriptions, as well as debug, edit, and autocomplete code. Additionally, generative AI technologies have been employed to assist in structuring sentences and performing grammatical checks. It is imperative to highlight that the conceptualization, ideation, and all prompts provided to the AI originate entirely from the authors' creative and intellectual efforts. All AI-generated content was reviewed and edited by the authors, who take full responsibility for the final content of the publication.

References

1. T. J. Cole, "The development of growth references and growth charts," *Ann. Hum. Biol.* **39**, 382–394 (2012).
2. R. A. I. Bethlehem et al., "Brain charts for the human lifespan," *Nature* **604-604**, 7906-525–533 (2022).
3. S. Rutherford et al., "Charting brain growth and aging at high spatial precision," *Elife* **11**, e72904 (2022).
4. R. A. I. Bethlehem et al., "A normative modelling approach reveals age-atypical cortical thickness in a subgroup of males with autism spectrum disorder," *Commun. Biol.* **3**(1), 486 (2020).
5. A. Lawry Aguila et al., "Deep normative modelling reveals insights into early-stage Alzheimer's disease using multi-modal neuroimaging data," *Alzheimers Res. Ther.* **17**, 107 (2025).
6. A. F. Marquand et al., "Conceptualizing mental disorders as deviations from normative functioning," *Mol. Psychiatry* **24**(10), 1415–1424 (2019).
7. R. Ge et al., "Normative modelling of brain morphometry across the lifespan with centileBrain: algorithm benchmarking and model optimisation," *Lancet Digital Health* **6**, e211–e221 (2024).
8. R. A. Rigby, D. M. Stasinopoulos, and P. W. Lane, "Generalized additive models for location, scale and shape," *J. R. Stat. Soc. Ser. C Appl. Stat.* **54**, 507–554 (2005).
9. E. Borghi et al., "Construction of the World Health Organization child growth standards: selection of methods for attained growth curves," *Stat. Med.* **25**, 247–265 (2006).
10. M. E. Kim et al., "White matter microstructure and macrostructure brain charts across the human lifespan," BioRxiv:2025.05.08.652953 (2025).
11. A. O. Muili et al., "Advancing brain health: harnessing the centilebrain model for enhanced diagnosis of mental health disorders," *Ann. Med. Surg.* **86**, 4944–4946 (2024).
12. M. Habes et al., "The brain chart of aging: machine-learning analytics reveals links between brain aging, white matter disease, amyloid burden, and cognition in the iSTAGING consortium of 10,216 harmonized MR scans," *Alzheimer's Dement.* **17**, 89–102 (2021).
13. S. Rutherford et al., "Evidence for embracing normative modeling," *Elife* **12**, e85082 (2023).
14. T. Lawn et al., "Normative modelling of molecular-based functional circuits captures clinical heterogeneity transdiagnostically in psychiatric patients," *Commun. Biol.* **7**, 689 (2024).
15. L. Sun et al., "Human lifespan changes in the brain's functional connectome," *Nat. Neurosci.* **28**, 891–901 (2025).
16. A. H. Zhu et al., "Lifespan reference curves for harmonizing multi-site regional brain white matter metrics from diffusion MRI," *Sci. Data* **12**(1), 1–18 (2025).
17. J. E. Villalón-Reina et al., "Lifespan normative modeling of brain microstructure," BioRxiv:2024.12.15.628527 (2026).
18. D. K. Jones, *Diffusion MRI: Theory, Methods, and Application*, Oxford University Press (2010).
19. D. S. Novikov et al., "Quantifying brain microstructure with diffusion MRI: theory and parameter estimation," *NMR Biomed.* **32**, e3998 (2018).
20. D. C. Alexander et al., "Imaging brain microstructure with diffusion MRI: practicality and applications," *NMR Biomed.* **32**, e3841 (2019).
21. F. C. Yeh, "Shape analysis of the human association pathways," *Neuroimage* **223**, 117329 (2020).
22. M. E. Kim et al., "Analytic bounds on GAMLSS model variability of normative white matter brain charts," BioRxiv:2026.01.28.702453 (2026).
23. M. E. Kim et al., "Scalable quality control on processing of large diffusion-weighted and structural magnetic resonance imaging datasets," *PLoS One* **20**, e0327388 (2025).
24. M. E. Kim et al., "Scalable, reproducible, and cost-effective processing of large-scale medical imaging datasets," *Proc. SPIE* **13411**, 134110O (2025).
25. L. Y. Cai et al., "PreQual: an automated pipeline for integrated preprocessing and quality assurance of diffusion weighted MRI images," *Magn. Reson. Med.* **86**, 456–470 (2021).
26. J. D. Tournier et al., "MRtrix3: a fast, flexible and open software framework for medical image processing and visualisation," *Neuroimage* **202**, 116137 (2019).
27. J. Wasserthal, P. Neher, and K. H. Maier-Hein, "TractSeg - fast and accurate white matter tract segmentation," *Neuroimage* **183**, 239–253 (2018).

28. E. Renauld et al., "Tractography analysis with the scilpy toolbox," *Apert. Neuro* **6** (2026).
29. M. E. Kim et al., "Empirical assessment of the assumptions of ComBat with diffusion tensor imaging," *J. Med. Imaging* **11**, 024011 (2024).
30. M. Kim and K. Schilling, "White matter microstructure and macrostructure brain charts across the human lifespan - models," (2025).
31. J. Carlos Alves Barata et al., "The moore–Penrose pseudoinverse: a tutorial review of the theory," *Braz. J. Phys.* **42**(1), 146–165 (2011).
32. G. W. Oehlert, "A note on the delta method," *Am. Stat.* **46**, 27–29 (1992).
33. T. J. Cole, "Sample size and sample composition for constructing growth reference centiles," *Stat. Methods Med. Res.* **30**, 488–507 (2020).
34. M. Grotheer et al., "White matter myelination during early infancy is linked to spatial gradients and myelin content at birth," *Nat. Commun.* **13**(1), 997 (2022).
35. A. M. Fjell and K. B. Walhovd, "Structural brain changes in aging: courses, causes and cognitive consequences," *Rev. Neurosci.* **21**, 187–221 (2010).
36. S. Kumar, P. R. O. Payne, and A. Sotiras, "Normative modeling using multimodal variational autoencoders to identify abnormal brain volume deviations in Alzheimer's disease," *Proc. SPIE* **12465**, 1246503 (2023).
37. S. Bogdanov et al., "Lifespan trajectories of asymmetry in white matter tracts," *Hum. Brain Mapp.* **47**, e70519 (2026).
38. S. Frangou et al., "Cortical thickness across the lifespan: data from 17,075 healthy individuals aged 3-90 years," *Hum. Brain Mapp.* **43**, 431–451 (2021).
39. A. Mousley et al., "Topological turning points across the human lifespan," *Nat. Commun.* **16**(1), 10055 (2025).
40. V. de Padua V. Alves et al., "Computed tomography-based measurements of normative liver and spleen volumes in children," *Pediatr. Radiol.* **53**(3), 378–386 (2022).
41. L. W. Remedios et al., "Lifespan pancreas morphology for control versus type 2 diabetes using AI on largescale clinical imaging," *Clin. Anat.* 1–12 (2026).

Michael E. Kim is a computer science PhD candidate at Vanderbilt University. He graduated from Hamilton College in 2021 with a degree in biochemistry and molecular biology and received his master's degree in computer science from Vanderbilt University in 2025. His current research interests include medical image analysis with applications to the clinical domain.

Biographies of the other authors are not available.