PNAS RESEARCH ARTICLE | POPULATION BIOLOGY OPEN ACCESS

Empirical validation of race-neutral normative brain morphometry models across ethnoracially diverse populations

Ruiyang Ge[a], Yuetong Yu[a], Faye New[b], Shalaila S. Haas[b], Nicole Sanford[a], Kevin Yu[a], Paul Allen[c], Seda Arslan[d,e], Mihai Avram[f], Stefan Borgwardt[f], Nicolas A. Crossley[g], Camilo de la Fuente-Sandoval[h], Masaki Fukunaga[i], Jia-Hong Gao[j], Alfonso Gonzalez-Valderrama[k,l], Ryota Hashimoto[m], Felice Iasevoli[n], Daniel Keeser[o,p,q], Kader Kubat[e,r], Veena Kumari[s,t], Junya Matsumoto[u], Urvakhsh M. Mehta[v], Kiyotaka Nemoto[w], Giuseppe Pontillo[x], Florian J. Raabe[o,y], Francisco Reyes-Madrigal[h], Neelabja Roy[y], Didenur Şahin-Çevik[e,r], Tuba Sahin-Ilikoglu[e,r], Timothea Toulopoulou[b,d,e,r,z], Elias Wagner[aa,bb], Guoyuan Yang[cc], Mariana Zurita[c], Paul M. Thompson[dd], and Sophia Frangou[a,b,1]

Affiliations are included on p. 7.

Edited by David C. Van Essen, Washington University in St. Louis School of Medicine, St. Louis, MO; received September 2, 2025; accepted March 18, 2026

Normative models of brain morphometry quantify individual deviations from typical anatomical patterns and hold promise for enhancing clinical decision-making. However, their clinical utility depends critically on demonstrating generalizability across diverse ethnoracial populations. We previously developed sex-specific, race-neutral normative models for cortical thickness, surface area, and subcortical volumes using brain scans from a large international sample of healthy individuals, as part of the CentileBrain Project, a global initiative to provide open-access, neuroimaging reference models. The primary aim of the present study was to empirically evaluate the generalizability and accuracy of these pretrained models across multiple ethnoracial groups. To this end, we tested model performance in independent samples of healthy individuals from Africa, Asia, Europe, and the Americas, with ethnoracial classification defined either by self-identification or genetic ancestry (N = 4,862). We further compared performance against normative models developed exclusively from a single-population Chinese cohort. Across all groups, as well as in the pooled sample, the pretrained CentileBrain models demonstrated consistently high accuracy, with relative mean absolute error values below 10% for subcortical volume and surface area and below 5% for cortical thickness. Model performance was highly concordant across self-identified and ancestry-defined groups. In a separate analysis, the CentileBrain models performed comparably to a population-specific model when applied to an independent ancestry-matched sample. These findings provide empirical support for the generalizability of race-neutral normative models developed on large and diverse samples and underscore their potential utility for individualized neuroimaging assessment across ethnoracially diverse populations.

normative models | human | neuroimaging | brain morphometry

Significance

Normative models of brain structure are increasingly used to assess brain health, development, and aging by comparing an individual's brain with age- and sex-expected patterns. Yet their performance across diverse populations is largely untested, limiting confidence in their broad application. Here, we demonstrate that race-neutral models of brain morphometry, trained on a large, ethnically diverse dataset, retain high accuracy in multiple independent samples from distinct ethnoracial groups defined by self-report or ancestry. These findings show that explicit ethnoracial stratification is not required when training data are sufficiently large and globally inclusive. By establishing robust cross-population performance, this work provides a scalable and scientifically grounded framework for developing equitable, generalizable brain-based tools for research and ultimately clinical use.

Normative modeling of neuroimaging-derived brain morphometry provides a framework for quantifying individual-level deviations from typical neurostructural patterns and offers potential clinical relevance in both research and translational contexts (1–7). In recent years, there has been a proliferation of such models for brain morphometry (8–14). However, the utility of normative models depends critically on their generalizability across populations.

A major challenge in the field is that most neuroimaging research to date has relied heavily on samples composed predominantly of individuals identifying as white (15), despite increasing calls to improve ethnoracial representation in biomedical research (16–18). In response to concerns about the limited generalizability of existing models, some normative modeling initiatives have focused on constructing population-specific frameworks using data from a single ethnoracial group, for example, normative models developed exclusively from Chinese cohorts (14). Moreover, although race and ethnicity often correlate with genetic ancestry (19, 20), these constructs are imprecise and may inadvertently reinforce historical and structural inequities if used uncritically (21). Growing scrutiny of race-based adjustments in clinical algorithms underscores this concern, as such adjustments have been found to offer limited benefit and may lead to adverse outcomes for some groups (22, 23).

In light of these concerns, a key unresolved question is whether race-neutral normative models—developed without race-based stratification—can provide a viable and broadly

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: sophia.frangou@mssm.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2521055123/-/DCSupplemental.

Published May 12, 2026.

applicable framework. The CentileBrain Project (https://www.centilebrain.org), a major initiative of the ENIGMA Consortium (24), that provides openly available, sex-specific race-neutral normative models based on a large, globally sourced sample, offers an opportunity to empirically evaluate this question. The CentileBrain race-neutral, sex-specific normative models of brain morphometry were derived using FreeSurfer-extracted measures of cortical thickness, surface area, and subcortical volumes from over 37,000 healthy individuals spanning early childhood to late adulthood. The development sample included individuals from Australia, East Asia, Europe, North America, and South Africa (13). Model construction was guided by extensive benchmarking to optimize algorithm performance, covariate selection, sample size, and site harmonization. The resulting models are openly available through the CentileBrain platform (https://centilebrain.org/) and are in use by more than 2,500 researchers over 60 countries.

In the current study, we tested the central premise of the CentileBrain initiative: that race-neutral normative models, when built on a large globally diverse sample, can generalize across populations not explicitly modeled. To this end, we evaluated the performance of the CentileBrain models of regional brain morphometry in independent datasets of healthy individuals from Africa, Asia, Europe, and the Americas, with ethnoracial identity defined either by self-report or high-confidence genetic ancestry, reflecting common classification approaches in biomedical and neuroimaging research. To further assess generalizability, we also compared the CentileBrain models to normative models developed exclusively from a single ethnoracial group. Across all analyses, the race-neutral models demonstrated consistently high accuracy, providing empirical support for their applicability across ethnoracially diverse populations.

Results

Neuroimaging Features. All analyses used anonymized data from healthy individuals with either self-reported or ancestry-defined (African, Admixed American, East Asian, or European) ethnoracial identities (*SI Appendix* and *SI Appendix*, Table S1). Neuroimaging features comprising regional cortical thickness, cortical surface area, and subcortical volumes were derived using FreeSurfer, based on the Desikan–Killiany and Aseg atlases, following rigorous quality control. These features were selected to align with those used in training the CentileBrain models, which were developed from entirely independent samples (13).

Concordance of Model Performance Across Ethnoracial Groups. The pretrained, sex-specific, race-neutral CentileBrain models were applied to these features separately in males and females within each ethnoracial group. Model performance was evaluated using two standard metrics of predictive accuracy—mean absolute error (MAE) and RMSE, which quantify the average absolute and squared deviations, respectively, between predicted and observed morphometric values. Results for self-reported and ancestry-defined ethnoracial groups presented in Datasets S1–S3 and Datasets S7–S9, and Datasets S4–S6 and Datasets S10–S12, respectively.

To assess the consistency of model performance across groups, we computed Pearson's correlation coefficients between group-specific MAE and RMSE values of each regional morphometric model and those derived from the original CentileBrain reference models. These correlations were uniformly high, ranging from 0.80 to 0.99, with all P-values < 0.001 and 95% CI ranging from [0.61, 0.94] to [0.99, 1] (Figs. 1 and 2 and *SI Appendix*, Figs. S1–S6). The figures illustrate the regional correspondence in prediction error patterns between each ethnoracial group and the reference models, underscoring the stability and robustness of model performance across ethnoracial definition methods and population subgroups.

Standardized Deviation Magnitude from Reference Model Performance. Building on the observed concordance in prediction error profiles, we further quantified the deviation of the regional morphometry model performance of each ethnoracial group relative to the CentileBrain reference models. Across all groups, whether defined by self-reported identity or genetic ancestry, the mean MAE and RMSE values for regional morphometric measures differed by no more than 0.5 SD from the corresponding values in the reference models. This narrow range of deviation supports the stability and consistent accuracy of the CentileBrain models when applied to previously unseen samples across ethnoracial classifications.

Normalized Prediction Error Across Brain Morphometry Domains. To facilitate comparison across ethnoracial groups and account for scale differences among brain morphometric features, we also computed the relative MAE (RMAE), defined as the MAE divided by the observed value for each regional morphometric measure. This normalization provides a dimensionless index of model accuracy, allowing for the evaluation of proportional error independent of absolute morphometric magnitude. RMAE values demonstrated consistently strong performance across all ethnoracial groups and morphometric domains. For subcortical volume and cortical surface area, mean RMAE values remained below 10% across all groups. For cortical thickness, mean RMAE values were even lower, remaining below 5% across all regions and groups (Fig. 3 and *SI Appendix*, Fig. S7 and Datasets S13–S18). These findings further support the generalizability of the CentileBrain models, indicating that prediction error remained proportionally low across brain regions and ethnoracial classifications.

Permutation-Based Evaluation of Model Generalizability. Across all 150 regional models, 12 external validation samples, and both sexes, prediction errors (MAE, RMSE, and RMAE) in the different ethnoracial samples were highly consistent with those observed in the CentileBrain reference models. The threshold of statistical significance was set as $P < 0.05$ following Holm family-wise error correction, which was applied separately across 450 comparisons (3 error measures × 150 brain regional models) for each ethnoracial sample. Of the total 10,800 comparisons undertaken, 93.4% were above the threshold (*SI Appendix*, Figs. S8 and S9). Among the limited number of comparisons below the threshold, deviations were sporadic and lacked spatial or metric-wise coherence, and the discrepancies between predicted and observed values were of small magnitude, consistently within one SD of the corresponding CentileBrain reference distribution. This sparse and unsystematic pattern of low-magnitude departures indicates that regional MAE, RMSE, and RMAE in the external ethnoracial validation samples are broadly comparable to, and do not systematically diverge from, CentileBrain reference performance.

Taken together, the convergence of standardized deviation magnitudes from the reference model performance, normalized prediction error profiles, and permutation-based statistical evaluations demonstrates that the CentileBrain normative models maintain stable and proportional predictive performance across ethnoracial groups and heterogeneous MRI datasets.

Normative Deviation Scores in Exemplar Regions. To visually illustrate the consistency of model predictions across ethnoracial groups, we focus on three exemplar brain regions: left thalamic

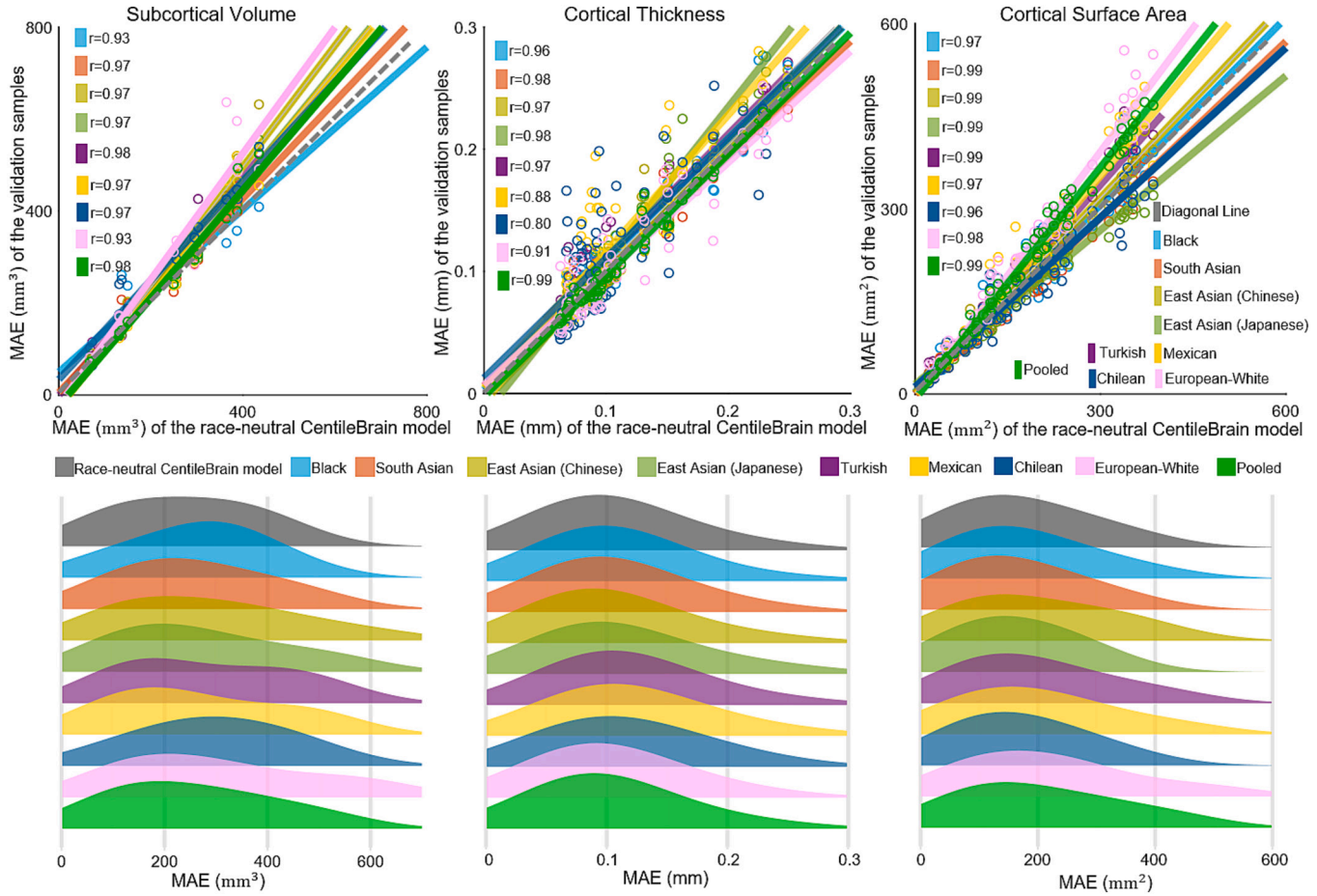

Fig. 1. *Top* panel: Scatterplot of the MAE values obtained from the self-reported ethnoracial groups and the corresponding values from the race-neutral CentileBrain model of female participants. *Lower* panel: Distributions of the MAE values across the 14 subcortical volumes and cortical thickness and surface area across the 68 cortical regions for the race-neutral CentileBrain model and each ethnoracial group of female participants. Each circle denotes a regional morphometric measure. Different colors indicate different ethnoracial groups. The Pearson's correlation coefficient (r) was computed between the MAE values of each ethnoracial group with those obtained in the race-neutral CentileBrain model. Additional information is provided in Datasets S1–S3. The results of RMSE were consistent with results of MAE, the results of RMSE are presented in *SI Appendix*, Fig. S1. In both sexes, the pattern identified was consistent for all region-specific models. The corresponding data for males are presented in *SI Appendix*, Figs. S2 and S3.

volume and left medial orbitofrontal cortical thickness and surface area. For each region, normative deviation scores (Z-scores) (*SI Appendix*, *Materials and Methods*) were computed using the CentileBrain reference models within each ethnoracial group. Across all groups, the mean Z-scores were close to zero (Fig. 4 and *SI Appendix*, Fig. S10), indicating minimal systematic deviation from normative expectations. These findings further support the stability and generalizability of the CentileBrain models across both morphometric modalities and population subgroups (Datasets S19–S21).

Model Performance in the Pooled Sample. To complement the group-specific analyses, we applied the CentileBrain models to the pooled study sample comprising all individuals across self-reported or ancestry-defined ethnoracial groups. Analyses in the pooled sample enable assessment of average model performance across heterogeneous populations and facilitate evaluation of consistency and potential bias in group-specific applications. Analyses in the pooled sample followed the same approach as those performed for individual ethnoracial groups, including evaluation of MAE, RMSE, and RMAE across all regional morphometric measures and both sexes. Model performance in the pooled sample was consistent with the patterns observed in the separate ethnoracial groups. MAE and RMSE values remained within the expected range based on the CentileBrain reference models, and RMAE values indicated low relative error across subcortical volumes, cortical surface area, and cortical thickness measures (Fig. 3 and *SI Appendix*, Fig. S7). These findings further support the robustness and scalability of the CentileBrain models when applied to heterogeneous populations.

Comparison with a Population-Specific Model. Comparing the accuracy of normative models derived from multiethnic datasets, such as CentileBrain, to those developed using data from a single ethnoracial group is critical for assessing their broader applicability. At present, the Chinese Brain Charts are the only available population-specific normative models for brain morphometry have been developed exclusively from Chinese individuals (N = 43,037 participants; age range = 0 to 100 y) (14) (also *SI Appendix*, *Materials and Methods*). To test their comparative performance, the Centilebrain and the Chinese Brain Charts models were applied to a new, previously unseen sample of Chinese individuals that had not been used in the development of either platform (N = 123; females = 77; age range = 47 to 76 y). Model performance was assessed using the RMAE and explained variance for regional predictions of cortical thickness, surface area, and subcortical volume. This is different from the preceding analyses where we tested the performance of the same platform

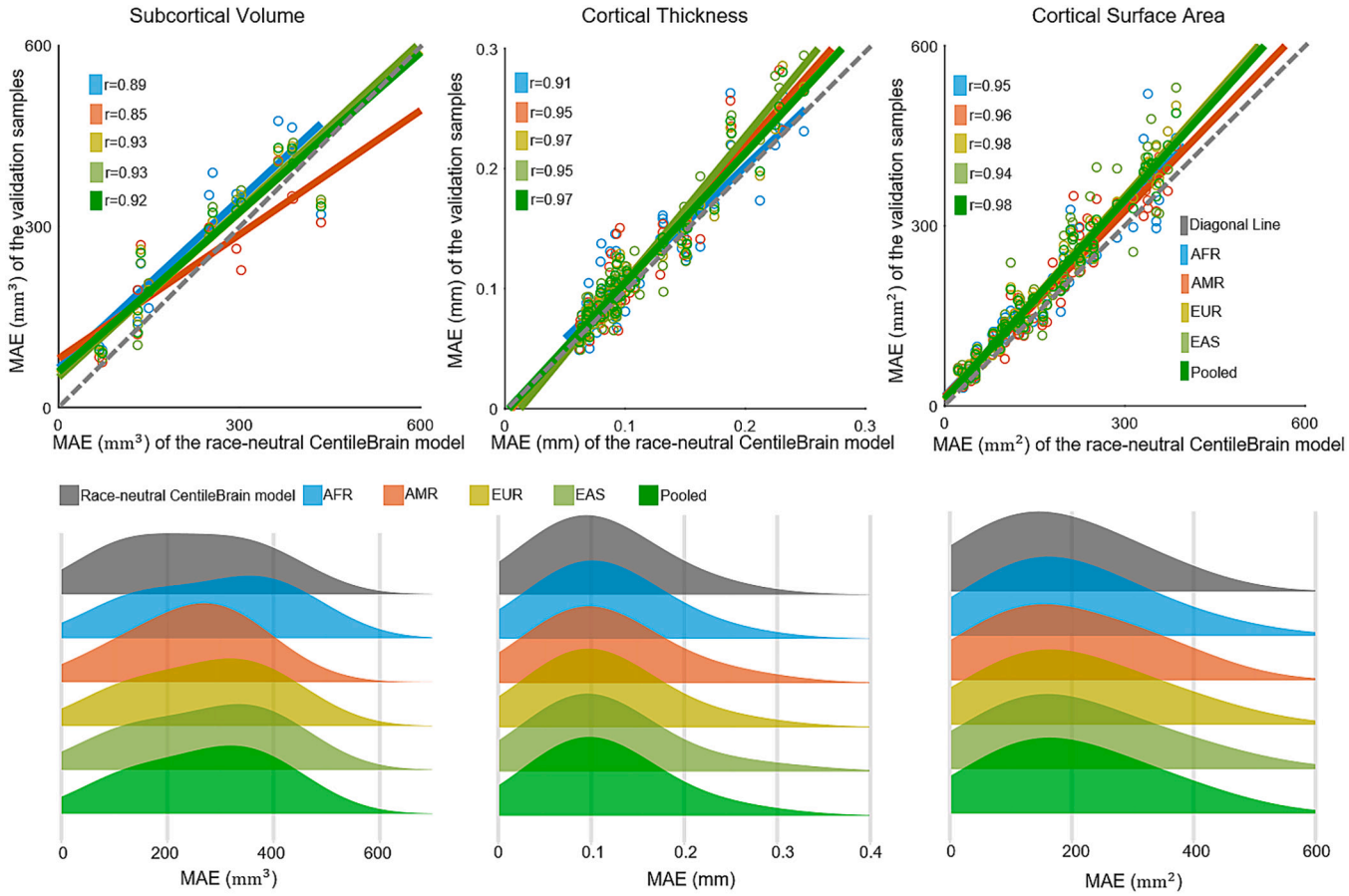

Fig. 2. *Top* panel: Scatterplot of the MAE values obtained from the genetic ancestry-determined ethnoracial groups and the corresponding values from the race-neutral CentileBrain model of female participants. *Lower* panel: Distributions of the MAE values across the 14 subcortical volumes and cortical thickness and surface area across the 68 cortical regions for the race-neutral CentileBrain model and each ethnoracial group of female participants. Each circle denotes a regional morphometric measure. Different colors indicate different ethnoracial groups. The Pearson's correlation coefficient (r) was computed between the MAE values of each ethnoracial group with those obtained in the race-neutral CentileBrain model. Additional information is provided in Datasets S4–S6. The results of RMSE were consistent with results of MAE, the results of RMSE are presented in *SI Appendix*, Fig. S4. In both sexes, the pattern identified was consistent for all region-specific models. The corresponding data for males are presented in *SI Appendix*, Figs. S5 and S6. AFR = African; AMR = Admixed American; EAS = East Asian; EUR = European.

(i.e., CentileBrain) in different ethnic populations. The models derived from CentileBrain and the Chinese Brain Charts produced comparable, results, with closely aligned RMAE values although explained variance was higher for the CentileBrain models across brain regions (*SI Appendix*, Fig. S11). These findings indicate that CentileBrain models, though developed from a globally sourced, multiethnic dataset, can match or exceed the performance of ancestry-specific frameworks, even in ancestry-matched samples.

Discussion

This study provides empirical support for the generalizability of race-neutral, sex-specific normative models of brain morphometry across a broad range of ethnoracial groups. Developed from a globally sourced sample including participants from five continents, the CentileBrain models achieved consistent accuracy when applied to previously unseen datasets defined by both self-identified race/ethnicity and genetically inferred ancestry. These findings help address a critical gap in the normative modeling literature relating to the lack of systematic evidence evaluating model performance across populations that vary in their ethnoracial composition.

The longstanding overrepresentation of white individuals in neuroimaging research (15–17) has raised fundamental concerns about the inclusiveness and translational relevance of derived models and metrics. While important efforts are now underway to improve ethnoracial diversity in neuroimaging cohorts (25–28), their impact remains constrained by uneven data availability and the absence of consensus on how race, ethnicity, and ancestry should be handled in statistical modeling (29). In this context, the present work represents a methodologically rigorous demonstration that models that intentionally exclude race as a covariate—but are trained on large, multiethnic samples—can generalize across diverse populations.

Our findings suggest that this race-neutral approach may offer a pragmatic and scientifically robust strategy for enhancing fairness and applicability without reifying socially constructed categories of human variation (30). While genetic ancestry classifications are grounded in shared patterns of genetic variation, they too represent simplified constructs that may obscure the continuous and admixed nature of human genetic diversity. Most individuals do not fall neatly into discrete ancestral groups but instead reflect varying degrees of admixture shaped by centuries of migration, intermarriage, and demographic shifts (30, 31). As such, even ancestry-based categories, though more biologically informed than race or ethnicity, can be reductive. To this point, we intentionally avoided analyses aiming to predict ethnoracial groups from brain morphometry data (32), as doing so risks implying biological determinism for categories that are socially and historically constructed. We also chose not to include contextual variables—such as childhood adversity, socioeconomic status, or exposure to structural inequalities—as

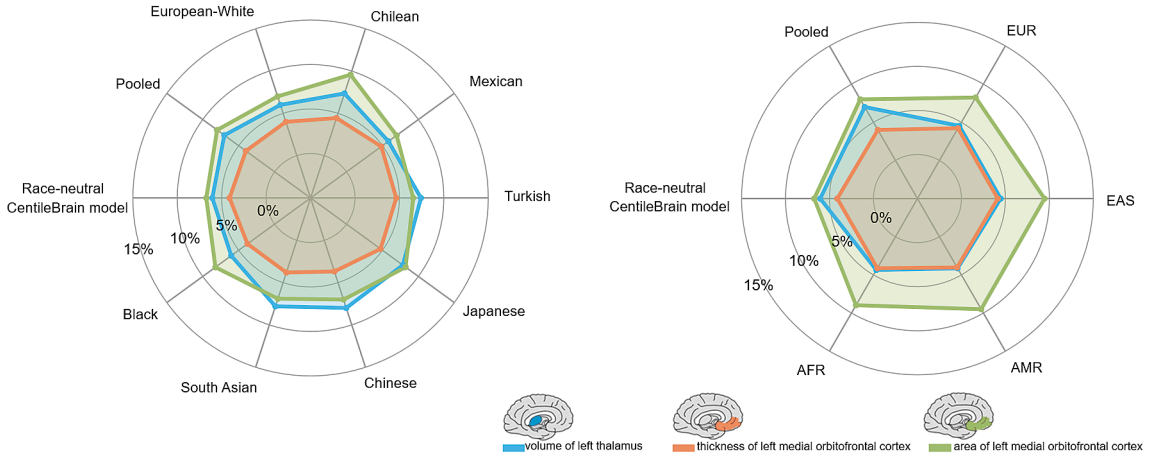

Fig. 3. *Left* panel: Spider plot of the RMAE values obtained from the self-reported ethnoracial groups and the corresponding values from the race-neutral CentileBrain model of female participants. *Right* panel: Spider plot of the RMAE values obtained from the genetic ancestry-determined ethnoracial groups and the corresponding values from the race-neutral CentileBrain model of female participants. We illustrate these findings for females using the left thalamic volume and left medial orbitofrontal cortical thickness and surface area as exemplars (the corresponding data in males are in *SI Appendix*, Fig. S7). AFR = African; AMR = Admixed American; EAS = East Asian; EUR = European.

covariates in model construction. These factors are well-documented contributors to individual variation in brain morphometry (33–35), but adjusting for them in the context of normative modeling would have limited our ability to assess whether the models generalize to populations with differing levels of environmental exposure. Including such variables may inadvertently normalize structural disadvantage, making it difficult to detect deviations from normative brain development that are themselves meaningful indicators of social and environmental inequities. This decision is consistent with broader concerns in clinical medicine, where the use of race-based adjustments in diagnostic algorithms has been increasingly challenged (22, 23). These adjustments, often justified as proxies for unmeasured contextual influences, can misattribute group-level differences to inherent biological factors (22). By resisting both race-based stratification and adjustment for upstream social determinants, our approach preserves the sensitivity of the models to detect meaningful population-level deviations—while avoiding assumptions that such differences are intrinsic to ethnoracial identity.

To facilitate fair cross-group comparisons when the CentileBrain Models were applied to the different ethnoracial groups, we introduced the RMAE as a scale-independent metric of predictive performance. This normalization allowed us to assess whether apparent discrepancies in prediction accuracy were driven by differences in model fit or by underlying group differences in brain morphometry. The low RMAE values observed across all morphometric domains and groups further affirm the broad applicability of the CentileBrain models and their resilience to ethnoracial variation in raw structural measures.

When comparing the Chinese Brain Charts model with the CentileBrain model in an independent Chinese sample, we found comparable R^2 values. This contrasts with prior reports of a performance gap between the Chinese Brain Charts and models derived from large, predominantly Western samples that excluded individuals of Asian ethnoracial background (14). Several factors may explain this finding. First, the CentileBrain training dataset includes Asian participants, which may enhance its generalizability to Chinese samples, as also indicated by our analyses across diverse ethnoracial groups (Figs. 1 and 2). Second, CentileBrain employs a distinct modeling algorithm (*SI Appendix*, *Materials and Methods*) that may better capture variance relevant to the target sample. While large, population-specific models such as the Chinese Brain Charts may confer advantages for characterizing individuals within a given population, their applicability is more limited in international consortia or in countries with highly mixed ethnoracial compositions, where CentileBrain models may offer broader utility.

Several limitations merit consideration. The original model development sample for CentileBrain, though globally diverse, was still skewed toward individuals identifying as white (13). This imbalance reflects the current state of data availability in neuroimaging and remains a barrier to full representation. Although the size and coverage of the validation cohorts are among the largest available for this purpose, some ethnoracial subgroups remain unrepresented due to data unavailability. These limitations highlight the need for sustained global investment in neuroimaging resources that capture underrepresented populations across both geographic and demographic axes. Our comparative analysis with population- specific normative models of brain morphometry was limited to a single platform. However, the findings suggest that race-neutral models developed from multiethnic samples can achieve prediction accuracy comparable to those trained exclusively on a single ethnoracial group. This result supports the broader viability of unified modeling approaches that do not require stratification by race or ethnicity, provided that development datasets are sufficiently large and globally inclusive.

In conclusion, the current study advances the field of normative brain morphometry by providing direct evidence that race-neutral normative models can generalize effectively across both self-identified and genetically inferred ethnoracial groups. These findings support the continued development and application of inclusive, publicly available modeling platforms such as CentileBrain and underscore the importance of moving toward population-spanning frameworks in precision neuroscience. Future work should focus on expanding these models to underrepresented populations, incorporating longitudinal data, and exploring interactions with environmental, developmental, and social determinants of brain health.

Materials and Methods

Samples. The data source of each sample is shown in *SI Appendix*, Table S1. All neuroimaging and associated demographic data were fully anonymized and used in accordance with the data sharing and governance policies of the

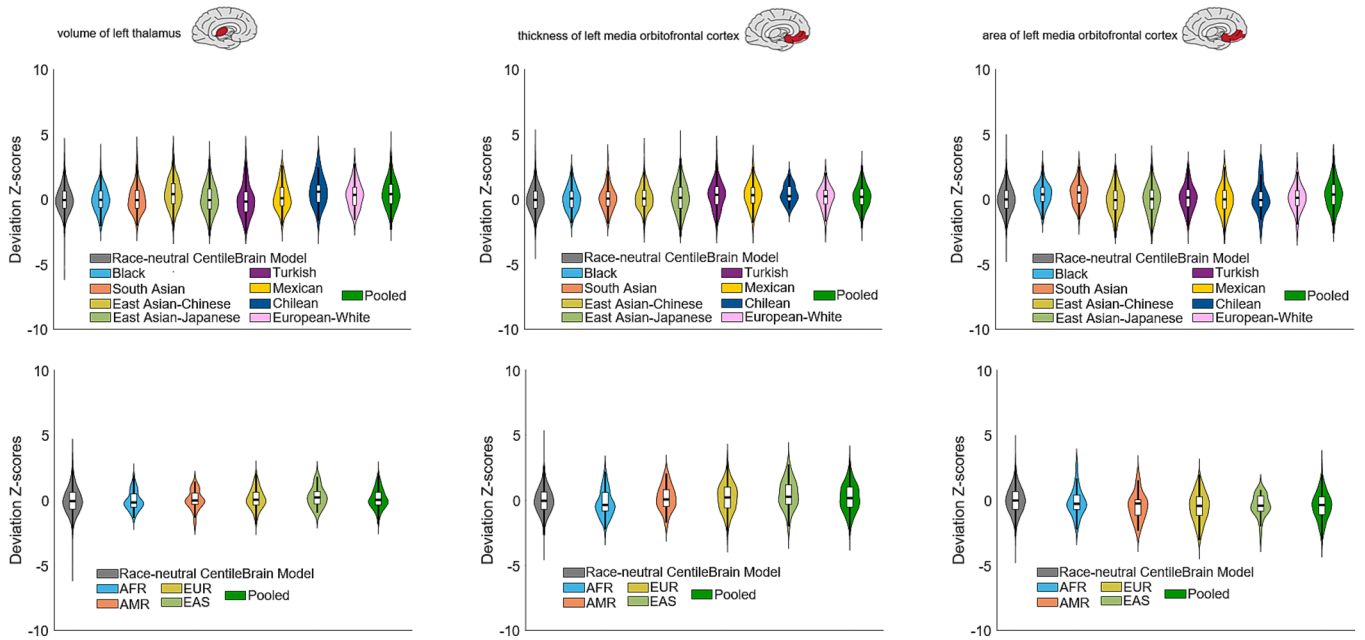

Fig. 4. Normative deviation scores (Z-scores) in each ethnoracial group and in the race-neutral CentileBrain model of female participants. We illustrate these findings for females using the left thalamic volume and left medial orbitofrontal cortical thickness and surface area as exemplars. The corresponding data in males are in *SI Appendix*, Fig. S10. AFR = African; AMR = Admixed American; EAS = East Asian; EUR = European.

originating databases. We accessed data from four data sharing repositories: the Adolescent Brain Cognitive Development study (ABCD; https://abcdstudy.org/), the UK Biobank Study (UKB; https://www.ukbiobank.ac.uk), the OpenNeuro Platform for the Pragmatic Language Study (https://openneuro.org/datasets/ds003481/versions/1.0.3), the Psychosis MRI Shared Data Resource (PsyShareD; https://psyshared.com/) the Clinical Deep Phenotyping Working Group (https://www.psych.mpg.de/2948741/cdp-working-group) and the Chinese Human Connectome Project (26) (see also *SI Appendix, Materials and Methods*). Additional samples were provided by Southwest University Adult Lifespan Dataset (36), the Southwest University Longitudinal Imaging Multimodal Study (37), the Cognitive Genetics Collaborative Research Organization (25), the Adolescent Program of Neuropsychiatric and Imaging Study (38) and the Brain Development in Adolescence and Early Adulthood in Twins (BRAID-TWINS) (39). The ancestry-defined samples and comprised individuals derived from the ABCD dataset where the "fastSTRUCTURE" algorithm (40) was used to define with four ancestry factors corresponding to 4 super-populations: African (AFR), Admixed American (AMR), East Asian (EAS), and European (EUR). We further selected participants if they could be exclusively allocated to only one of the four super-population groups based on their estimated posterior probability of the factor for that group being higher than 0.99. Details on the selection of ABCD participants based on ancestry are provided in the Supplementary Materials

Neuroimaging Data. Parcellated neuroimaging data–derived using the standard processing pipelines of the FreeSurfer image analysis suite (https://surfer.nmr.mgh.harvard.edu/) applied to whole-brain T1-weighted structural MRI scans–were either generated by the research team or downloaded from publicly available repositories or obtained directly from the data owners. Only parcellations based on T1-weighted images that had passed quality control procedures implemented by the original data sources were included. In addition, all FreeSurfer-derived outputs underwent secondary quality control by the study team using the standardized ENIGMA protocols (https://enigma.ini.usc.edu/protocols/imaging-protocols) to ensure anatomical plausibility and exclude processing artifacts. These procedures yielded regional measures of cortical thickness and cortical surface area based on the Desikan-Killiany atlas (41) and subcortical volumes based on the Aseg atlas (42) across participants in all ethnoracial groups.

CentileBrain Normative Models of Brain Morphometry. The CentileBrain Project (https://centilebrain.org/), which is nested within the Lifespan Working Group (https://enigma.ini.usc.edu/ongoing/enigma-lifespan/) of the ENIGMA, Consortium (https://enigma.ini.usc.edu/) provides 150 openly accessible, sex-specific, race-neutral normative models for regional cortical thickness and cortical surface area based on the Desikan-Killiany atlas (40) and for regional subcortical volumes based on the Aseg atlas (41) derived using the standard FreeSurfer pipelines (13). Models were developed using multivariable fractional polynomial regression in a large international sample of 37,407 healthy individuals (aged 3 to 90 y) from 87 datasets (see also *SI Appendix, Materials and Methods*). ComBat–Generalized Additive Models (ComBat-GAM) (43) were applied to harmonize the new data from the multiple ethnoracial groups across all sites in a single step, enabling the resulting harmonized dataset to be entered directly into the pretrained normative model without additional recalibration or parameter adjustments for each site.

Statistical Procedures.

Concordance of model performance across ethnoracial groups. For each ethnoracial group–defined by self-report or genetic ancestry–the pretrained, sex-specific, race-neutral CentileBrain models were applied separately to male and female participants. Model performance was quantified using the MAE and RMSE for each regional morphometric measure. To evaluate consistency with the CentileBrain models, Pearson's correlation coefficients were computed in Matlab (version R2021a) between group-specific and CentileBrain reference MAE or RMSE vectors, and statistical significance was tested against the null hypothesis of zero correlation. CI for correlations were estimated using the Fisher *r*-to-*z* transformation, with correlation coefficients first transformed to Fisher's *z* values, confidence limits computed in the *z* domain, and the results back-transformed to the correlation scale for reporting. Statistical significance was tested against the null hypothesis of zero correlation.

Standardized deviation magnitude from reference model performance. Differences in mean MAE and RMSE between each ethnoracial group and the CentileBrain models were expressed in SD units with the CentileBrain model's SD as the scaling factor: Standardized Difference = $(\text{mean}_{\text{group}} - \text{mean}_{\text{CentileBrain}})/\text{sd}_{\text{CentileBrain}}$, where $\text{mean}_{\text{group}}$ and $\text{mean}_{\text{CentileBrain}}$ indicate mean MAE or RMSE values for each ethnoracial group and CentileBrain reference model, respectively, and $\text{sd}_{\text{CentileBrain}}$ indicates SD of the CentileBrain model's MAE or RMSE values across brain regions.

Normalized prediction error across brain morphometry domains. RMAE was calculated as MAE divided by the observed value for each regional morphometric measure, producing a dimensionless index. Between-group differences in RMAE were assessed using multivariate ANOVA across cortical thickness, cortical surface area, and subcortical volumes, with post hoc comparisons FDR corrected for multiple testing at $P_{FDR} < 0.05$.

Model performance in the pooled sample. In the pooled datasets including all participants across self-reported or ancestry-defined groups, the analyses followed the same procedures as in the group-specific analyses.

Permutation-based evaluation of model generalizability. To test whether the CentileBrain models generalized to each external ethnoracial sample, we compared subject-level prediction errors between the training dataset and each validation sample using a permutation-based approach robust to unequal sample sizes (see detailed information in *SI Appendix*, Materials and Methods). For each of the 150 regional models and each error metric (MAE, RMSE, RMAE), we computed Welch's t-statistic for the observed difference in mean prediction error. Subject-level errors were then pooled, group labels randomly permuted while preserving original sample sizes, and Welch's t-statistics recomputed across 10,000 permutations to generate the empirical null distribution. Two-sided *P*-values reflected the proportion of permuted statistics exceeding the observed statistic, providing a distribution-free assessment of whether model performance differed from CentileBrain reference expectations. The script can be found online at https://github.com/CentileBrain/centilebrain/tree/main/ethnoracial_test.

Comparison with a population-specific model. For each normative model (CentileBrain and the Chinese Brain Charts), predictions for each participant in the independent Chinese-ancestry sample (UKB study, N = 123) were generated for each regional brain feature (e.g., thickness of the left medial orbitofrontal cortex, volume of the right thalamus). For each brain feature, predictive accuracy for the CentileBrain and Chinese Brain Charts models was quantified by the RMAE values and explained variance for regional predictions of cortical thickness, surface area, and subcortical volume. Explained variance was selected for this comparison because it provides a direct measure of the proportion of variance in the observed data explained by each model, facilitating a standardized comparison of predictive strength between independently developed platforms. This procedure yielded two RMAE values and two explained variance per feature, one from each model. Subsequently, statistical comparisons were conducted using paired-sample *t* tests and FDR correction for multiple testing. For the Chinese Brain Charts, we applied the released model parameters directly to the independent sample without retraining, as only the model parameters (and not the training data) were available. We do not consider this a limitation, given that the independent sample used for model comparison comprised less than 0.2% of the original Chinese training cohort, and inclusion or otherwise of these cases would be unlikely to affect performance in a stable model.

Data, Materials, and Software Availability. The pretrained CentileBrain Models are freely available at https://centilebrain.org/ while the deviation Z-scores of all samples used here can be access through https://doi.org/10.6084/m9.figshare.31100953. The original imaging data can be accessed through a number of repositories with a range of licensing conditions. Specifically, the ABCD dataset can be accessed through the US National Data Archive (https://nda.nih.gov/); the CHCP dataset can be accessed through the Science Data Bank website (https://doi.org/10.11922/sciencedb.01374); the Psy-ShareD dataset can be accessed by request at https://psyshared.com/; the SALD and SLIM datasets can be accessed through the International Data-sharing Initiative (https://fcon_1000.projects.nitrc.org/indi/retro/sald.html and https://fcon_1000.projects.nitrc.org/indi/retro/southwestuni_qiu_index.html); the UKB dataset can be access through the UK Biobank data-access protocol (https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access). Data from BRAID-TWINS cohort (39), the Cognitive Genetics Collaborative Research Organization (25) and the Clinical Deep Phenotyping Working Group (https://www.psych.mpg.de/2948741/cdp-working-group) and South American datasets (38) can be made available through the cited original sources.

ACKNOWLEDGMENTS. Computations of the COCORO data were performed at the Research Center for Computational Science, Okazaki, Japan (Projects: NIPS, 15-IMS-C137, 16-IMS-C135, 17-IMS-C152, 18-IMS-C162, 19-IMS-C181, 20-IMS-C162, 21-IMS-C179, and 22-IMS-C195) supported by Japan Agency for Medical Research and Development under Grant Nos. JP21uk1024002 and JP24dk0307132, the Intramural Research Grant (6-1)for Neurological and Psychiatric Disorders of National Center of Neurology and Psychiatry, and Japan Society for the Promotion of Science Grant-in-Aid for Scientific Research (C) JP23K07001. The Psychosis MRI Shared Data Resource (Psy-ShareD) project is funded by the UK Medical Research Council under Grant No. MR/X010651/. The PIENSA data were collected with the support of the Sistema Nacional de Investigadoras e Investigadores/Secretaría de Ciencia, Humanidades, Tecnología e Innovación, Mexico. Data collection at National Institute of Mental Health and NeuroSciences was supported by the Wellcome Trust/Department of Biotechnology India Alliance (Grant No. IA/E/12/1/500755).

Author affiliations: [a]Djavad Mowafaghian Centre for Brain Health, University of British Columbia, Vancouver, BC V6T 1Z3, Canada; [b]Department of Psychiatry, Icahn School of Medicine, Mount Sinai, NY 10029; [c]Institute of Psychiatry, Psychology and Neuroscience, London SE5 8AF, United Kingdom; [d]Department of Psychology, Bilkent University, Ankara 06800, Türkiye; [e]National Magnetic Resonance Research Center, Bilkent University, Ankara 06800, Türkiye; [f]Translational Psychiatry, Department of Psychiatry and Psychotherapy, University of Lübeck, Lübeck 23538, Germany; [g]Department of Psychiatry, School of Medicine, Pontificia Universidad Católica de Chile, Santiago 8330077, Chile; [h]Laboratory of Experimental Psychiatry, Instituto Nacional de Neurología y Neurocirugía, Ciudad de México 14269, Mexico; [i]Section of Brain Function Information, National Institute for Physiological Sciences, Okazaki 444-8585, Japan; [j]IDG McGovern Institute for Brain Research, Center for MRI Research, Peking University, Beijing 100871, China; [k]School of Medicine, Universidad Finis Terrae, Santiago 7501015, Chile; [l]Psychiatric Institute José Horwitz B., Santiago 8420000, Chile; [m]Department of Pathology of Mental Diseases, National Institute of Mental Health, National Center of Neurology and Psychiatry, Tokyo 187-8551, Japan; [n]Department of Neuroscience, University School of Naples "Federico II", Naples 80131, Italy; [o]Department of Psychiatry and Psychotherapy, LMU University Hospital, LMU Munich, Munich 80336, Germany; [p]NeuroImaging Core Unit Munich, LMU University Hospital, LMU Munich, Munich 80336, Germany; [q]Munich Center for Neurosciences, Ludwig-Maximilians-Universität München, Munich 82152, Germany; [r]Department of Neuroscience, Bilkent University Aysel Sabuncu Brain Research Center, Ankara 06800, Türkiye; [s]Centre for Cognitive and Clinical Neuroscience, Brunel University of London, Uxbridge UB8 3PH, United Kingdom; [t]Department of Psychology, College of Health, Medicine and Life Sciences, Brunel University of London, Uxbridge UB8 3PH, United Kingdom; [u]Department of Pathology of Mental Diseases, National Institute of Mental Health, National Center of Neurology and Psychiatry, Tokyo 187–8551, Japan; [v]Department of Psychiatry, National Institute of Mental Health and Neurosciences, Bangalore 560029, India; [w]Department of Medical Informatics and Management and Psychiatry, Institute of Medicine, University of Tsukuba, Ibaraki 305-857, Japan; [x]Department of Advanced Biomedical Sciences, University of Naples "Federico II", Naples 80131, Italy; [y]Max Planck Institute of Psychiatry, Munich 80804, Germany; [z]First Department of Psychiatry, National and Kapodistrian University of Athens, Athens 10679, Greece; [aa]Department of Psychiatry, Psychotherapy, and Psychosomatics, Medical Faculty University of Augsburg, Augsburg 86159, Germany; [bb]Evidence-Based Psychiatry and Psychotherapy, Faculty of Medicine, University of Augsburg, Augsburg 86159, Germany; [cc]Advanced Research Institute of Multidisciplinary Sciences, School of Medical Technology, School of Life Science, Beijing Institute of Technology, Beijing 100081, China; and [dd]Imaging Genetics Center, Mark & Mary Stevens Institute for Neuroimaging & Informatics, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033

Author contributions: S.F. designed research; S.F. performed research; R.G., Y.Y., F.N., S.S.H., N.S., K.Y., P.A., S.A., M.A., S.B., N.A.C., C.d.l.F.-S., M.F., J.-H.G., A.G.-V., R.H., F.I., D.K., K.K., V.K., J.M., U.M.M., K.N., G.P., F.J.R., F.R.-M., N.R., D.Ş.-Ç, T.S.-I., T.T., E.W., G.Y., M.Z., P.M.T., and S.F. analyzed data; R.G., F.N., S.S.H., N.S., K.Y., P.A., M.A., N.A.C., C.d.l.F.-S., M.F., J.-H.G., A.G.-V., R.H., F.I., D.K., K.K., V.K., J.M., U.M.M., K.N., G.P., F.J.R., F.R.-M., N.R., D.Ş.-Ç, T.S.-I., T.T., E.W., G.Y., M.Z., and P.M.T. manuscript Preparation; S.A. contributed to writing the paper; and S.F. wrote the paper.

1. T. Wolfers *et al.*, Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models. *JAMA Psychiatry* **75**, 1146–1155 (2018).
2. M. Zabihi *et al.*, Dissecting the heterogeneous cortical anatomy of Autism Spectrum Disorder using normative models. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **4**, 567–578 (2019).
3. M. Zabihi *et al.*, Fractionating autism based on neuroanatomical normative modeling. *Transl. Psychiatry* **10**, 1–10 (2020).
4. J. Lv *et al.*, Individual deviations from normative models of brain structure in a large cross-sectional schizophrenia cohort. *Mol. Psychiatry.* **26**, 3512–3523 (2021).
5. R. Kjelkenes *et al.*, Deviations from normative brain white and gray matter structure are associated with psychopathology in youth. *Dev. Cogn. Neurosci.* **58**, 101173 (2022).
6. X. Shan *et al.*, Mapping the heterogeneous brain structural phenotype of autism spectrum disorder using the normative model. *Biol. Psychiatry* **91**, 967–976 (2022).
7. S. Haas *et al.*, Normative modeling of brain morphometry in clinical high risk for psychosis. *JAMA Psychiatry* **81**, 77–88 (2024).
8. O. Potvin, A. Mouiha, L. Dieumegarde, S. Duchesne, Initiative AsDN, normative data for subcortical regional volumes over the lifetime of the adult human brain. *Neuroimage* **137**, 9–20 (2016).
9. O. Potvin, L. Dieumegarde, S. Duchesne, Normative morphometric data for cerebral cortical areas over the lifetime of the adult human brain. *Neuroimage* **156**, 315–339 (2017).
10. H.-M. Dong *et al.*, Charting brain growth in tandem with brain templates at school age. *Sci. Bull.* **65**, 1924–1934 (2020).
11. R. A. Bethlehem *et al.*, Brain charts for the human lifespan. *Nature* **604**, 1–11 (2022).
12. S. Rutherford *et al.*, Charting brain growth and aging at high spatial precision. *eLife* **11**, e72904 (2022).
13. R. Ge *et al.*, Normative modelling of brain morphometry across the lifespan with CentileBrain: Algorithm benchmarking and model optimisation. *Lancet Digit. Health* **6**, e211–e221 (2024).
14. L. Sun *et al.*, Population-specific brain charts reveal Chinese-Western differences in neurodevelopmental trajectories. bioRxiv [Preprint] (2025). https://www.biorxiv.org/content/10.1101/2025.06.17.659820 (Accessed 1 July 2025).

15. D. R. Weinberger, K. Dzirasa, L. L. Crumpton-Young, Missing in action: African ancestry brain research. *Neuron* **107**, 407–411 (2020).
16. E. B. Falk *et al.*, What is a representative brain? Neuroscience meets population science. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17615–17622 (2013).
17. G. M. Babulal *et al.*, Perspectives on ethnic and racial disparities in Alzheimer's disease and related dementias: Update and areas of immediate need. *Alzheimer's Dement.* **15**, 292–312 (2019).
18. A. Palk, J. Illes, P. M. Thompson, D. J. Stein, Ethical issues in global neuroimaging genetics collaborations. *Neuroimage* **221**, 117208 (2020).
19. Y. Banda *et al.*, Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the genetic epidemiology research on adult health and aging (GERA) cohort. *Genetics* **200**, 1285–1295 (2015).
20. J. C. Denny, Chapter 13: Mining electronic health records in the genomics era. *PLoS Comput. Biol.* **8**, e1002823 (2012).
21. R. S. Cooper, J. S. Kaufman, R. Ward, Race and genomics. *N. Engl. J. Med.* **348**, 1166 (2003).
22. D. A. Vyas, L. G. Eisenstein, D. S. Jones, Hidden in plain sight–Reconsidering the use of race correction in clinical algorithms. *N. Engl. J. Med.* **383**, 874–882 (2020).
23. S. R. Davidson *et al.*, Race adjustment of pulmonary function tests in the diagnosis and management of COPD: A scoping review. *Int. J. Chron. Obstruct. Pulmon. Dis.* **19**, 969–980 (2024).
24. P. M. Thompson *et al.*, ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. *Transl. Psychiatry* **10**, 1–28 (2020).
25. D. Koshiyama *et al.*, Neuroimaging studies within Cognitive Genetics Collaborative Research Organization aiming to replicate and extend works of ENIGMA. *Hum. Brain Mapp.* **43**, 182–193 (2022).
26. J. Ge *et al.*, Increasing diversity in connectomics with the Chinese Human Connectome Project. *Nat. Neurosci.* **26**, 163–172 (2023).
27. Q. Xu *et al.*, CHIMGEN: A Chinese imaging genetics cohort to enhance cross-ethnic and cross-geographic brain research. *Mol. Psychiatry* **25**, 517–529 (2020).
28. P. Prado *et al.*, The BrainLat project, a multimodal neuroimaging dataset of neurodegeneration from underrepresented backgrounds. *Sci. Data* **10**, 889 (2023).
29. J. P. A. Ioannidis, N. R. Powe, C. Yancy, Recalibrating the use of race in medical research. *JAMA* **325**, 623–624 (2021).
30. A. Flanagin, T. Frey, S. L. Christiansen, Updated guidance on the reporting of race and ethnicity in medical and science journals. *JAMA* **326**, 621–627 (2021).
31. K. L. Korunes, A. Goldberg, Human genetic admixture. *PLoS Genet.* **17**, e1009374 (2021).
32. S. Rutherford *et al.*, Measuring racial fairness of reference classes with normative modeling. arXiv [Preprint] (2024). https://doi.org/10.48550/arXiv.2407.19114 (Accessed 9 June 2025).
33. N. H. Brito, K. G. Noble, Socioeconomic status and structural brain development. *Front. Neurosci.* **8**, 276 (2014).
34. T. Frodl *et al.*, Childhood adversity impacts on brain subcortical structures relevant to depression. *J. Psychiatr. Res.* **86**, 58–65 (2017).
35. A. Zugman *et al.*, Country-level gender inequality is associated with structural differences in the brains of women and men. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2218782120 (2023).
36. D. Wei *et al.*, Structural and functional brain scans from the cross-sectional Southwest University adult lifespan dataset. *Sci. Data* **5**, 1–10 (2018).
37. W. Liu *et al.*, Longitudinal test-retest neuroimaging data from healthy young adults in Southwest China. *Sci. Data* **4**, 1–9 (2017).
38. L. S. Kegeles *et al.*, An imaging-based risk calculator for prediction of conversion to psychosis in clinical high-risk individuals using glutamate ^{1}H MRS. *Schizophr. Res.* **226**, 70–73 (2020).
39. D. Şahin-Çevik *et al.*, Psychotic-like experiences and working memory connectivity in adolescents and young adults: A study on Turkish twins. *Schizophr. Res.* **284**, 167–174 (2025).
40. A. Raj, M. Stephens, J. K. J. G. Pritchard, fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics* **197**, 573–589 (2014).
41. R. S. Desikan *et al.*, An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).
42. B. Fischl *et al.*, Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341–355 (2002).
43. R. Pomponio *et al.*, Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. *Neuroimage* **208**, 116450 (2020).