Motion Impact Score for Detecting Spurious Brain-Behavior Associations

Authors

Benjamin P. Kay[1], David F. Montez[1], Scott Marek[2], Brenden Tervo-Clemmens[3], Joshua S. Siegel[2], Babatunde Adeyemo[1], Timothy O. Laumann[3], Athanasia Metoki[1], Roselyne J. Chauvin[1], Andrew N. Van[1,4], Samuel R. Krimmel[1], Ryland L. Miller[1], Dillan J. Newbold[1,16], Annie Zheng[1], Nicole A. Seider[1], Kristen M. Scheidter[1], Julia Monk[1], Eric Feczko[5,6], Anita Randolph[5,6], Oscar Miranda-Dominguez[5,6], Lucille A. Moore[5,7], Anders J. Perrone[5,7], Gregory M. Conan[5,7], Eric A. Earl[5,7], Stephen M. Malone[8], Michaela Cordova[7], Olivia Doyle[7], Benjamin J. Lynch[5,9], James C. Wilgenbusch[5,7], Thomas Pengo[10], Alice M. Graham[7], Jarod L. Roland[11], Evan M. Gordon[12], Abraham Z. Snyder[1,12], Deanna M. Barch[2,,12, 13], Damien A. Fair[5,6,14], Nico U.F. Dosenbach[1,4,12,13,15]

[1]Department of Neurology, Washington University School of Medicine, St. Louis, MO, USA

[2]Department of Psychiatry, Washington University School of Medicine, St Louis, MO, USA

[3]Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

[4]Department of Biomedical Engineering, Washington University in St Louis, St Louis, MO, USA

[5]Masonic Institute for the Developing Brain, University of Minnesota Medical School, Minneapolis, MN, USA

[6]Department of Pediatrics, University of Minnesota Medical School, Minneapolis, MN, USA

[7]Department of Psychiatry, Oregon Health and Science University, Portland, OR, USA

[8]Department of Psychology, University of Minnesota, Minneapolis, MN, USA

[9]Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, MN, USA

[10]University of Minnesota Informatics Institute, University of Minnesota, Minneapolis, MN, USA

[11]Department of Neurological Surgery, Washington University School of Medicine, St Louis, MO, USA

[12]Department of Radiology, Washington University School of Medicine, St. Louis, MO, USA

[13]Department of Psychological and Brain Sciences, Washington University in St Louis, St Louis, MO, USA

[14]Institute of Child Development, University of Minnesota Medical School, Minneapolis, MN, USA

[15]Department of Pediatrics, Washington University School of Medicine, St Louis, MO, USA

[16]Department of Neurology, NYU Langone Medical Center, New York, NY, USA

Abstract

Between-participant differences in head motion introduce systematic bias to resting state fMRI brain-wide association studies (BWAS) that is not completely removed by denoising algorithms. Researchers who study traits, or phenotypes associated with in-scanner head motion (e.g. psychiatric disorders) need to know if trait-functional connectivity (FC) correlations are biased by residual motion artifact in order to avoid reporting false positive results. We devised an adaptable method, Split Half Analysis of Motion Associated Networks (SHAMAN), to assign a motion impact score to specific trait-FC correlations. The SHAMAN approach distinguishes between motion artifact causing false positive vs false negative bias. SHAMAN was > 95% specific at sample sizes of n = 100 and above. SHAMAN was 95% sensitive to detection of false positive motion impact score at sample sizes of n = 3,000 but only 59% sensitive to detection of false negative motion impact score, making it most useful for large BWAS. We computed motion impact scores for trait-FC correlations with 45 demographic, biophysical, cognitive, and personality traits from n = 7,270 participants in the Adolescent Brain Cognitive Development (ABCD) Study. After standard denoising with ABCD-BIDS and without motion censoring, 60% (27/45) of traits had significant ($p < 0.05$) false positive motion impact scores and 36% (16/45) of traits had false negative motion impact scores. Censoring at framewise displacement (FD) < 0.2 mm reduced the proportion of traits with significant false positive motion impact scores from 60% to 2% (1/45) but did not decrease the number of traits with significant false negative motion impact scores.

Introduction

Head motion is the largest source of artifact in structural and functional MRI (fMRI) signals[1–20]. The technical challenge posed by motion cannot be overstated and has motivated the creation of behavioral interventions [9,21–24] and real-time motion tracking software to reduce the amount of in-scanner head motion [5,25]. Even with highly compliant participants, involuntary sub-millimeter head movements systematically alter fMRI data [6,14,17,19,20]. Unfortunately, non-linear characteristics of MRI physics make removal of motion artifact during post-processing difficult [26]. Compared to task fMRI, resting-state functional connectivity (FC) is especially vulnerable to motion artifact because the timing of the underlying neural processes is unknown [2,4,6,9,10,12,14,19,20,27–30]. The effect of motion on FC is spatially systematic, causing decreased long-distance connectivity and increased short-range connectivity, most notably in the default mode network [19,20,27,28].

Results from early studies of children, older adults, and patients with neurological or psychiatric disorders have been spuriously related to motion [6,19,29,30]. For example, motion artifact systematically decreases FC between distant brain regions [28] leading some investigators to conclude that autism decreases long-distance FC when, in fact, their results were due to increased head motion in autistic study participants. These cautionary findings have motivated the creation of numerous approaches to mitigate motion artifact including global signal regression [28,31], motion parameter regression [29], spectral filtering, respiratory filtering [25,32,33], principal component analysis [1,11], independent component analysis [7,15,34], multi-echo pulse sequences [8], despiking of high-motion frames [13], and combinations thereof [3,4,16,25]. However, given the complexity of these approaches, it is difficult to be certain that enough motion artifact has been removed to avoid spurious (false positive) interpretation of trait-FC correlations and detect real trait-FC correlations (avoid false negatives).

It is increasingly common for brain-wide association studies (BWAS) involving many thousands of participants (e.g. HCP, ABCD, UK Biobank) to provide data that have already been processed to remove motion [35–38]; however, even in these cases there is some choice as to how much data to retain or to censor. In such large studies, obtaining the raw data and re-applying a different motion processing method is computationally expensive. Excluding high-motion fMRI frames (timepoints) from analysis by censoring is a post-hoc approach shown to further reduce residual motion artifact [14,20,27,29]. Power et al. [39] note a natural tension between the need to remove some motion-contaminated volumes to reduce spurious findings (false positive inference) but not so many volumes as to bias the sample distribution of a trait by systematically excluding individuals with high motion who may exhibit important variance in the trait of interest (e.g., low scores on attention measures associated with greater motion).

This difficulty in censoring threshold selection arises in part because most approaches for quantifying motion are agnostic to the hypothesis under study [6,14,19,20,29]. However, some traits or groups of participants are more strongly correlated with motion than others. For example, study participants with attention-deficit hyperactivity disorder or autism have higher in-scanner head motion than neurotypical participants [5,6,9,30]. Even when much of the overall signal variance associated with motion has been removed, inferences about such motion-correlated traits may still be significantly impacted by motion artifact [6,9,18,20,29,30]. Therefore, in addition to standard approaches for quantifying motion, methods for quantifying trait-specific motion artifact in FC are needed.

Approaches for quantifying the association of specific trait-FC correlations with motion include measuring changes in distance-dependent correlations at different levels of motion censoring [27,28], measuring spatial (i.e. across edges) similarity of trait-FC effects with motion-FC effects [29,30], and measuring differences in trait-FC effects between groups matched according to levels of motion [6,19]. However, these measures do not establish a threshold for acceptable or unacceptable levels of trait-specific motion. Nielsen et al. [12] proposed using support vector machines to test whether head motion is significantly predicted by FC, but the method applies only to high-order multivariate models of trait-FC effects. Siegel et al. [18] proposed conceptualizing the relationship between motion and trait-FC correlations by comparing within-participant and between-participant variance in the trait-FC correlations explained by motion. Siegel's original method was limited because it required repeated resting-state fMRI (rs-fMRI) scans of the same participant on different days, and it used a simple correlation measure that could not account for covariates or distinguish between motion artifacts contributing to false positive vs false negative results.

Thus, we developed a novel method for detecting trait-specific motion impact score that operates on one or more rs-fMRI scans per participant and can be adapted to model covariates. We capitalize on Siegel et al.'s [18] observation that traits (e.g. weight, intelligence) are stable over the timescale of an MRI scan whereas motion is a state that varies from second to second. The proposed **S**plit **H**alf **A**nalysis of **M**otion **A**ssociated **N**etworks (SHAMAN) capitalizes on the relative stability of traits over time by measuring difference in the correlation structure between split high- and low-motion halves of each participant's fMRI timeseries. When the trait-FC correlations are independent of motion, the difference in each half of the connectivity will be not-significant because traits are stable over time. A significant difference is detected only when state-dependent differences in motion have an impact on the trait's connectivity. Motion-induced differences that enhance the trait-FC correlations contribute to false-positive inference, whereas motion-induced differences that reduce the trait-FC correlations are associated with false-negative inference. Permutation of the timeseries and non-parametric combining [40] across pairwise connections yields a motion impact score and a p-value distinguishing significant from not-significant impacts of motion on trait-FC correlations.

Recently the Adolescent Brain Cognitive Development (ABCD) Study collected up to 20 minutes of rs-fMRI data on 11,874 children ages 9-10 years [21,41,42] with extensive demographic, biophysical, and behavioral data [43,44]. Such large data sets have made it possible to quantify reproducibility in resting-state fMRI, and they have revealed that the true effect sizes of brain-wide association studies (BWAS) are smaller than previously thought due to sampling variability [45]. Failure to consider head motion is another source of inconsistent results. Thus, we first characterized the effectiveness of standard denoising approaches at reducing motion artifact. Then we considered the residual trait-specific impact of head motion on FC in the high-quality ABCD data after denoising and varying levels of post-hoc motion censoring.

Results

<u>The Effect of Residual Motion is Large Even After Denoising and Censoring</u>

In order to characterize the impact of residual head motion on trait-FC correlations, we first performed preliminary analyses to quantify how much residual motion was left in the data after denoising. ABCD-BIDS is the default denoising algorithm for pre-processed ABCD data [25,36]. It includes global signal regression, respiratory filtering, spectral (low-pass) filtering, despiking, and regressing out the motion parameter timeseries. The relative performance of ABCD-BIDS was evaluated by comparing how much of the

between-participant variability in the fMRI timeseries (averaged across regions of interest) was explained by head motion (framewise displacement, FD) before and after applying ABCD-BIDS (Figure S1). After minimal processing [38] (i.e. motion-correction by frame realignment only), 9% of signal variance was explained by head motion (r = 0.29). After further denoising using ABCD-BIDS (respiratory filtering, motion timeseries regression, despiking/interpolation of high-motion frames), less than 1% of signal variance was explained by head motion (r = 0.03). Therefore, ABCD-BIDS achieved a relative reduction in the proportion of signal variance related to motion of 98.9% compared to minimal processing alone (see Methods).

However, even after denoising with ABCD-BIDS, it was still possible to detect large between-participant differences due to head motion. The average FC matrix across all participants is shown in Figure 1a,b. The residual motion-FC effect was quantified by regressing each participant's FD (averaged over all resting-state scans) against their FC to generate a motion-FC matrix (Figure 1c,d). The average FC matrix and motion-FC matrix were compared by computing the correlation between their edges (pairwise connections). The motion-FC matrix had a strong, negative correlation of r = -0.63 with the average connectivity matrix. This strong, negative correlation persisted even after motion censoring at FD < 0.2 mm (r = -0.55). Thus, across all functional connections, as a participant moved more, the magnitude of that participant's functional connectivity tended to decrease.

The decrease in FC due to head motion was larger than the increase or decrease in FC related to traits of interest (Figure 1e,f). The largest motion-FC effect size for a single connection was |r| = 0.10. More conservatively, the 98th percentile for the motion-FC effect size was |r| = 0.04. While these effect sizes were small in absolute terms, Marek et al.[45] have shown that the largest 1% of univariate trait-FC effect sizes in ABCD are on the order of |r| = 0.06. Of the 45 variables we evaluated, parental income bracket had the largest trait-FC effect size at |r| = 0.06 (98th percentile |r| = 0.02). With motion censoring at FD < 0.2 mm the largest motion-FC effect was |r| = 0.08 (98th percentile |r| = 0.03) and the largest effect size for parental income was |r| = 0.06 (98th percentile |r| = 0.02). Therefore, even with motion denoising and censoring, the motion-FC effect was large in relation to trait-FC effects.

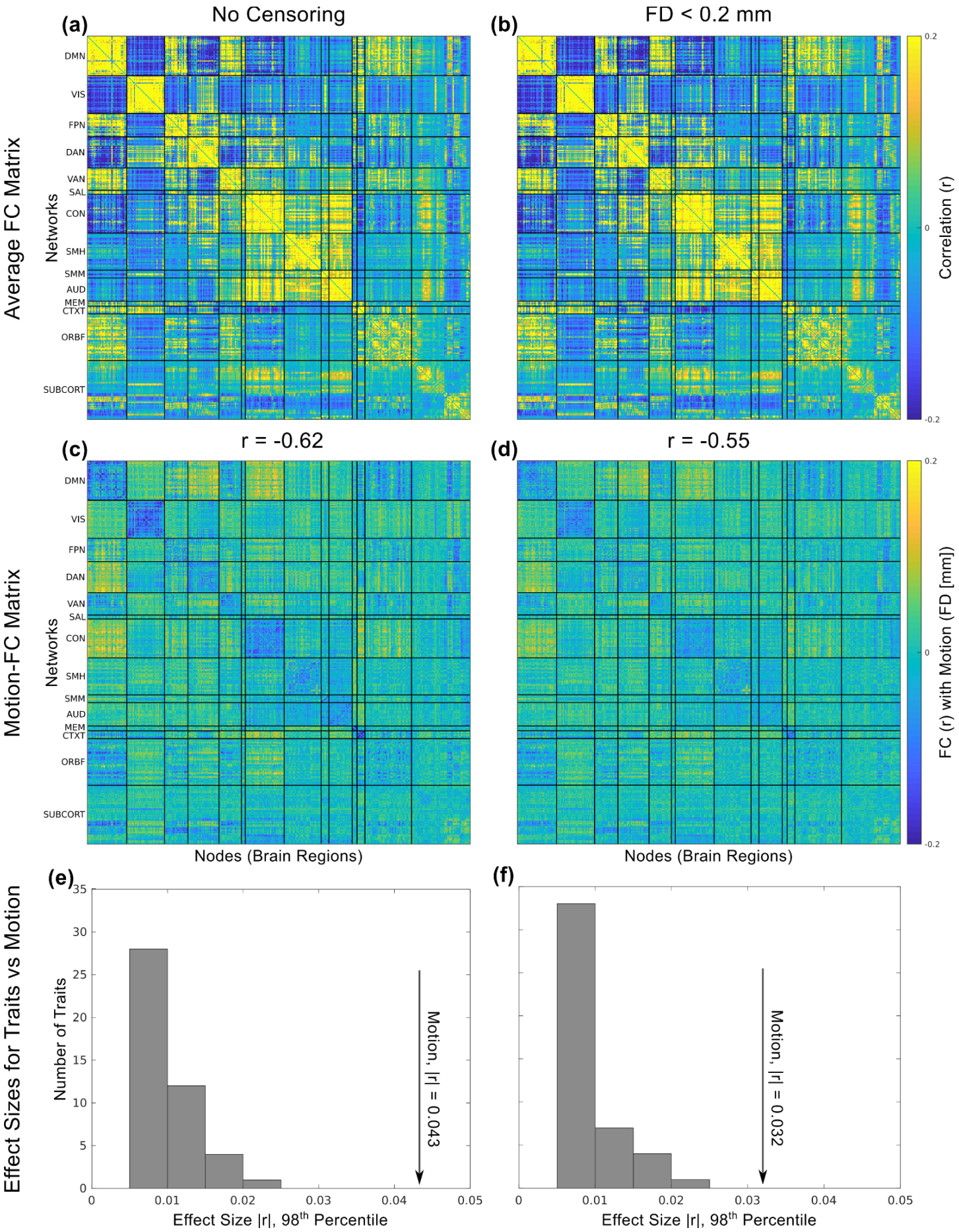
(a)
No Censoring
(b)
FD < 0.2 mm
Average FC Matrix
Networks
DMN
VIS
FPN
DAN
VAN
SAL
CON
SMH
SMM
AUD
MEM
CTXT
ORBF
SUBCORT
Correlation (r)
0.2
0
-0.2
(c)
r = -0.62
(d)
r = -0.55
Motion-FC Matrix
FC (r) with Motion (FD [mm])
Nodes (Brain Regions)
Nodes (Brain Regions)
(e)
(f)
Effect Sizes for Traits vs Motion
Number of Traits
Motion, |r| = 0.043
Motion, |r| = 0.032
Effect Size |r|, 98th Percentile
Effect Size |r|, 98th Percentile
0
5
10
15
20
25
30
35
0.01
0.02
0.03
0.04
0.05

Figure 1: Average connectivity matrix, motion-connectivity matrix, and effect sizes. (a) The average (across participants) correlation matrix for functional connectivity (FC), after denoising with ABCD-BIDS without frame censoring; **(b)** and with frame censoring at framewise displacement (FD) < 0.2 mm. **(c)** The motion-FC correlation matrix, i.e. correlation matrix for functional connectivity (FC) with framewise displacement (FD) without frame censoring **(d)** and with censoring at FD < 0.2. **(a,c)** The upper triangular parts of the FC connectivity matrices were edge-for-edge correlated at r = -0.63. **(b,d)** After motion censoring at FD < 0.2 mm, the average FC and motion-FC correlation matrices were edge-for-edge correlated at r = -0.55. **(e)** The effect size of the motion-FC correlation (arrow) is plotted relative to the empirical distribution of effect sizes for the 45 traits in this study without motion censoring and **(f)** with motion censoring at framewise displacement (FD) < 0.2 mm. Effect sizes were computed as the largest (98th percentile) normalized difference in functional connectivity (correlation, |r|) across edges for each trait. Networks: DMN = Default Mode Network, VIS = Visual, FPN = Frontoparietal Network, DAN = Dorsal Attention Network, VAN = Ventral Attention Network, SAL = Salience, CON = Cingulo-Opercular Network, SMH = Somatomotor Hand, SMM = Somatomotor Mouth, AUD = Auditory, MEM = Parietal Memory Network, CTXT = Context, ORBF = Orbitofrontal, SUBCORT = subcortical regions.

Most Traits Were Significantly Correlated with Head Motion

Surprisingly, many seemingly unrelated traits were shown to be correlated with head motion in the Human Connectome Project (HCP). [18] In ABCD, 87% (39/45) of traits or variables evaluated showed significant ($p < 0.05$) correlations between the trait (e.g. BMI) and head motion (FD, in mm; Table S1). WISC-V (Wechsler Intelligence Scale for Children, 5th edition) matrix reasoning subscore [46] had the highest trait-FD correlation at $r = -0.12$, $p < 0.001$, and Child Behavioral Checklist (CBCL) somatization subscore had the least trait-FD correlation at $r = 0.0004$ (not significant).

BMI [47] and matrix reasoning subscore [46] were selected as exemplar traits. Both exhibited significant trait-FD correlations (Figure S2). Children with higher BMI exhibited more in-scanner head motion ($r = 0.09$, $p < 0.001$). Children with higher matrix reasoning scores had lower head motion ($r = -0.12$, $p < 0.001$).

SHAMAN Detected the Impact of Motion on Trait-FC Correlations in Simulated Data

To clarify and validate the principle of the SHAMAN approach, a generative model was used to simulate fMRI timeseries data with experimentally-controlled amounts of "BRAIN" trait-FC correlations, “MOTION” trait-FC correlations, and relationship between them. Simulation of non-linear relationships between the trait-FC correlations and motion caused denoising to fail and for SHAMAN to detect a significant impact of residual motion on trait-FC correlations (motion impact score) as depicted in Figure 2. The SHAMAN algorithm is described further in the Methods, and simulation methods are described further in the Simulation Supplement. Briefly, the FC matrix for a simulated participant contained both "BRAIN" and "MOTION" correlations (Figure 2a), reflecting the failure of denoising to completely remove the "MOTION" correlations. The participant's fMRI timeseries was split into high- and low-motion halves based on the FD timeseries. More "MOTION" was visible in the FC matrix from the high-motion half of the timeseries compared to the low-motion half, whereas the amount of “BRAIN” did not vary over time and was therefore was equal between halves (Figure 2b). When the low-motion trait-FC matrix was subtracted from the high-motion trait-FC matrix, the identical "BRAIN" trait-FC correlations canceled out in the difference FC matrix (Figure 2d) whereas the “MOTION” did not. Finally, the "BRAIN" trait was regressed against the difference FC matrices to generate a matrix of motion impact score (Figure 2e).

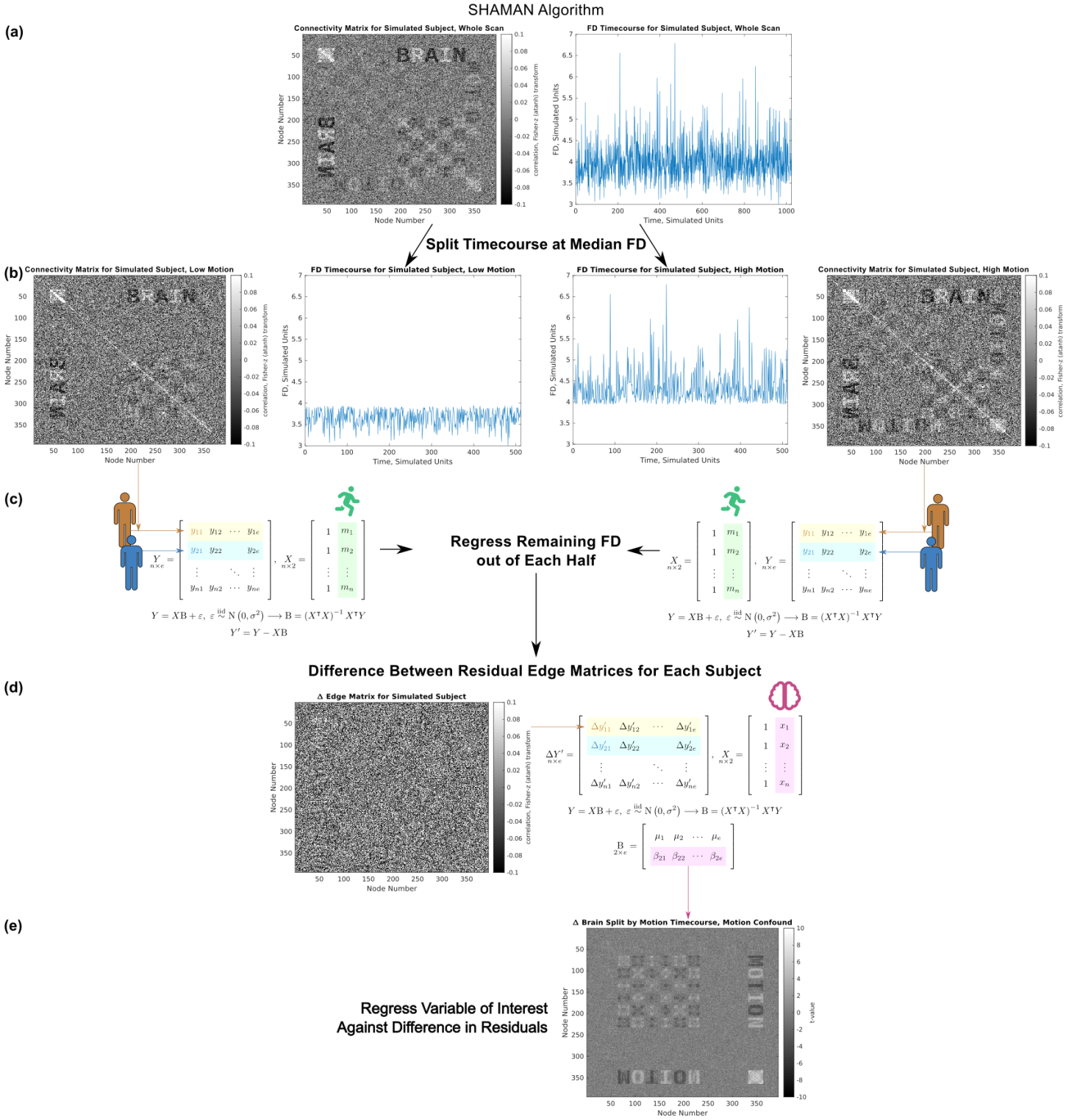

Figure 2: Diagram of the SHAMAN (Split-Half Analysis of Motion-Associated Networks) algorithm using simulated data. The simulated timeseries of brain signal generates the "BRAIN" edge matrix, and the simulated timeseries of motion generates the "MOTION" edge matrix. **(a)** A single participant's simulated functional connectivity (FC) matrix and motion (framewise displacement, FD) timeseries. **(b)** Each participant's fMRI timeseries data is split into high- and low-motion halves using the FD timeseries. High- and low-motion FC matrices and corresponding motion timeseries are shown for the simulated participant. **(c)** Between-participant variation in motion is regressed out of each half edge matrix to generate residual FC matrices. The difference between the high- and low-motion residual FC matrices is shown for the simulated participant. **(d)** The trait of interest is regressed against the difference FC matrices. **(e)** Motion impact score specific to the trait in (d).

Residual Motion Inflates Functional Connectivity Associations with Biophysical Traits

We found that residual motion artifact inflated the effect size of many traits, especially the biophysical traits of BMI, age, sex, weight, and height. Trait-FC correlations were calculated separately for each of these traits on fMRI data after denoising with ABCD-BIDS using conventional methods with mean framewise displacement (FD) as a covariate to control for between-participant differences in motion (e.g., FC ~ BMI + FD). Without frame censoring, all five biophysical traits had significant false positive motion impact scores, meaning that their trait-FC correlations were significantly ($p < 0.05$) greater due to the impact of residual head motion. Three biophysical traits, BMI, sex, and weight, also had significant false negative motion impact scores, meaning that motion inflated trait-FC correlations in some networks but suppressed them in other networks (Tables S2, S3).

BMI was selected as an exemplar trait to illustrate a false positive motion impact score. The FC matrix for the trait-FC correlations of BMI (FD as a covariate, without motion censoring) was compared to its motion impact score at each edge in Figure 3 and projected onto the cortical surface using root mean square connectivity [48] in Figure 4. Many of the network pairs (e.g. on the diagonal: default mode, visual, dorsal attention, and cingulo-opercular) had a negative trait-FC correlation and a negative motion impact. Since these trait-FC correlations and corresponding motion impacts were both in the same direction (negative in this case), the impact of motion was to inflate the size of the trait-FC correlations. An overall false positive motion impact (Stouffer's Z[40]) score for BMI of 91 was calculated using omnibus combining across the entire connectivity matrix to control for multiple comparisons. An omnibus significance of $p < 0.001$ was obtained using permutation.

Note that regressing against the inverse of BMI, a trait for which high values were associated with low in-scanner head motion, caused the trait-FC correlations of inverse-BMI and the motion impact to both be positive instead of negative (data not shown). Since the trait-FC correlations and corresponding motion impact were still both in the same direction (positive in this counterexample), the overall motion impact score was still false positive.

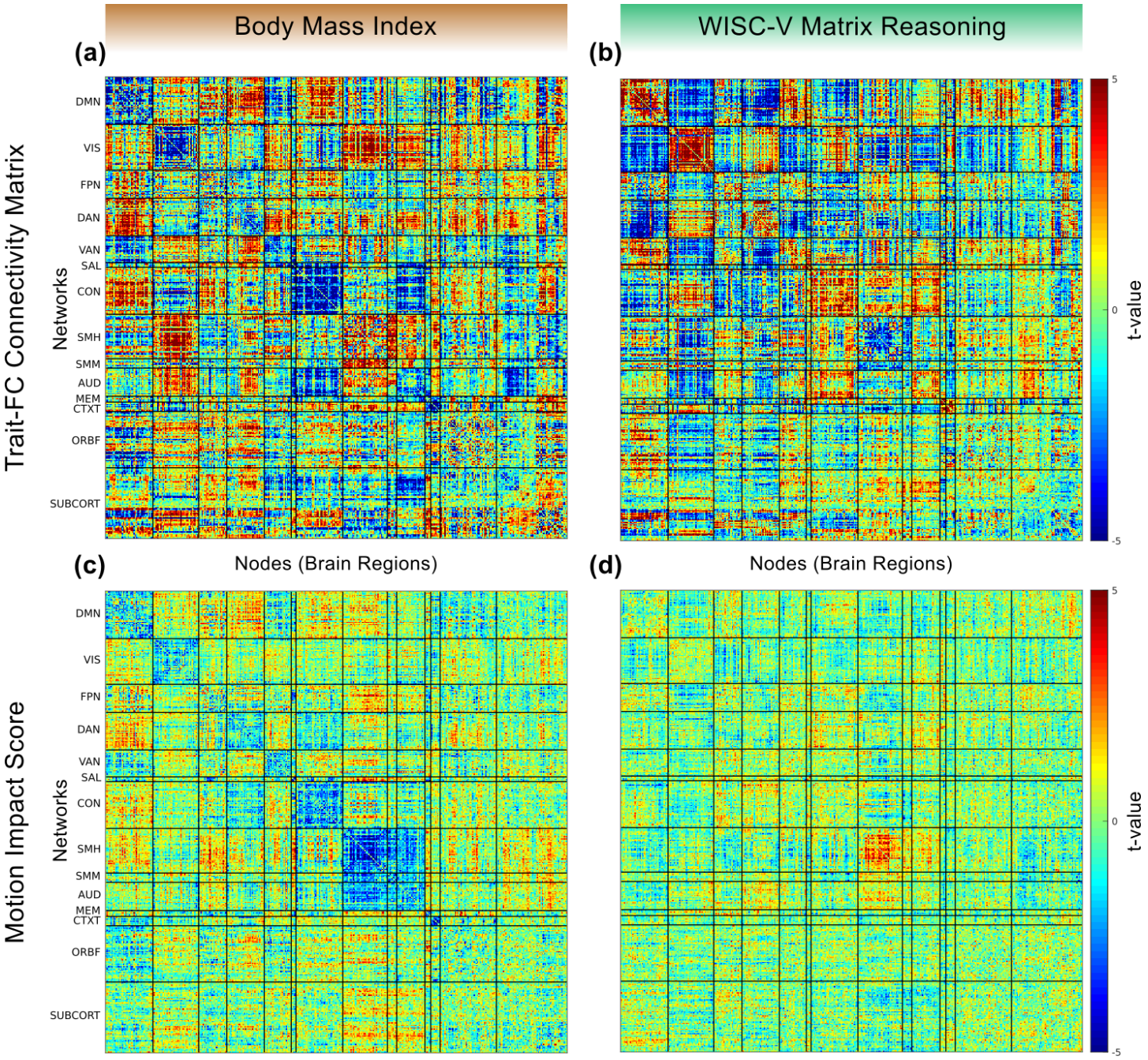

Figure 3: Trait-specific impact of motion on functional connectivity matrices. Data are shown for two exemplar traits, body mass index (BMI, left) and WISC-V matrix reasoning score (right). **(a)** Resting state functional connectivity (RSFC) matrix for BMI. **(b)** Connectivity matrix for matrix reasoning. **(c)** Motion impact score for BMI. The overall (over the entire matrix, omnibus Stouffer's Z) false positive motion impact score for BMI was 91, $p < 0.001$. **(d)** Motion impact score for WISC-V. The omnibus false negative motion impact score for WISC-V was 55, $p < 0.001$. All measures were computed after motion processing with ABCD-BIDS and without frame censoring. Connectivity matrices were computed with framewise displacement (FD) as a covariate. Networks: DMN = Default Mode Network, VIS = Visual, FPN = Frontoparietal Network, DAN = Dorsal Attention Network, VAN = Ventral Attention Network, SAL = Salience, CON = Cingulo-Opercular Network, SMH = Somatomotor Hand, SMM = Somatomotor Mouth, AUD = Auditory, MEM = Parietal Memory Network, CTXT = Context, ORBF = Orbitofrontal, SUBCORT = subcortical regions.

Residual Motion Can Inflate and Suppress Associations with Demographic and Behavioral Trait

WISC-V matrix reasoning score was selected as an exemplar trait to illustrate a false negative motion impact score (Figures 3, 4). The overall false negative motion impact score for WISC-V of 55 was significant at (omnibus) $p < 0.001$. The false positive motion impact score for WISC-V was not significant. The trait-FC correlation matrix for WISC-V (FD as a covariate, without motion censoring) and its respective motion impact scores are shown in Figure 3 and projected onto the cortical surface in Figure 4. In networks where the motion impact was significant (e.g. sensorimotor hand), it trended in the opposite direction (positive vs negative) of the corresponding trait-FC correlation, thus suppressing the trait-FC correlation.

In total, 53% (20/38) of demographic and behavioral traits had significant false positive motion impact scores, and 32% had significant false negative motion impact scores, prior to frame censoring (Table S2, S3). Broken down further by category, there was a false positive impact of motion for 20% (1/5) of demographic traits, 60% (6/10) of cognitive traits, and 59% (13/22) of personality traits. There was a false negative impact of motion for 67% (4/6) of demographic traits, 30% (3/10) of cognitive traits, and 22% (5/22) of personality traits. Thus, demographic and behavioral traits were less impacted by residual motion than biophysical traits.

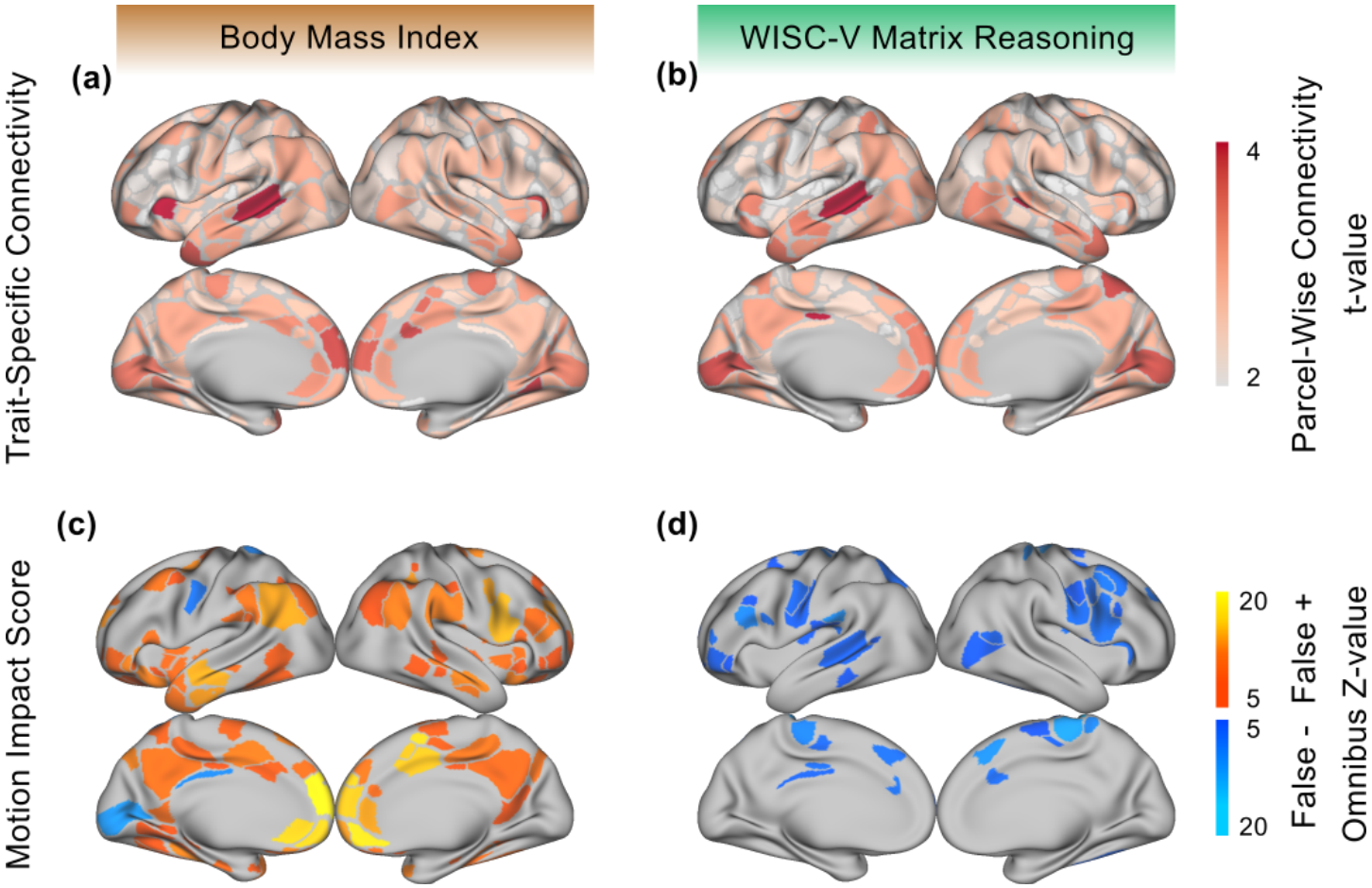

Figure 4: Trait-specific impact of motion on functional connectivity (FC). Top: Parcel-wise FC, computed as the root mean square (RMS) of connectivity values for each parcel/node in the connectivity matrix [48] for **(a)** body mass index (BMI) and **(b)** WISC-V matrix reasoning score. **Bottom:** Motion impact score (omnibus Stouffer's Z [40,49,50], higher = more motion) for **(c)** BMI and **(d)** WISC-V. Spurious inflation of the trait-FC correlation by motion is labeled False + (positive, in orange). Suppression of trait-FC correlations by motion is labeled False - (negative, in blue).

Frame Censoring after Denoising Reduces False Positives due to Motion Artifact

Frame censoring is a simple, post-hoc approach to address residual motion artifact after motion processing; however, censoring also has the potential to bias sample proportions through exclusion of high-motion participants [14,20,27,29]. The SHAMAN algorithm was used to quantify the tradeoff between reduction in motion impact score and sampling bias due to the exclusion of high motion participants (Tables S2-S4).

Prior to frame censoring, 7,270 participants had at least 8 minutes of rs-fMRI data and were not missing data on any of the 45 trait traits studied. Frame censoring at FD < 0.2 mm (filtered for respiratory motion)[25,32,33] and excluding participants with < 8 minutes of data remaining excluded 11% (818/7270) of participants. Exclusion of these participants caused the average values of only two traits, gender and number of MRI runs completed, to shift by more than 1% (Figure S3) compared to the uncensored sample of participants. On the other hand, censoring at FD < 0.2 reduced the number of traits with false positive motion impact scores from 60% (27/45) to just one physical trait, height, see Figure 5 (Tables S2, S3).

More stringent censoring at FD < 0.1 mm (filtered) did not reduce the number of traits with false positive motion impact score any further, but exclusion of 36% (2712/7270) participants at FD < 0.1 shifted the average values of 24% (11/45) of traits by > 1%, compared to no censoring (Table S4). For example, boys moved their heads 0.07 mm more, on average (mean FD), than girls ($r = 0.10$, $p < 0.001$). At a very stringent motion censoring threshold of FD < 0.1, the study population shifted from majority boys (52.6%) to majority girls (51.6%; Figure S3). Motion censoring at FD < 0.1 did reduce false negative motion impact from 44% (20/45) at FD < 0.2 to 20% (9/45) at FD < 0.1.

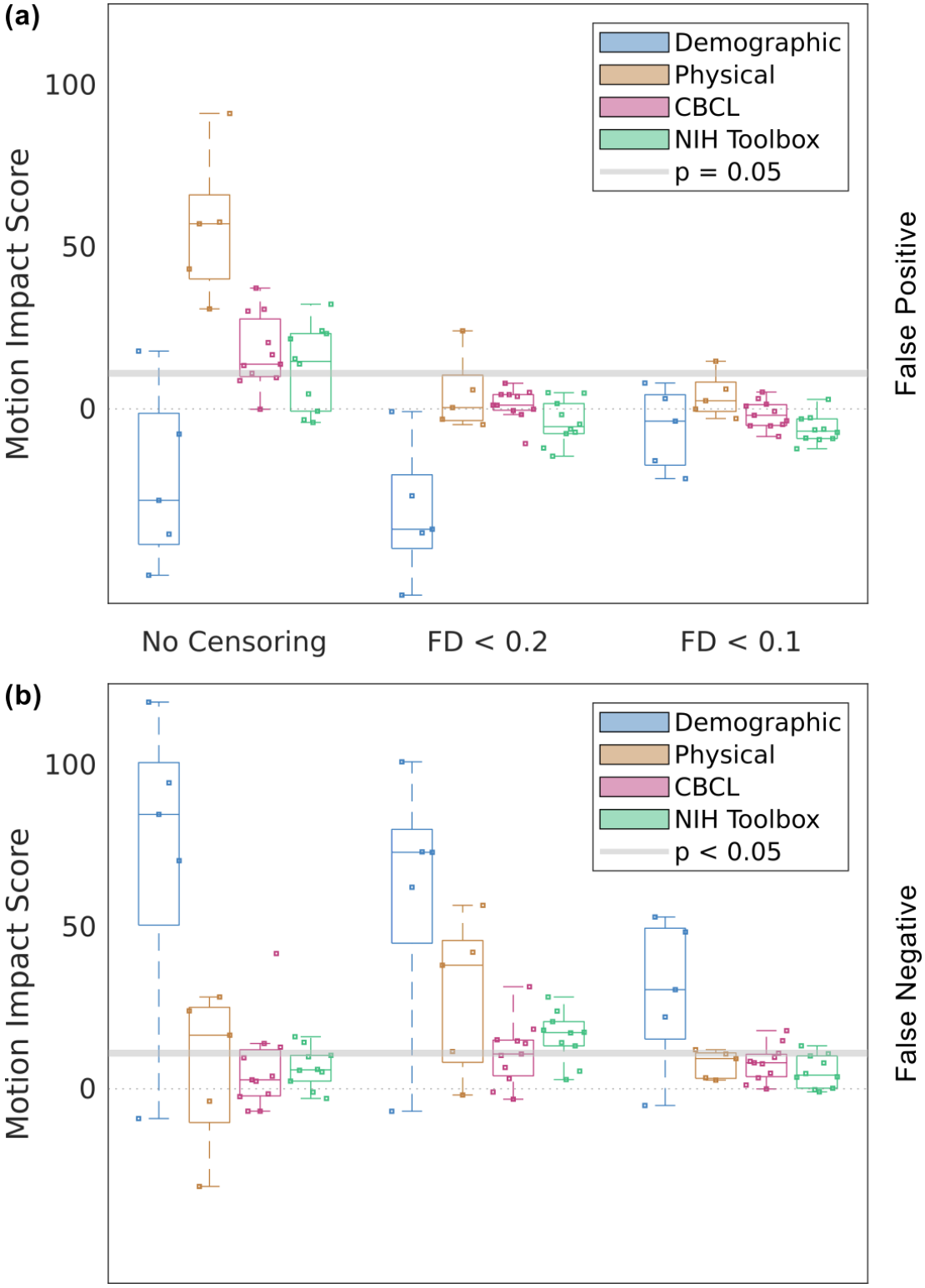

Figure 5: Effects of frame censoring on trait-specific motion distortion. Motion impact score (omnibus Stouffer's Z, higher = more motion) for traits of a given class at different levels of motion censoring. The significance threshold is indicated by a gray line. **(a)** False positive motion impact score. **(b)** False negative motion impact score. Trait classes are demographic (number of people living in household, combined parental income bracket, parents' marital status, highest parental education level, and residential deprivation index), physical (sex assigned at birth, age in months, height, weight, and body mass index), measures from the Child Behavior Check List (CBCL), and measures from the NIH cognitive toolbox.

Motion Impact Was Distributed Across Brain Regions

Motion impact scores were computed at multiple levels of granularity: for individual pairwise connections (edgewise), combined (using omnibus Stouffer's Z) across edges in a region or node (parcelwise), or combined across the whole brain. Significance at the whole-brain level could theoretically be driven by extremely high motion impact scores concentrated in a few regions or by high motion impact scores distributed across multiple brain regions. To quantify the regional distribution of motion impact score, the whole-brain motion impact score was recalculated after iteratively excluding the brain region with the highest motion impact score until the whole-brain motion impact score was no longer significant. The number of regions excluded was the number of regions significantly contributing to the impact of motion across the brain. We found that, especially for false positive motion impact scores, significant impacts of motion were distributed across many brain regions. BMI and matrix reasoning were selected as exemplar traits to visualize the number of motion-impacted regions needed to detect a significant impact of motion across the whole brain (Figure S4). For BMI, 39% (155/394) of regions contributed to false positive motion impact score whereas < 1% of regions contributed to false negative motion impact score. For matrix reasoning, 19% (76/394) of regions contributed to false negative motion impact score; the whole-brain false positive motion impact score for matrix reasoning was not significant.

Trait-FD Correlation Did Not Independently Predict Trait-FC Correlation

Calculating motion impact score is computationally intensive and might be unnecessary if spurious trait-FC correlations could be predicted from the trait-FD correlation alone. A post-hoc analysis was performed to test whether trait-FD correlation predicted motion impact score. Large trait-FD correlation failed to predict false positive motion impact score ($r = 0.02$, not significant) and therefore did not detect spurious trait-FC correlations without the aid of SHAMAN (Figure S5). This was true even when limiting analysis to behavioral traits, which tend to be of greatest interest to researchers (33 traits, $r = 0.15$, not significant). Large trait-FD correlation was correlated with false negative motion impact score ($r = 0.33$, $p = 0.03$).

Testing for Trait-specific Motion Effects Requires Thousands of Participants

Our analyses were conducted on a large sample ($n = 7{,}270$). We performed bootstrapping (sampling with replacement) at different sample sizes to assess the sensitivity, or statistical power, of SHAMAN to detect significant motion impact scores in our exemplar traits, BMI and matrix reasoning, as a function of sample size. As expected, power increased as the sample size increased (Figure S6). A sample size of about $n = 2{,}000$ participants was powered ($\alpha = 0.05$) to detect a false positive motion impact score for BMI in 89% of simulations. At $n = 3{,}000$, false positive motion impact score was successfully detected 97% of the time, par for expectation with a significance threshold of $\alpha = 0.05$. The false negative motion impact score for matrix reasoning was more difficult to detect, with successful detection in only 92% of simulations at $n = 6{,}000$. The need for thousands of participants to avoid false positive inference is consistent with recent fMRI studies.[45]

Motion Impact Score is Specific to Motion Artifact

To be useful, SHAMAN must also be specific, producing not-significant motion impact scores when trait-FC correlations are not biased by residual motion artifact. The randomly-assigned participant ID was used as a trait to simulate a trait-FC correlation not biased by residual motion artifact. Participant ID was not significantly correlated with FD ($r = 0.0006$). We performed bootstrapping with participant ID at different sample sizes to assess SHAMAN's specificity. At the smallest sample size of $n = 100$ participants the rate of incorrect detection of a significant motion impact score (either false positive or false negative) with the random participant ID was 3.9%, par for expectation with a significance threshold of $\alpha = 0.05$ (Figure S6). The rate of incorrect detection was zero at each of the larger sample sizes tested up to $n = 6{,}000$ participants.

Discussion

The Impact of Motion on Brain-Wide Association Studies is Widespread and Trait Dependent

It has been proposed that head motion likely affects inference about some trait-FC correlations more than others [6,19,27–30], but only a few studies [12,18] have explored the differential impact of motion on specific trait-FC correlations within a statistical framework. We build upon this prior work [18] by employing a novel metric, motion impact score, to quantify the impact of residual in-scanner head motion on trait-FC correlations for 45 demographic, biophysical, intelligence, and personality traits within the ABCD study. Similar to findings in the Human Connectome Project (HCP data, [18] we find that, even after denoising with ABCD-BIDS a large proportion (60%) of trait-FC correlations in the ABCD data (n = 7,270) are significantly inflated by residual head motion when motion censoring is not used, and that post-hoc motion censoring is an effective strategy for mitigating residual motion.

Motion May Negatively Affect the Signal-to-Noise-Ratio (SNR)

We find that, after denoising, the motion-FC connectivity matrix has a large effect size (max |r| = 0.10) and is strongly anti-correlated with the between-participant average connectivity matrix (r = -0.63), thus potentially impacting the interpretation of many traits whose FC correlations are shaped by the average connectivity matrix. The most parsimonious explanation for this motion-connectivity anti-correlation is that data from participants with higher motion have a lower signal to noise ratio (SNR). Decreased SNR could occur due to the biophysical characteristics of head motion increasing the proportion of noise relative to brain signal captured by the MRI scanner. It is also possible that denoising algorithms inadvertently remove more brain signal from high-motion data than they do from low-motion data. Designers of denoising algorithms may wish to evaluate not only how much motion is removed from the data, but also how the SNR of the data are affected by processing.

Other factors contributing to the observed motion-FC connectivity matrix likely include distance-dependent attenuation of connectivity overlapping the average connectivity matrix. [39] Propensity of an individual to move is itself a heritable trait,[51–53] and some [54] have even proposed that intrinsic differences in FC related to movement are encoded in the connectivity matrix. While our findings do not exclude the possibility that motion has an intrinsic neurobiological basis, we find it improbable that motion's trait-FC effect would have a larger effect size than any other trait-FC effect.

Biophysical Traits Were Most Likely to be Confounded by Motion Distortions

While we find that many trait-FC relationships (60%) are spuriously inflated by motion, it is remarkable that all five biophysical trait-FC relationships (BMI, age, sex, weight, and height) are significantly inflated by motion. Biophysical traits are also the only class for which we detect significant false positive motion impact after censoring. As seen in Figure 3 illustrating BMI, the largest deviations in the trait-FC connectivity matrices for biophysical traits cluster along the diagonal and have a large degree of similarity to both the average (across participants) FC matrix and the FD-FC connectivity matrix. It is possible that the biophysical traits are directly related to motion (e.g. larger head sizes are capable of greater displacement), affect signal to noise ratio in a similar way to motion (e.g. excess adipose tissue attenuates radiofrequency energy, decreasing SNR), or that biophysical traits serve as a proxy for other behavioral traits related to movement. If the relationship between biophysical traits and SNR or behavioral covariates could be modeled then their true trait-FC relationships could be identified.

Motion Censoring Reduces False Positives

Motion censoring at FD < 0.2 mm (filtered for respiratory artifact) is very effective at mitigating false positive motion impact score (Figure 5, Tables S2, S3) and results in relatively little sample bias due to participant exclusion (Figure S3, Table S4). Siegel et al. [18] also observed decreased residual impact of motion after framewise censoring in the HCP data. Therefore, motion censoring at FD < 0.2 mm (filtered) is an excellent way to avoid making false positive inferences about trait-FC correlations.

Even more stringent censoring at FD < 0.1 (filtered) does not further reduce false positive motion impact score. However, censoring at FD < 0.1 does improve false negative motion impact score (FIgure 5, Tables S2, S3), potentially offsetting concerns about loss of power from excluding data and participants. The potential for sampling bias [39] and loss of power from motion censoring might be further offset by oversampling from high-motion groups (e.g. patients with psychiatric disorders) to increase the amount of usable data per high-motion participant and number of high-motion participants with usable data after censoring. For example, a study could recruit more participants from the high-motion group, scan them for longer, and employ real-time adaptive quality assurance strategies such as FIRMM [5] to ensure collection of sufficient data to maintain the desired power after frame censoring.

Motion Reduces Statistical Power

Statistical power is the ability to reject the null hypothesis when it is actually false, or the ability to avoid false negative inference. We show that motion predisposes to false negative inference about trait-FC correlations for many traits, even after motion censoring. Therefore, motion reduces statistical power for specific traits in a way that is not accounted for by conventional power analysis methods. While researchers may be more attentive to false positive results, false negative results and statistical power are also important. For example, failure to detect an association between lead exposure and neurotoxicity would contribute to incorrect policy decisions about lead mitigation, increased lead exposure, and adverse public health outcomes. False negative results are especially damaging when they arise in large brain wide associations studies (BWAS, e.g., ABCD), which are assumed to be adequately powered.

SHAMAN Facilitates Region of Interest (ROI) Analyses

It is common to select a priori regions or nodes of interest on which to perform FC analyses. ROI analyses may also be beneficial for excluding spatially-structured motion artifacts. Using non-parametric combining, SHAMAN can assign a motion impact score to a single node or a subset of nodes, thereby providing insight into whether a priori regions of interest are impacted by residual motion artifact, and facilitating analysis of a few nodes which do not have a significant motion impact score even when the whole brain does have a significant motion impact score. While ROI analysis is typically used to limit the number of statistical tests or comparisons, as the brain can be subdivided into thousands of combinations of ROIs, the number of possible statistical tests/comparisons is vast. Therefore, caution is advised when interpreting “cherry picked” regions of interest with a low motion impact score, especially for traits where motion impact score is widely distributed over a large proportion of nodes (e.g., BMI: 39% of nodes).

SHAMAN Can Quantify Statistical Significance of Motion Distortion

The problem of estimating motion artifact is intractable because the precise relationship between head motion and fMRI signal is not known[26]. Denoising algorithms must therefore make assumptions about the motion-fMRI relationship. For example, regressing the motion parameters out of the fMRI signal[29] makes a strong, linear assumption about the motion-fMRI relationship. Detecting residual motion after denoising is especially challenging because the same assumptions used by the denoising algorithm cannot be reused. For example, regressing the motion parameters out of the fMRI signal a second time will not remove additional motion artifact. SHAMAN approaches this problem by making the weak but reasonable assumption that the high-motion half of the data will have the same correlation structure as the low-motion half except for artifact from motion. Therefore, while quantifying residual motion precisely is impossible, SHAMAN detects when the impact of motion on a trait-FC relationship is significant. In theory the residual impact of motion might be significantly non-zero but small enough that it would not cause false inference. However, consistent with prior work, [45] we show that the motion-FC effect is similar in magnitude to the effect size of most trait-FC correlations. Therefore, a significant motion impact score is likely to cause false inference.

SHAMAN Predicts Residual Motion Impact Better Than Trait-motion Correlation Alone

This study was motivated by the hypothesis that trait-motion correlations for traits (e.g. BMI) correlated with head motion would be more likely to be spuriously related to motion. Are strong trait-FD correlations sufficient to predict spurious trait-FC correlations? Traits highly correlated with motion were more likely to have their trait-FC relationships obscured by motion, limiting statistical power and creating a bias toward false negative results. However, strong trait-FD correlations did not predict which trait-FC relationships would be spuriously inflated by motion, biasing toward false positive results. Therefore, SHAMAN provides vital information about the risk of false-positive inference due to residual trait-specific motion impact beyond the trait-motion correlation alone.

Detecting the Impact of Residual Trait-Specific Motion Requires Thousands of Participants

Marek et al. [45] showed that thousands of participants are needed to measure trait-FC correlations reproducibly. Similarly, we find that 2,000-3,000 participants are needed to detect the false positive impact of motion on BMI-FC correlations, and 6,000 participants are needed to detect the false negative impact of motion on matrix reasoning-FC correlations (Figure S6). We postulate that false negative motion impact requires larger sample sizes for detection because affected trait-FC correlations would have smaller effect sizes. Our findings therefore reinforce the benefit of large sample sizes for BWAS. When large samples are not feasible (e.g. in studies of rare diseases), extrapolation of our findings suggests framewise motion censoring is a prudent strategy to avoid false positive inferences. Researchers should also exercise caution when interpreting trait-FC correlations for biophysical traits (e.g. BMI) because their relationship to SNR makes them particularly susceptible to spurious inflation by motion artifact.

Quantifying Motion Impact Helps To Avoid False Inference

Many Traits of interest to human population neuroscience are significantly correlated with head motion (87%). Motion denoising algorithms successfully remove much of the effect of motion, but trait-specific motion effects remain likely to distort inferences drawn from functional connectivity. The risk of false positives is highest for biophysical traits (e.g., BMI) and false negatives are most likely for cognitive/behavioral traits (e.g. matrix reasoning). Frame censoring is an effective strategy to reduce trait-specific motion distortions and can increase the likelihood of finding a significant, reproducible effect. Spatial masks generated by our novel SHAMAN method can be used to avoid the most motion-impacted regions during ROI selection. Given that residual head motion is most likely to falsely suppress true associations between FC and cognitive traits, more aggressive motion suppression and denoising techniques should lead to the discovery of new cognition-FC relationships.

Methods

Standard Methods

Behavioral

The Adolescent Brain Cognitive Development (ABCD) study participants are well-phenotyped with demographic, physical, mental health [43], and cognitive [44] batteries. For the purpose of this report we selected 45 broadly-interesting traits for which complete data was available on a majority of participants: NIH Toolbox, Wechsler Intelligence Scale for Children (WISC-V)[46], Cognitive Behavioral Checklist (CBCL), Prodromal Psychosis Scale (PPS), Behavioral Inhibition/Avoidance Scales (BIS/BAS), and Urgency Premeditation Perseverance Sensation Impulsive Scale (UPPS-P), see Table 1. We also selected 5 physical (BMI, age, sex, weight, and height) and 5 demographic (number of people living in household, combined parental income bracket, parents' marital status, highest parental education level, and residential deprivation index) traits commonly used as covariates, as well as 2 study-related variables (number of MRI scans completed and drowsiness), to see if any of these would be highly correlated with motion. Data were downloaded from the NIMH Data Archive (ABCD Release 2.0), and the traits of interest were extracted using the ABCDE software we have developed and which we have made available here: https://gitlab.com/DosenbachGreene/abcde.

Exemplar Traits

Two exemplar traits were selected from the ABCD study *a posteriori* to illustrate the SHAMAN method. Body mass index (BMI) is measured in units of kg/m^2 with a normative range of about 14-22 for children ages 9-10 years [47]. The Wechsler Intelligence Scale for Children, 5th edition (WISC-V) matrix reasoning subscore has a mean of 10, standard deviation of 3, and maximum value of 19 [46], with higher scores indicating superior performance on the test.

MR Imaging

Functional magnetic resonance imaging (fMRI) was acquired at 21 sites using a protocol harmonized for 3 Tesla GE, Philips, and Siemens scanners with multi-channel receive coils [21]. In addition to anatomical and task-fMRI, each participant had up to four 5-minute-long resting-state scans (TR = 800 ms, 20 minutes total). Participants with less than 8 minutes of resting-state data, the minimum duration needed for high-quality estimation of connectivity [55], were excluded from analysis. A subset of sites using Siemens scanners used FIRMM motion tracking software [5] that allows extending the scan on the basis of on-line measurement of motion.

Following acquisition, fMRI data were processed using standardized methods including correction for field distortion, frame-by-frame motion co-registration, alignment to standard stereo-tactic space, and extraction of the cortical ribbon [38]. Resting-state data were further processed to remove respiratory and motion artifact by temporal bandpass filtering, global signal regression, and regression against the rigid-body motion parameters using the ABCD-BIDS motion processing pipeline, a derivative of the Human Connectome Project (HCP) processing pipeline [37] described by Fair et al [25,36]. Processing dependencies include FSL [56] and FreeSurfer [57].

Parcellation

It is possible to compute functional connectivity between each voxel or vertex. However, this approach is burdened by a high proportion of unstructured noise and large computer memory requirements. We therefore adopted a ROI-based approach based on the 333 cortical parcels described by Gordon et al. [58] augmented by the 61 subcortical parcels described by Seitzman et al [59] for a total of 394 parcels, or nodes.

Quantifying Motion

Motion in fMRI studies is typically estimated using spatial co-registration of each fMRI volume (or frame) to a reference frame, or temporal analysis of variance [14]. In this study we quantify motion using framewise displacement, FD (L1-norm), in mm after filtering for respiratory artifact [32,33] because it is the default metric used in the ABCD-BIDS processing pipeline [25,36].

Assessing the Performance of ABCD-BIDS

The primary aim of this study was to quantify *trait-specific* residual motion, but we also quantified trait-agnostic residual motion for comparison to our trait-specific findings and to assess the overall performance of the ABCD-BIDS motion processing tool (Figure S1). Multiple methods have been developed to quantify trait-agnostic residual motion after the fMRI data have been denoised or “processsed.” [39] We employed a straightforward approach of measuring the proportion of between-participant variation in fMRI signal variance explained by motion. FD was averaged over each participant’s resting-state fMRI scans. The fMRI signal variance within each parcel was computed for each participant before and after ABCD-BIDS (without frame censoring). The fMRI signal variance was averaged across parcels to generate a single measure of fMRI signal variance for each participant. Finally, the correlation (r) between FD and fMRI signal variance across participants was computed before and after ABCD-BIDS. The relative reduction in the proportion of signal variance related to motion achieved by ABCD-BIDS was computed as $1 - (r_{after}/r_{before})^2$.

Correlation of Demographic & Behavioral Traits with Motion

We compare continuous traits of interest with motion (FD, in mm) using the linear product-moment correlation coefficient, r. This quantity reflects the degree to which in-scanner head motion is related to a trait of interest and is reported in Figure S2.

Frame Censoring

Frame censoring excludes spurious variance at the cost of data loss [60]. Therefore, determining the optimal frame censoring threshold is an empirical question. We considered the effect of no motion censoring (i.e. all frames included in analysis), censoring at a stringent FD cutoff of 0.2 mm, and censoring at a very stringent FD cutoff of 0.1 mm. Participants with less than 8 minutes of resting-state data were excluded from further analysis at each censoring threshold, including the no motion censoring threshold. This resulted in a different number of participants at each censoring threshold (Figure S3). The 8 minute cutoff was selected due the difficulty of estimating stable values for functional connectivity with shorter scan duration [55].

Functional Connectivity

We begin with the standard approaches for computing and making inferences about functional connectivity, which are familiar to all fMRI researchers. To avoid ambiguity, the methods are briefly summarized here, and diagrammed in supplemental Figure S7 for reference. By convention, each brain region or parcel is referred to as a “node” and the functional connections between nodes, which are referred to as “edges,” are computed as the pairwise linear correlation coefficients between nodes. A vector of edges is typically visualized in an edge, correlation, or “connectivity” matrix. As correlations are constrained to vary from -1 to 1, the correlation coefficients are atanh (FIsher Z) transformed prior to regression. Inferences about trait-FC effects are made by modeling the atanh-transformed edges as a linear combination of some participant-wise trait of interest (e.g. matrix reasoning score) and some covariates (e.g. FD) using least-squares regression. The estimate of the trait-FC effect at each edge is divided by the standard error of the estimate to obtain a t-value at each edge.

Methods for SHAMAN

Rationale

We begin by assuming ***traits*** and their effect on resting-state functional connectivity are stable over time whereas ***states*** vary with time [61]. Here a trait is a variable such as height, intelligence score, or favorite color, which is stable over the timescale of an fMRI scan. Extrinsic variables like parental income are also considered traits for this purpose. State is a variable such as instantaneous respiratory rate or head motion that varies during an fMRI scan. Although it has been observed that the propensity to move is a trait which is stable over months to years [52], here we consider second-to-second variations in the FD timeseries.

It follows from this assumption that if we split an individual participant’s fMRI BOLD timeseries in half, we will draw the same individual- and group-level inferences on a trait from each half of the data (to within sampling error). If trait-specific connectivity is not associated with head motion then split-half inference will be invariant to the amount of motion in each half of the data. Thus, we can detect trait-specific motion-associated connectivity by comparing inferences when the data is split according to the FD timeseries vs split at random.

Algorithm Steps

The SHAMAN (**S**plit-**H**alf **A**nalysis of **M**otion-**A**ssociated **N**etworks) algorithm is diagrammed in Figure 2. Its steps are enumerated below.

1. An individual participant’s resting-state BOLD timeseries from one or more fMRI scans are concatenated (Figure 2a).
2. The individual participant’s fMRI BOLD timeseries is split in half (Figure 2b). One half (left) contains the half of the timeseries with the lowest FD (motion) values, and the other half (right) contains the half with the highest FD values.

3. A connectivity/edge matrix is computed separately for each half of the individual participant's data (Figure 2b). See Figure S8 for an illustration using real connectivity data.
4. Average (over time) FD is computed for each half of the data (Figure 2C).
5. Steps 1-4 are repeated for each participant in the study.
6. The between-participant effect of motion is accounted for by regressing (across participants) mean FD out of each half of the connectivity data (Figure 2C).
7. The difference in residuals between low- and high-motion halves from Step 6 is computed for each participant (Figure 2D).
8. The trait/variable of interest is regressed against the difference in residuals from Step 7 (Figure 2D), optionally in the presence of covariates.
9. The resultant connectivity matrix is the motion impact for that trait (Figure 2E).

If motion processing works perfectly, or if participants do not move at all, then each half of the data will yield identical connectivity matrices, and the difference in residuals in Step 7 above (Figure 2D) will be zero to within sampling error. Subsequently, the result of regression in Steps 8 and 9 will also be zero (Figure 2E).

If spurious effects of motion remain in the data then the difference in residuals in Step 7 (Figure 2D) will not be all zeros. If the trait/variable of interest is associated with head motion then the regression in Steps 8 and 9 will reveal the trait-specific motion impact (Figure 2E).

Statistical Inference

The SHAMAN algorithm reveals the trait-specific impact of motion. This connectivity matrix will be zero under the null hypothesis of zero motion impact. To detect significant motion impact, we test the omnibus alternative hypothesis that one or more edges in this matrix is non-zero. To avoid making strong parametric assumptions about the data, we use a permutation scheme followed by the non-parametric combining method described by Winkler et al [40].

The permutation scheme is diagrammed in Figure S9. In Step 2 of the algorithm (Figure S9b) the data are split in half at random, without respect to the FD timeseries. The rest of the algorithm proceeds as before. The algorithm is repeated many times with different random splits (permutations) of the data to obtain a null distribution for the motion-associated connectivity matrix (Figure S9e).

Once a null distribution is generated, we use non-parametric combining [40] to test the null hypothesis that all edges in the motion-associated connectivity matrix are zero. Each edge is treated as a separate "modality" to avoid assumptions about exchangeability or clustering of edges. We selected Stouffer's combining function [49,50] because Stouffer's Z-score is well-known and has an intuitive interpretation. The non-parametric combining process yields a single, omnibus Stouffer's Z-score for the entire connectivity matrix, which we call the motion impact score. Comparing the motion impact score to its null distribution yields an omnibus p-value that controls for family-wise error rate across the entire motion-associated connectivity matrix. When the p-value is *small* a significant amount of motion-associated connectivity is present. When the motion impact score (Stouffer's Z-score) is *large*, a large amount of motion associated connectivity is present.

The SHAMAN p-value Relates to Motion, Not Broader Significance

The p-value generated by the SHAMAN algorithm tests the omnibus null hypothesis that no edge (pairwise connection) in a variable/trait's connectivity matrix is significantly obscured by or inflated by head motion. It does *not* test the hypothesis that the variable itself has a significant relationship with connectivity. On the contrary, a small p-value from SHAMAN (or a correspondingly large motion score) indicates that variable's relationship with FC may be spuriously associated with head motion.

False Positive vs False Negative Motion Impact Score

The main objective of SHAMAN is to detect how motion artifact impacts trait-FC correlations for specific traits. When motion impact has the same sign as a the trait-FC correlation (both positive or both negative) then motion inflates the size of the trait-FC correlation and would tend to cause false positive inference (concluding the trait-FC correlation is significant it is not). When motion impact has the opposite sign of the trait-FC correlation then motion diminishes the size of the trait-FC correlation and would tend to cause false negative inference (failing to detect a significant trait-FC correlation when it really is significant). We can then compute separate motion scores for motion associated connectivity predisposing to false positive or false negative inferences by counting only those edges whose motion impact is the same (for false positive) or opposite (for false negative) sign as the trait-FC correlation. Distinguishing between false positive and false negative error requires us to know with some degree of certainty whether the trait-FC correlation is positive or negative. We use a heuristic approach in which the trait-FC correlation at a given edge is positive if its t-value is > 2 and negative if its t-value is < -2. If the t-value is near zero then we ignore the motion impact at that edge.

Node Analysis

We can quantify motion impact within individual nodes/parcels of the brain network. This may be advantageous when planning a region of interest analysis or to help visualize motion-associated connectivity in anatomical space. Instead of performing omnibus non-parametric combining over the entire connectivity matrix, we perform non-parametric combining to obtain a Stouffer's Z-score and p-value over the edges connected to a single node or subset of nodes.

The process for node analysis is best illustrated In Figure S4. First, an omnibus p-value is computed across all edges. Then a Stouffer's Z-score (motion impact score) is computed for each individual node using the edges connected to that node. The node with the largest motion score is excluded by deleting its connecting edges from the connectivity matrix. The whole process is then repeated until all nodes have been excluded. The thick red/black line in Figure S4 shows the omnibus p-value increasing (becoming less significant) from left to right as each successive node is excluded. Portions of the line below $p = 0.05$ are highlighted in red. The point at which the line intersects with $p = 0.05$ corresponds to the number of nodes affected by motion-associated connectivity.

Node analysis is also used to visualize motion-associated connectivity anatomically in Figure 4c,d. The Stouffer's Z-score, or "motion Z-score," is computed for each individual node as above. The resultant motion scores are visualized on the cortical surface.

Acknowledgements

ABCD Acknowledgement

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9-10 and follow them over 10 years into early adulthood. The ABCD Study is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, U24DA041147, U01DA041093, and U01DA041025. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/scientists/workgroups/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators.

The ABCD data repository grows and changes over time. The ABCD data used in this report came from Anual Release 2.0, DOI 10.15154/1503209.

Grant Support

This work was supported by NIH grants MH100019 (SM), MH121518 (SM), MH123091 (AZ), NS115672 (AZ), NS110332 (DJN), MH100019 (NAS), MH129493 (DMB), NS123345 (BPK), NS098482 (BPK), NS098577 (AZS), DA041148 (DAF), DA04112 (DAF), MH115357 (DAF), MH096773 (DAF and NUFD), MH122066 (EMG, DAF, and NUFD), MH121276 (EMG, DAF, and NUFD), MH124567 (EMG, DAF, and NUFD), NS129521 (EMG, DAF, and NUFD), and NS088590 (NUFD); by the National Spasmodic Dysphonia Association (EMG); by Mallinckrodt Institute of Radiology pilot funding (EMG); by the McDonnell Center for Systems Neuroscience (AZ); by the Andrew Mellon Predoctoral Fellowship from the Dietrich School of Arts & Sciences, University of Pittsburgh (BTC); and by the Extreme Science and Engineering Discovery Environment (XSEDE) Bridges at the Pittsburgh Supercomputing Center through allocation TG-IBN200009 (BTC).

Declaration of Interest

DAF and NUFD have a financial interest in NOUS Imaging Inc. and may financially benefit if the company is successful in marketing FIRMM motion-monitoring software products. DAF and NUFD may receive royalty income based on FIRMM technology developed at Washington University School of Medicine and Oregon Health and Sciences University and licensed to NOUS Imaging Inc. DAF and NUFD are co-founders of NOUS Imaging Inc. These potential conflicts of interest have been reviewed and are managed by Washington University School of Medicine, Oregon Health and Sciences University and the University of Minnesota. The other authors declare no competing interests.

Supplemental Tables

Variable	Correlation w/FD	SHAMAN p-Value			Shaman Motion Score		
		Type I + II	Type I	Type II	Type I + II	Type I	Type II
bmi	0.09	0	0	0.038	118.80	91.19	16.56
weight	0.054	0	0	0.001	88.04	57.69	28.33
age_months	-0.121	0	0	1	32.47	57.16	-30.13
male	0.102	0	0	0.016	93.85	43.19	24.07
pps_yes_num	0.072	0	0	0.963	23.89	37.63	-15.45
cbcl_withdep	0.059	0	0	0.382	29.13	37.29	2.34
nihtbx_reading	-0.057	0	0	0.057	27.87	32.33	10.34
height	-0.047	0.001	0	0.779	22.32	30.89	-3.82
cbcl_social	0.097	0	0	0.046	27.98	30.78	14.00
cbcl_total	0.089	0.004	0	0.884	17.28	30.20	-6.89
upps_neg_urgency	0.073	0.001	0	0.685	17.70	26.97	-2.61
nihtbx_pattern	-0.089	0	0	0.24	19.99	24.15	5.21
nihtbx_picture	-0.073	0.006	0.002	0.736	14.92	23.26	-2.98
sleepy	0.022	0.007	0.002	0.295	16.77	22.87	3.61
mri_runs	-0.095	0	0.014	0.049	30.19	21.90	16.03
nihtbx_total	-0.11	0.001	0.006	0.416	20.55	21.64	2.38
pps_severity	0.068	0.021	0.021	0.863	12.38	20.68	-9.44
cbcl_thought	0.063	0.036	0.003	0.62	11.30	20.49	-1.57
home_roster	-0.032	0.127	0.006	0.955	6.80	17.86	-9.17
upps_perserverance	0.061	0.07	0.011	0.833	8.66	17.20	-4.86
cbcl_attention	0.116	0.008	0.041	0.345	14.75	16.74	3.90
nihtbx_cryst	-0.062	0	0.017	0.021	21.29	15.49	16.08
nihtbx_fluid	-0.111	0.009	0.041	0.228	16.42	13.88	5.71
cbcl_aggressive	0.068	0.007	0.031	0.038	16.74	13.81	12.86
cbcl_external	0.088	0.012	0.026	0.082	14.81	13.39	9.58
upps_planning	0.077	0.198	0.027	0.873	4.90	12.35	-5.44
cbcl_anxdep	0.01	0.099	0.037	0.34	7.84	11.02	2.77
cbcl_rulebreak	0.089	0	0.12	0	36.20	9.67	41.74
cbcl_internal	0.021	0.228	0.059	0.728	4.73	8.75	-2.40
bis	0.007	0.299	0.086	0.739	3.19	7.83	-2.49
bas_fs	0.068	0.004	0.142	0	16.27	7.16	20.39
upps_seeking	0.005	0.142	0.14	0.316	6.22	6.19	2.91
nihtbx_flanker	-0.045	0.049	0.196	0.052	10.17	4.64	9.89
upps_pos_urgency	0.09	0.048	0.44	0.167	9.93	1.55	9.23
cbcl_somatic	0	0.807	0.549	0.945	-3.82	-0.08	-6.90
nihtbx_cardsort	-0.085	0.29	0.554	0.199	3.69	-0.68	5.99
nihtbx_picvocab	-0.051	0.072	0.751	0.033	8.85	-3.35	14.35
nihtbx_list	-0.053	0.709	0.838	0.62	-2.07	-4.16	-1.03
bas_rr	0.055	0.57	0.833	0.241	-0.35	-4.28	4.26
subject_id	0	0.999	0.966	1	-11.38	-7.33	-15.70
residential_deprivation	0.08	0	0.802	0	78.02	-7.74	70.41
wiscv_matrix	-0.124	0	0.813	0	45.81	-8.02	54.53
bas_drive	0.095	0.002	0.915	0.001	21.71	-10.28	36.77
parents_married	-0.071	0	0.995	0	81.33	-28.21	84.70
parental_education	-0.096	0	1	0	82.26	-38.63	94.45
parental_income	-0.114	0	1	0	108.50	-51.30	119.35

Table S1: Relationship between selected variables in the ABCD study with head motion (framewise displacement, FD) averaged over each participant's resting state fMRI scans. The SHAMAN (omnibus) p-values and motion scores (omnibus Stouffer's Z-scores) are also given for false positive/Type I Error, false negative/Type II Error, and both/either error type. Results in this table reflect data after motion-reduction with ABCD-BIDS and with*out* additional motion censoring. Variables are sorted by Type I motion score. See Figure S2.

Variable	SHAMAN p-Value								
	No Censoring			FD < 0.2 mm			FD < 0.1 mm		
	Type I + II	Type I	Type II	Type I + II	Type I	Type II	Type I + II	Type I	Type II
male	0	0	0.016	0	0.679	0	0.035	0.531	0.077
age_months	0	0	1	0.096	0.487	0.067	0.338	0.363	0.314
height	0.001	0	0.779	0.001	0	0.678	0.037	0.014	0.353
weight	0	0	0.001	0	0.259	0	0.019	0.238	0.091
bmi	0	0	0.038	0	0.736	0	0.145	0.682	0.137
nihtbx_pattern	0	0	0.24	0.213	0.947	0.019	0.126	0.317	0.038
nihtbx_reading	0	0	0.057	0.017	0.207	0.001	0.603	0.94	0.298
cbcl_total	0.004	0	0.884	0.106	0.233	0.054	0.322	0.668	0.095
cbcl_withdep	0	0	0.382	0.224	0.252	0.321	0.135	0.221	0.122
cbcl_social	0	0	0.046	0.036	0.301	0.027	0.532	0.846	0.251
pps_yes_num	0	0	0.963	0.028	0.117	0.55	0	0	0.415
upps_neg_urgency	0.001	0	0.685	0.034	0.089	0.057	0.566	0.671	0.5
sleepy	0.007	0.002	0.295	0.234	0.536	0.08	0.604	0.679	0.367
nihtbx_picture	0.006	0.002	0.736	0.442	0.857	0.195	0.808	0.977	0.523
cbcl_thought	0.036	0.003	0.62	0.216	0.646	0.064	0.226	0.594	0.07
home_roster	0.127	0.006	0.955	0.847	0.582	0.919	0.411	0.132	0.818
nihtbx_total	0.001	0.006	0.416	0.228	0.994	0.004	0.853	0.997	0.306
upps_perserverance	0.07	0.011	0.833	0.289	0.311	0.282	0.913	0.957	0.799
mri_runs	0	0.014	0.049	0.05	0.165	0.187	0.054	0.545	0.009
nihtbx_cryst	0	0.017	0.021	0	0.233	0	0.165	0.76	0.029
pps_severity	0.021	0.021	0.863	0.492	0.581	0.491	0.005	0.007	0.218
cbcl_external	0.012	0.026	0.082	0.045	0.437	0.008	0.288	0.864	0.078
upps_planning	0.198	0.027	0.873	0.072	0.058	0.1	0.337	0.413	0.214
cbcl_aggressive	0.007	0.031	0.038	0.03	0.523	0.005	0.076	0.737	0.015
cbcl_anxdep	0.099	0.037	0.34	0.097	0.093	0.153	0.312	0.313	0.309
nihtbx_fluid	0.009	0.041	0.228	0.093	0.997	0.002	0.606	0.983	0.224
cbcl_attention	0.008	0.041	0.345	0.326	0.973	0.037	0.401	0.937	0.056
cbcl_internal	0.228	0.059	0.728	0.489	0.463	0.621	0.474	0.478	0.436
bis	0.299	0.086	0.739	0.872	0.738	0.881	0.438	0.455	0.372
cbcl_rulebreak	0	0.12	0	0	0.235	0	0.146	0.786	0.028
upps_seeking	0.142	0.14	0.316	0.383	0.601	0.294	0.805	0.952	0.586
bas_fs	0.004	0.142	0	0.2	0.525	0.063	0.215	0.366	0.137
nihtbx_flanker	0.049	0.196	0.052	0.071	0.657	0.007	0.785	0.917	0.545
upps_pos_urgency	0.048	0.44	0.167	0.147	0.782	0.116	0.734	0.836	0.641
cbcl_somatic	0.807	0.549	0.945	0.686	0.582	0.756	0.517	0.424	0.535
nihtbx_cardsort	0.29	0.554	0.199	0.063	0.911	0.006	0.464	0.904	0.114
nihtbx_picvocab	0.072	0.751	0.033	0.006	0.429	0	0.252	0.714	0.07
residential_deprivation	0	0.802	0	0	1	0	0	0.995	0
wiscv_matrix	0	0.813	0	0	0.999	0	0.235	0.971	0.034
bas_rr	0.57	0.833	0.241	0.331	0.929	0.077	0.16	0.567	0.044
nihtbx_list	0.709	0.838	0.62	0.623	0.959	0.324	0.885	0.994	0.596
bas_drive	0.002	0.915	0.001	0.052	0.815	0.017	0.202	0.405	0.234
subject_id	0.999	0.966	1	0.521	0.39	0.7	0.505	0.563	0.507
parents_married	0	0.995	0	0	0.999	0	0	0.418	0.031
parental_income	0	1	0	0	1	0	0	0.958	0
parental_education	0	1	0	0	1	0	0	0.675	0.003

Table S2: Summary of motion-associated connectivity for selected variables in the ABCD study after motion-reduction with ABCD-BIDS and with (varying from left to right) no motion censoring, censoring at framewise displacement (FD) < 0.2 mm, and FD < 0.1. Variables are sorted by (omnibus) p-value without motion censoring. P-values corresponding to false positive/Type I Error and false negative/Type II Error are shown separately. See also Table S3.

Variable	SHAMAN Motion Score								
	No Censoring			FD < 0.2 mm			FD < 0.1 mm		
	Type I + II	Type I	Type II	Type I + II	Type I	Type II	Type I + II	Type I	Type II
bmi	118.8	91.2	16.6	56.7	-4.8	56.6	6.6	-3.0	9.3
weight	88.0	57.7	28.3	48.5	5.9	42.2	14.1	6.1	10.8
age_months	32.5	57.2	-30.1	8.3	0.4	11.5	3.0	2.5	3.4
male	93.9	43.2	24.1	45.2	-3.2	38.1	12.5	0.0	12.1
pps_yes_num	23.9	37.6	-15.4	11.7	11.8	-0.8	27.8	31.7	2.0
cbcl_withdep	29.1	37.3	2.3	4.9	4.4	3.2	7.1	5.2	7.7
nihtbx_reading	27.9	32.3	10.3	13.8	4.9	17.5	-1.1	-7.2	3.7
height	22.3	30.9	-3.8	17.8	24.1	-1.9	11.2	14.7	2.7
cbcl_social	28.0	30.8	14.0	11.8	3.9	14.8	-0.1	-5.2	4.8
cbcl_total	17.3	30.2	-6.9	7.8	4.4	10.3	3.1	-1.9	8.0
upps_neg_urgency	17.7	27.0	-2.6	11.7	8.9	10.3	-0.2	-1.8	0.5
nihtbx_pattern	20.0	24.1	5.2	4.8	-7.2	13.3	7.6	3.0	10.8
nihtbx_picture	14.9	23.3	-3.0	1.2	-4.7	5.5	-4.3	-9.1	0.2
sleepy	16.8	22.9	3.6	4.5	0.2	8.9	-0.8	-1.9	2.4
mri_runs	30.2	21.9	16.0	10.9	8.4	6.6	11.0	-0.4	17.1
nihtbx_total	20.6	21.6	2.4	4.5	-12.0	18.1	-5.1	-12.3	3.6
pps_severity	12.4	20.7	-9.4	0.8	-1.2	0.6	19.7	20.8	6.9
cbcl_thought	11.3	20.5	-1.6	5.1	-1.7	10.8	4.8	-0.7	9.7
home_roster	6.8	17.9	-9.2	-4.6	-0.8	-6.9	1.8	8.0	-5.1
upps_perserverance	8.7	17.2	-4.9	3.5	3.1	3.6	-6.6	-8.1	-3.9
cbcl_attention	14.7	16.7	3.9	3.2	-10.7	14.0	2.0	-8.5	11.0
nihtbx_cryst	21.3	15.5	16.1	22.3	5.0	28.3	6.5	-3.1	13.3
nihtbx_fluid	16.4	13.9	5.7	8.1	-14.6	20.7	-0.9	-9.0	4.7
cbcl_aggressive	16.7	13.8	12.9	12.1	-0.1	18.4	10.1	-3.6	17.9
cbcl_external	14.8	13.4	9.6	10.4	1.2	15.1	3.6	-5.2	8.5
upps_planning	4.9	12.4	-5.4	9.0	9.7	7.6	2.7	1.6	4.7
cbcl_anxdep	7.8	11.0	2.8	8.4	7.9	6.6	3.4	3.2	3.4
cbcl_rulebreak	36.2	9.7	41.7	25.4	5.1	31.5	7.6	-4.8	14.8
cbcl_internal	4.7	8.7	-2.4	1.1	1.2	-1.0	0.8	0.9	1.2
bis	3.2	7.8	-2.5	-5.0	-2.6	-5.0	1.3	1.1	2.0
bas_fs	16.3	7.2	20.4	5.3	-0.2	10.7	5.1	2.3	6.5
upps_seeking	6.2	6.2	2.9	1.9	-1.0	3.3	-3.9	-7.6	-0.6
nihtbx_flanker	10.2	4.6	9.9	9.4	-1.8	14.3	-3.8	-6.5	-0.3
upps_pos_urgency	9.9	1.6	9.2	6.5	-5.2	10.1	-2.9	-5.9	-2.3
cbcl_somatic	-3.8	-0.1	-6.9	-2.1	-0.5	-3.2	0.3	1.5	0.0
nihtbx_cardsort	3.7	-0.7	6.0	9.8	-6.2	17.3	1.1	-6.2	8.0
nihtbx_picvocab	8.9	-3.3	14.3	16.5	1.7	24.0	4.2	-2.7	10.2
nihtbx_list	-2.1	-4.2	-1.0	-1.1	-7.6	2.8	-5.2	-9.5	-0.9
bas_rr	-0.4	-4.3	4.3	2.8	-7.3	9.9	6.7	-0.4	10.8
subject_id	-11.4	-7.3	-15.7	0.3	2.0	-1.9	0.1	-0.5	0.2
residential_deprivation	78.0	-7.7	70.4	56.5	-37.1	73.0	35.2	-21.5	48.4
wiscv_matrix	45.8	-8.0	54.5	23.1	-21.0	40.8	4.6	-11.2	14.9
bas_drive	21.7	-10.3	36.8	9.7	-6.1	19.0	5.1	2.2	5.7
parents_married	81.3	-28.2	84.7	58.2	-26.8	62.2	30.9	3.2	22.2
parental_education	82.3	-38.6	94.4	56.0	-38.2	73.1	27.9	-3.8	30.6
parental_income	108.5	-51.3	119.3	81.9	-57.5	100.9	48.7	-16.0	53.0

Table S3: Summary of motion-associated connectivity for selected variables in the ABCD study after motion-reduction with ABCD-BIDS and with (varying from left to right) no motion censoring, censoring at framewise displacement (FD) < 0.2 mm, and FD < 0.1. Variables are sorted by motion score (omnibus Stouffer's Z-score) without censoring. Motion scores for false positive/Type I Error and false negative/Type II Error are shown separately. See also Table S2.

Variable	Percent Difference in Mean Compared to No Censoring	
	FD < 0.2 mm	FD < 0.1 mm
male	-1.98%	-4.19%
parents_married	0.77%	2.87%
age_months	0.98%	2.43%
mri_runs	1.25%	1.95%
parental_income	0.64%	1.94%
residential_deprivation	-0.44%	-1.52%
bmi	-0.34%	-1.33%
wiscv_matrix	0.55%	1.20%
cbcl_total	-0.51%	-1.10%
upps_pos_urgency	-0.37%	-1.08%
cbcl_external	-0.51%	-1.03%
weight	-0.23%	-0.97%
bas_drive	-0.63%	-0.92%
upps_neg_urgency	-0.43%	-0.92%
cbcl_attention	-0.40%	-0.88%
pps_yes_num	-0.37%	-0.86%
parental_education	0.24%	0.84%
nihtbx_pattern	0.32%	0.81%
nihtbx_fluid	0.35%	0.81%
nihtbx_total	0.33%	0.80%
upps_perserverance	-0.38%	-0.74%
sleepy	-0.19%	-0.73%
cbcl_social	-0.34%	-0.71%
upps_planning	-0.38%	-0.68%
cbcl_thought	-0.33%	-0.67%
cbcl_rulebreak	-0.37%	-0.65%
nihtbx_cardsort	0.31%	0.64%
bas_rr	-0.19%	-0.59%
bas_fs	-0.33%	-0.59%
cbcl_aggressive	-0.28%	-0.59%
nihtbx_picture	0.21%	0.55%
cbcl_withdep	-0.29%	-0.51%
pps_severity	-0.21%	-0.47%
nihtbx_reading	0.19%	0.45%
nihtbx_cryst	0.18%	0.45%
nihtbx_list	0.21%	0.44%
cbcl_internal	-0.22%	-0.37%
home_roster	0.15%	0.35%
nihtbx_picvocab	0.11%	0.29%
nihtbx_flanker	0.12%	0.29%
upps_seeking	0.04%	0.25%
bis	-0.07%	-0.14%
height	0.15%	0.13%
cbcl_anxdep	-0.13%	-0.07%
cbcl_somatic	0.03%	-0.03%

Table S4: Sampling bias at different levels of motion censoring. The percentage difference in sample mean (compared to no censoring) is shown for selected variables in ABCD after moderate censoring at a framewise displacement (FD) of 0.2 mm or stringent censoring at FD < 0.1. Variables are sorted by absolute percentage difference at FD < 0.1 mm. Two variables (gender and number of MRI runs completed) are biased by more than 1% at FD < 0.2 mm, and 11 variables are biased by > 1% at FD < 0.1 mm.

Supplemental Figures

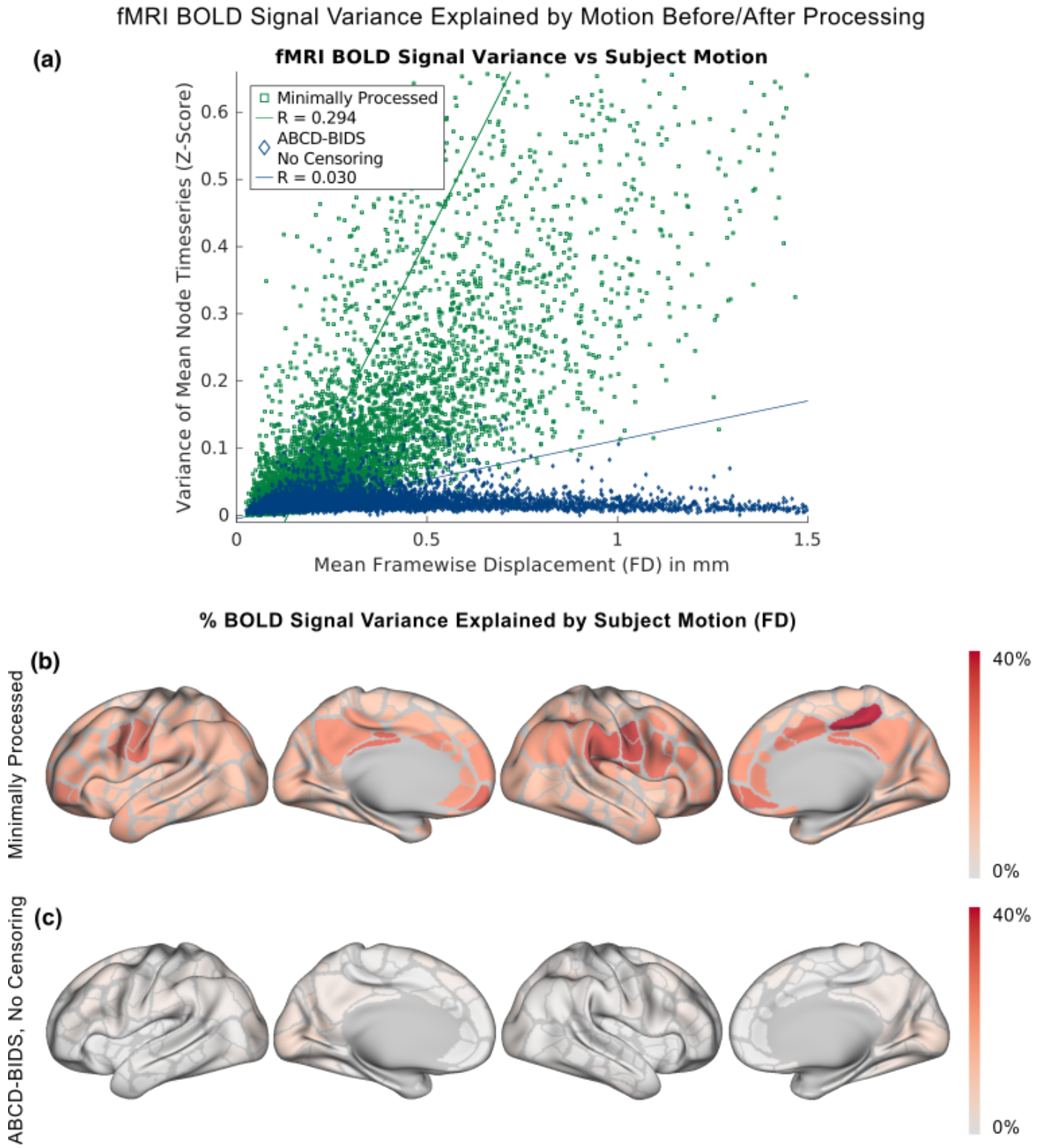

Figure S1: Proportion of fMRI BOLD signal variance explained by head motion before and after processing with the ABCD-BIDS (DCAN Labs) motion reduction algorithm. The comparison is made prior to any framewise motion censoring. **(a)** Scatterplot of variance (Z-score) of the mean node timeseries *vs.* mean framewise displacement (FD, in mm) for each child (n = 9,652). **(b)** Proportion of variance of mean node (parcel) timeseries explained by motion visualized on the cortical surface before and **(c)** after ABCD-BIDS.

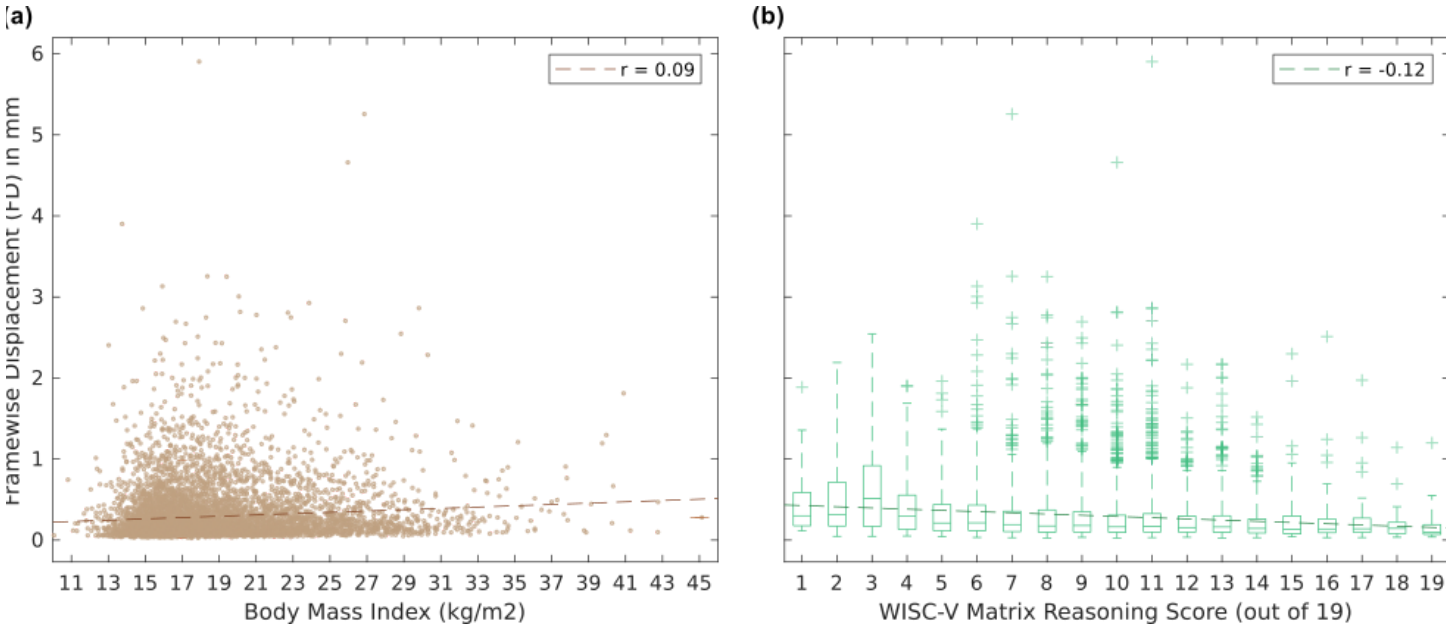

Figure S2: Correlation between trait variables and in-scanner head motion. (a) Body mass index (BMI, kg/m^2, mean 16.8, standard deviation 3.1), r = 0.09. **(b)** Wechsler Intelligence Scale for Children 5th Edition (WISC-V)[46] matrix reasoning score (mean 10, standard deviation 3, maximum 19), r = -0.12. Both correlations were significant at $p < 0.001$. Each point represents one participant in the ABCD study. Head motion was measured as framewise displacement (FD, in mm) averaged over resting-state scans. The best fit is shown as a dashed line.

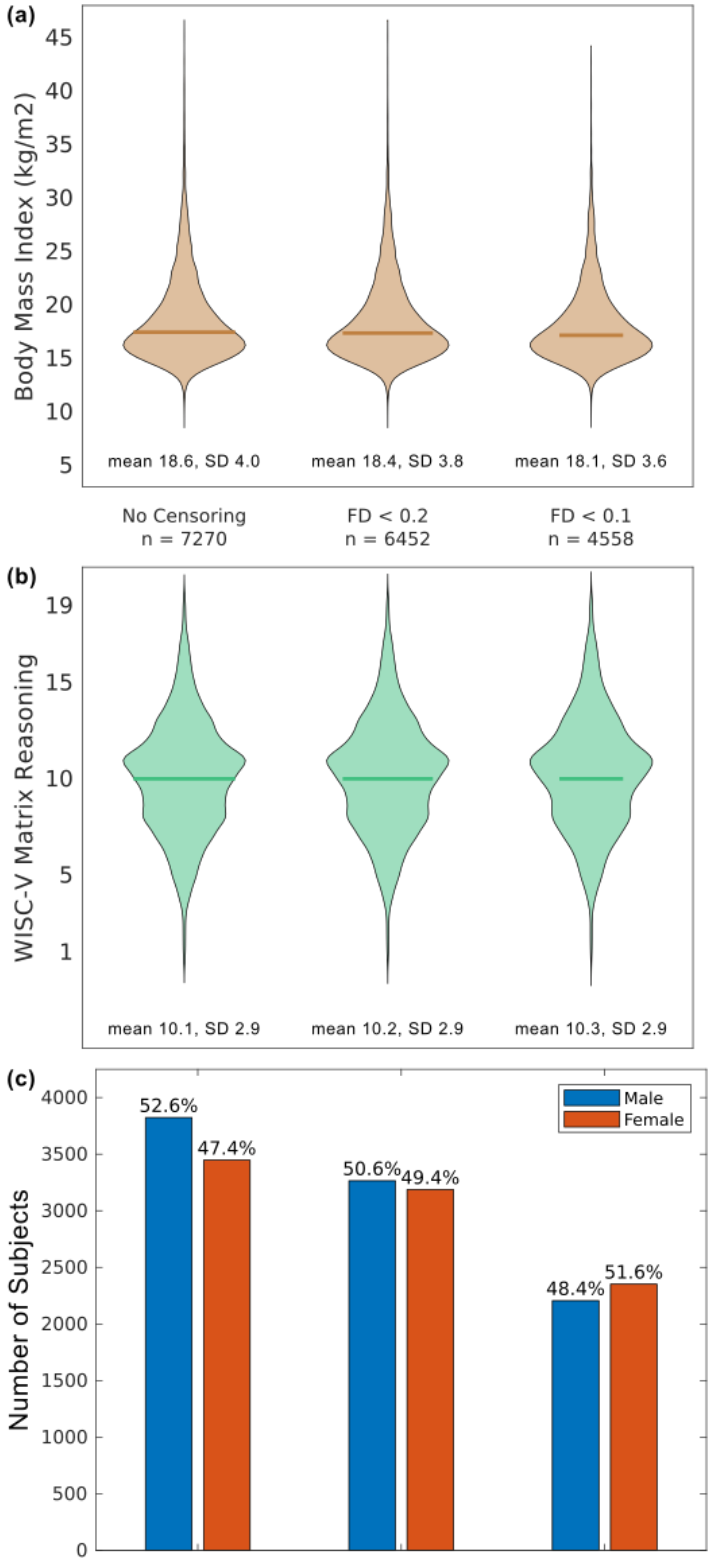

Figure S3: Sampling bias at different levels of framewise censoring. Sample distributons are shown for **(a)** body mass index (BMI), **(b)** WISC-V matrix reasoning score, and **(c)** sex assigned at birth. Mean (solid line) and standard deviation (SD) for (a) BMI and (b) matrix reasoning were similar across all levels of motion censoring, but the proportion of boys to girls shifted significantly from 52.6% without motion censoring to 51.6% with very stringent censoring at framewise displacement (FD) < 0.1 mm.

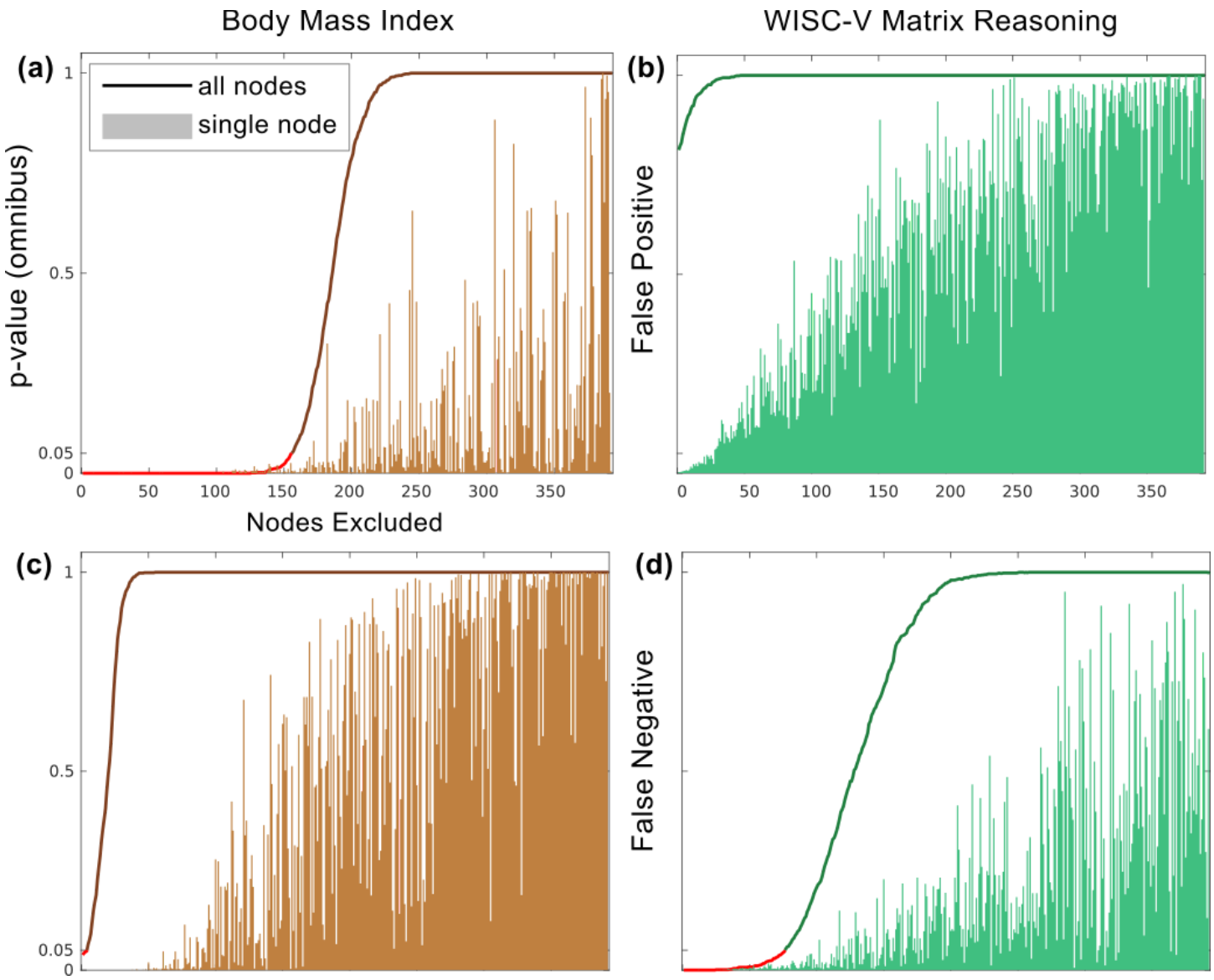

Figure S4: Number of nodes contributing to significance of motion impact. Top: Nodes driving false positive motion impact for **(a)** body mass index (BMI) and **(b)** WISC-V matrix reasoning score. **Bottom:** Nodes driving false motion impact for **(c)** BMI and **(d)** matrix reasoning. Reading each plot from left to right, the node with the highest motion score (lowest omnibus p-value) was iteratively excluded, and the omnibus p-value for edges across all remaining nodes is plotted as a solid line. The line is red for $p < 0.05$ (significant motion associated connectivity) and black for $p > 0.05$ (not significant). The omnibus p-value for each single node excluded is shown in the background.

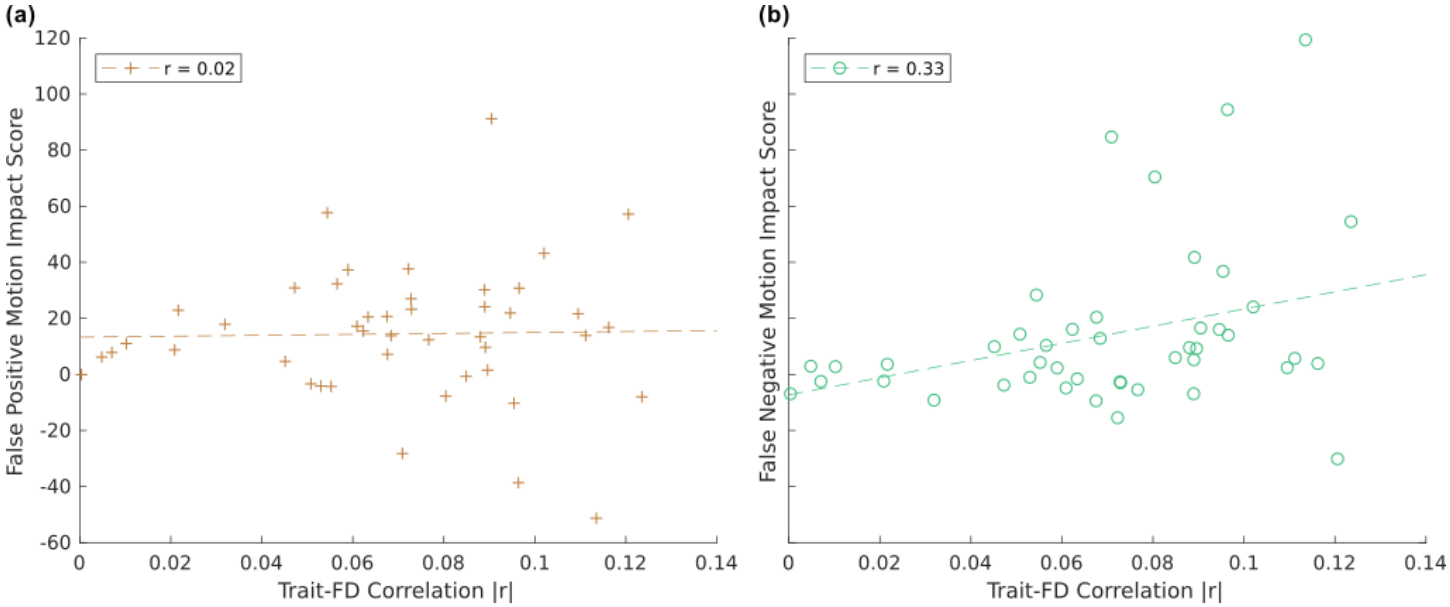

Figure S5: Correlation between SHAMAN motion impact score and in-scanner head motion. (a) False positive and **(b)** false negative motion impact scores (omnibus Stouffer's Z-score) for selected variables from the ABCD study are plotted vs the correlation (absolute value) of those variables with average participant head motion (framewise displacement, FD). Results reflect fMRI data after motion-reduction with ABCD-BIDS and without additional motion censoring. See Table S1.

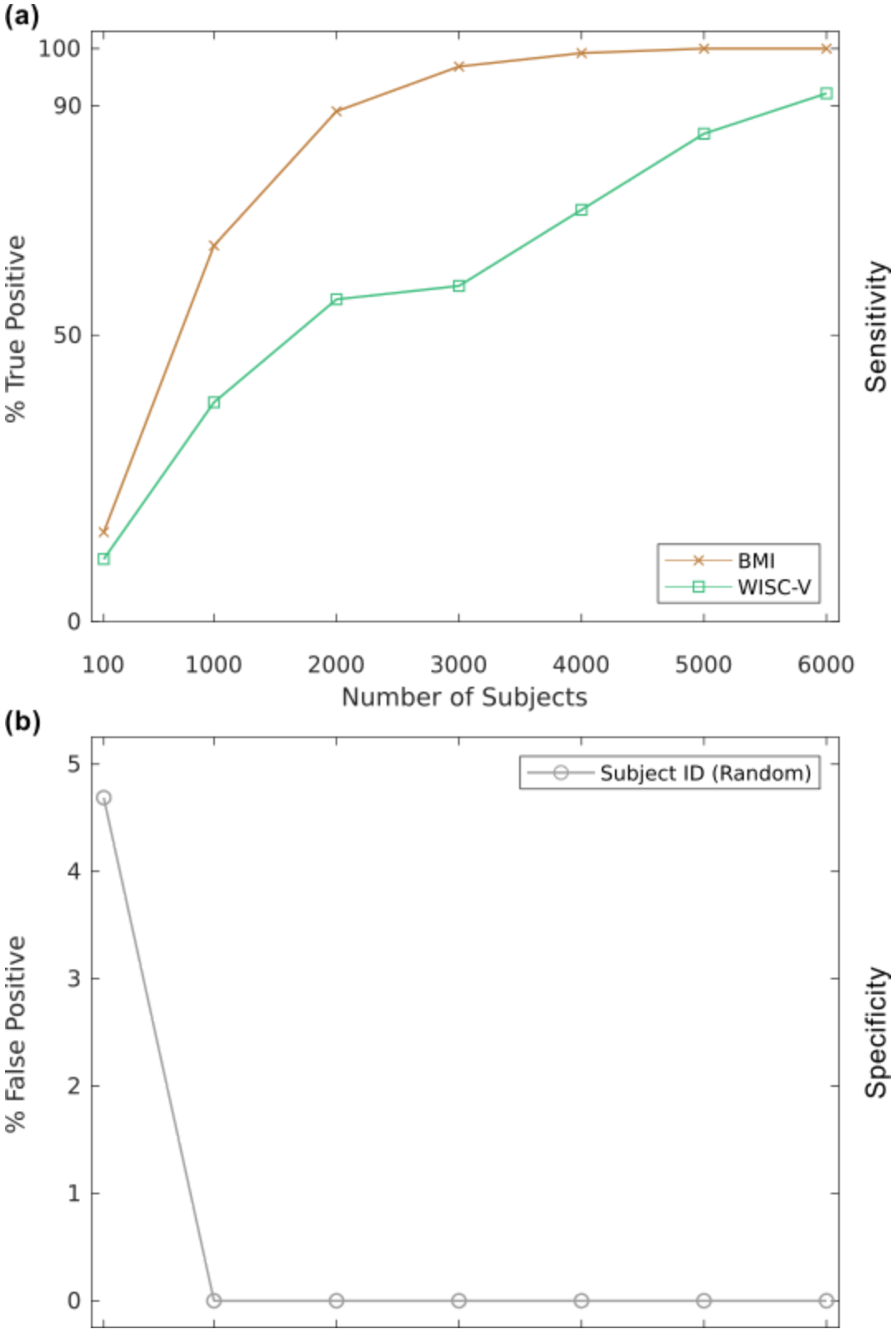

Figure S6: Statistical Power of SHAMAN Method. (a) True positive detection rate of SHAMAN computed using bootstrapped subsamples of different sizes (numbers of participants) with body mass index as an exemplar variable for significant motion associated connectivity predisposing to false positive inference (Type I Error) and WISC-V matrix reasoning score as an exemplar for false negative inference (Type II Error). **(b)** False positive detection rate of SHAMAN for either Type I or II Error using a random variable generated from the unique participant identifier.

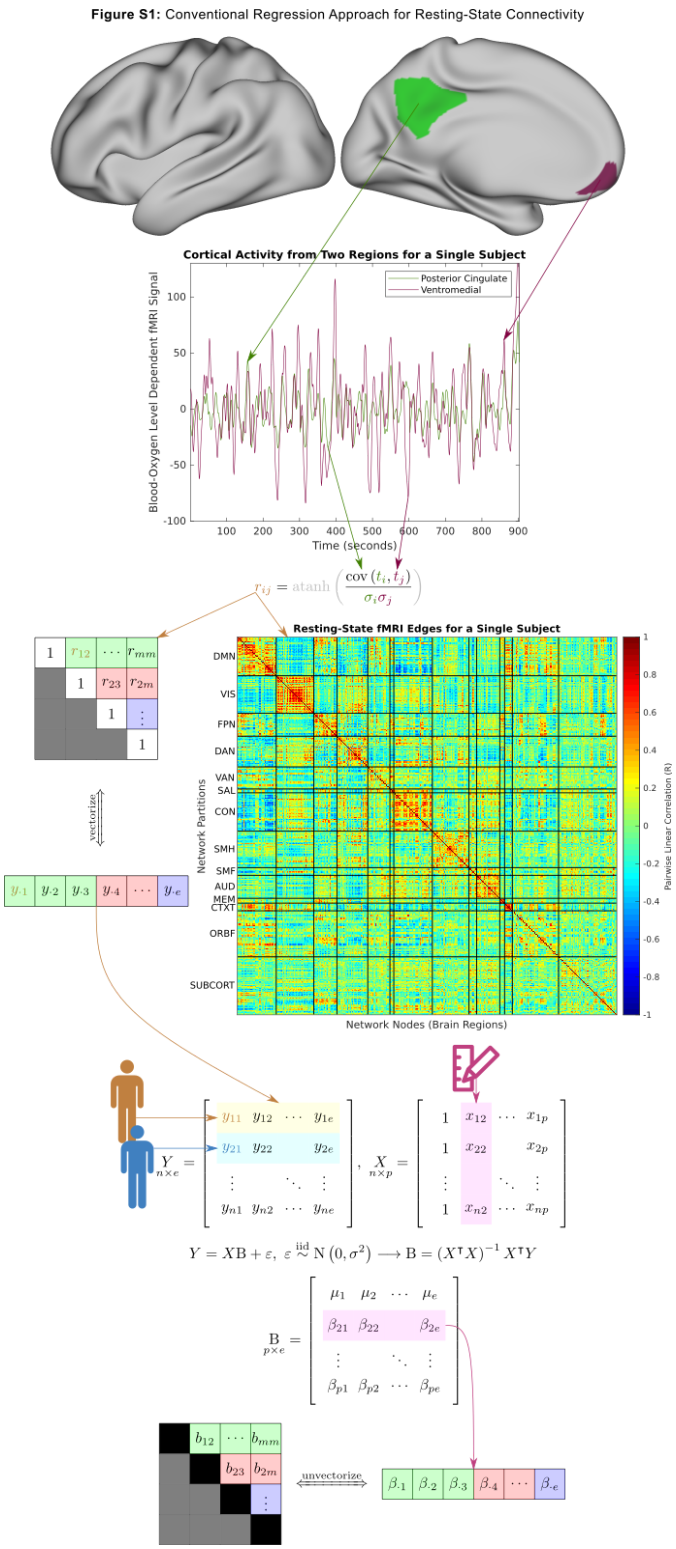

Figure S7: Conventional regression model for resting-state functional connectivity.

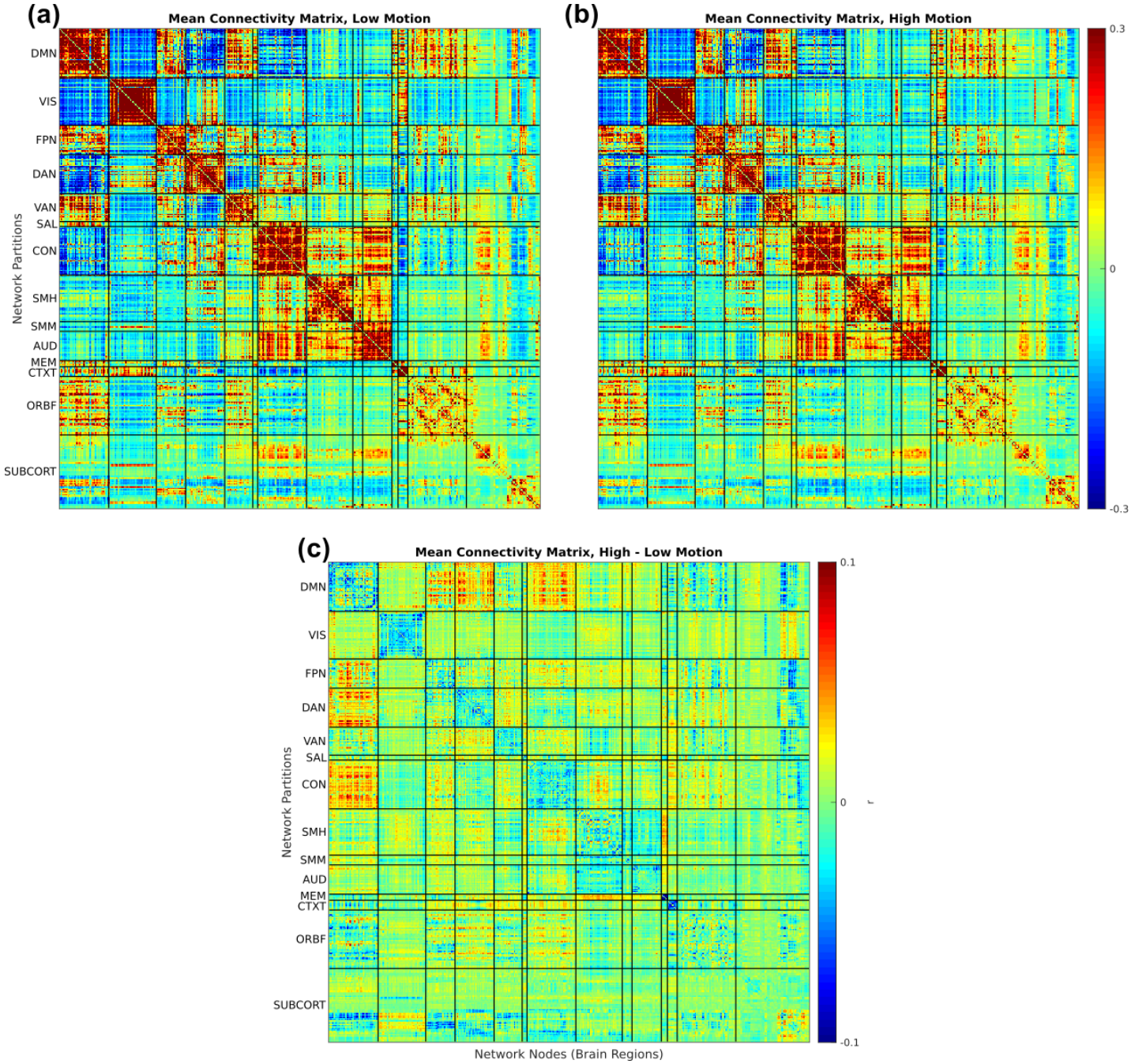

Figure S8: Mean (average) connectivity matrix computed from the low-motion **(a)** and high-motion **(b)** halves of each participant's resting-state fMRI data. The difference (rescaled to show visual detail) is shown in **(c)**. Figures are generated from data after denoising with ABCD-BIDS but with*out* motion censoring.

Figure S2: SHAMAN Algorithm (Permuted Null Model)

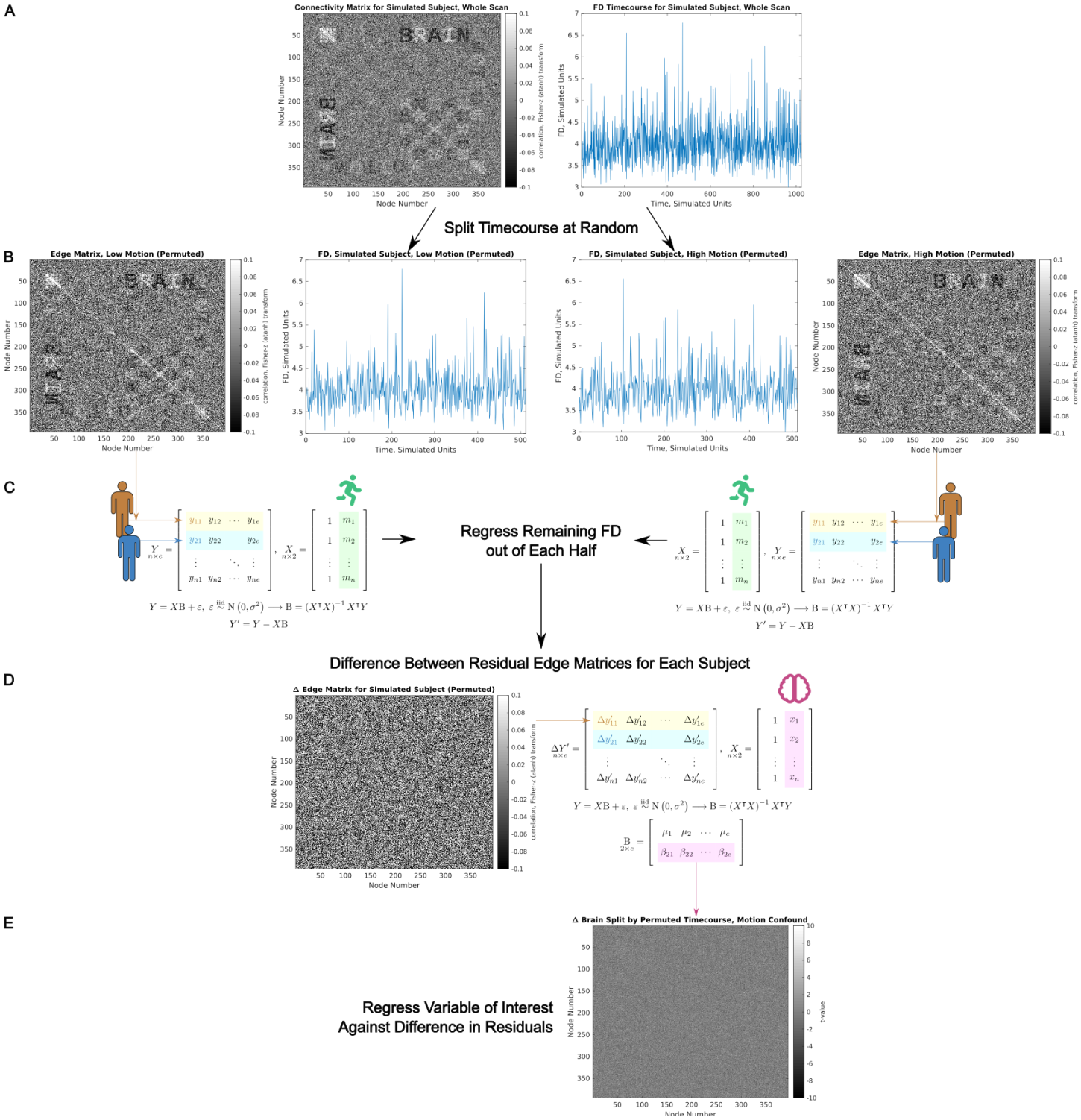

Figure S9: Null model permutation scheme for SHAMAN algorithm depicted in Figure 8. In **(B)** the data is split in half at random, without respect to the FD timeseries. The rest of the algorithm proceeds as before. The algorithm is repeated many times for each random split (permutation) to generate a null distribution for the motion-associated connectivity matrix in **(E)**.

Simulation Supplement

Generative Model

In order to test the validity of the SHAMAN algorithm we devised a generative model for fMRI BOLD timeseries data for which the parameters of brain signal and motion can be experimentally controlled. Such a model necessarily requires simplifying assumptions since the true biophysical properties of the brain are not known. Several fMRI data simulators exist [62], but none allows for realistic simulation of head motion artifact. Therefore, for our simulator we worked backwards from the assumptions of the conventional, massively univariate, resting-state functional connectivity regression model (Figure S4) in which between-participant variables of interest are predictors (the design matrix) and atanh (Fisher Z) transformed linear correlation coefficients between brain regions are the observed data.

We generated two basis correlation matrices with the same dimensions (394 x 394 nodes) as the Gordon-Laumann-Seitzman [58,59] brain parcellation. Figure SA.1a shows the basis correlation matrix for the simulated "brain" signal and Figure SA.1b shows the matrix for simulated head motion. The Cholesky decomposition of each basis correlation matrix was used to simulate the timeseries of brain and motion signals.

Figure SA.1: Basis Correlation Matrices for Simulation

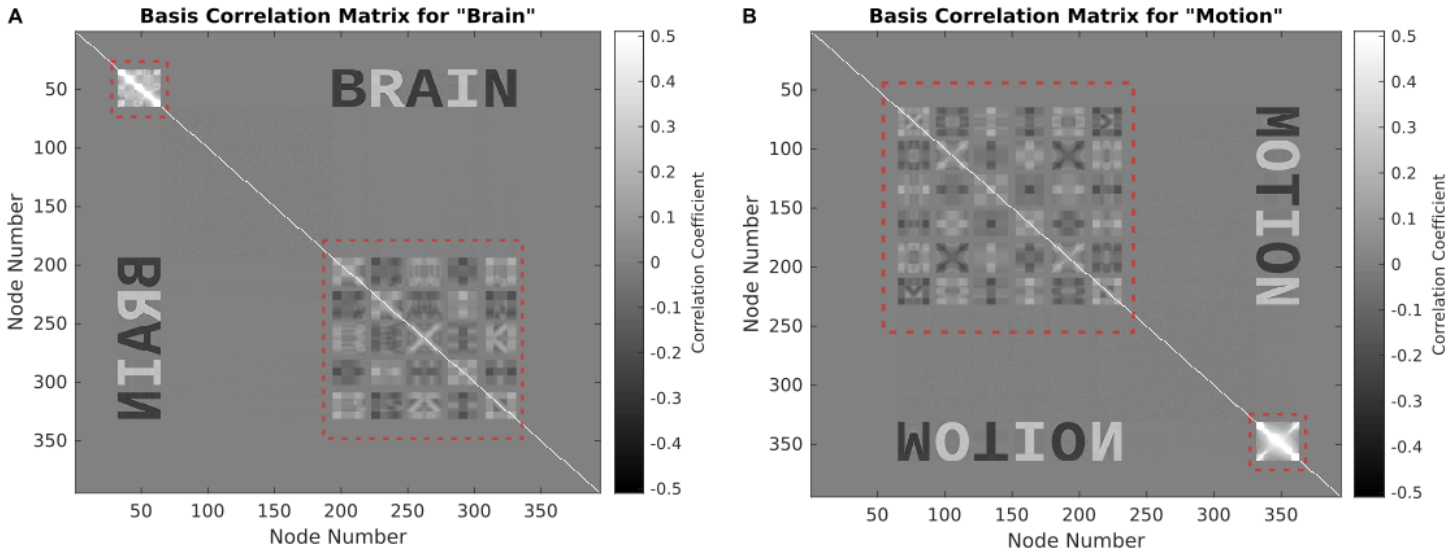

Figure SA.1: Basis correlation matrices. **(a)** "Brain" signal. **(b)** "Motion" artifact. The dimensions of each matrix are 394 x 394 nodes. The portions of the matrices outlined in the red dashed boxes were algorithmically added to make each matrix positive semi-definite.

The process for generating simulated data is diagrammed in Figure SA.2. Normally-distributed random data were multiplied by the Cholesky decomposition of the respective basis correlation matrix to generate simulated "brain" and "motion" timeseries for each simulated participant. We simulated 394 nodes x 1,024 fMRI frames (time points) for 1,024 simulated participants.

The variance, across the 394 nodes, at each of 1,024 time points in the motion timeseries was used as the FD timeseries for each participant.

The brain and motion timeseries were mixed together in controlled proportions. Working backwards, the 1,024 participant x 1 predictor "brain" column of the across-participants design matrix (i.e. the trait variable) was drawn at random from the standard normal distribution and then shifted to obtain all positive numbers. The

1,024 x 1 “motion” column of the design matrix was taken as the mean FD timeseries of each participant. The brain and motion timeseries for each participant were then mixed by addition in the time domain according to their proportions (brain, motion) of variance in the design matrix. For example, if a participant’s row in the design matrix were (1.6329, 0.2216) then the mixed timeseries was generated by mixed = sqrt(1.6329) * brain + sqrt(0.2216) * motion. An additional step (detailed below) was used to correct the variance of the mixed timeseries to satisfy the assumptions of linear regression on correlation matrices.

Finally, the 394 x 394 correlation matrices for each participant were computed from their mixed timeseries. Correlation coefficients were transformed from the domain of (-1, 1) to connectivity over (-∞, ∞) using the atanh (Fisher’s Z) transform. The design matrix and connectivity matrices were used as inputs for the conventional resting-state connectivity regression model in Figure S4.

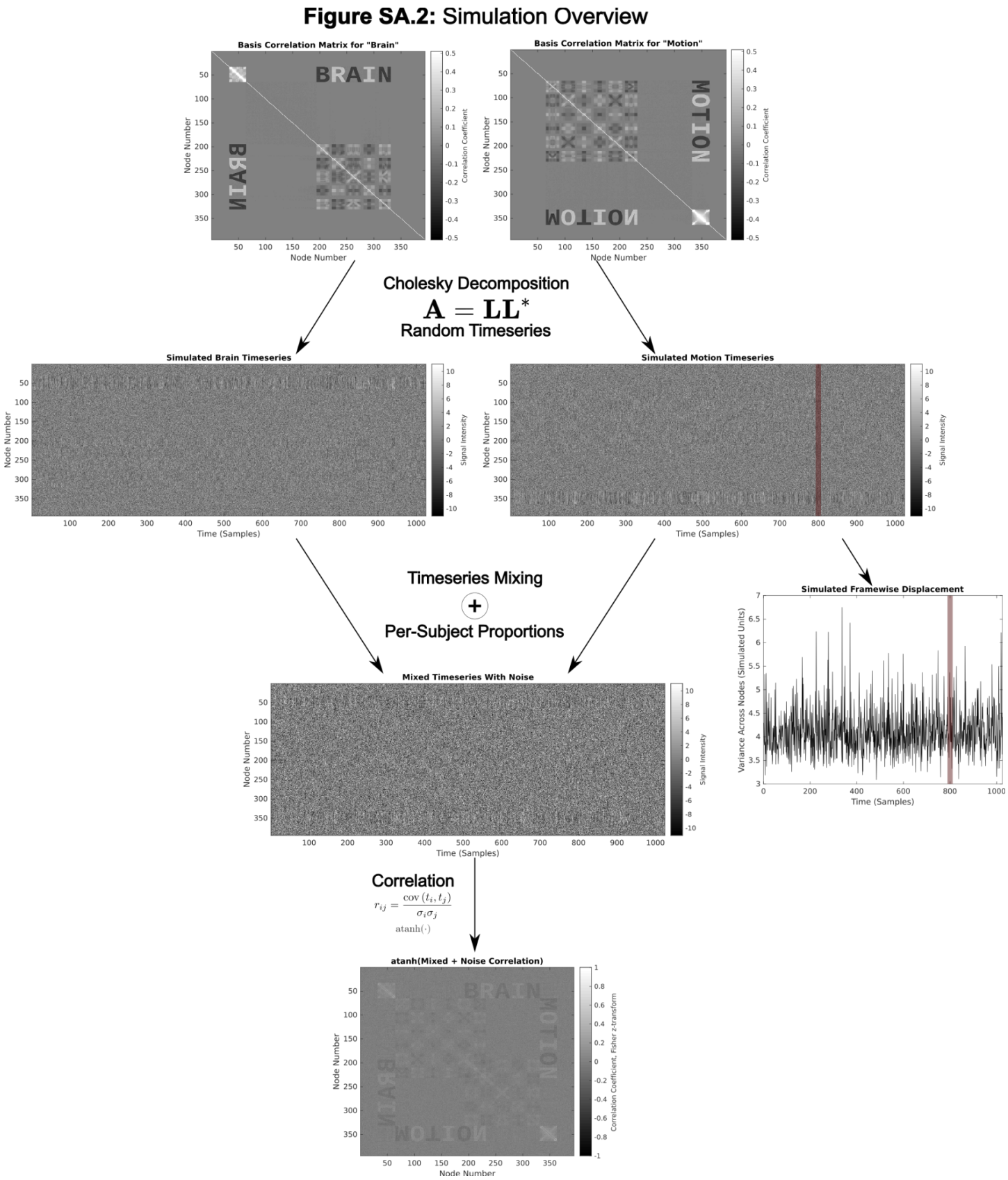

Figure SA.2: Method for generating simulated fMRI timeseries and connectivity matrices. 394 x 394 basis correlation matrices for “brain” and “motion” were Cholesky-decomposed to generate random 394 x 1,024 timeseries for each of 1,024 participants. The simulated FD (framewise displacement) timeseries was computed from the across-node variance of the motion timeseries at each point in time. The brain predictor variable was generated randomly. Brain and motion timeseries were mixed according to their proportions in the

design matrix as diagrammed in Figure SA.4. Finally, each participant's connectivity matrix was generated as the atanh (Fizher Z transform) of the correlation matrix of the mixed timeseries.

Simulation Results

Simulation results are summarized in Figure SA.3. First, as a negative control, we simulated FD timeseries but did not add any motion into the simulated fMRI data. This simulation depicts a scenario in which head motion occurred during scanning, but motion denoising algorithms were able to perfectly remove it from the data. No motion was seen in the mean connectivity matrix nor the brain-specific connectivity matrix from regression, nor did SHAMAN detect any trait-specific motion impact (results not shown).

As a second negative control, we simulated cases where motion *was* mixed into the data in equal proportion to brain signal, and across-participant motion was linearly separable from brain signal. This linear separability occurred for simulations where the proportions of brain and motion signal across participants were uncorrelated (i.e. trait/variable is not correlated with motion at all), and also for simulations where the proportions of brain and motion signal in the design matrix were linearly related to the true proportions of brain and motion signal mixed into the data (i.e. FD is a highly-accurate linear function of motion-induced fMRI signal). For these simulations we did observe motion in the mean connectivity matrix (Figure SA.4a) but there was no motion in the trait-specific connectivity matrix (Figure SA.4c). SHAMAN correctly did not detect any motion impact (Figure SA.4e).

As a positive control, we simulated cases where motion was mixed into the data in equal proportion to brain signal, but the FD timeseries was not a linear function of the proportion of mixed motion signal. In Figure SA.4, second column, we achieved this by drawing the motion-signal timeseries, x, from a random normal distribution correlated with the simulated brain signal $r = 0.5$. We used x to compute FD, but then transformed the motion-signal timeseries by $f(x) = 1 + x^2$ before mixing (i.e. FD is related to motion but not a pure linear function of it). As expected, we observed motion in the mean connectivity matrix (Figure SA.4b). We also observed residual motion artifact in the brain/trait-specific connectivity matrix from regression (Figure SA.4d), even when modeling mean FD as a covariate in regression. SHAMAN correctly detected the presence of this trait-specific motion impact (Figure SA.4f).

Figure SA.3: Simulation Results

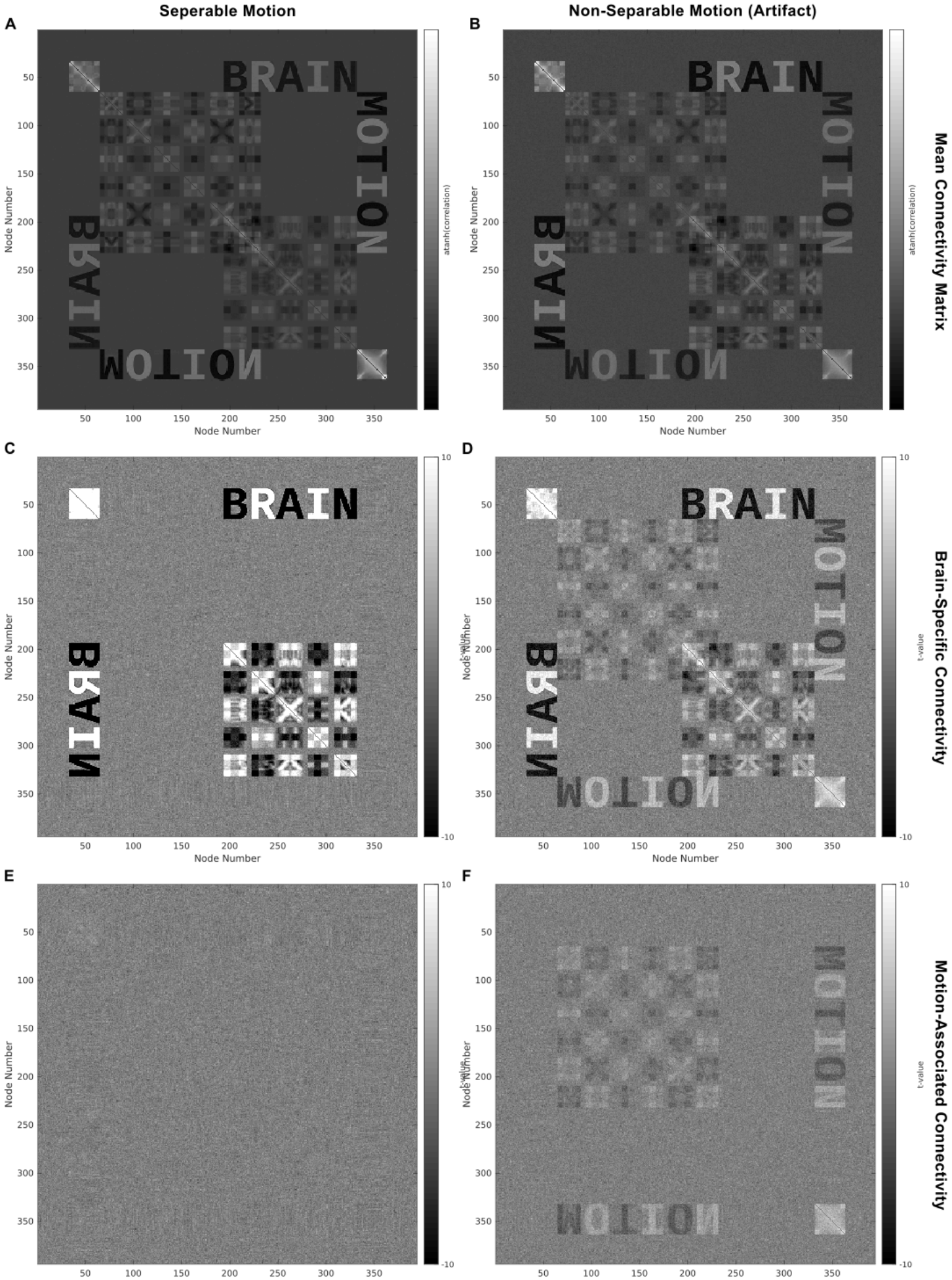

Figure SA.3: Simulation results when motion is linearly separable (left) and when motion is not linearly separable (right). In both cases **(a, b)** motion is present in the mean connectivity matrix. The linear regression model for "brain" with motion (participants' mean FD) as a covariate successfully removes the motion in **(c)** but not in **(d)**. SHAMAN does not detect motion-associated connectivity in **(e)** but correctly detects the artifact in **(f)** ($p < 0.001$).

Variance Correction

Proper timeseries mixing requires an additional step of variance correction to account for assumptions in the conventional model for resting-state fMRI regression. This model, diagrammed in Figure S4, assumes that connectivity matrices (derived from correlation matrices) are linearly superimposable for the purposes of linear regression. Correlation is computed by taking the covariance of two timeseries and dividing by the product of the standard deviations of the timeseries. Thus, regardless of the total amount of variation in each timeseries, the correlation is constrained to vary between -1 and 1. Therefore, in order to make correlation matrices (or their atanh-transformed connectivity matrices) linearly superimposable, we simulated data with the same total variance across participants.

Furthermore, we assume that (after any time-series/autoregressive filtering during pre-processing) computing correlation does not depend on the temporal ordering of the fMRI timeseries. This assumption allows us to "censor" high-motion frames from a timeseries, or to concatenate timeseries from different scans together. Therefore, in addition to generating data with the same total variance across participants, we simulated data with the same total variance across time.

The process for mixing the simulated timeseries to satisfy these assumptions is diagrammed in Figure SA.4. First the brain and motion timeseries were added together to generate a provisionally-mixed timeseries. Then the target variance to use across all participants was computed by summing the columns of the target design matrix and finding the row with the maximum value corresponding to the participant with the highest, or "maximum" variance. A random variance-correction timeseries was generated for each participant such that, when added to the provisionally-mixed timeseries, the resultant variance-corrected timeseries had constant across-node variance over time equal to the maximum variance.

Finally, an equal amount of normally-distributed noise was added to the variance-corrected timeseries for each participant to obtain the final, properly-mixed timeseries used to compute each participant's correlation and connectivity matrices. This additional noise models unstructured noise from the fMRI scan (e.g. thermal noise), and also allows for the across-node variance of the mixed timeseries to vary with time while still satisfying the assumptions of the linear model.

Figure SA.4: Details of Timeseries Mixing

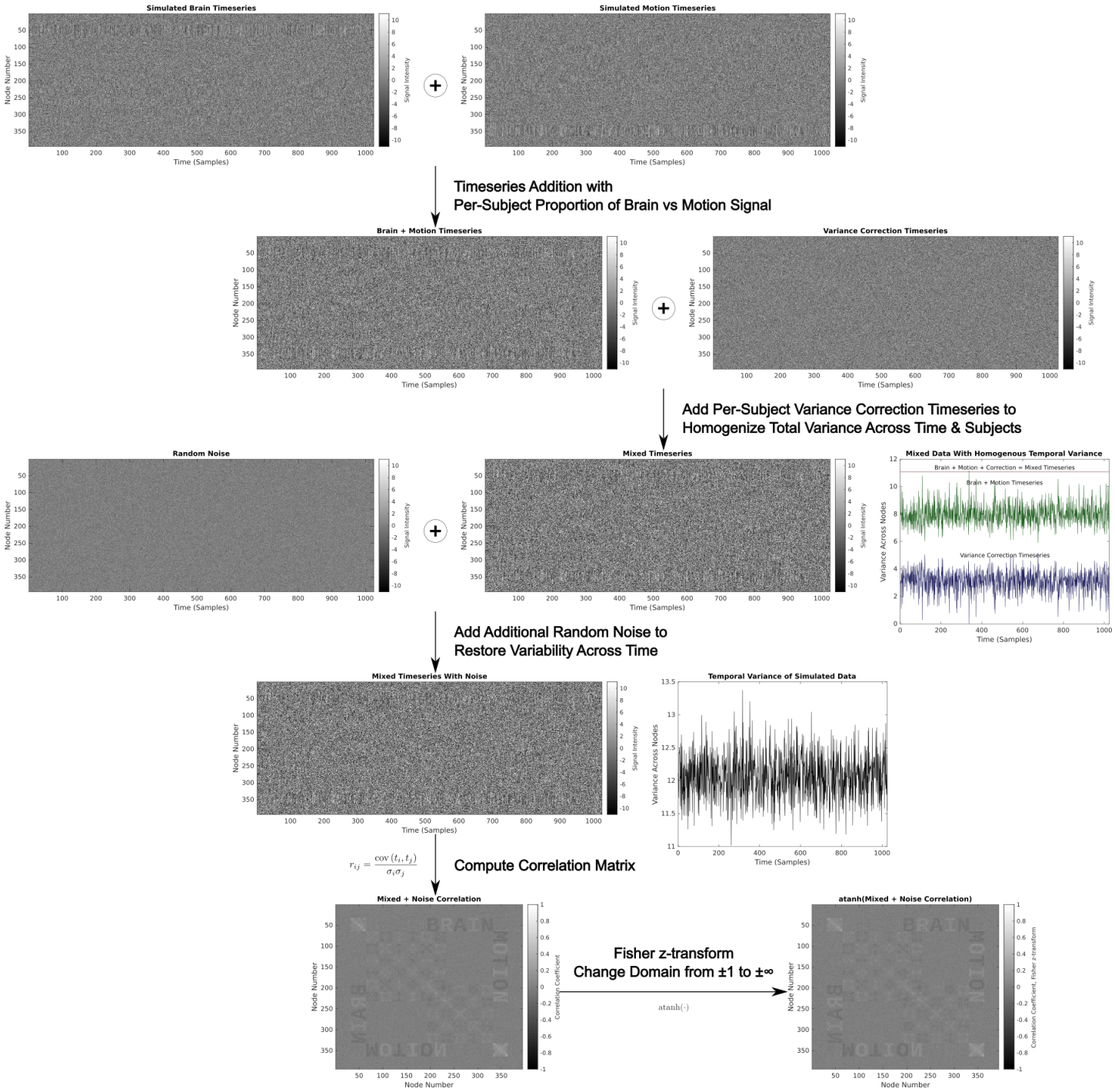

Figure SA.4: Method for proper timeseries mixing of brain and motion signals in controlled proportions. A provisional mixing of the brain and motion timeseries is obtained by addition in the time domain. A random variance-correction timeseries is then added such that, at each time point, the across-node variance of the variance-corrected timeseries is equal to the maximum variance of any participant. Finally, an equal amount of normally distributed noise is added to the variance-corrected timeseries to simulate thermal noise and to allow for the across-node variance of the final, properly-mixed timeseries to vary with time. The mixed timeseries are used to generate connectivity matrices for each participant as in Figure SA.2.

References

1. Behzadi Y, Restom K, Liau J, Liu TT. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*. 2007;37(1):90-101. doi:10.1016/j.neuroimage.2007.04.042
2. Bolton TAW, Kebets V, Glerean E, et al. Agito ergo sum: Correlates of spatio-temporal motion characteristics during fMRI. *NeuroImage*. 2020;209:116433. doi:10.1016/j.neuroimage.2019.116433
3. Burgess GC, Kandala S, Nolan D, et al. Evaluation of Denoising Strategies to Address Motion-Correlated Artifacts in Resting-State Functional Magnetic Resonance Imaging Data from the Human Connectome Project. *Brain Connect*. 2016;6(9):669-680. doi:10.1089/brain.2016.0435
4. Ciric R, Wolf DH, Power JD, et al. Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *NeuroImage*. 2017;154:174-187. doi:10.1016/j.neuroimage.2017.03.020
5. Dosenbach NUF, Koller JM, Earl EA, et al. Real-time motion analytics during brain MRI improve data quality and reduce costs. *NeuroImage*. 2017;161:80-93. doi:10.1016/j.neuroimage.2017.08.025
6. Fair DA, Nigg JT, Iyer S, et al. Distinct neural signatures detected for ADHD subtypes after controlling for micro-movements in resting state functional connectivity MRI data. *Front Syst Neurosci*. 2013;6. doi:10.3389/fnsys.2012.00080
7. Griffanti L, Salimi-Khorshidi G, Beckmann CF, et al. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *NeuroImage*. 2014;95:232-247. doi:10.1016/j.neuroimage.2014.03.034
8. Kundu P, Brenowitz ND, Voon V, et al. Integrated strategy for improving functional connectivity mapping using multiecho fMRI. *Proc Natl Acad Sci*. 2013;110(40):16187-16192. doi:10.1073/pnas.1301725110
9. Makowski C, Lepage M, Evans AC. Head motion: the dirty little secret of neuroimaging in psychiatry. *J Psychiatry Neurosci*. 2019;44(1):62-68. doi:10.1503/jpn.180022
10. Mowinckel AM, Espeseth T, Westlye LT. Network-specific effects of age and in-scanner subject motion: A resting-state fMRI study of 238 healthy adults. *NeuroImage*. 2012;63(3):1364-1373. doi:10.1016/j.neuroimage.2012.08.004
11. Muschelli J, Nebel MB, Caffo BS, Barber AD, Pekar JJ, Mostofsky SH. Reduction of motion-related artifacts in resting state fMRI using aCompCor. *NeuroImage*. 2014;96:22-35. doi:10.1016/j.neuroimage.2014.03.028
12. Nielsen AN, Greene DJ, Gratton C, Dosenbach NUF, Petersen SE, Schlaggar BL. Evaluating the Prediction of Brain Maturity From Functional Connectivity After Motion Artifact Denoising. *Cereb Cortex*. 2019;29(6):2455-2469. doi:10.1093/cercor/bhy117
13. Patel AX, Kundu P, Rubinov M, et al. A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series. *NeuroImage*. 2014;95:287-304. doi:10.1016/j.neuroimage.2014.03.012
14. Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*. 2012;59(3):2142-2154. doi:10.1016/j.neuroimage.2011.10.018
15. Pruim RHR, Mennes M, van Rooij D, Llera A, Buitelaar JK, Beckmann CF. ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *NeuroImage*. 2015;112:267-277. doi:10.1016/j.neuroimage.2015.02.064
16. Salimi-Khorshidi G, Douaud G, Beckmann CF, Glasser MF, Griffanti L, Smith SM. Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*. 2014;90:449-468. doi:10.1016/j.neuroimage.2013.11.046

17. Satterthwaite TD, Wolf DH, Loughead J, et al. Impact of in-scanner head motion on multiple measures of functional connectivity: Relevance for studies of neurodevelopment in youth. *NeuroImage*. 2012;60(1):623-632. doi:10.1016/j.neuroimage.2011.12.063
18. Siegel JS, Mitra A, Laumann TO, et al. Data Quality Influences Observed Links Between Functional Connectivity and Behavior. *Cereb Cortex*. 2017;27(9):4492-4502. doi:10.1093/cercor/bhw253
19. Van Dijk KRA, Sabuncu MR, Buckner RL. The influence of head motion on intrinsic functional connectivity MRI. *NeuroImage*. 2012;59(1):431-438. doi:10.1016/j.neuroimage.2011.07.044
20. Yan CG, Cheung B, Kelly C, et al. A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *NeuroImage*. 2013;76:183-201. doi:10.1016/j.neuroimage.2013.03.004
21. Casey BJ, Cannonier T, Conley MI, et al. The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Dev Cogn Neurosci*. 2018;32:43-54. doi:10.1016/j.dcn.2018.03.001
22. Greene DJ, Koller JM, Hampton JM, et al. Behavioral interventions for reducing head motion during MRI scans in children. *NeuroImage*. 2018;171:234-245. doi:10.1016/j.neuroimage.2018.01.023
23. Pagliaccio D, Luby J, Gaffrey M, et al. Anomalous functional brain activation following negative mood induction in children with pre-school onset major depression. *Dev Cogn Neurosci*. 2012;2(2):256-267. doi:10.1016/j.dcn.2011.11.008
24. Vanderwal T, Kelly C, Eilbott J, Mayes LC, Castellanos FX. Inscapes : A movie paradigm to improve compliance in functional magnetic resonance imaging. *NeuroImage*. 2015;122:222-232. doi:10.1016/j.neuroimage.2015.07.069
25. Fair DA, Miranda-Dominguez O, Snyder AZ, et al. Correction of respiratory artifacts in MRI head motion estimates. *NeuroImage*. 2020;208:116400. doi:10.1016/j.neuroimage.2019.116400
26. Friston KJ, Williams S, Howard R, Frackowiak RSJ, Turner R. Movement-Related effects in fMRI time-series: Movement Artifacts in fMRI. *Magn Reson Med*. 1996;35(3):346-355. doi:10.1002/mrm.1910350312
27. Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*. 2014;84:320-341. doi:10.1016/j.neuroimage.2013.08.048
28. Power JD, Plitt M, Gotts SJ, et al. Ridding fMRI data of motion-related influences: Removal of signals with distinct spatial and physical bases in multiecho data. *Proc Natl Acad Sci*. 2018;115(9). doi:10.1073/pnas.1720985115
29. Satterthwaite TD, Elliott MA, Gerraty RT, et al. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*. 2013;64:240-256. doi:10.1016/j.neuroimage.2012.08.052
30. Tyszka JM, Kennedy DP, Paul LK, Adolphs R. Largely Typical Patterns of Resting-State Functional Connectivity in High-Functioning Adults with Autism. *Cereb Cortex*. 2014;24(7):1894-1905. doi:10.1093/cercor/bht040
31. Murphy K, Fox MD. Towards a consensus regarding global signal regression for resting state functional connectivity MRI. *NeuroImage*. 2017;154:169-173. doi:10.1016/j.neuroimage.2016.11.052
32. Gratton C, Dworetsky A, Coalson RS, et al. Removal of high frequency contamination from motion estimates in single-band fMRI saves data without biasing functional connectivity. *NeuroImage*. 2020;217:116866. doi:10.1016/j.neuroimage.2020.116866
33. Havsteen I, Ohlhues A, Madsen KH, Nybing JD, Christensen H, Christensen A. Are Movement Artifacts in Magnetic Resonance Imaging a Real Problem?—A Narrative Review. *Front Neurol*. 2017;8:232. doi:10.3389/fneur.2017.00232
34. Griffanti L, Douaud G, Bijsterbosch J, et al. Hand classification of fMRI ICA noise components.

NeuroImage. 2017;154:188-205. doi:10.1016/j.neuroimage.2016.12.036
35. Alfaro-Almagro F, Jenkinson M, Bangerter NK, et al. Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage*. 2018;166:400-424. doi:10.1016/j.neuroimage.2017.10.034
36. Feczko E, Conan G, Marek S, et al. *Adolescent Brain Cognitive Development (ABCD) Community MRI Collection and Utilities*. Neuroscience; 2021. doi:10.1101/2021.07.09.451638
37. Glasser MF, Sotiropoulos SN, Wilson JA, et al. The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*. 2013;80:105-124. doi:10.1016/j.neuroimage.2013.04.127
38. Hagler DJ, Hatton SeanN, Cornejo MD, et al. Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *NeuroImage*. 2019;202:116091. doi:10.1016/j.neuroimage.2019.116091
39. Power JD, Schlaggar BL, Petersen SE. Recent progress and outstanding issues in motion correction in resting state fMRI. *NeuroImage*. 2015;105:536-551. doi:10.1016/j.neuroimage.2014.10.044
40. Winkler AM, Webster MA, Brooks JC, Tracey I, Smith SM, Nichols TE. Non-parametric combination and related permutation tests for neuroimaging: NPC and Related Permutation Tests for Neuroimaging. *Hum Brain Mapp*. 2016;37(4):1486-1511. doi:10.1002/hbm.23115
41. Jernigan TL, Brown SA. Introduction. *Dev Cogn Neurosci*. 2018;32:1-3. doi:10.1016/j.dcn.2018.02.002
42. Volkow ND, Koob GF, Croyle RT, et al. The conception of the ABCD study: From substance use to a broad NIH collaboration. *Dev Cogn Neurosci*. 2018;32:4-7. doi:10.1016/j.dcn.2017.10.002
43. Barch DM, Albaugh MD, Avenevoli S, et al. Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: Rationale and description. *Dev Cogn Neurosci*. 2018;32:55-66. doi:10.1016/j.dcn.2017.10.010
44. Luciana M, Bjork JM, Nagel BJ, et al. Adolescent neurocognitive development and impacts of substance use: Overview of the adolescent brain cognitive development (ABCD) baseline neurocognition battery. *Dev Cogn Neurosci*. 2018;32:67-79. doi:10.1016/j.dcn.2018.02.006
45. Marek S, Tervo-Clemmens B, Calabro FJ, et al. Reproducible brain-wide association studies require thousands of individuals. *Nature*. 2022;603(7902):654-660. doi:10.1038/s41586-022-04492-9
46. Wechsler D, Pearson Education I, Psychological Corporation. *WISC-V: Wechsler Intelligence Scale for Children*. NCS Pearson, Inc. : PsychCorp; 2014.
47. Grummer-Strawn LM, Reinold C, Krebs NF, Centers for Disease Control and Prevention (CDC). Use of World Health Organization and CDC growth charts for children aged 0-59 months in the United States. *MMWR Recomm Rep Morb Mortal Wkly Rep Recomm Rep*. 2010;59(RR-9):1-15.
48. Laumann TO, Gordon EM, Adeyemo B, et al. Functional System and Areal Organization of a Highly Sampled Individual Human Brain. *Neuron*. 2015;87(3):657-670. doi:10.1016/j.neuron.2015.06.037
49. Stouffer SA, Suchman EA, Devinney LC, Star SA, Williams Jr. RM. *The American Soldier: Adjustment during Army Life. (Studies in Social Psychology in World War II), Vol. 1*. Princeton Univ. Press; 1949:xii, 599.
50. van Zwet WR, Oosterhoff J. On the Combination of Independent Test Statistics. *Ann Math Stat*. 1967;38(3):659-680.
51. Couvy-Duchesne B, Blokland GAM, Hickie IB, et al. Heritability of head motion during resting state functional MRI in 462 healthy twins. *NeuroImage*. 2014;102:424-434. doi:10.1016/j.neuroimage.2014.08.010
52. Engelhardt LE, Roe MA, Juranek J, et al. Children's head motion during fMRI tasks is heritable and stable over time. *Dev Cogn Neurosci*. 2017;25:58-68. doi:10.1016/j.dcn.2017.01.011
53. Hodgson K, Poldrack RA, Curran JE, et al. Shared Genetic Factors Influence Head Motion

During MRI and Body Mass Index. *Cereb Cortex*. Published online October 15, 2016:cercor;bhw321v1. doi:10.1093/cercor/bhw321
54. Zeng LL, Wang D, Fox MD, et al. Neurobiological basis of head motion in brain imaging. *Proc Natl Acad Sci*. 2014;111(16):6058-6062. doi:10.1073/pnas.1317424111
55. Birn RM, Molloy EK, Patriat R, et al. The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *NeuroImage*. 2013;83:550-558. doi:10.1016/j.neuroimage.2013.05.099
56. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. *NeuroImage*. 2012;62(2):782-790. doi:10.1016/j.neuroimage.2011.09.015
57. Fischl B. Automatically Parcellating the Human Cerebral Cortex. *Cereb Cortex*. 2004;14(1):11-22. doi:10.1093/cercor/bhg087
58. Gordon EM, Laumann TO, Adeyemo B, Huckins JF, Kelley WM, Petersen SE. Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cereb Cortex*. 2016;26(1):288-303. doi:10.1093/cercor/bhu239
59. Seitzman BA, Gratton C, Marek S, et al. A set of functionally-defined brain regions with improved representation of the subcortex and cerebellum. *NeuroImage*. 2020;206:116290. doi:10.1016/j.neuroimage.2019.116290
60. Raut RV, Mitra A, Snyder AZ, Raichle ME. On time delay estimation and sampling error in resting-state fMRI. *NeuroImage*. 2019;194:211-227. doi:10.1016/j.neuroimage.2019.03.020
61. Gratton C, Laumann TO, Nielsen AN, et al. Functional Brain Networks Are Dominated by Stable Group and Individual Factors, Not Cognitive or Daily Variation. *Neuron*. 2018;98(2):439-452.e5. doi:10.1016/j.neuron.2018.03.035
62. Ellis CT, Baldassano C, Schapiro AC, Cai MB, Cohen JD. Facilitating open-science with realistic fMRI simulation: validation and application. *PeerJ*. 2020;8:e8564. doi:10.7717/peerj.8564