Archival Report

Biological Psychiatry: CNNI

Representing Brain-Behavior Associations by Retaining High-Motion Minoritized Youth

Jivesh Ramduny, Lucina Q. Uddin, Tamara Vanderwal, Eric Feczko, Damien A. Fair, Clare Kelly, and Arielle Baskin-Sommers

ABSTRACT

BACKGROUND: Population neuroscience datasets provide an opportunity for researchers to estimate reproducible effect sizes for brain-behavior associations because of their large sample sizes. However, these datasets undergo strict quality control to mitigate sources of noise, such as head motion. This practice often excludes a disproportionate number of minoritized individuals.

METHODS: We used motion-ordering and motion-ordering+resampling (bagging) to test whether these methods preserve functional magnetic resonance imaging (fMRI) data in the Adolescent Brain Cognitive Development (ABCD) Study (N = 5733). For the 2 methods, brain-behavior associations were computed as the partial Spearman's rank correlations (R_s) between functional connectivity and cognitive performance (NIH Cognition Toolbox) as well as externalizing and internalizing psychopathology (Child Behavior Checklist) while adjusting for participant sex assigned at birth and head motion.

RESULTS: Black and Hispanic youth exhibited excess head motion relative to data collected from White youth and were discarded disproportionately when conventional approaches were used. Motion-ordering and bagging methods retained more than 99% of Black and Hispanic youth. Both methods produced reproducible brain-behavior associations across low-/high-motion racial/ethnic groups based on motion-limited fMRI data.

CONCLUSIONS: The motion-ordering and bagging methods are 2 feasible approaches that can enhance sample representation for testing brain-behavior associations and that result in reproducible effect sizes in diverse populations.

https://doi.org/10.1016/j.bpsc.2025.01.014

Population neuroscience datasets made available through consortia efforts provide unprecedented opportunities for researchers to relate individual differences in brain functions (e.g., functional connectivity) to individual differences in behavior (e.g., general psychopathology, cognitive ability) with sufficient statistical power (1–3). These datasets have accelerated scientific discovery, improved reproducible research practices, and democratized the field by reducing the resources required to do research (4–7). However, despite the power that comes with the overall sample sizes available in consortia datasets, minoritized individuals tend to be disproportionally discarded through the standard preprocessing protocols typically used for magnetic resonance imaging (MRI) data analysis. As a result, studies on brain-behavior relationships have made relatively limited progress in identifying reproducible effects across samples with diverse sociodemographic characteristics (8–10).

To minimize variations in MRI data acquisition across consortia sites, researchers have applied strict quality control (QC) strategies to mitigate potential biases. One QC strategy is to address the potential impact of participant head motion (1). Head motion can significantly inflate functional connectivity estimates across a diverse range of populations (11–17). This issue is exacerbated in studies of youth due to heightened in-scanner motion relative to adult populations and a robust negative correlation between head motion and age (11,18–21). Researchers exclude high-motion participants irrespective of their racial or ethnic backgrounds by using a head motion threshold. However, these practices inadvertently tend to exclude minoritized youth disproportionately, due to their lower adherence to mainstream protocols and a lack of culturally informed strategies to engage and retain racial/ethnic minorities in neuroimaging research, particularly individuals from underprivileged backgrounds (8,22,23).

The goals of the current study were to test 1) the utility of 2 head motion mitigation methods (24) that attempt to retain high-motion minoritized youth using their motion-limited functional MRI (fMRI) data and 2) the possibility to obtain reproducible brain-behavior associations using the motion-limited fMRI data from high-motion youth. We sought to extend previous work on head motion mitigation methods (13,14,24–27) by using a sample that was larger and that had greater variability in participant racial representation than previous studies, with head motion estimates across multiple fMRI scans and well-studied behaviors (previous work focused on age).

ISSN: 2451-9022

We used data from the Adolescent Brain Cognitive Development (ABCD) Study given its adequate sample size and distribution of race/ethnicity (Black [14%], Hispanic [22%], and White [64%]) (28). We utilized motion-ordering and motion-ordering+resampling (bagging) to estimate functional connectivity to maximize the inclusion of high-motion minoritized youth. Motion-ordering is ranking time points from least motion to most motion in motion-limited timeseries data. For motion-ordering, we computed the functional connectivity matrix of each participant using the top time points from (ordered) motion-limited timeseries data. Bagging is randomly resampling the motion-ordered time points several times. For bagging, we generated 500 bootstrapped samples with replacement for each participant using the top time points from (ordered) motion-limited timeseries data. Then, we computed the bagged functional connectivity matrix of each participant iteratively. To test whether these head motion mitigation methods retained high-motion youth, we looked at the proportion of participants included by race. Based on previous research (24), it was expected that the 2 proposed methods would retain a maximum number of high-motion participants across the 3 racial/ethnic groups as well as generate reproducible brain-behavior associations when the high-motion youth were included.

To assess the validity of the 2 head motion mitigation methods, we examined the reproducibility of the brain-behavior associations as a function of *n* across the 3 racial/ethnic groups (1,24). We focused on brain-behavior relationships, specifically relationships between whole-brain functional connectivity and cognitive performance and psychopathology, because these have been extensively studied using consortia datasets and present important opportunities to advance biological psychiatry (1,2,29–32). We used the edge level because methods such as connectome-based predictive modeling, network-based statistics, and canonical correlation analysis prioritize the most relevant edges in the functional connectome for computing brain-behavior relationships (33–36). To test whether these head motion mitigation methods supported reproducible brain-behavior relationships, we examined 2 metrics. First, we examined confidence intervals as a metric of sampling variability within participants from each racial/ethnic group. As sample sizes increase, variability tends to decrease and stabilize (1), which is then reflected in tighter confidence intervals. Second, we computed area under the curve (AUC) (24) as a metric of brain-behavior effect size. If the standard, motion-ordered, and bagged brain-behavior relationships are identical, then the confidence intervals will be identical, leading to a rate of change in AUC = 0%. In practice, it is unlikely to achieve an AUC = 0% due to individual variations that are present in standard, motion-ordered, and bagged functional connectivity estimates. Instead, we focused on obtaining comparable brain-behavior relationships using the standard, motion-ordered, and bagging methods, where the rates of change in AUC are within a statistically acceptable range.

We cannot hope to fully understand the biological bases of clinical conditions if we systematically exclude minoritized individuals due to biased preprocessing methods. Therefore, it is crucial to identify methods that retain as many high-motion minoritized youth as possible to develop a more inclusive understanding of the brains and behaviors of the populations represented in our data (37–40).

METHODS AND MATERIALS

Participants

Complete demographic, behavioral, and imaging data from 5733 youth were obtained from the ABCD Study Baseline Release (age 9–10 years) (28). Youth who belonged to one of the 3 racial/ethnic groups (Black, Hispanic, and White) were included (Table 1); approximately 13% of the ABCD sample participants who belonged to other racial/ethnic categories were not included due to their limited sample sizes (Supplemental Methods 1). The current sample reflects participants who passed the MRI QC criteria described by the ABCD Data Analysis and Informatics Center (DAIC) (https://wiki.abcdstudy.org/release-notes/imaging/quality-control.html; imgincl_t1w_include=1; imgincl_rsfmri_include = 1) (41). The DAIC recommendations minimize data QC failure from biasing brain-behavior associations estimated from low-/high-motion participants. A total of 1881 participants were not included based on the DAIC QC protocols, which represents ∼16% of the ABCD sample. We further cross-checked the DAIC recommendations with the ABCD-BIDS (Brain Imaging Data Structure) Community Collection (ABCC) (42) to ensure that the participants had existing preprocessed resting-state fMRI data. There were no statistical differences in race/ethnicity or sex assigned at birth when we compared participants in our final sample with the participants who were not included in the analysis (Supplemental Methods 1).

Table 1. Demographic, Behavioral, and fMRI Characteristics Derived From the ABCD Study NIMH Data Archive Release 4.0

	Racial/Ethnic Groups		
	Black	Hispanic	White
Scans	Baseline	Baseline	Baseline
Sample Size	810	1247	3676
Age, Years	9–10	9–10	9–10
ADI	1.1–124.5	1.1–124.6	1.1–125.7
Sex, Female/Male	53%/47%	50%/50%	48%/52%
NIH Cognition Toolbox	46–108	51–109	54–117
CBCL-Externalizing	33–84	33–79	33–83
CBCL-Internalizing	33–80	33–88	33–93
Sites	21	21	21
Frames	196–1560	542–1560	380–1570
tSNR	11.7–49.8	11.2–53.0	10.8–55.6
Mean FD, mm	0.04–2.1	0.04–2.2	0.03–2.4

Values are presented as *n*, range, or %. Participant sex denotes youth's sex assigned at birth. NIH Cognition Toolbox indexed cognitive performance. CBCL indexed externalizing and internalizing behaviors. Frames indicate the number of volumes acquired across 4 resting-state scans. Note that the CBCL externalizing and internalizing and NIH Cognition Toolbox scores were *T*-standardized.

ABCD, Adolescent Brain Cognitive Development; ADI, Area Deprivation Index; CBCL, Child Behavior Checklist; FD, framewise displacement; fMRI, functional magnetic resonance imaging; NIMH, National Institute of Mental Health; tSNR, temporal signal-to-noise ratio.

Behavioral Data

We focused on cognitive performance, externalizing psychopathology, and internalizing psychopathology (Supplemental Methods 2). The NIH Cognition Toolbox (http://nihtoolbox.org/) was used to measure general cognitive performance based on the fluid and crystallized cognition composite scores of the youth (43–45). For each youth, the total composite score on the NIH Cognition Toolbox was obtained, which was *T*-standardized. The Child Behavior Checklist (CBCL) is a parent-report assessment that was used to measure externalizing and internalizing behaviors (46). The CBCL externalizing and internalizing scores of the youth were obtained from their respective syndrome scales, which were subsequently *T*-standardized. There were significant differences in the distributions of the NIH Toolbox and CBCL externalizing scores between the low-motion and high-motion youth (Figure S1).

MRI Preprocessing

All ABCD MRI data were processed with the ABCC (42) (https://github.com/DCAN-Labs/abcd-hcp-pipeline) (Supplemental Methods 4 and 5). Standard fMRI blood oxygen level–dependent preprocessing was further carried out using the ABCD-BIDS Pipeline (41,47) (Supplemental Methods 6). To identify low-motion youth in the ABCD Study, we applied a head motion threshold of mean framewise displacement (FD) $<$0.20 mm to the participants' resting-state fMRI data across a maximum of 4 runs (14,18) (Supplemental Methods 7). Applying a head motion threshold of mean FD $<$0.20 mm retained 3342 participants, corresponding to $\sim$58% of the total sample as follows: Black n = 410, Hispanic n = 666, and White n = 2266. This threshold provided a reasonable balance between the sample size requirement necessary to achieve adequate statistical power for brain-behavior associations and good data quality standards to mitigate motion artifacts in such studies (13,14,41). Traditionally, a threshold $<$0.10 mm is regarded as strict, whereas a threshold $>$0.50 mm is considered liberal (1,13,18,48). There is no one-threshold-fits-all populations because head motion varies by several factors including age (e.g., infants, adolescents, adults) and mental health diagnosis (11,13,21,49,50).

Functional Connectivity With Standard Method

For each youth, a whole-brain parcellation scheme was applied to extract their resting-state fMRI timeseries concatenated across a maximum of 4 runs. Functional connectivity was derived from the fMRI timeseries by computing and Fisher *z*-transforming the Pearson correlation coefficient (r) between all possible pairs of regions of interest using an extended Gordon parcellation scheme (41,47,51) to construct their 352 $\times$ 352 functional connectivity matrices. This pipeline is referred to as the standard method (Supplemental Methods 8 and 9).

Functional Connectivity With Motion-Ordered and Bagging Methods

The motion-ordered and bagging methods were performed by applying a frame-censoring procedure known as scrubbing (14,16,52,53) to identify and remove motion-corrupted time points (T) in the fMRI timeseries (Figures 1 and 2) (Supplemental Methods 10 and 11). Each T was identified with a head motion threshold of FD $>$0.20 mm (18). For each T, 1 preceding ($T-1$) and 2 succeeding ($T+1$, $T+2$) time points were censored in the fMRI timeseries to minimize the presence of residual motion and micro head movements. For each participant, we created subsets of the scrubbed time points that matched a threshold referred to as minimum time point (minTP) to prevent participants from having a varying number of motion-limited time points in their timeseries. The scrubbed timeseries were ranked by their lowest FD values, and the top minTP-matched time points were selected. For the motion-ordered method, the functional connectivity matrix of each participant was computed using the minTP-matched time points (24). For bagging, 500 bootstrapped samples of size TP were generated with replacement for each participant using the minTP-matched time points (24). Iteratively, the bagged functional connectivity matrix of each participant was recomputed. A mean bagged functional connectivity matrix was also generated for each participant by averaging the 500 connectivity matrices. Both minTP and TP had identical sizes to prevent either parameter from influencing the brain-behavior associations.

Brain-Behavior Associations With Standard, Motion-Ordered, and Bagging Methods

The standard, motion-ordered, and bagged functional connectivity matrices for each participant were used to compute partial Spearman's rank correlations (R_s) between connectivity patterns and behavior for cognitive performance (NIH Cognition Toolbox) and externalizing and internalizing psychopathology (CBCL) at the edge level (Supplemental Methods 3 and 12). Participant sex assigned at birth and mean FD were treated as covariates when the brain-behavior correlations were computed.

The brain-behavior relationships were obtained from the strongest edge (R_s) in the functional connectome after false discovery rate (FDR) correction (q = .05) (54) across 61,776 edges (Figure S2). There was no edge that produced a significant association with CBCL internalizing scores after FDR correction. Therefore, the brain-behavior analyses focused only on the NIH Cognition Toolbox and CBCL externalizing scores. We established the reproducibility of standard brain-behavior associations for the 3 racial/ethnic groups across different sample sizes (1,24). Each sample size was bootstrapped without replacement over 500 iterations, where we randomly selected samples of low-motion youth at 11 logarithmically spaced intervals. For each sample size, the mean brain-behavior R_s and 95% CIs were plotted as a function of n.

Then, we performed the motion-ordered and bagged brain-behavior associations by computing R_s between connectivity patterns and NIH Toolbox and CBCL externalizing scores using the scrubbed timeseries of the low-motion racial/ethnic groups. We used the same edge that shared the highest correlation strength with the standard method for the motion-ordered and bagging procedures. We used sampling variability to assess the reproducibility of the brain-behavior R_s using motion-ordered and bagged functional connectivity matrices. For both methods, subsamples of participants were selected randomly for each low-motion racial/ethnic

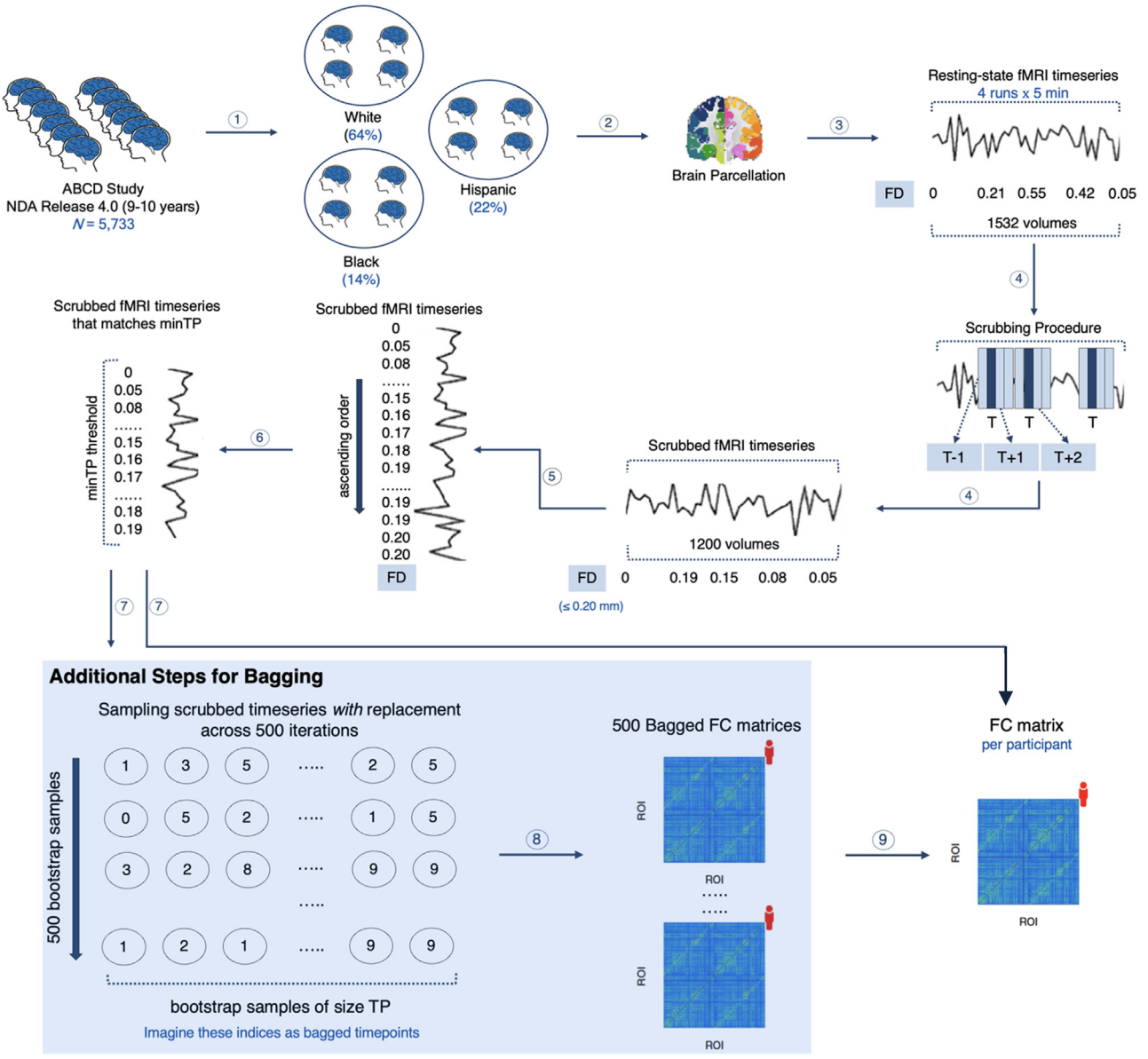

Figure 1. Motion-ordered and bagging methods. 1) The participants in the ABCD (Adolescent Brain Cognitive Development) Study were categorized into 3 racial/ethnic groups: Black, Hispanic, and White. 2) A whole-brain parcellation was applied to extract the resting-state functional magnetic resonance imaging (fMRI) timeseries of the youth concatenated across a maximum of 4 runs. 3) The framewise displacement (FD) was computed for every time point in the fMRI timeseries to quantify the amount of in-scanner motion for one volume relative to its preceding one. 4) Scrubbing was applied to identify and remove motion-corrupted time points (T) if their FD was >0.20 mm. For each T, 1 preceding ($T - 1$) and 2 succeeding ($T + 1$, $T + 2$) time points were further censored. 5) For each youth, their scrubbed timeseries were ranked by their lowest FD values (0→0.20 mm). 6) A minimum time point (minTP) threshold was imposed to prevent youth from having different number of time points in their timeseries. 7) For the motion-ordered method, the functional connectivity (FC) of each participant was computed between all pairs of regions of interest (ROIs) using Pearson correlation coefficients from the minTP-matched time points. For the bagging method, 500 bootstrapped timeseries samples of size TP were generated with replacement for each participant using the minTP-matched time points. 8) Iteratively, the FC between all pairs of ROIs was computed using Pearson correlation coefficients for the 500 bootstrapped samples. 9) A mean bagged FC matrix also was generated for each youth by averaging the 500 connectivity matrices. [The motion-ordered and bagging methods are adapted from previous work (24).] The analysis code for this study can be found at https://github.com/JRam02/inclusivity. NDA, National Institute of Mental Health Data Archive.

group without replacement over 500 iterations. The mean brain-behavior R_s and 95% CIs were plotted as a function of n. Then, we estimated the absolute ΔAUC to quantify the rate of change in AUC between the standard, motion-ordered, and bagging methods. Finally, we evaluated the agreement of the standard and motion-ordered/bagged brain-behavior relationships using Lin's concordance coefficient (ρ_c) (55,56) across the 3 racial/ethnic groups. By definition, ρ_c measures the

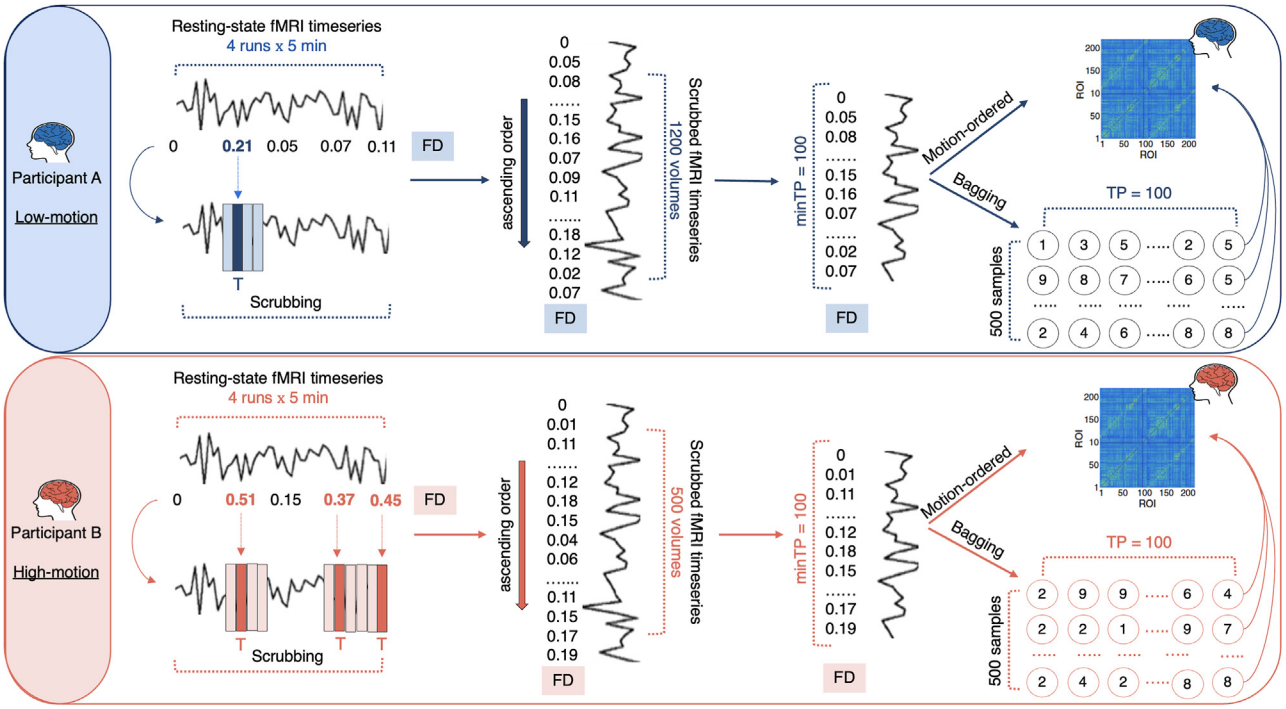

Figure 2. Example of 2 participants' resting-state functional magnetic resonance imaging (fMRI) timeseries data using the motion-ordered and bagging methods. Top panel: Participant A is identified as low-motion, which means that their fMRI data likely have fewer motion-corrupted time points. Bottom panel: Participant B is identified as high-motion, which means that their fMRI data likely have more motion-corrupted time points. In both cases, all the motion-corrupted time points (*T*) are censored on the basis that their framewise displacement (FD) was >0.20 mm using a procedure known as scrubbing. For each identified *T*, its preceding and 2 succeeding time points are also removed. The scrubbed timeseries are then ranked by the lowest FD values such that the smallest value will be 0 and the largest value will be 0.20 mm. A minimum time point (minTP) threshold is applied to select the same number of least motion-corrupted time points in the timeseries. In the current study, this minTP threshold (e.g., 100 time points) is chosen based on the extent to which we can retain the maximum number of low-/high-motion participants across the 3 racial/ethnic groups. If a participant has fewer scrubbed time points than the minTP threshold, they are not included in the subsequent analysis. Once the minTP-matched time points of all low-/high-motion participants are selected, the motion-ordered and bagging methods are applied. For the motion-ordered method, the functional connectivity matrix of each participant is computed using the minTP-matched time points. For bagging, 500 bootstrapped samples are generated with replacement for each participant using minTP-matched time points of size TP. Here, TP corresponds to the number of time points being bootstrapped with replacement from the minTP-matched time points. In the current study, minTP and TP have identical sizes. For each participant, the bagged functional connectivity matrix is computed in an iterative manner. ROI, region of interest. .

agreement between 2 variables from a set of bivariate data ranging from −1 (perfect discordance) to 1 (perfect concordance) (Supplemental Methods 14). It can be interpreted as a weighted version of the Pearson correlation coefficient, with ρ_c <0.20 denoting poor concordance and ρ_c > 0.80 representing excellent concordance (57).

RESULTS

Disproportionate Exclusion of High-Motion Minoritized Youth

We observed that a strict head motion threshold drastically reduced the sample sizes across racial/ethnic groups relative to a liberal head motion threshold (Figure 3A). However, there was a disproportionate decline in the sizes of the Black and Hispanic groups regardless of the mean FD threshold. This was due to a significant difference in mean FD across the 3 racial/ethnic groups [see (58)]. Minoritized youth exhibited greater head motion than White youth (Figure 3B) (Kruskal-Wallis test, H_2 = 61.9, p < .001; Black median mean FD = 0.20 mm, Hispanic median mean FD = 0.19 mm, and White median mean FD = 0.16 mm). We observed a gradual disproportionate decrease in the temporal signal-to-noise ratio of the Black and Hispanic youth when liberal mean FD thresholds were applied (Figure 3A). There was also a significant difference in the number of scrubbed time points across the 3 racial/ethnic groups. Minoritized youth had fewer scrubbed time points in their motion-limited fMRI timeseries than White youth (Figure 3A) (Kruskal-Wallis test, H_2 = 34.7, p < .001; Black median scrubbed time points = 936, Hispanic median scrubbed time points = 940, and White median scrubbed time points = 1016). Because most low-/high-motion participants had at least 100 scrubbed time points, we set the minTP and TP thresholds to 100 time points to maximize participant inclusivity across the 3 racial/ethnic groups.

Standard Brain-Behavior Associations From Low-Motion Racial and Ethnic Groups

We examined the reproducibility of the standard brain-behavior relationships as a function of *n* ranging from 25 (typical) to maximum *n* in each group (large = 250). We observed tightening of the 95% CI as the sample sizes of the

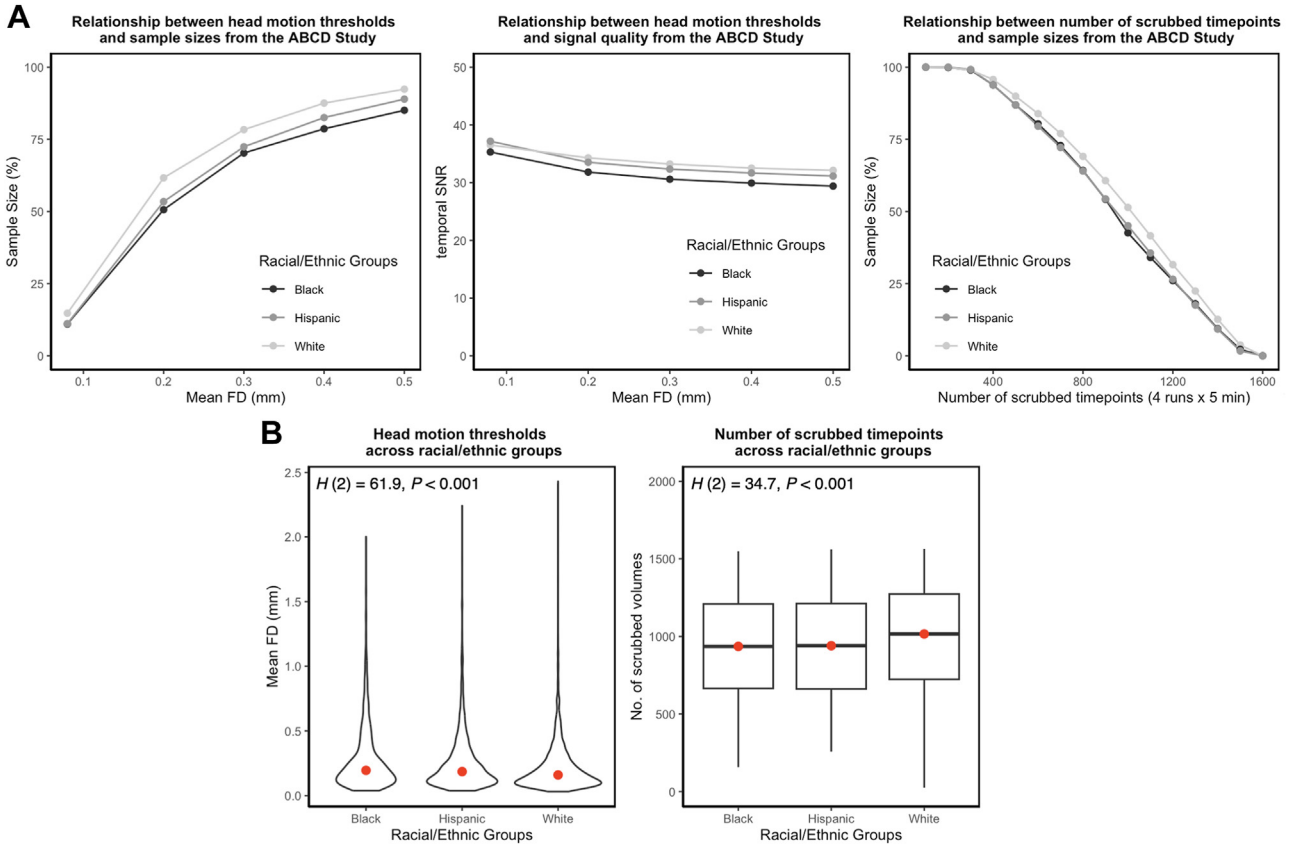

Figure 3. Impact of head motion across racial and ethnic groups in the ABCD (Adolescent Brain Cognitive Development) Study. **(A)** Left: Relationship between head motion thresholds and sample sizes. Head motion was quantified by mean framewise displacement (FD) such that mean FD ∈ (0.08, 0.20, 0.30, 0.40, 0.50). Middle: Relationship between head motion thresholds and signal quality indexed by the temporal signal-to-noise ratio. Mean FD ∈ (0.08, 0.20, 0.30, 0.40, 0.50). Right: Relationship between number of scrubbed time points and sample sizes. Number of scrubbed time points ∈ (100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500). **(B)** Left: There was a significant difference in head motion across the 3 racial/ethnic groups. Right: There was a significant difference in the number of scrubbed time points across the 3 racial/ethnic groups.

racial/ethnic groups grew from typical to large for both the NIH Cognition Toolbox (Table 2) and CBCL externalizing scores (Table 3). For the NIH Cognition Toolbox, Black $|R_s|$ = 0.12, Hispanic $|R_s|$ = 0.053, and White $|R_s|$ = 0.10. For CBCL externalizing scores, Black $|R_s|$ = 0.057, Hispanic $|R_s|$ = 0.080, and White $|R_s|$ = 0.097.

Motion-Ordered and Bagged Brain-Behavior Associations From Low-Motion Racial and Ethnic Groups

Next, we assessed the reproducibility of the motion-ordered and bagged brain-behavior relationships as a function of *n* ranging from 25 (typical) to maximum *n* of each group (large). We observed tightening of the 95% CI as the sample sizes of the racial/ethnic groups increased from typical to large for the NIH Cognition Toolbox (Table 2) and CBCL externalizing scores (Table 3). The brain-behavior relationships were comparable to those obtained from the standard method. For the NIH Cognition Toolbox, Black $|R_s|$ = 0.13, Hispanic $|R_s|$ = 0.066, and White $|R_s|$ = 0.11 using the motion-ordered method, and Black $|R_s|$ = 0.12, Hispanic $|R_s|$ = 0.066, and White $|R_s|$ = 0.11 using bagging. For CBCL externalizing scores, Black $|R_s|$ = 0.056, Hispanic $|R_s|$ = 0.065, and White $|R_s|$ = 0.071 using the motion-ordered method and Black $|R_s|$ = 0.049, Hispanic $|R_s|$ = 0.067, and White $|R_s|$ = 0.070 using bagging. For the NIH Cognition Toolbox, $\Delta AUC_{motion\text{-}ordered}$ was as follows: Black ΔAUC = 1.03%, Hispanic ΔAUC = 4.96%, and White ΔAUC = 0.22%. For the NIH Cognition Toolbox, $\Delta AUC_{bagging}$ was as follows: Black ΔAUC = 0.26%, Hispanic ΔAUC = 3.80%, and White ΔAUC = 2.14%. For CBCL externalizing scores, $\Delta AUC_{motion\text{-}ordered}$ was as follows: Black ΔAUC = 2.75%, Hispanic ΔAUC = 2.86%, and White ΔAUC = 2.40%. For CBCL externalizing scores, $\Delta AUC_{bagging}$ was as follows: Black ΔAUC = 3.42%, Hispanic ΔAUC = 1.93%, and White ΔAUC = 2.70%.

Retaining High-Motion Youth for Inclusive and Reproducible Brain-Behavior Associations

Next, we tested whether motion-ordered and bagging methods can maximize participant inclusion for generating inclusive and reproducible brain-behavior associations. Instead of discarding all the high-motion youth based on

Table 2. Correlations Between Functional Connectivity and NIH Cognition Toolbox Obtained From the Standard, Motion-Ordered, and Bagging Methods as a Function of Sample Size Across the 3 Low-Motion Racial/Ethnic Groups

	Brain-Behavior Relationships for NIH Cognition Toolbox		
Sample Size, *n*	Standard	Motion-Ordered	Bagging
Low-Motion White Youth			
25	0.10 [−0.32 to 0.49]	0.10 [−0.31 to 0.50]	0.11 [−0.28 to 0.49]
39	0.10 [−0.22 to 0.42]	0.088 [−0.23 to 0.39]	0.098 [−0.22 to 0.43]
62	0.097 [−0.14 to 0.35]	0.11 [−0.15 to 0.34]	0.094 [−0.17 to 0.37]
97	0.093 [−0.098 to 0.29]	0.11 [−0.077 to 0.29]	0.11 [−0.088 to 0.28]
152	0.11 [−0.057 to 0.26]	0.10 [−0.050 to 0.26]	0.11 [−0.038 to 0.23]
238	0.095 [−0.019 to 0.21]	0.11 [−0.011 to 0.21]	0.10 [−0.012 to 0.23]
374	0.10 [0.006 to 0.19]	0.099 [0.009 to 0.20]	0.11 [0.016 to 0.19]
586	0.10 [0.033 to 0.18]	0.11 [0.028 to 0.17]	0.11 [0.038 to 0.17]
920	0.10 [0.051 to 0.15]	0.11 [0.057 to 0.15]	0.11 [0.057 to 0.15]
1444	0.10 [0.073 to 0.13]	0.11 [0.074 to 0.14]	0.11 [0.076 to 0.14]
2266	0.10 [0.10 to 0.10]	0.11 [0.11 to 0.11]	0.11 [0.11 to 0.11]
AUC	250.01	250.55	244.67
Low-Motion Black Youth			
25	0.13 [−0.29 to 0.51]	0.12 [−0.23 to 0.52]	0.13 [−0.33 to 0.50]
33	0.12 [−0.22 to 0.42]	0.13 [−0.19 to 0.44]	0.11 [−0.22 to 0.42]
44	0.12 [−0.17 to 0.39]	0.14 [−0.14 to 0.42]	0.13 [−0.17 to 0.42]
58	0.11 [−0.12 to 0.33]	0.12 [−0.16 to 0.38]	0.14 [−0.092 to 0.39]
77	0.12 [−0.096 to 0.30]	0.13 [−0.078 to 0.33]	0.12 [−0.11 to 0.34]
101	0.12 [−0.060 to 0.29]	0.13 [−0.050 to 0.28]	0.12 [−0.050 to 0.28]
134	0.12 [−0.017 to 0.25]	0.13 [−0.005 to 0.27]	0.11 [−0.029 to 0.24]
177	0.12 [0.003 to 0.23]	0.13 [0.013 to 0.24]	0.12 [0.007 to 0.22]
234	0.12 [0.034 to 0.21]	0.12 [0.031 to 0.21]	0.12 [0.033 to 0.20]
310	0.12 [0.067 to 0.18]	0.13 [0.073 to 0.18]	0.12 [0.069 to 0.18]
410	0.12 [0.12 to 0.12]	0.13 [0.13 to 0.13]	0.12 [0.12 to 0.12]
AUC	85.33	86.21	85.55
Low-Motion Hispanic Youth			
25	0.049 [−0.34 to 0.48]	0.086 [−0.36 to 0.51]	0.062 [−0.39 to 0.46]
35	0.056 [−0.29 to 0.38]	0.063 [−0.27 to 0.41]	0.080 [−0.26 to 0.40]
48	0.056 [−0.24 to 0.35]	0.060 [−0.23 to 0.33]	0.064 [−0.22 to 0.35]
67	0.044 [−0.19 to 0.25]	0.058 [−0.14 to 0.28]	0.068 [−0.15 to 0.29]
93	0.056 [−0.14 to 0.25]	0.064 [−0.11 to 0.25]	0.064 [−0.11 to 0.25]
129	0.055 [−0.11 to 0.21]	0.066 [−0.091 to 0.22]	0.062 [−0.086 to 0.21]
179	0.052 [−0.065 to 0.17]	0.065 [−0.056 to 0.18]	0.064 [−0.066 to 0.18]
249	0.055 [−0.047 to 0.15]	0.066 [−0.032 to 0.16]	0.066 [−0.044 to 0.16]
345	0.054 [−0.035 to 0.13]	0.066 [−0.005 to 0.15]	0.064 [−0.010 to 0.14]
480	0.056 [0.004 to 0.11]	0.067 [0.021 to 0.11]	0.065 [0.018 to 0.11]
666	0.054 [0.054 to 0.054]	0.066 [0.066 to 0.066]	0.066 [0.064 to 0.067]
AUC	123.47	117.35	118.78

Values are presented as mean R_s [95% CI]. The standard method corresponded to the brain-behavior associations derived from the full fMRI timeseries of low-motion youth with a mean FD <0.20 mm. The motion-ordered method corresponded to the brain-behavior associations derived from the scrubbed fMRI timeseries ranked and thresholded by their 100 least motion-corrupted time points (0 < FD < 0.20 mm) to construct the functional connectivity matrices of low-motion youth. The bagging method corresponded to the brain-behavior associations derived from the scrubbed fMRI timeseries ranked and thresholded by their 100 least motion-corrupted time points (0 < FD < 0.20 mm), from which 100 time points were bootstrapped across 500 iterations to construct the functional connectivity matrices of low-motion youth. The sample sizes were bootstrapped at 11 logarithmically spaced *n* intervals: Black $n \in$ (25, 33, 44, 58, 77, 101, 134, 177, 234, 310, 410); Hispanic $n \in$ (25, 35, 48, 67, 93, 129, 179, 249, 345, 480, 666); and White $n \in$ (25, 39, 62, 97, 152, 238, 374, 586, 920, 1444, 2266). For each low-motion racial/ethnic group and behavior, we computed a total of 11 intervals × 500 bootstrapped *n* samples = 5500 standard and motion-ordered correlations. For each low-motion racial/ethnic group and behavior, we computed a total of 11 intervals × 500 bootstrapped *n* samples × 500 bootstrapped time points = 2.75 million bagged correlations. For each racial/ethnic group, the brain-behavior R_s and 95% CIs are shown as a function of sample size. The AUC for the brain-behavior associations are also displayed.

AUC, area under the curve; FD, framewise displacement; fMRI, functional magnetic resonance imaging.

common practices in brain-behavior association studies, such as applying a mean FD <0.20 mm, we retained all low-/high-motion youth who had usable time points in their fMRI timeseries. To maximize participant inclusion, we retained all low-/high-motion youth who had a minimum of 100 least motion-corrupted time points in their scrubbed fMRI timeseries

Table 3. Correlations Between Functional Connectivity and CBCL Externalizing Scores Obtained From the Standard, Motion-Ordered, and Bagging Methods as a Function of Sample Size Across the 3 Low-Motion Racial/Ethnic Groups

	Brain-Behavior Relationships for CBCL Externalizing		
Sample Size, *n*	Standard	Motion-Ordered	Bagging
Low-Motion White Youth			
25	0.097 [−0.29 to 0.46]	0.064 [−0.34 to 0.49]	0.071 [−0.32 to 0.44]
39	0.10 [−0.19 to 0.41]	0.086 [−0.22 to 0.39]	0.067 [−0.26 to 0.40]
62	0.094 [−0.13 to 0.33]	0.063 [−0.20 to 0.30]	0.067 [−0.20 to 0.30]
97	0.092 [−0.099 to 0.28]	0.066 [−0.12 to 0.25]	0.068 [−0.12 to 0.27]
152	0.098 [−0.050 to 0.25]	0.072 [−0.078 to 0.22]	0.066 [−0.082 to 0.22]
238	0.095 [−0.033 to 0.22]	0.069 [−0.049 to 0.18]	0.073 [−0.042 to 0.19]
374	0.099 [0.007 to 0.19]	0.072 [−0.027 to 0.16]	0.070 [−0.023 to 0.16]
586	0.10 [0.031 to 0.17]	0.072 [−0.006 to 0.14]	0.070 [0.005 to 0.15]
920	0.098 [0.053 to 0.14]	0.070 [0.022 to 0.12]	0.070 [0.019 to 0.12]
1444	0.099 [0.068 to 0.13]	0.071 [0.038 to 0.10]	0.072 [0.041 to 0.10]
2266	0.099 [0.099 to 0.099]	0.071 [0.071 to 0.071]	0.071 [0.071 to 0.072]
AUC	244.82	250.70	251.44
Low-Motion Black Youth			
25	0.055 [−0.35 to 0.45]	0.078 [−0.31 to 0.48]	0.039 [−0.36 to 0.43]
33	0.060 [−0.28 to 0.38]	0.063 [−0.30 to 0.42]	0.045 [−0.30 to 0.39]
44	0.056 [−0.23 to 0.35]	0.052 [−0.23 to 0.36]	0.050 [−0.22 to 0.31]
58	0.054 [−0.18 to 0.27]	0.056 [−0.18 to 0.29]	0.051 [−0.18 to 0.28]
77	0.048 [−0.15 to 0.23]	0.047 [−0.15 to 0.26]	0.046 [−0.16 to 0.25]
101	0.055 [−0.12 to 0.22]	0.055 [−0.10 to 0.22]	0.044 [−0.14 to 0.22]
134	0.061 [−0.065 to 0.19]	0.052 [−0.088 to 0.17]	0.053 [−0.094 to 0.19]
177	0.059 [−0.048 to 0.16]	0.055 [−0.055 to 0.16]	0.050 [−0.055 to 0.17]
234	0.063 [−0.023 to 0.15]	0.053 [−0.030 to 0.14]	0.053 [−0.026 to 0.14]
310	0.059 [0.007 to 0.11]	0.051 [−0.004 to 0.11]	0.052 [−0.005 to 0.10]
410	0.059 [0.059 to 0.059]	0.054 [0.054 to 0.054]	0.051 [0.049 to 0.053]
AUC	82.69	84.96	85.52
Low-Motion Hispanic Youth			
25	0.061 [−0.34 to 0.43]	0.065 [−0.33 to 0.45]	0.053 [−0.37 to 0.45]
35	0.079 [−0.22 to 0.43]	0.058 [−0.29 to 0.38]	0.065 [−0.25 to 0.37]
48	0.089 [−0.20 to 0.34]	0.066 [−0.18 to 0.35]	0.073 [−0.21 to 0.35]
67	0.081 [−0.15 to 0.31]	0.067 [−0.17 to 0.31]	0.073 [−0.16 to 0.32]
93	0.086 [−0.091 to 0.27]	0.065 [−0.11 to 0.27]	0.067 [−0.11 to 0.26]
129	0.083 [−0.063 to 0.23]	0.064 [−0.082 to 0.22]	0.065 [−0.085 to 0.22]
179	0.082 [−0.036 to 0.21]	0.063 [−0.056 to 0.19]	0.073 [−0.039 to 0.19]
249	0.077 [−0.016 to 0.16]	0.064 [−0.034 to 0.15]	0.069 [−0.015 to 0.16]
345	0.082 [0.007 to 0.14]	0.068 [−0.008 to 0.14]	0.068 [−0.008 to 0.14]
480	0.082 [0.037 to 0.13]	0.066 [0.021 to 0.11]	0.067 [0.018 to 0.11]
666	0.082 [0.082 to 0.082]	0.066 [0.066 to 0.066]	0.066 [0.065 to 0.068]
AUC	113.75	117.00	115.94

Values are presented as mean R_s [95% CI]. The standard method corresponded to the brain-behavior associations derived from the full fMRI timeseries of low-motion youth with a mean FD <0.20 mm. The motion-ordered method corresponded to the brain-behavior associations derived from the scrubbed fMRI timeseries ranked and thresholded by their 100 least motion-corrupted time points ($0 < FD < 0.20$ mm) to construct the functional connectivity matrices of low-motion youth. The bagging method corresponded to the brain-behavior associations derived from the scrubbed fMRI timeseries ranked and thresholded by their 100 least motion-corrupted time points ($0 < FD < 0.20$ mm), from which 100 time points were bootstrapped across 500 iterations to construct the functional connectivity matrices of low-motion youth. The sample sizes were bootstrapped at 11 logarithmically spaced *n* intervals: Black $n \in$ (25, 33, 44, 58, 77, 101, 134, 177, 234, 310, 410); Hispanic $n \in$ (25, 35, 48, 67, 93, 129, 179, 249, 345, 480, 666); and White $n \in$ (25, 39, 62, 97, 152, 238, 374, 586, 920, 1444, 2266). For each low-motion racial/ethnic group and behavior, we computed a total of 11 intervals × 500 bootstrapped *n* samples = 5500 standard and motion-ordered correlations. For each low-motion racial/ethnic group and behavior, we computed a total of 11 intervals × 500 bootstrapped *n* samples × 500 bootstrapped time points = 2.75 million bagged correlations. For each racial/ethnic group, the brain-behavior R_s and 95% CIs are shown as a function of sample size. The AUC for the brain-behavior associations are also displayed.

AUC, area under the curve; CBCL, Child Behavior Checklist; FD, framewise displacement; fMRI, functional magnetic resonance imaging.

(Supplemental Methods 13). This procedure helped to retain a total of 5732 participants, corresponding to ∼99.98% of the total sample size: Black *n* = 810, Hispanic *n* = 1247, and White *n* = 3675. This procedure maximized the representation of minoritized youth in subsequent brain-behavior analyses (loss in Black *n* = 0%, loss in Hispanic *n* = 0%, and loss in White *n* = 0.03%).

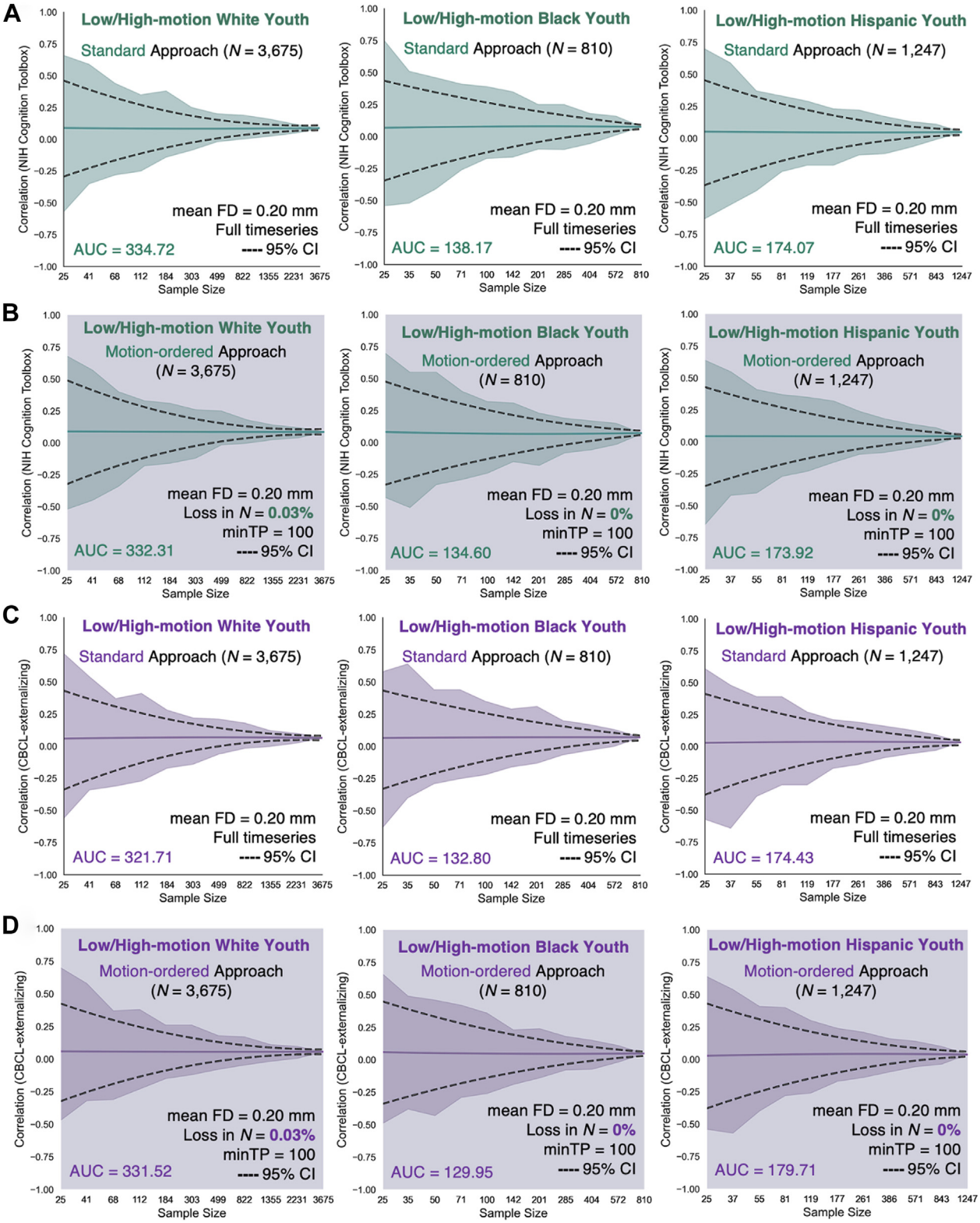

Figure 4. Correlations between functional connectivity and **(A, B)** NIH Cognition Toolbox and **(C, D)** Child Behavior Checklist (CBCL) externalizing scores obtained from the standard and motion-ordered methods as a function of sample size across the 3 low-/high-motion racial/ethnic groups. The standard method corresponded to the brain-behavior associations derived from the full functional magnetic resonance imaging (fMRI) timeseries of low-/high-motion

We reexamined the reproducibility of the motion-ordered and bagged brain-behavior associations as a function of n ranging from 25 (typical) to maximum n of each group (large). We computed standard brain-behavior associations to compare the performances of the motion-ordered and bagging methods when including the high-motion youth. We observed similar tightening of the 95% CIs as sample size increased with the standard and motion-ordered methods (Figure 4) in addition to the standard and bagging methods (Figure 5) for the NIH Cognition Toolbox and CBCL externalizing scores. The standard, motion-ordered, and bagged brain-behavior associations remained comparable even when the high-motion youth were included. From the standard method and NIH Cognition Toolbox, Black $|R_s|$ = 0.076, Hispanic $|R_s|$ = 0.047, and White $|R_s|$ = 0.090. From the motion-ordered method and NIH Cognition Toolbox, Black $|R_s|$ = 0.071, Hispanic $|R_s|$ = 0.044, and White $|R_s|$ = 0.086. From bagging and the NIH Cognition Toolbox, Black $|R_s|$ = 0.072, Hispanic $|R_s|$ = 0.043, and White $|R_s|$ = 0.085. From the standard method and CBCL externalizing scores, Black $|R_s|$ = 0.069, Hispanic $|R_s|$ = 0.034, and White $|R_s|$ = 0.065. From the motion-ordered method and CBCL externalizing scores, Black $|R_s|$ = 0.049, Hispanic $|R_s|$ = 0.037, and White $|R_s|$ = 0.056. From bagging and CBCL externalizing scores, Black $|R_s|$ = 0.045, Hispanic $|R_s|$ = 0.041, and White $|R_s|$ = 0.055. The $\Delta AUC_{motion\text{-}ordered}$ for the NIH Cognition Toolbox were as follows: Black ΔAUC = 2.58%, Hispanic ΔAUC = 0.09%, and White ΔAUC = 0.72%. For the NIH Cognition Toolbox, $\Delta AUC_{bagging}$ was as follows: Black ΔAUC = 3.03%, Hispanic ΔAUC = 3.87%, and White ΔAUC = 0.55%. For CBCL externalizing scores, $\Delta AUC_{motion\text{-}ordered}$ was as follows: Black ΔAUC = 2.15%, Hispanic ΔAUC = 3.03%, and White ΔAUC = 3.05%. For CBCL externalizing scores, $\Delta AUC_{bagging}$ was as follows: Black ΔAUC = 2.00%, Hispanic ΔAUC = 1.47%, and White ΔAUC = 0.37%.

Concordance of Standard and Motion-Ordered/Bagged Brain-Behavior Relationships From Low-/High-Motion Youth

At n = 25 (typical) and n = 250 (large), we observed strong concordance between the standard and motion-ordered brain-behavior relationships for the NIH Cognition Toolbox ($0.76 < \rho_c < 0.83$) and CBCL externalizing scores ($0.76 < \rho_c < 0.83$) derived from the low-motion racial/ethnic groups (Figures 6 and 7). We observed similar concordance patterns at typical and large sample sizes when the high-motion youth were included for the NIH Cognition Toolbox ($0.76 < \rho_c < 0.82$) and CBCL externalizing scores ($0.69 < \rho_c < 0.80$) (Figures 6 and 7). Additionally, the concordance of the standard and bagged brain-behavior associations was consistent with those derived from the motion-ordered method at typical and large sample sizes (Figures S3 and S4). Finally, the effect sizes produced by the standard, motion-ordered, and bagging methods were comparable using a single fMRI scan (Supplemental Methods 14; Figures S5–S9) and the weakest edge in the functional connectome (Figures S9–S13) with and without including the high-motion youth at typical and large sample sizes.

DISCUSSION

In modern neuroimaging research and specifically in the subfield of biological psychiatry, the quest for reproducible brain-behavior relationships has underscored a need for large sample sizes that can provide adequate statistical power. At the same time, consortia datasets undergo rigorous data QC to mitigate the impact of noise, such as that produced by participant head motion, with strict thresholds for data inclusion. This is also a feature of research that uses neuroimaging data from smaller, community-based samples, which often excludes high-motion participants despite the sample size. Unfortunately, applying strict head motion thresholds drastically reduces sample size, particularly in minoritized developmental cohorts. Here, replicating and extending previous work (24), we demonstrated the utility of motion-ordering and motion-ordering+resampling (bagging) methods for preserving fMRI data that contain high levels of head motion. Although minoritized youth exhibited greater head motion than White youth, both methods retained more than 99% of all minoritized youth and produced inclusive and reproducible effect sizes across races/ethnicities. The motion-ordered and bagging methods enhance sample representation in brain-behavior association studies and help generate reproducible effect sizes across sociodemographically diverse cohorts.

We showed that fMRI data that contain high head motion can be retained and used to compute meaningful brain-behavior associations for general cognitive performance and externalizing psychopathology. Applying the motion-ordered and bagging methods produced brain-behavior correlations comparable to those produced by the standard motion correction method when high-motion datasets collected from minoritized youth were retained. The effect sizes from the standard versus motion-ordered and bagging methods differed by a maximum of 7%. Some people may consider this a small gain, but it is important to note that the effect sizes of the proposed methods were obtained by including more than 99% of Black, Hispanic, and White youth. This remained true when a single fMRI scan and the weakest edge in the functional connectome was used. This extends previous work (24) showing that we may not need to discard thousands of participants in consortia datasets, particularly from among

youth who have been retained for the analyses without imposing an initial head motion threshold of mean framewise displacement (FD) <0.20 mm. The motion-ordered method corresponded to the brain-behavior associations derived from the scrubbed fMRI timeseries of low-/high-motion youth who were retained based on the assumption that they had a minimum of 100 least motion-corrupted time points. Their scrubbed fMRI timeseries were ranked by their lowest FD values, and 100 least motion-corrupted time points ($0 < FD < 0.20$ mm) were selected to construct the functional connectivity matrices of the youth. The sample sizes were bootstrapped at 11 logarithmically spaced n intervals: Black $n \in$ (25, 35, 50, 71, 100, 142, 201, 285, 404, 572, 810); Hispanic $n \in$ (25, 37, 55, 81, 119, 177, 261, 386, 571, 843, 1247); and White $n \in$ (25, 41, 68, 112, 184, 303, 499, 822, 1355, 2231, 3675). For each low-/high-motion racial/ethnic group and behavior, we computed a total of 11 intervals × 500 bootstrapped n samples = 5500 standard and motion-ordered correlations. Solid teal and purple lines show the mean correlations from the 500 bootstrapped samples for a given sample size. Teal and purple shadings denote the minimum and maximum correlations across 500 bootstrapped samples for a given sample size. Black dotted lines represent the lower and upper bounds of the 95% CIs for a given sample size. The areas under the curve (AUCs) for the brain-behavior associations are also displayed. minTP, mininum time point.

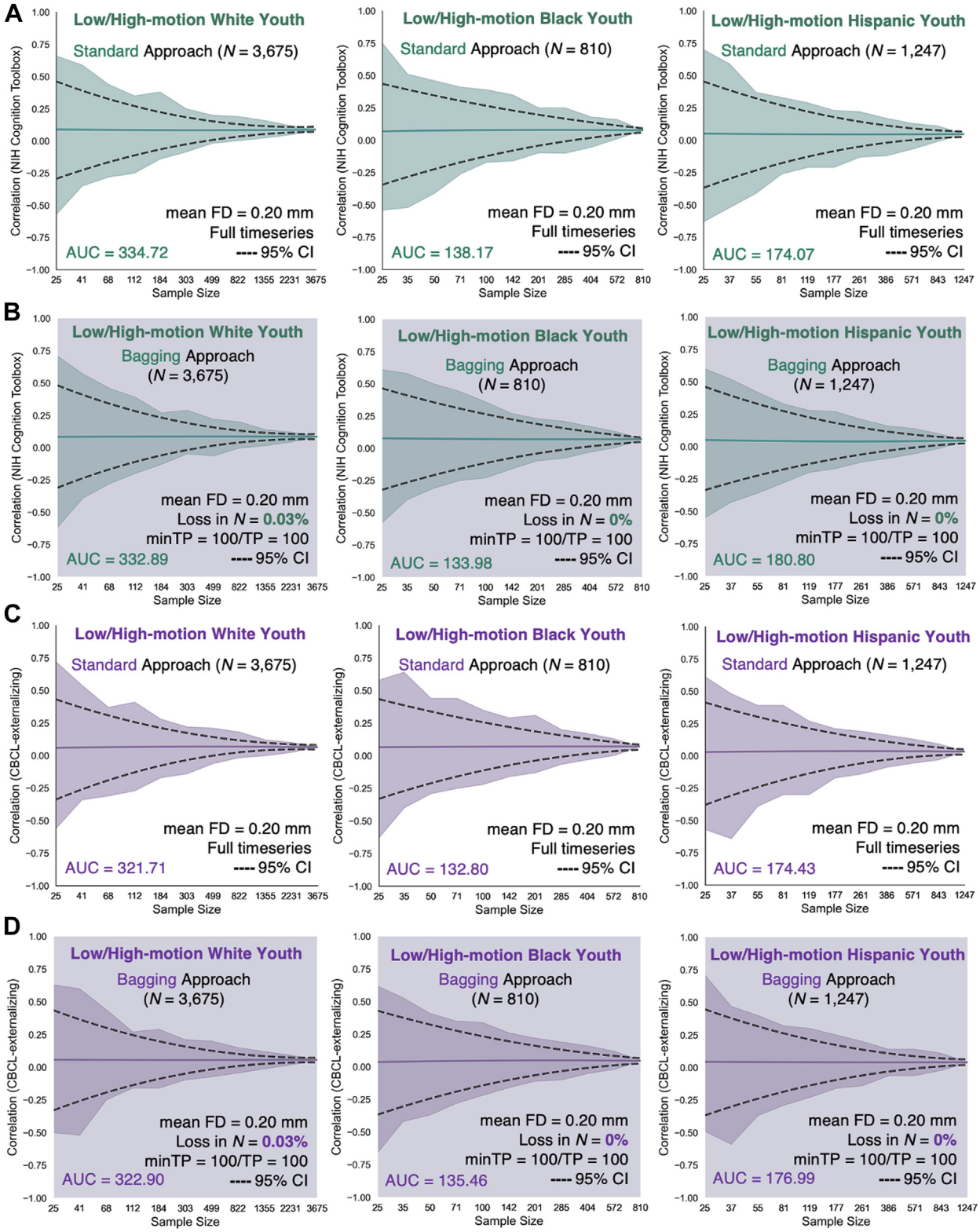

Figure 5. Correlations between functional connectivity and **(A, B)** the NIH Cognition Toolbox and **(C, D)** Child Behavior Checklist (CBCL) externalizing scores obtained from the standard and bagging methods as a function of sample size across the 3 low-/high-motion racial/ethnic groups. The standard method corresponded to the brain-behavior associations derived from the full functional magnetic resonance imaging (fMRI) timeseries of low-/high-motion

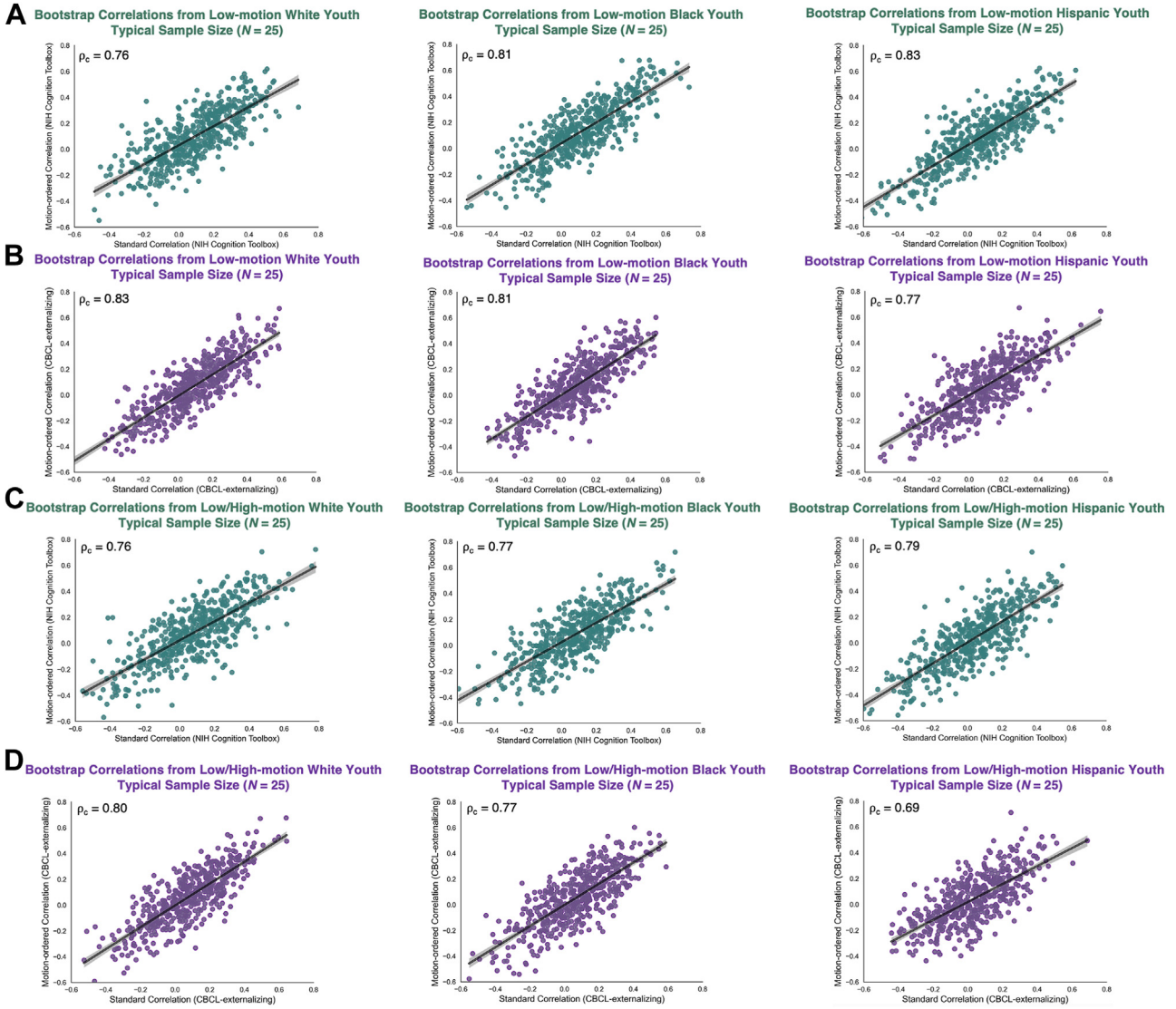

Figure 6. Bootstrapped brain-behavior correlations obtained from the standard and motion-ordered methods at typical sample size (n = 25) with **(A, B)** and without **(C, D)** excluding the high-motion youth across the 3 racial/ethnic groups. The teal and purple data points correspond to bootstrapped samples at n = 25 obtained from 500 iterations, and these samples were used to compute the associations between functional connectivity and the NIH Cognition Toolbox and Child Behavior Checklist (CBCL) externalizing scores. Lin's concordance correlation coefficient (ρ_c) was computed to assess the reproducibility of the brain-behavior associations obtained from the standard and motion-ordered methods. The low-/high-motion youth who were retained based on the assumption that they had a minimum of 100 least motion-corrupted time points in their functional magnetic resonance imaging timeseries.

youth who have been retained for the analyses without imposing an initial head motion threshold of mean framewise displacement (FD) <0.20 mm. The bagging method corresponded to the brain-behavior associations that were derived from the scrubbed fMRI timeseries of low-/high-motion youth who were retained based on the assumption that they had a minimum of 100 least motion-corrupted time points. Their scrubbed fMRItimeseries were ranked by their lowest FD values, and 100 least motion-corrupted time points (0 < FD < 0.20 mm) were selected, from which 100 time points were bootstrapped across 500 iterations to construct the functional connectivity matrices of the youth. The sample sizes were bootstrapped at 11 logarithmically spaced n intervals: Black $n \in$ (25, 35, 50, 71, 100, 142, 201, 285, 404, 572, 810); Hispanic $n \in$ (25, 37, 55, 81, 119, 177, 261, 386, 571, 843, 1247); and White $n \in$ (25, 41, 68, 112, 184, 303, 499, 822, 1355, 2231, 3675). For each low-/high-motion racial/ethnic group and behavior, we computed a total of 11 intervals × 500 bootstrapped n samples = 5500 standard correlations. For each low-/high-motion racial/ethnic group and behavior, we computed a total of 11 intervals × 500 bootstrapped n samples × 500 bootstrapped time points = 2.75 million bagged correlations. The solid teal and purple lines show the mean correlations from the 500 bootstrapped samples for a given sample size. Teal and purple shadings denote the minimum and maximum correlations across 500 bootstrapped samples for a given sample size. The black dotted lines represent the lower and upper bounds of the 95% CIs for a given sample size. The areas under the curve (AUCs) for the brain-behavior associations are also displayed. minTP, minimum time point.

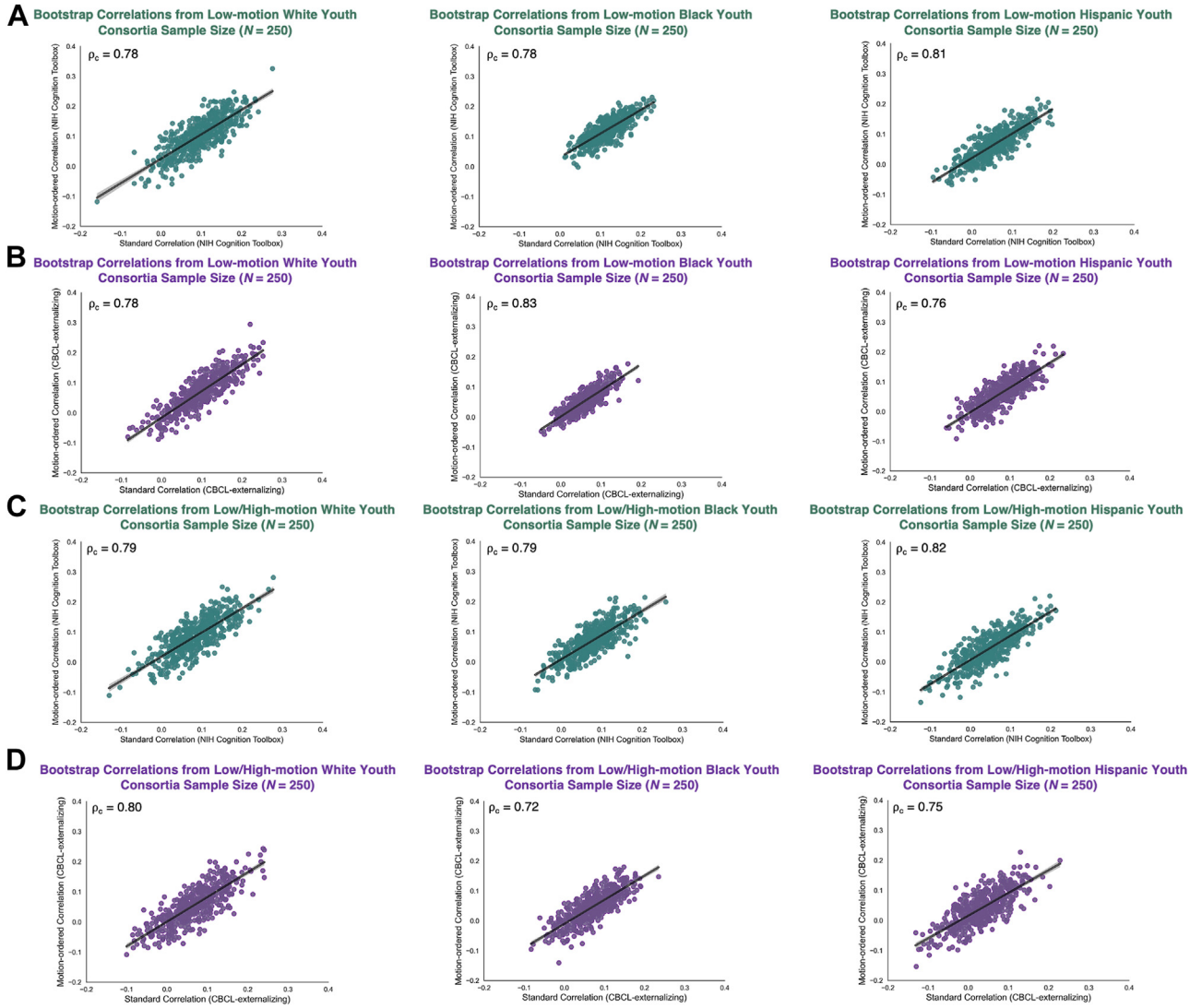

Figure 7. Bootstrapped brain-behavior correlations obtained from the standard and motion-ordered methods at large sample size (n = 250) with **(A, B)** and without **(C, D)** excluding the high-motion youth across the 3 racial/ethnic groups. The teal and purple data points correspond to bootstrapped samples at n = 250 obtained from 500 iterations, and these samples were used to compute the associations between functional connectivity and NIH Cognition Toolbox and Child Behavior Checklist (CBCL) externalizing scores. Lin's concordance correlation coefficient (ρ_c) was computed to assess the reproducibility of the brain-behavior associations obtained from the standard and motion-ordered methods. The low-/high-motion youth were retained based on the assumption that they had a minimum of 100 least motion-corrupted time points in their functional magnetic resonance imaging timeseries.

minoritized individuals, to produce meaningful brain-behavior associations. In practice, motion-ordering may be preferred for brain-behavior association studies due to its considerably lower computational demands than bagging (24). However, bagging may be particularly beneficial when deriving parameters/features for brain-behavior predictive models that use short acquisition scans (59,60). Regardless of which method researchers opt to use, individuals do not need to be excluded; instead, they can be retained if they have enough motion-limited timeseries data.

We focused on generating inclusive and reproducible brain-behavior relationships by retaining a large proportion of data collected from minoritized youth. Researchers are also interested in using brain-based models to predict behavior. However, predictive models are often trained on samples that lack diversity, resulting in poorer prediction of behaviors in minoritized individuals. Models trained on resting-state functional connectivity data have been found to predict a broad range of behaviors more accurately in White Americans than in African Americans. Cognitive scores were found to be underpredicted in African Americans compared with White Americans in the ABCD Study, although the predictive performance of the minoritized youth improved when the models were trained only on African Americans (61). In contrast, externalizing and internalizing behaviors were overpredicted in African Americans compared with White Americans (61).

Brain-behavior models can also fail to predict behaviors from individuals who do not fit common neurocognitive profiles (62). Not only has model failure—misclassification of neurocognitive scores as low or high from functional connectivity—been shown to be generalizable across datasets, behaviors, and participant groups, but it can also be associated with sociodemographic variables (62). High-scoring participants who belong to non-White groups have been shown to be misclassified frequently as low scoring and vice versa for White (not Hispanic/Latino) participants (62). Ultimately, before moving to prediction, it is important for researchers to enhance sample representation so that the data used to train these predictive models are inclusive. Our methods provide a way for improving the retention of minoritized individuals and securing a diverse sample.

Before concluding, it is important to note some limitations of the current study. First, we used ~50% of the ABCD Baseline Release. We followed DAIC and ABCC procedures to minimize potential artifacts from confounding brain-behavior associations, especially when retaining high-motion youth. However, investment in examining participant retention and potential bias across all stages of preprocessing is an important avenue for future research. Second, the proposed methods are effective for functional connectivity applications that do not rely on the continuity of the fMRI timeseries. The ranking of the motion-limited time points may not be desirable for traditional general linear models and intersubject correlations (63). Similarly, the 2 methods are not ideal for questions that involve examining the test-retest reliability or spatial reproducibility of functional connectomes. While we retained high-motion youth using 100 motion-limited time points, achieving highly reliable and reproducible functional connectomes typically requires much longer scan durations (36,64–70). Third, we need to replicate the motion-ordered and bagged brain-behavior associations in existing datasets that have a good representation of minoritized identities. The Human Connectome Project Development and Philadelphia Neurodevelopmental Cohort are large datasets that could be used to replicate the utility of these methods and extend the work to a broader age range to examine the impact of these methods on developmental change. Finally, the motion-ordered and bagging methods were developed to compute brain-behavior associations at the edge level as a function of sample size. Given that edges in the functional connectome share variable behavioral relevance (36), some may not exhibit any meaningful relationship with a behavioral domain. There was no edge in the functional connectome that shared a significant association with internalizing. Future work could expand the proposed methods to examine their abilities to produce inclusive and reproducible brain-behavior associations at different levels of functional organization (e.g., component, network) (1).

Conclusions

We can generate inclusive and reproducible brain-behavior relationships using motion-limited fMRI data from low- and high-motion youth across races and ethnicities. It is our hope that these methods will reduce the tension between data quality standards and sample size requirements and by doing so, mitigate resource wastage and the unethical practice of discarding large numbers of participants who have given their time, effort, and trust to the research process. Ultimately, retaining high-motion minoritized populations offers the opportunity to uncover biological and clinical variability that might otherwise be missed due to underrepresentation, thereby advancing our ability to identify neural mechanisms that are relevant to these populations.

ACKNOWLEDGMENTS AND DISCLOSURES

The ABCD Study is supported by the National Institutes of Health and additional federal partners under the following award numbers: U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, and U24DA041147. The full list of federal supporters is available at https://abcdstudy.org/federal-partners.html. The complete list of participating sites and study investigators can be found at https://abcdstudy.org/consortium_members/. The ABCD Consortium investigators designed and implemented the study and/or provided the data but did not necessarily participate in the analysis or writing of this article. This article reflects the views of the authors and may not reflect the opinions or views of the National Institutes of Health or ABCD Consortium investigators.

Additional support for this work was made possible from the National Institute of Environmental Health Sciences (Grant Nos. R01-ES032295, R01-ES031074, and R21DA057592). This work also obtained support from the Yale Kavli Institute for Neuroscience and the Wu Tsai Institute at Yale University. This work used computational resources from the Masonic Institute for the Developing Brain, Neuroimaging Genomics Data Resource, and Minnesota Supercomputing Institute at the University of Minnesota.

JR was responsible for formal analysis and visualization. JR and AB-S were responsible for conceptualization. JR, CK, and AB-S were responsible for methodology and validation. JR, LQU, TV, and AB-S were responsible for investigation and writing the original draft of the article. JR, LQU, TV, CK, and AB-S were responsible for reviewing and editing the article. JR, EF, and DAF were responsible for data curation. EF and DAF were responsible for resources. AB-S was responsible for supervision, project administration, and funding acquisition.

We thank the ABCD JEDI Workgroup 3 (Responsible Use of ABCD Study Data) for discussions regarding the motivations for and findings of this study.

Data used in the preparation of this article were obtained from the ABCD Study (http://abcdstudy.org/), which is held in the National Institute of Mental Health Data Archive. The ABCD Study is a multisite longitudinal study designed to recruit more than 10,000 9- to 10-year-old children and follow them over 10 years into early adulthood. The ABCD Study is openly available following access permission granted to one or multiple National Institute of Mental Health Data Archive Collections (https://nda.nih.gov/nda/access-data-info). The ABCD data repository grows and changes over time (https://nda.nih.gov/). The ABCD data used in this report came from ABCD Collection 3165 and were preprocessed using the ABCC from the DCAN (Developmental Cognition and Neuroimaging) Labs (https://nda.nih.gov/edit_collection.html?id=3165).

The analysis code for the motion-ordered and bagging methods can be found at https://github.com/JRam02/bagging. The analysis code specific to this study can be found at https://github.com/JRam02/inclusivity. The code for processing the ABCD Study by the ABCC can be found at https://github.com/DCAN-Labs/abcd-hcp-pipeline. The MRI data analysis code can be found at https://github.com/ABCD-STUDY/nda-abcd-collection-3165.

A previous version of this article was published as a preprint on bioRxiv: https://doi.org/10.1101/2024.06.22.600221.

DAF has a financial interest in Turing Medical Inc. and may benefit financially if the company is successful in marketing FIRMM motion monitoring software products. DAF may receive royalty income based on FIRMM technology developed at Washington University School of Medicine and Oregon Health & Science University and licensed to Turing Medical Inc. DAF

is a cofounder of Turing Medical Inc. These potential conflicts of interest have been reviewed and are managed by Washington University School of Medicine, Oregon Health & Science University, and the University of Minnesota. All other authors report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Department of Psychology, Yale University, New Haven, Connecticut (JR, AB-S); Kavli Institute for Neuroscience, Yale University, New Haven, Connecticut (JR); Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, California (LQU); Department of Psychiatry and Biobehavioral Sciences, University of California Los Angeles, Los Angeles, California (LQU); Department of Psychology, University of California Los Angeles, Los Angeles, California (LQU); Department of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada (TV); BC Children's Hospital Research Institute, Vancouver, British Columbia, Canada (TV); Masonic Institute for the Developing Brain, University of Minnesota, Minneapolis, Minnesota (EF, DAF); Department of Pediatrics, University of Minnesota, Minneapolis, Minnesota (EF, DAF); Institute of Child Development, University of Minnesota, Minneapolis, Minnesota (DAF); School of Psychology, Trinity College Dublin, Dublin, Ireland (CK); Department of Psychiatry, School of Medicine, Trinity College Dublin, Dublin, Ireland (CK); Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland (CK); Department of Psychiatry, Yale University, New Haven, Connecticut (AB-S); and Wu Tsai Institute, Yale University, New Haven, Connecticut (AB-S).

Address correspondence to Jivesh Ramduny, Ph.D., at jivesh.ramduny@yale.edu.

Received Oct 25, 2024; revised Jan 26, 2025; accepted Jan 27, 2025.

Supplementary material cited in this article is available online at https://doi.org/10.1016/j.bpsc.2025.01.014.

REFERENCES

1. Marek S, Tervo-Clemmens B, Calabro FJ, Montez DF, Kay BP, Hatoum AS, *et al.* (2022): Reproducible brain-wide association studies require thousands of individuals. Nature 603:654–660.
2. Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, *et al.* (2015): Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. Nat Neurosci 18:1664–1671.
3. Rosenberg MD, Casey BJ, Holmes AJ (2018): Prediction complements explanation in understanding the developing brain. Nat Commun 9:589.
4. Markiewicz CJ, Gorgolewski KJ, Feingold F, Blair R, Halchenko YO, Miller E, *et al.* (2021): The OpenNeuro resource for sharing of neuroscience data. eLife 10:e71774.
5. Poldrack RA, Gorgolewski KJ (2014): Making big data open: Data sharing in neuroimaging. Nat Neurosci 17:1510–1517.
6. Milham MP, Craddock RC, Son JJ, Fleischmann M, Clucas J, Xu H, *et al.* (2018): Assessment of the impact of shared brain imaging data on the scientific literature. Nat Commun 9:2818.
7. Poline J-B, Breeze JL, Ghosh S, Gorgolewski K, Halchenko YO, Hanke M, *et al.* (2012): Data sharing in neuroimaging research. Front Neuroinform 6:9.
8. Ricard JA, Parker TC, Dhamala E, Kwasa J, Allsop A, Holmes AJ (2023): Confronting racially exclusionary practices in the acquisition and analyses of neuroimaging data [published correction appears in Nat Neurosci 2023; 26:2251]. Nat Neurosci 26:4–11.
9. Kopal J, Uddin LQ, Bzdok D (2023): The end game: Respecting major sources of population diversity. Nat Methods 20:1122–1128.
10. United States Census Bureau (2021): American Community Survey. https://www.census.gov/programs-surveys/acs. Accessed December 4, 2023
11. Frew S, Samara A, Shearer H, Eilbott J, Vanderwal T (2022): Getting the nod: Pediatric head motion in a transdiagnostic sample during movie- and resting-state fMRI. PLOS One 17:e0265112.
12. Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE (2014): Methods to detect, characterize, and remove motion artifact in resting state fMRI. Neuroimage 84:320–341.
13. Yan CG, Cheung B, Kelly C, Colcombe S, Craddock RC, Di Martino A, *et al.* (2013): A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. Neuroimage 76:183–201.
14. Yan C-G, Craddock RC, Zuo X-N, Zang Y-F, Milham MP (2013): Standardizing the intrinsic brain: Towards robust measurement of inter-individual variation in 1000 functional connectomes. Neuroimage 80:246–262.
15. Van Dijk KRA, Sabuncu MR, Buckner RL (2012): The influence of head motion on intrinsic functional connectivity MRI. Neuroimage 59:431–438.
16. Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE (2012): Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59:2142–2154.
17. Kelly C, Biswal BB, Craddock RC, Castellanos FX, Milham MP (2012): Characterizing variation in the functional connectome: Promise and pitfalls. Trends Cogn Sci 16:181–188.
18. Satterthwaite TD, Wolf DH, Loughead J, Ruparel K, Elliott MA, Hakonarson H, *et al.* (2012): Impact of in-scanner head motion on multiple measures of functional connectivity: Relevance for studies of neurodevelopment in youth. Neuroimage 60:623–632.
19. Vanderwal T, Kelly C, Eilbott J, Mayes LC, Castellanos FX (2015): Inscapes: A movie paradigm to improve compliance in functional magnetic resonance imaging. Neuroimage 122:222–232.
20. Vanderwal T, Eilbott J, Castellanos FX (2019): Movies in the magnet: Naturalistic paradigms in developmental functional neuroimaging. Dev Cogn Neurosci 36:100600.
21. Engelhardt LE, Roe MA, Juranek J, DeMaster D, Harden KP, Tucker-Drob EM, Church JA (2017): Children's head motion during fMRI tasks is heritable and stable over time. Dev Cogn Neurosci 25:58–68.
22. George S, Duran N, Norris K (2014): A systematic review of barriers and facilitators to minority research participation among African Americans, Latinos, Asian Americans, and Pacific Islanders. Am J Public Health 104:e16–e31.
23. Habibi A, Sarkissian AD, Gomez M, Ilari B (2015): Developmental brain research with participants from underprivileged communities: Strategies for recruitment, participation, and retention. Mind Brain Educ 9:179–186.
24. Ramduny J, Vanderwal T, Kelly C (2024): Data rescue in high-motion youth cohorts for robust and reproducible brain-behavior relationships. bioRxiv https://doi.org/10.1101/2024.06.04.597447.
25. Parkes L, Fulcher B, Yücel M, Fornito A (2018): An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. Neuroimage 171:415–436.
26. Satterthwaite TD, Ciric R, Roalf DR, Davatzikos C, Bassett DS, Wolf DH (2019): Motion artifact in studies of functional connectivity: Characteristics and mitigation strategies. Hum Brain Mapp 40:2033–2051.
27. Ciric R, Wolf DH, Power JD, Roalf DR, Baum GL, Ruparel K, *et al.* (2017): Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. Neuroimage 154:174–187.
28. Casey BJ, Cannonier T, Conley MI, Cohen AO, Barch DM, Heitzeg MM, *et al.* (2018): The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. Dev Cogn Neurosci 32:43–54.
29. Mantwill M, Gell M, Krohn S, Finke C (2022): Brain connectivity fingerprinting and behavioural prediction rest on distinct functional systems of the human connectome. Commun Biol 5:261.
30. Sripada C, Angstadt M, Taxali A, Clark DA, Greathouse T, Rutherford S, *et al.* (2021): Brain-wide functional connectivity patterns support general cognitive ability and mediate effects of socioeconomic status in youth. Transl Psychiatry 11:571.
31. Kurashige H, Kaneko J, Yamashita Y, Osu R, Otaka Y, Hanakawa T, *et al.* (2019): Revealing relationships among cognitive functions using functional connectivity and a large-scale meta-analysis database. Front Hum Neurosci 13:457.

32. Zhu J, Qiu A (2022): Interindividual variability in functional connectivity discovers differential development of cognition and transdiagnostic dimensions of psychopathology in youth. Neuroimage 260:119482.
33. Shen X, Finn ES, Scheinost D, Rosenberg MD, Chun MM, Papademetris X, Constable RT (2017): Using connectome-based predictive modeling to predict individual behavior from brain connectivity. Nat Protoc 12:506–518.
34. Serin E, Zalesky A, Matory A, Walter H, Kruschwitz JD (2021): NBS-Predict: A prediction-based extension of the network-based statistic. Neuroimage 244:118625.
35. Smith SM, Nichols TE, Vidaurre D, Winkler AM, Behrens TEJ, Glasser MF, *et al.* (2015): A positive-negative mode of population covariation links brain connectivity, demographics and behavior. Nat Neurosci 18:1565–1567.
36. Noble S, Spann MN, Tokoglu F, Shen X, Constable RT, Scheinost D (2017): Influences on the test–retest reliability of functional connectivity MRI and its relationship with behavioral utility. Cereb Cortex 27:5415–5429.
37. Tiego J, Martin EA, DeYoung CG, Hagan K, Cooper SE, Pasion R, *et al.* (2023): Precision behavioral phenotyping as a strategy for uncovering the biological correlates of psychopathology. Nat Ment Health 1:304–315.
38. Dhamala E, Yeo BTT, Holmes AJ (2023): One size does not fit all: Methodological considerations for brain-based predictive modeling in psychiatry. Biol Psychiatry 93:717–728.
39. Kotov R, Krueger RF, Watson D, Achenbach TM, Althoff RR, Bagby RM, *et al.* (2017): The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. J Abnorm Psychol 126:454–477.
40. Xia CH, Ma Z, Ciric R, Gu S, Betzel RF, Kaczkurkin AN, *et al.* (2018): Linked dimensions of psychopathology and connectivity in functional brain networks. Nat Commun 9:3003.
41. Hagler DJ Jr, Hatton S, Cornejo MD, Makowski C, Fair DA, Dick AS, *et al.* (2019): Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. Neuroimage 202:116091.
42. Feczko E, Conan G, Marek S, Tervo-Clemmens B, Cordova M, Doyle O, *et al.* (2024): Adolescent Brain Cognitive Development (ABCD) community MRI collection and utilities. bioRxiv https://doi.org/10.1101/2021.07.09.451638.
43. Hodes RJ, Insel TR, Landis SC, , NIH Blueprint for Neuroscience Research (2013): The NIH toolbox: Setting a standard for biomedical research. Neurology 80(suppl 3):S1.
44. Gershon RC, Wagster MV, Hendrie HC, Fox NA, Cook KF, Nowinski CJ (2013): NIH toolbox for assessment of neurological and behavioral function. Neurology 80(suppl 3):S2–S6.
45. Bleck TP, Nowinski CJ, Gershon R, Koroshetz WJ (2013): What is the NIH Toolbox, and what will it mean to neurology? Neurology 80:874–875.
46. Achenbach TM, Rescorla LA (2001): Manual for the ASEBA School-Age Forms & Profiles: An Integrated System of Multi-informant Assessment. Burlington, VT: University of Vermont.
47. Miranda-Domínguez O, Perrone A, Earl E, Feczko E, Fair DA (2021): DCAN BOLD Processing. OSF. Available at: https://osf.io/huz54/. Accessed December 4, 2023.
48. Fair DA, Miranda-Dominguez O, Snyder AZ, Perrone A, Earl EA, Van AN, *et al.* (2020): Correction of respiratory artifacts in MRI head motion estimates. Neuroimage 208:116400.
49. Durston S, Tottenham NT, Thomas KM, Davidson MC, Eigsti IM, Yang Y, *et al.* (2003): Differential patterns of striatal activation in young children with and without ADHD. Biol Psychiatry 53:871–878.
50. Nebel MB, Lidstone DE, Wang L, Benkeser D, Mostofsky SH, Risk BB (2022): Accounting for motion in resting-state fMRI: What part of the spectrum are we characterizing in autism spectrum disorder? Neuroimage 257:119296.
51. Gordon EM, Laumann TO, Adeyemo B, Huckins JF, Kelley WM, Petersen SE (2016): Generation and evaluation of a cortical area parcellation from resting-state correlations. Cereb Cortex 26:288–303.
52. Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE (2013): Steps toward optimizing motion artifact removal in functional connectivity MRI; a reply to Carp. Neuroimage 76:439–441.
53. Power JD, Schlaggar BL, Petersen SE (2015): Recent progress and outstanding issues in motion correction in resting state fMRI. Neuroimage 105:536–551.
54. Benjamini Y, Hochberg Y (1995): Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300.
55. Lin LI (1989): A concordance correlation coefficient to evaluate reproducibility. Biometrics 45:255–268.
56. Shinn M, Hu A, Turner L, Noble S, Preller KH, Ji JL, *et al.* (2023): Functional brain networks reflect spatial and temporal autocorrelation. Nat Neurosci 26:867–878.
57. Altman DG (1999): Practical Statistics for Medical Research. New York: Chapman & Hall/CRC.
58. Cosgrove KT, McDermott TJ, White EJ, Mosconi MW, Thompson WK, Paulus MP, *et al.* (2022): Limits to the generalizability of resting-state functional magnetic resonance imaging studies of youth: An examination of ABCD study® baseline data. Brain Imaging Behav 16:1919–1925.
59. O'Connor D, Lake EMR, Scheinost D, Constable RT (2021): Resample aggregating improves the generalizability of connectome Predictive Modeling. Neuroimage 236:118044.
60. Wei L, Jing B, Li H (2020): Bootstrapping promotes the RSFC-behavior associations: An application of individual cognitive traits prediction. Hum Brain Mapp 41:2302–2316.
61. Li J, Bzdok D, Chen J, Tam A, Ooi LQR, Holmes AJ, *et al.* (2022): Cross-ethnicity/race generalization failure of behavioral prediction from resting-state functional connectivity. Sci Adv 8:eabj1812.
62. Greene AS, Shen X, Noble S, Horien C, Hahn CA, Arora J, *et al.* (2022): Brain-phenotype models fail for individuals who defy sample stereotypes. Nature 609:109–118.
63. Nastase SA, Gazzola V, Hasson U, Keysers C (2019): Measuring shared responses across subjects using intersubject correlation. Soc Cogn Affect Neurosci 14:667–685.
64. Ramduny J, Kelly C (2024): Connectome-based fingerprinting: Reproducibility, precision, and behavioral prediction. Neuropsychopharmacology 50:114–123.
65. Noble S, Scheinost D, Constable RT (2019): A decade of test–retest reliability of functional connectivity: A systematic review and meta-analysis. Neuroimage 203:116157.
66. Greene DJ, Marek S, Gordon EM, Siegel JS, Gratton C, Laumann TO, *et al.* (2020): Integrative and network-specific connectivity of the basal ganglia and thalamus defined in individuals. Neuron 105:742–758.e6.
67. Marek S, Greene DJ (2021): Precision functional mapping of the subcortex and cerebellum. Curr Opin Behav Sci 40:12–18.
68. Marek S, Siegel JS, Gordon EM, Raut RV, Gratton C, Newbold DJ, *et al.* (2018): Spatial and temporal organization of the individual human cerebellum. Neuron 100:977–993.e7.
69. Shah LM, Cramer JA, Ferguson MA, Birn RM, Anderson JS (2016): Reliability and reproducibility of individual differences in functional connectivity acquired during task and resting state. Brain Behav 6:e00456.
70. Gordon EM, Laumann TO, Gilmore AW, Newbold DJ, Greene DJ, Berg JJ, *et al.* (2017): Precision functional mapping of individual human brains. Neuron 95:791–807.e7.