Motion Impact Score for Detecting Spurious Brain-Behavior Associations

Authors

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

[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 Biomedical Engineering, Washington University in St Louis, St Louis, MO, USA

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

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

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

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

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

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

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

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

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

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

[14]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) effects 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 effects. The SHAMAN approach distinguishes between motion artifact causing overestimation or underestimation of trait-FC effects. SHAMAN was > 95% specific at sample sizes of n = 100 and above. SHAMAN was powered to detect motion overestimation scores 80% of the time at sample sizes of n = 5,000 but could detect motion underestimation scores only 50% of the time at n = 5000, making it most useful for researchers seeking to avoid overestimating trait-FC effects in large BWAS. We computed motion impact scores for trait-FC effect 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, 42% (19/45) of traits had significant ($p < 0.05$) motion overestimation scores and 38% (17/45) of traits had significant motion underestimation scores. Censoring at framewise displacement (FD) < 0.2 mm reduced the proportion of traits with significant motion overestimation scores from 42% to 2% (1/45) but did not decrease the number of traits with significant motion underestimation 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 has been shown to be 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 [6,19,29–31]. 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 [28,32], motion parameter regression [29], spectral filtering, respiratory filtering [25,33,34], principal component analysis [1,11], independent component analysis [7,15,35], 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 over- or underestimating trait-FC effects.

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 [36–39]; 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. [40] and Pham et al. [41] 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 effects 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 effects by comparing within-participant and between-participant variance in the trait-FC effects 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 artifact causing over- or under-estimation of the trait-FC effects.

Thus, we developed a novel method for computing a 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 effects 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. A direction (positive or negative) of the motion impact score that is aligned with the direction of the trait-FC effect is consistent with motion causing overestimation of the trait-FC effect, a "motion overestimation score." A motion impact score opposite the direction of the trait-FC effect is consistent with motion causing underestimation of the trait-FC effect, a "motion underestimation score." Permutation of the timeseries and non-parametric combining [42,43] across pairwise connections yields a motion impact score and a p-value distinguishing significant from not-significant impacts of motion on trait-FC effects.

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 [21,44,45] with extensive demographic, biophysical, and behavioral data [46,47]. 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 [48]. 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. We performed supplementary analyses on the Human Connectome Project [49,50] to demonstrate the generalizability of our results to other denoising methods and data sets.

Results

The Effect of Residual Motion is Large Even After Denoising and Censoring

In order to characterize the impact of residual head motion on trait-FC effects, we first performed preliminary analyses to quantify how much residual motion was left in the data after denoising. Of the 11,874 children recruited into ABCD, n = 9,652 children with at least 8 minutes of rs-fMRI data were included in this portion of the analysis. ABCD-BIDS is the default denoising algorithm for pre-processed ABCD data [25,37]. 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) in a linear model before and after applying ABCD-BIDS (Figure S1). After minimal processing [39] (i.e. motion-correction by frame realignment only), 9% of signal variance was explained by head motion (Pearson 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). We also used Spearman correlation to relax the assumption of linearity, obtaining Speaman r = 0.61 after minimal processing and r = 0.27 after ABCD-BIDS. Therefore, ABCD-BIDS achieved a relative reduction in the proportion of signal variance related to motion of 98.9% (Pearson, linear) or 79.9% (Spearman) 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 effect matrix with units of change in FC per mm FD (Figure 1c,d). The average FC matrix and motion-FC effect matrix were compared by computing the correlation between their edges (pairwise connections). The motion-FC effect matrix had a strong, negative correlation of Pearson r = -0.63 (Spearman r = -0.58) with the average FC matrix. This strong, negative correlation persisted even after motion censoring at FD < 0.2 mm (Pearson r = -0.55, Spearman r = -0.51). Thus, across all functional connections, connection strength tended to be larger in participants who moved more compared to participants who moved less.

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.[48] 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.

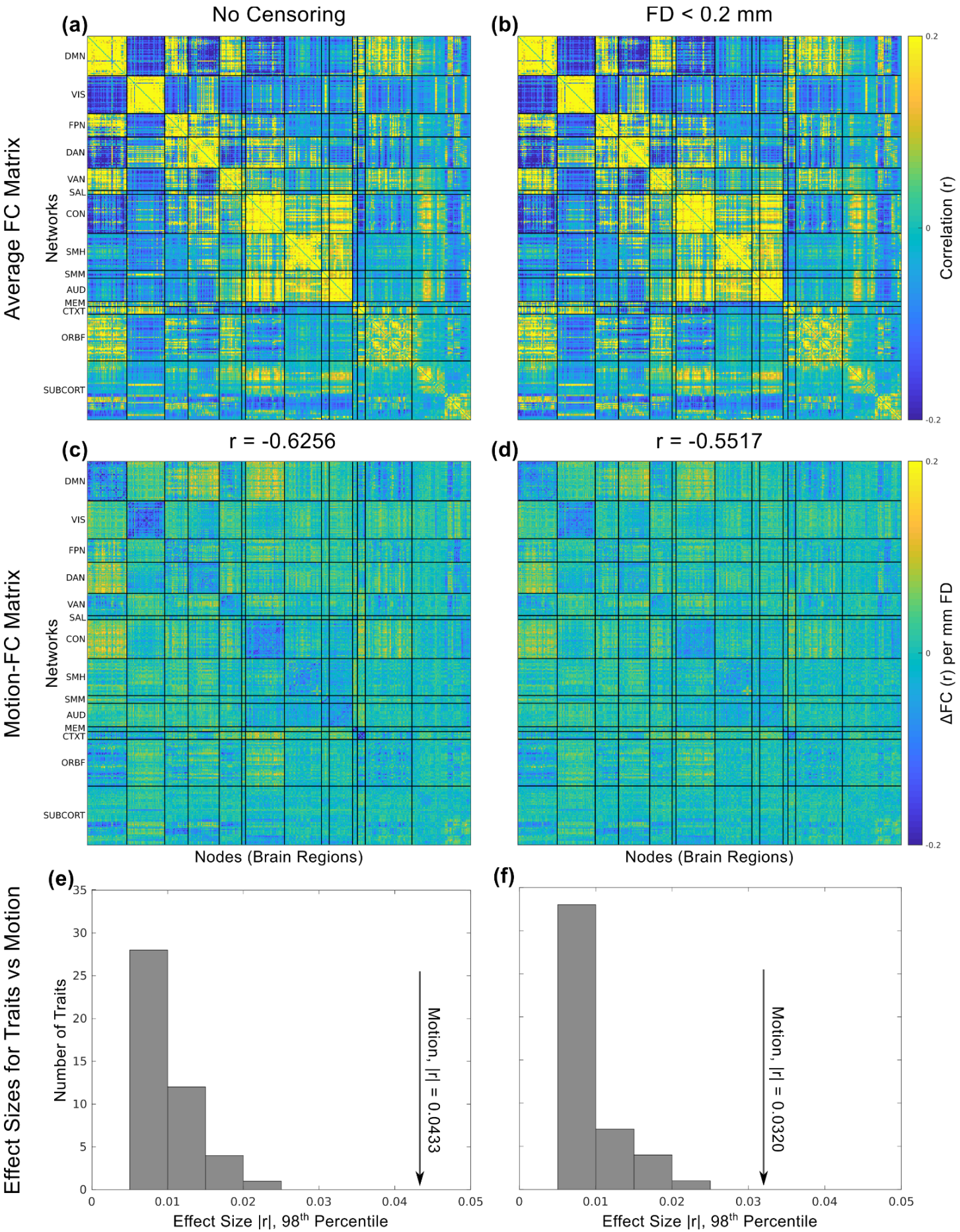
(a)
No Censoring
(b)
FD < 0.2 mm
Average FC Matrix
Networks
DMN
VIS
FPN
DAN
VAN
SAL
CON
SMH
SMM
AUD
MEM
CTXT
ORBF
SUBCORT
Correlation (r)
0.2
0
-0.2
(c)
r = -0.6256
(d)
r = -0.5517
Motion-FC Matrix
ΔFC (r) per mm FD
Nodes (Brain Regions)
Nodes (Brain Regions)
(e)
(f)
Effect Sizes for Traits vs Motion
Number of Traits
Motion, |r| = 0.0433
Motion, |r| = 0.0320
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-FC effect 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 effect matrix obtained by regressing average FD against FC for each participant, FC ~ 1 + FD. The motion-FC effect matrix has units of change in FC per mm FD. **(d)** Motion-FC effect matrix with censoring at FD < 0.2. **(a,c)** The upper triangular parts of the FC connectivity matrices were edge-for-edge correlated at Pearson r = -0.6256 (Spearman r = -0.5822). **(b,d)** After motion censoring at FD < 0.2 mm, the average FC and motion-FC effect matrices were edge-for-edge correlated at Pearson r = -0.5517 (Spearman r = -0.5059). **(e)** The effect size of the motion-FC effect (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

Of the 11,874 total children recruited into ABCD, and 9,652 children with at least 8 minutes of rs-fMRI data, n = 7,270 children without missing data for any of the 45 traits we examined were included in this and subsequent analyses.

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 [51] had the highest trait-FD correlation at Pearson r = -0.12, Spearman r = -0.12, $p < 0.001$, and Child Behavioral Checklist (CBCL) somatization subscore had the least trait-FD correlation at Pearson r = 0.0004, Spearman r = -0.02 (not significant).

BMI [52] and matrix reasoning subscore [51] were selected as exemplar traits. Both exhibited significant trait-FD correlations (Figure S2). Children with higher BMI exhibited more in-scanner head motion (Pearson r = 0.09, Spearman r = 0.13, $p < 0.001$). Children with higher matrix reasoning scores had lower head motion (Pearson r = -0.12, Spearman r = -0.12, $p < 0.001$).

SHAMAN Detected the Impact of Motion on Trait-FC Effects 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 effects, “MOTION” trait-FC effects, and relationship between them. Simulation of non-linear relationships between the trait-FC effects and motion caused denoising to fail and for SHAMAN to detect a significant impact of residual motion on trait-FC effects (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" effects (Figure 2a), reflecting the failure of denoising to completely remove the "MOTION" effects. 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 effect matrix, the identical "BRAIN" trait-FC effects 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). The motion impact score approximates how estimates of the trait-FC effects are changed by head motion (Figure SA3.D). A motion overestimation score approximates how much the trait-FC effect is overestimated due to motion. A motion underestimation score approximates how much the trait-FC effect is underestimated due to motion .

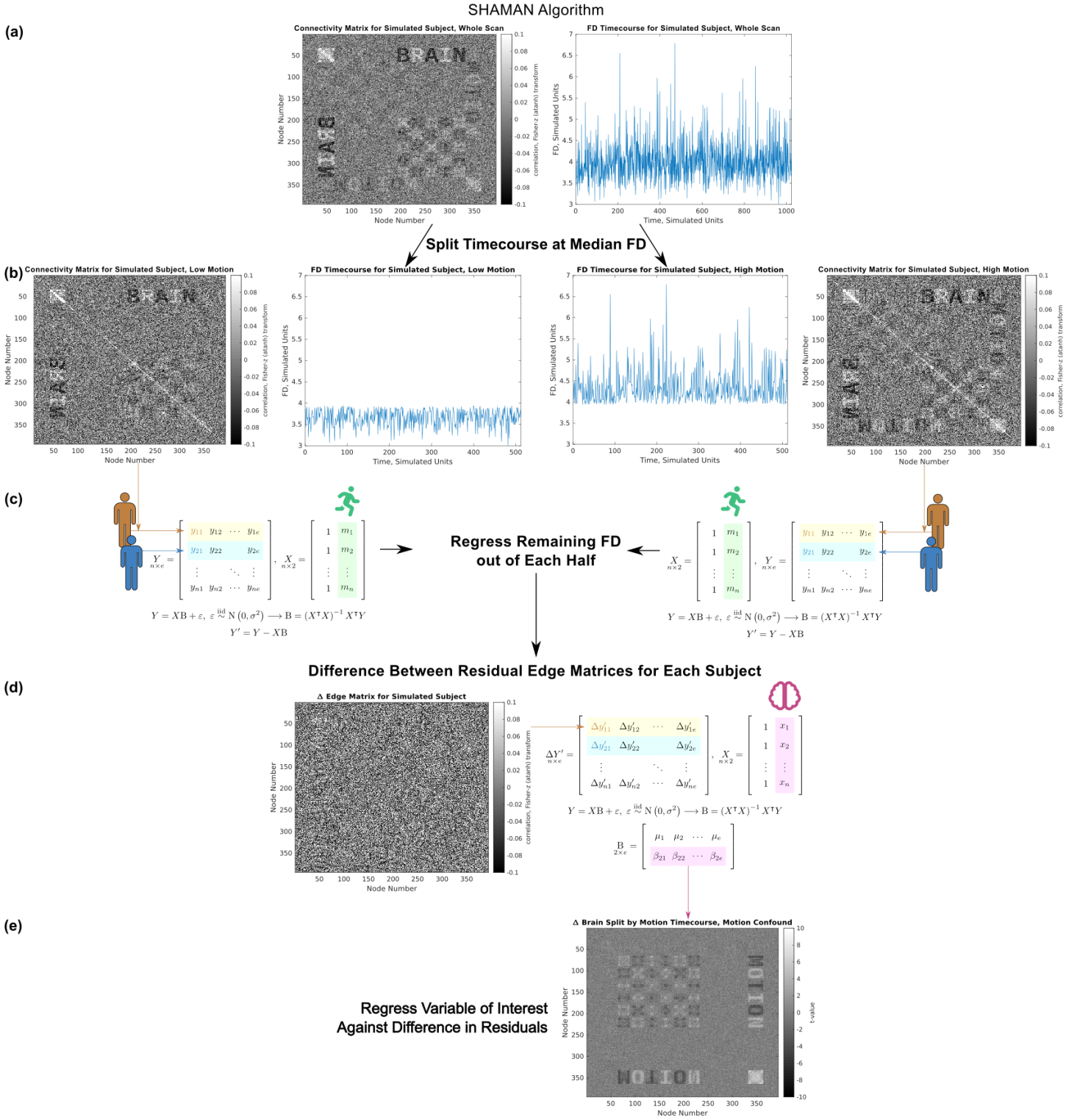

Figure 2: Diagram of the SHAMAN (Split-Half Analysis of Motion-Associated Networks) algorithm using simulated data. The simulated timeseries of brain signal generates the "BRAIN" edge matrix, and the simulated timeseries of motion generates the "MOTION" edge matrix. **(a)** A single participant's simulated 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 effects 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 motion overestimation scores, meaning that their trait-FC effects were significantly ($p < 0.05$) greater due to the impact of residual head motion. Three biophysical traits, BMI, sex, and weight, also had significant motion underestimation scores, meaning that motion inflated trait-FC effects in some networks but suppressed them in other networks (Tables S2, S3).

BMI was selected as an exemplar trait to illustrate overestimation of the trait-FC effect due to motion. The trait-FC effects of BMI (FD as a covariate, without motion censoring) were compared to its motion impact score at each edge in Figure 3 and projected onto the cortical surface using root mean square connectivity [53] 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 effect and a negative motion impact. Since these trait-FC effects 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 effects. An overall motion overestimation score (Stouffer's Z[42]) score for BMI of 92 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 negation of BMI, a trait for which high values were associated with low in-scanner head motion, caused the trait-FC effects of inverse-BMI and the motion impact to both be positive instead of negative (data not shown). Since the trait-FC effects and corresponding motion impact were still both in the same direction (positive in this counterexample), the overall motion impact score was still a motion overestimation score.

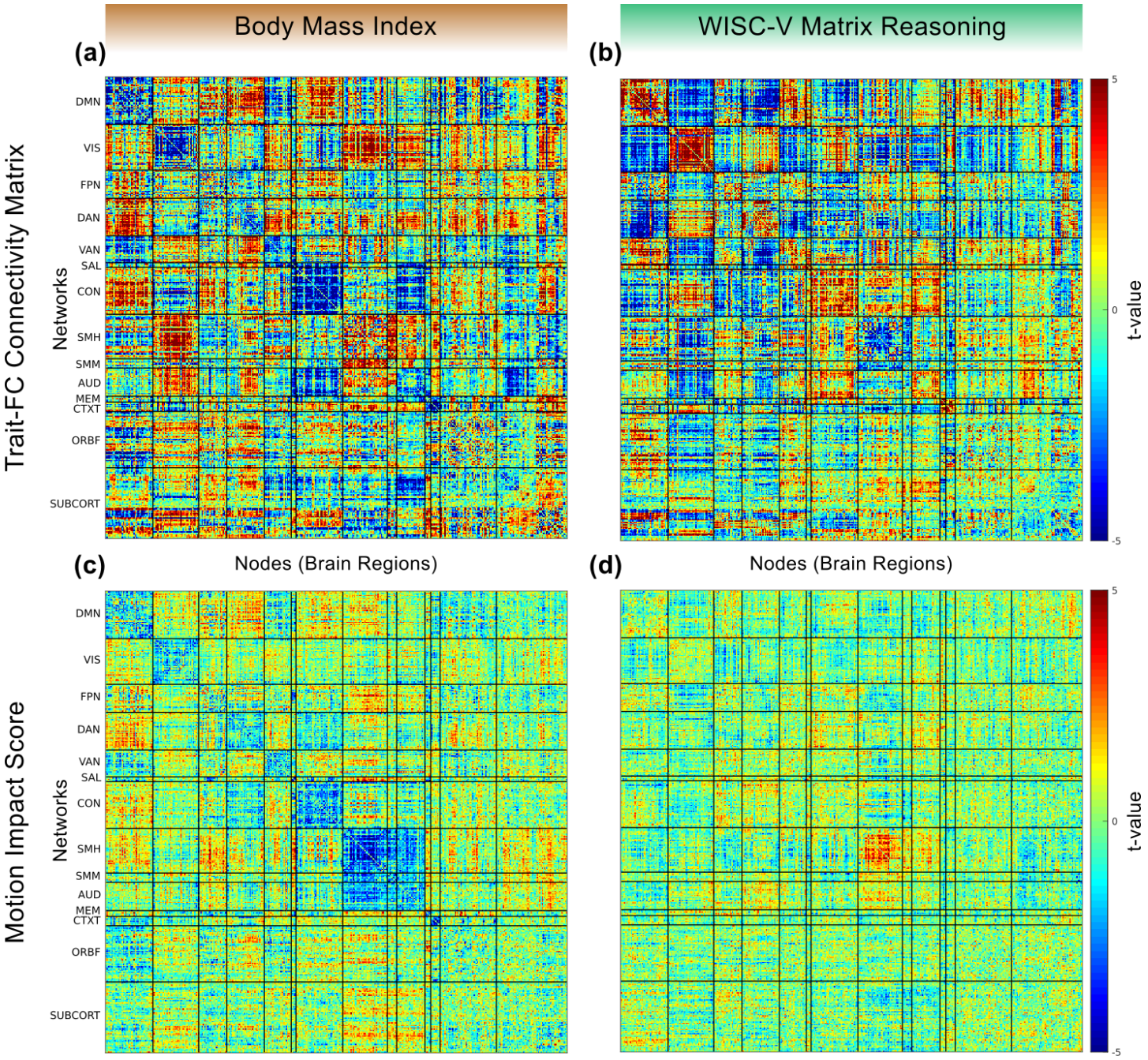

Figure 3: Trait-specific impact of motion on functional connectivity matrices. Data are shown for two exemplar traits, body mass index (BMI, left) and WISC-V matrix reasoning score (right). **(a)** Resting state functional connectivity (RSFC) effect matrix for BMI. **(b)** Trait-FC effect matrix for matrix reasoning. **(c)** Motion impact score for BMI. The overall (over the entire matrix, omnibus Stouffer's Z) motion overestimation score for BMI was 92, $p < 0.001$. **(d)** Motion impact score for WISC-V. The omnibus motion underestimation score for WISC-V was 44, $p < 0.001$. 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 Motion underestimation score (Figures 3, 4). The overall motion underestimation score for WISC-V of 44 was significant at (omnibus) $p < 0.001$. The motion overestimation score for WISC-V was not significant. The trait-FC effects 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 effect, thus suppressing the trait-FC effect.

In total, 34% (13/38) of demographic and behavioral traits had significant motion overestimation scores, and 34% had significant motion underestimation impact scores, prior to frame censoring (Table S2, S3). Broken down further by category, there was a significant motion overestimation score for 20% (1/5) of demographic traits, 50% (5/10) of cognitive traits, and 32% (7/22) of personality traits. There was a significant motion underestimation score for 67% (4/6) of demographic traits, 70% (7/10) of cognitive traits, and 14% (3/22) of personality traits. Thus, demographic and behavioral traits were less impacted by residual motion than biophysical traits.

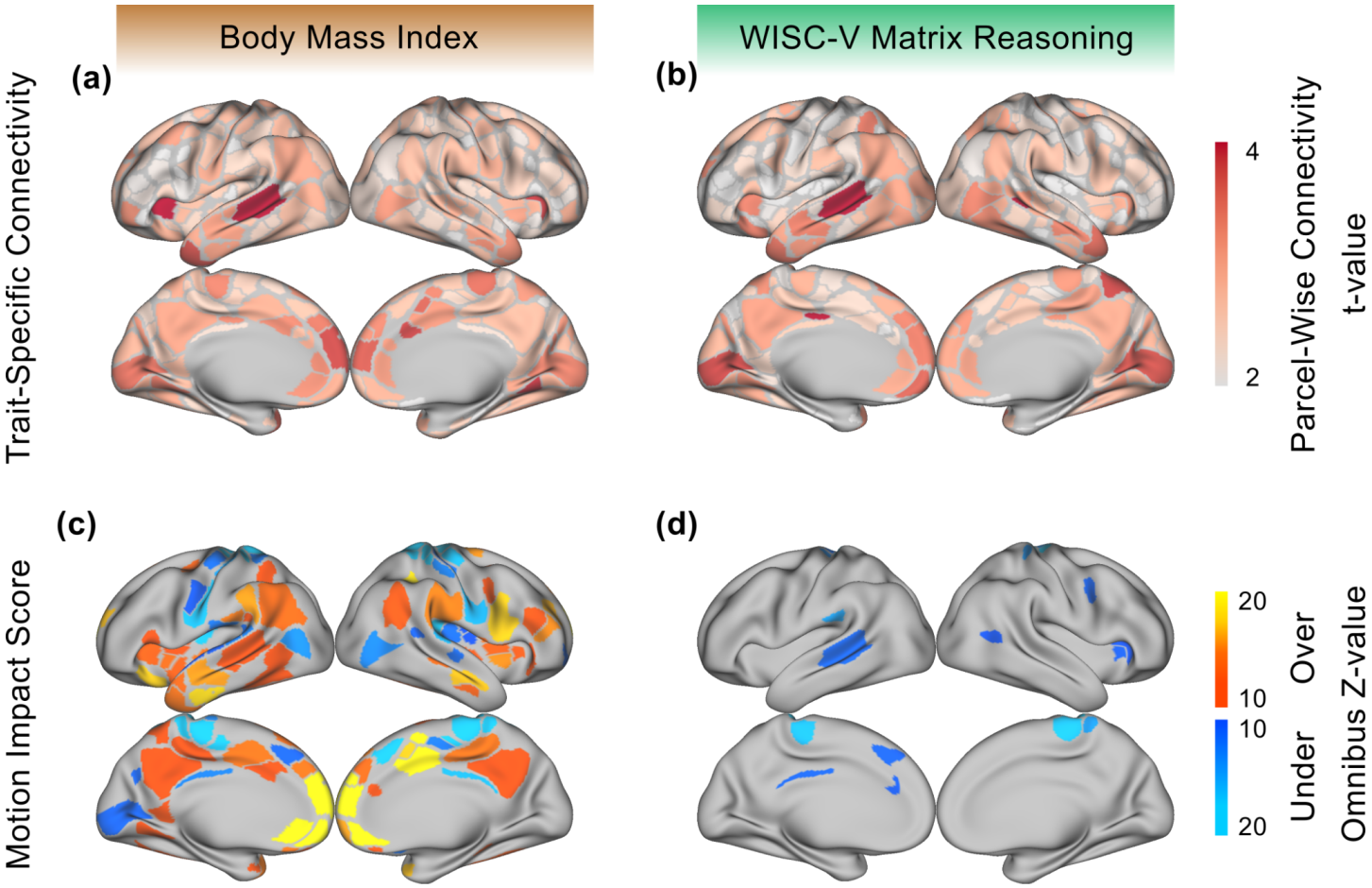

Figure 4: Trait-specific impact of motion on functional connectivity (FC). Top: Parcel-wise FC, computed as the root mean square (RMS) of connectivity values for each parcel/node in the trait-FC effect matrix [53] for **(a)** body mass index (BMI) and **(b)** WISC-V matrix reasoning score. **Bottom:** Motion impact score (omnibus Stouffer's Z [42,54,55], higher = more motion) for **(c)** BMI and **(d)** WISC-V. Motion overestimation scores are labeled "Over" in orange and motion underestimation scores are labeled "Under" 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 [14,20,27,29,31,41]. 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)[25,33,34] 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 S4) compared to the uncensored sample of participants. On the other hand, censoring at FD < 0.2 reduced the number of traits with significant motion overestimation scores from 42% (19/45) to just one physical trait, height, see Figure 5 (Tables S2, S3). Similar results using the ABCD data were obtained in our supplementary analyses using DVARS, an alternative to FD for quantifying in-scanner head motion calculated from the root mean square variance of the difference between successive fMRI timepoints. The overall trend of significant motion impact scores without motion censoring that improved with motion censoring at FD < 0.2 mm (especially for biophysical traits) was also observed in data from the Human Connectome Project (HCP).

More stringent censoring at FD < 0.1 mm (filtered) did not reduce the number of traits with significant motion overestimation scores 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 S4). Boxplots showing trait distributions at different motion censoring thresholds for all 45 traits can be found in Figure S5. Motion censoring at FD < 0.1 did reduce significant motion underestimation scores from 38% (17/45) with no censoring and at FD < 0.2 to 20% (9/45) at FD < 0.1.

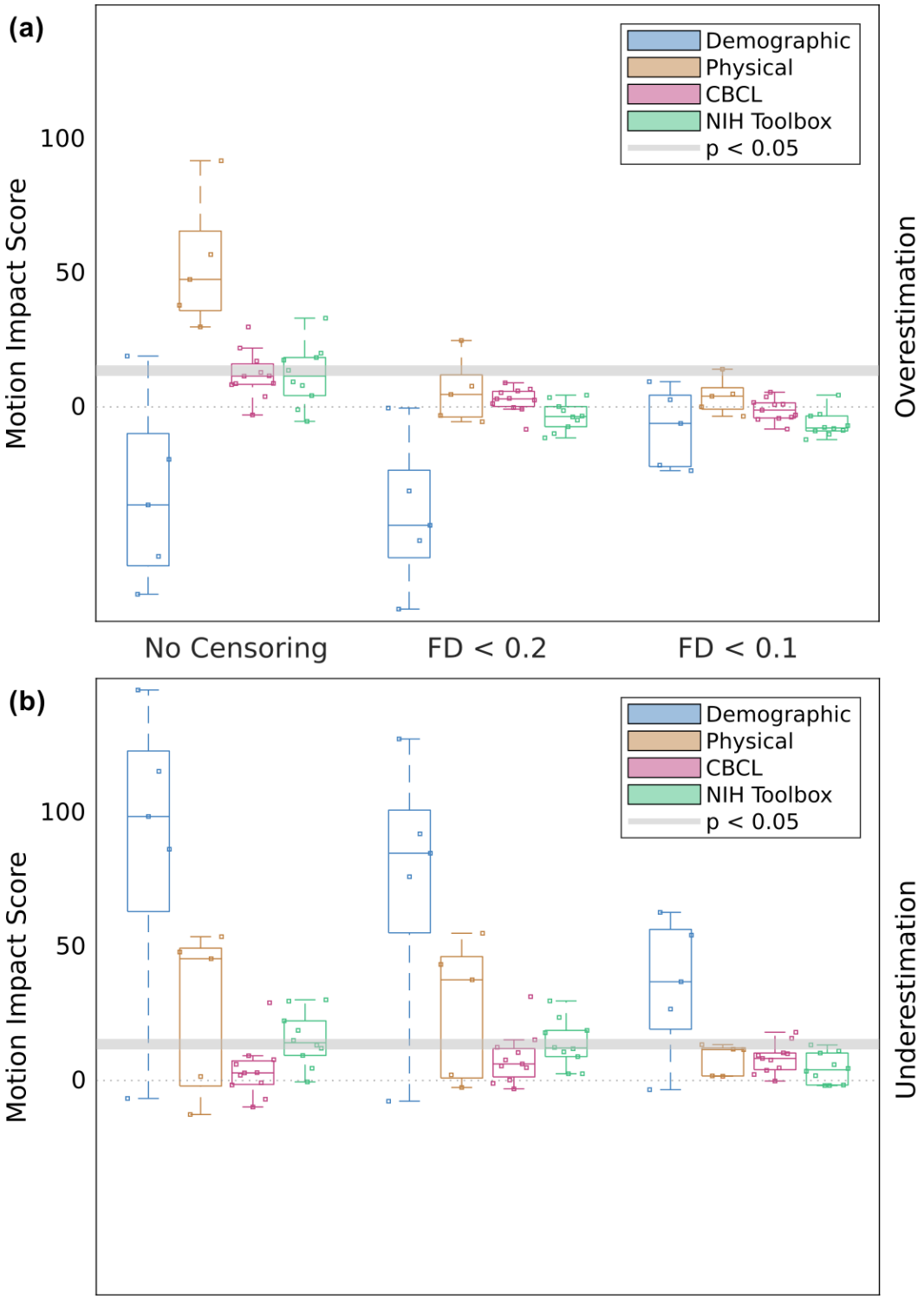

Figure 5: Effects of frame censoring on motion impact score. Motion impact score (omnibus Stouffer's Z, higher = more motion) for traits of a given category at different levels of motion censoring. The cutoff score for significance at $p < 0.05$ is different for each trait. The range of significance thresholds is indicated by a gray band. **(a)** Motion overestimation score. **(b)** Motion underestimation score. Categories are 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. Boxplots indicate the minimum, 25th percentile, median, 75th percentile, and maximum motion impact score.

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 motion overestimation 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 S6). For BMI, 39% (155/394) of regions contributed to motion overestimation score whereas < 1% of regions contributed to motion underestimation score. For matrix reasoning, 19% (76/394) of regions contributed to motion overestimation score; the whole-brain motion overestimation score for matrix reasoning was not significant.

Trait-FD Correlation Did Not Predict Motion Overestimation Score

Calculating motion impact score is computationally intensive and might be unnecessary if overestimation of spurious trait-FC effects could be detected using the trait-motion (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 motion overestimation score ($r = -0.05$, $p = 0.74$) and therefore did not detect spurious trait-FC correlations without the aid of SHAMAN (Figure S7). This was true even when limiting analysis to behavioral traits, which tend to be of greatest interest to researchers (33 traits, $r = 0.05$, $p = 0.74$). Large trait-FD correlation was correlated with motion underestimation score ($r = 0.36$, $p = 0.015$).

Data Quantity Affected Motion Impact Score

Motion impact scores without motion censoring were computed for 7,270 participants, of whom 1,291 did not complete all four 5-minute-long resting-state fMRI scans. The number of fMRI frames before censoring was not related to FD overall (Pearson $r = -0.013$, Spearman $r = 0.155$, $p = 0.20$), but there was a significant negative association when restricting analysis to participants with mean FD > 1.0 mm (Pearson $r = -0.237$, Spearman $r = -0.188$, $p < 0.001$). Thus, participants with very high amounts of motion had less fMRI data even before motion censoring. A supplementary analysis clamping the number of frames/timepoints to 600 (8 minutes of data) per participant was performed to test whether the ability to detect significant motion impact scores was dependent on participants having unequal amounts of rs-fMRI data. The results are shown in Table S5. Clamping the number of timepoints to 600, without motion censoring or exclusion of participants, reduced the number of traits with significant motion overestimation scores from 42% (19/45) to 20% (9/45). All 5 of the physical traits still had significant motion overestimation scores after clamping. Clamping the number of frames decreased the number of traits with significant motion underestimation scores from 38% (17/45) to 20% (9/45). Clamping the number of frames was not as effective as motion censoring at FD < 0.2 mm, which reduced the number of traits with significant motion overestimation scores to 2% (1/45).

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 S8). A sample size of about $n = 5,000$ participants was powered to detect a motion overestimation score for BMI in 88% of simulations at a significant threshold of ($\alpha = 0.05$). The motion underestimation score for matrix reasoning was more difficult to detect, with

successful detection in only 55% of simulations at n = 5,000. The need for thousands of participants to avoid false positive inference is consistent with recent fMRI studies.[48]

The motion blocks permutation method (Figure S12), which was used throughout this study, constrains the set of possible permutations to those permutations that preserve temporal continuity within the shuffled data blocks. The motion blocks permutation method thus accounts for weak autocorrelations in the fMRI timeseries at the cost of statistical power. A supplementary power analysis using unconstrained permutation was less conservative, detecting motion overestimation score in 89% of simulations at n = 2,000 and in 97% of simulations at n = 3,000. Unconstrained permutation detected motion underestimation score in 85% of simulations at n = 5,000.

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 effects are not biased by residual motion artifact. The randomly-assigned participant ID was used as a trait to simulate a trait-FC effect 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. The conservative, motion blocks permutation method did not falsely detect a significant motion impact score at any sample size (Figure S8). The less conservative, unconstrained permutation method falsely detected a motion overestimation score in 4.7% of simulations at n = 100 participants, par for expectation with a significance threshold of α = 0.05. 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 effects 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 effects 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 effects for 45 demographic, biophysical, intelligence, and personality traits within the ABCD study. Similar to prior findings in the Human Connectome Project (HCP) data, [18] we find that, even after denoising with ABCD-BIDS a large proportion (42%) of trait-FC effects 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 Impact May Arise from Heterogeneous Sources

The SHAMAN method distinguishes between transient (state) and persistent (trait) changes in fMRI signal related to a nuisance timeseries (head motion), but it does not distinguish between different biophysical sources of transient fMRI signal. Such sources could include not only artifactual signal from proton spin effects, but also fMRI activation in motor cortex, and even the transition between different FC states [56] . While these latter sources of fMRI signal might not be considered artifactual *per se*, all such sources of motion-related transient fMRI signal change have the potential to affect reproducibility in the context of a BWAS if motion-related differences in FC between participants are not removed or controlled for.

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

We find that, after denoising, the motion-FC effect matrix has a large effect size (max |r| = 0.10) and is strongly anti-correlated with the between-participant average FC matrix (r = -0.63), thus potentially impacting the interpretation of many traits whose FC effects are shaped by the average FC 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. We also

observe that participants with very high motion (mean FD > 1.0 mm) completed fewer fMRI scans leading to a shorter duration of fMRI data, higher sampling variability, and thus lower SNR. 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.

As previously shown, other factors contributing to the observed motion-FC effect matrix may include distance-dependent attenuation of connectivity by motion artifact. [40] Propensity of an individual to move is itself a stable, heritable trait,[57–59] and some [60] have even proposed that intrinsic differences in FC related to movement are encoded in the FC 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 (42%) 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 motion underestimation scores 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 motion-FC effecty 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 the Risk of False Positive Inference

Motion censoring at FD < 0.2 mm (filtered for respiratory artifact) is very effective at mitigating motion overestimation score (Figure 5, Tables S2, S3) and results in relatively little selection bias due to participant exclusion (Figure S4, S5, Table S4) in this neurotypical cohort. Siegel et al. [18] also observed decreased residual impact of motion after framewise censoring in the HCP data. Even more stringent censoring at FD < 0.1 (filtered) does not further reduce motion overestimation score, has little effect on motion underestimation score, and introduces greater selection bias, for example, reversing the ratio of boys to girls in the censored data compared to the uncensored data. The degree of selection bias would likely be even greater for children with developmental disorders such as autism [31]. Therefore, motion censoring at FD < 0.2 mm (filtered) is a reasonable strategy to avoid making false positive inferences about trait-FC effects.

The potential for selection bias [31,40] and loss of power from motion censoring might be further addressed through a combination of approaches including: recruiting more participants from high-motion groups (e.g. children with autism), scanning them for longer, adopting behavioral strategies to reduce in-scanner head motion,[22] employing real-time adaptive quality assurance strategies such as FIRMM [5] to ensure collection of sufficient data to maintain the desired power after frame censoring, adaptively selecting an optimal motion censoring threshold [41], and employing statistical methods to account for selection bias [31].

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 effects (motion underestimation score) 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, a subset of nodes, or the whole brain, reflecting a continuum [43] from high specificity (at the level of the ROI) and low power to low specificity and high power (across the whole brain) to detect motion impact. ROI analysis of motion impact score can provide insight into whether a priori regions of interest are impacted by residual motion artifact, and facilitate 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, [48] we show that the motion-FC effect is similar in magnitude to the effect size of most trait-FC effects. Therefore, a significant motion overestimation 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 (FD) 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 effects? Traits highly correlated with motion were more likely to have significant motion underestimation scores. 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. [48] showed that thousands of participants are needed to measure trait-FC effects reproducibly. Similarly, we find that 2,000-3,000 participants are needed to detect the motion overestimation score on the BMI-FC effect, and 6,000 participants are needed to detect the motion underestimation score on the matrix reasoning-FC effect (Figure S8). We postulate that motion underestimation scores require larger sample sizes for detection because affected trait-FC effects would tend to 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 effects for biophysical traits (e.g. BMI) because they are 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

Ethics

This project used resting-state functional MRI, demographic, biophysical, and behavioral data from 11,572 9–10 year old participants from the ABCD 2.0 release. [61] The ABCD Study obtained centralized institutional review board (IRB) approval from the University of California, San Diego. Each of the 21 sites also obtained local IRB approval. Ethical regulations were followed during data collection and analysis. Parents or caregivers provided written informed consent, and children gave written assent.

Behavioral

The Adolescent Brain Cognitive Development (ABCD) study participants are well-phenotyped with demographic, physical, mental health [46], and cognitive [47] 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)[51], 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 [52]. 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 [51], 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 [62], 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 [39]. 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 [38] described by Fair et al [25,37]. Processing dependencies include FSL [63] and FreeSurfer [64].

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. [65] augmented by the 61 subcortical parcels described by Seitzman et al [66] 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 [33,34] because it is the default metric used in the ABCD-BIDS processing pipeline [25,37].

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.” [40] 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 Pearson linear correlation (r) between FD and fMRI signal variance across participants was computed before and after ABCD-BIDS. Spearman correlation was also computed to relax the assumption of linearity. With each correlation measure, 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 [67]. 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 S4). The 8 minute cutoff was selected due the difficulty of estimating stable values for functional connectivity with shorter scan duration [62].

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 Figure S9 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, effect, 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 [68]. 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 [58], 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 (see Figure S13 for distribution of median FD) 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 S10 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 [42].

The permutation scheme is diagrammed in Figure S11 and S12 and described in greater detail in the following section. In Step 2 of the algorithm (Figure S11b) 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 S11e).

Once a null distribution is generated, we use non-parametric combining [42] 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 [54,55] 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.

Motion Blocks Permutation

Functional MRI timeseries are known to be weakly non-stationary and autocorrelated [69], therefore the assumption of exchangeability and unconstrained permutation of timepoints may not be appropriate. Specifically, consecutive timepoints could be randomly assigned to opposite halves of the data, making it more likely that the randomly (during permutation) split halves of the timeseries data would be similar. This would have the effect of making the difference between the randomly split halves smaller, thereby shrinking the null distribution toward zero and overestimating the significance of the motion impact score.

The problem of inference on weakly-dependent data is well-known in the statistical literature. Politis & Romano [70] describe the block and stationary bootstrap procedures for use on weakly-dependent data. Essentially, rather than resample individual timepoints with replacement, the Politis & Romano procedure resamples blocks of consecutive timepoints with replacement, where the block lengths are drawn from a geometric distribution. However, resampling the data into randomly-sized blocks makes it difficult to optimally separate the high- and low-motion halves of the data, since the randomly-sized blocks might not conform to the natural divisions of the motion timeseries.

Therefore, we devised a scheme we call “motion blocks” permutation that allows for optimal separation of high- and low-motion data while accounting for weak temporal dependency within the timeseries. The motion blocks permutation scheme is depicted in Figure S12. For the un-permuted split, the data are divided into high- and low-motion halves as usual, resulting in “blocks” of consecutive high- and low-motion data points. For the

permuted split, the blocks, rather than individual data points, are randomly assigned to each half of the data. This strategy preserves the same amount of weak temporal dependence in the permuted data as in the un-permuted data. Performance of the “motion blocks” and “unconstrained” permutation method are compared in Tables S6-8 and Figure S8. The motion blocks method was more conservative but had lower power.

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 FC effect 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.

Motion Overestimation vs Underestimation Score

The main objective of SHAMAN is to detect how motion artifact impacts trait-FC effect for specific traits. When motion impact has the same sign as the trait-FC effect (both positive or both negative) then motion causes overestimation of the size of the trait-FC effect and would tend to cause false positive inference (concluding the trait-FC effect is significant it is not). When motion impact has the opposite sign of the trait-FC effect then motion causes underestimation of the size of the trait-FC effect and would tend to cause false negative inference (failing to detect a significant trait-FC effect when it really is significant). We can then compute separate motion impact scores, a motion overestimation score and a motion underestimation score, by counting only those edges whose motion impact is the same (overestimation) or opposite (underestimation) sign as the trait-FC effect. Distinguishing between over- and under-estimation requires us to know with some degree of certainty whether the trait-FC effect is positive or negative. We use a heuristic approach in which the trait-FC effect 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 S6. 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 S6 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.

Data Availability

Participant level data from ABCD are openly available pursuant to consortium-level data access rules. The ABCD data repository grows and changes over time (https://nda.nih.gov/abcd). The ABCD data used in this study came from ABCD Annual Release 2.0 (https://doi.org/10.15154/1503209).

Code Availability

Analysis code specific to this study can be found at https://github.com/DosenbachGreene/shaman. Code for processing ABCD and UKB data can be found at https://github.com/DCAN-Labs/abcd-hcp-pipeline.

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. Nebel MB, Lidstone DE, Wang L, Benkeser D, Mostofsky SH, Risk BB. Accounting for motion in resting-state fMRI: What part of the spectrum are we characterizing in autism spectrum disorder? *NeuroImage*. 2022;257:119296. doi:10.1016/j.neuroimage.2022.119296
32. 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
33. 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
34. 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
35. 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
36. 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
37. 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
38. 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
39. 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
40. 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
41. Phạm DĐ, McDonald DJ, Ding L, Nebel MB, Mejia AF. Less is more: balancing noise reduction and data retention in fMRI with data-driven scrubbing. *NeuroImage*. 2023;270:119972. doi:10.1016/j.neuroimage.2023.119972
42. 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
43. Noble S, Mejia AF, Zalesky A, Scheinost D. Improving power in functional magnetic resonance imaging by moving beyond cluster-level inference. *Proc Natl Acad Sci*. 2022;119(32):e2203020119. doi:10.1073/pnas.2203020119
44. Jernigan TL, Brown SA. Introduction. *Dev Cogn Neurosci*. 2018;32:1-3. doi:10.1016/j.dcn.2018.02.002
45. 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
46. 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
47. 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
48. 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
49. Elam JS, Glasser MF, Harms MP, et al. The Human Connectome Project: A retrospective. *NeuroImage*. 2021;244:118543. doi:10.1016/j.neuroimage.2021.118543
50. Van Essen DC, Smith SM, Barch DM, Behrens TEJ, Yacoub E, Ugurbil K. The WU-Minn Human Connectome Project: An overview. *NeuroImage*. 2013;80:62-79. doi:10.1016/j.neuroimage.2013.05.041
51. Wechsler D, Pearson Education I, Psychological Corporation. *WISC-V: Wechsler Intelligence Scale for Children*. NCS Pearson, Inc. : PsychCorp; 2014.
52. 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.

53. 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
54. 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.
55. van Zwet WR, Oosterhoff J. On the Combination of Independent Test Statistics. *Ann Math Stat*. 1967;38(3):659-680.
56. Abreu R, Leal A, Figueiredo P. Identification of epileptic brain states by dynamic functional connectivity analysis of simultaneous EEG-fMRI: a dictionary learning approach. *Sci Rep*. 2019;9(1):638. doi:10.1038/s41598-018-36976-y
57. 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
58. 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
59. 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
60. 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
61. Jernigan T, Jernigan T. Adolescent Brain Cognitive Development Study (ABCD) - Annual Release 2.0. Published online 2019. doi:10.15154/1503209
62. 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
63. 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
64. Fischl B. Automatically Parcellating the Human Cerebral Cortex. *Cereb Cortex*. 2004;14(1):11-22. doi:10.1093/cercor/bhg087
65. 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
66. 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
67. 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
68. 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
69. Guan S, Jiang R, Bian H, et al. The Profiles of Non-stationarity and Non-linearity in the Time Series of Resting-State Brain Networks. *Front Neurosci*. 2020;14:493. doi:10.3389/fnins.2020.00493
70. Politis DN, Romano JP. The Stationary Bootstrap. *J Am Stat Assoc*. 1994;89(428):1303-1313. doi:10.1080/01621459.1994.10476870
71. 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
72. Smyser CD, Snyder AZ, Neil JJ. Functional connectivity MRI in infants: Exploration of the functional organization of the developing brain. *NeuroImage*. 2011;56(3):1437-1452.

doi:10.1016/j.neuroimage.2011.02.073
73. Afyouni S, Nichols TE. Insight and inference for DVARS. *NeuroImage*. 2018;172:291-312. doi:10.1016/j.neuroimage.2017.12.098
74. Smith SM, Beckmann CF, Andersson J, et al. Resting-state fMRI in the Human Connectome Project. *NeuroImage*. 2013;80:144-168. doi:10.1016/j.neuroimage.2013.05.039

Supplemental Tables

Variable	Correlation w/FD		SHAMAN p-Value			Shaman Motion Score		
	Pearson	Spearman	Impact	Over	Under	Impact	Over	Under
male	0.102	0.117	0	0	0	108.00	37.95	47.95
age_months	-0.121	-0.153	0	0	0.99	31.74	47.58	-12.67
height	-0.047	-0.047	0.002	0.001	0.574	24.28	29.87	1.50
weight	0.054	0.076	0	0	0	100.21	56.88	45.45
bmi	0.09	0.135	0	0	0	147.29	91.86	53.67
home_roster	-0.032	-0.049	0.121	0.016	0.946	9.31	18.99	-6.69
parental_income	-0.114	-0.118	0	1	0	128.00	-69.87	145.70
parents_married	-0.071	-0.097	0	1	0	90.48	-36.56	98.52
parental_education	-0.096	-0.11	0	1	0	97.61	-55.74	115.37
residential_deprivation	0.08	0.092	0	0.999	0	84.44	-19.55	86.29
mri_runs	-0.095	0.009	0.004	0.007	0.063	27.19	23.39	15.52
sleepy	0.022	0.032	0.271	0.14	0.62	7.90	13.40	-1.02
nihtbx_total	-0.11	-0.128	0	0.013	0.001	26.24	17.50	22.28
nihtbx_cryst	-0.062	-0.068	0	0.037	0	28.17	13.68	29.62
nihtbx_fluid	-0.111	-0.133	0.012	0.1	0.014	15.99	9.37	14.96
nihtbx_picvocab	-0.051	-0.054	0.039	0.71	0.013	13.13	-1.00	18.71
nihtbx_flanker	-0.045	-0.052	0.068	0.149	0.106	10.69	8.00	9.34
nihtbx_list	-0.053	-0.056	0.815	0.97	0.745	-1.78	-5.37	-0.52
nihtbx_cardsort	-0.085	-0.102	0.327	0.354	0.342	4.74	4.25	4.57
nihtbx_pattern	-0.089	-0.116	0.002	0.003	0.023	18.32	18.45	13.21
nihtbx_picture	-0.073	-0.091	0.003	0.002	0.056	18.89	20.08	12.04
nihtbx_reading	-0.057	-0.066	0	0	0	37.10	33.13	30.11
wiscv_matrix	-0.124	-0.119	0	0.898	0	45.82	-6.84	54.99
cbcl_internal	0.021	0.011	0.321	0.137	0.797	4.61	8.33	-1.67
cbcl_external	0.088	0.073	0.086	0.097	0.253	9.72	8.70	6.06
cbcl_total	0.089	0.087	0.013	0	0.537	14.84	21.97	1.98
cbcl_anxdep	0.01	-0.006	0.16	0.058	0.436	8.20	11.52	2.83
cbcl_withdep	0.059	0.054	0.001	0.001	0.119	25.16	29.85	9.25
cbcl_somatic	0	-0.025	0.976	0.874	0.999	-7.09	-2.97	-9.89
cbcl_social	0.097	0.086	0.098	0.021	0.44	10.99	17.11	2.94
cbcl_thought	0.063	0.056	0.201	0.042	0.738	6.61	12.90	-0.86
cbcl_attention	0.116	0.12	0.76	0.385	0.975	-0.99	3.88	-6.98
cbcl_rulebreak	0.089	0.097	0	0.06	0.001	25.93	11.64	29.03
cbcl_aggressive	0.068	0.063	0.113	0.115	0.167	9.64	8.81	7.77
pps_yes_num	0.072	0.111	0.001	0.001	0.993	19.65	34.78	-18.13
pps_severity	0.068	0.111	0.15	0.035	0.973	8.49	19.48	-12.92
bis	0.007	0.015	0.51	0.19	0.904	2.23	6.77	-3.57
bas_rr	0.055	0.07	0.861	0.876	0.791	-2.68	-3.15	-1.72
bas_drive	0.095	0.078	0.02	0.682	0.009	15.19	-0.89	21.61
bas_fs	0.068	0.062	0.042	0.067	0.047	11.62	10.22	11.59
upps_perserverance	0.061	0.054	0.205	0.051	0.636	7.45	11.42	0.39
upps_planning	0.077	0.052	0.668	0.321	0.955	0.31	4.47	-5.34
upps_seeking	0.005	0.006	0.247	0.225	0.518	5.76	6.01	1.99
upps_pos_urgency	0.09	0.087	0.381	0.775	0.321	4.11	-2.70	5.11
upps_neg_urgency	0.073	0.072	0.053	0.01	0.74	11.66	17.43	-0.68

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 overestimation scores, underestimation scores, and impact from both/either type of score. Results in this table reflect data after motion-reduction with ABCD-BIDS and with*out* additional motion censoring.

Variable	SHAMAN p-Value								
	No Censoring			FD < 0.2 mm			FD < 0.1 mm		
	Impact	Over	Under	Impact	Over	Under	Impact	Over	Under
male	0	0	0	0	0.731	0	0.041	0.577	0.087
age_months	0	0	0.99	0.348	0.315	0.491	0.401	0.339	0.492
height	0.002	0.001	0.574	0.012	0	0.768	0.085	0.03	0.514
weight	0	0	0	0	0.243	0	0.022	0.35	0.116
bmi	0	0	0	0	0.828	0	0.108	0.769	0.108
home_roster	0.121	0.016	0.946	0.873	0.649	0.955	0.396	0.148	0.8
parental_income	0	1	0	0	1	0	0	0.99	0
parents_married	0	1	0	0	1	0	0	0.461	0.02
parental_education	0	1	0	0	1	0	0	0.779	0.001
residential_deprivation	0	0.999	0	0	1	0	0	0.996	0
mri_runs	0.004	0.007	0.063	0.014	0.129	0.151	0.083	0.679	0.029
sleepy	0.271	0.14	0.62	0.265	0.584	0.086	0.659	0.733	0.423
nihtbx_total	0	0.013	0.001	0.311	1	0.011	0.9	0.999	0.39
nihtbx_cryst	0	0.037	0	0.004	0.389	0	0.228	0.837	0.04
nihtbx_fluid	0.012	0.1	0.014	0.376	0.997	0.032	0.81	0.989	0.52
nihtbx_picvocab	0.039	0.71	0.013	0.024	0.591	0.006	0.328	0.799	0.096
nihtbx_flanker	0.068	0.149	0.106	0.203	0.745	0.072	0.901	0.977	0.756
nihtbx_list	0.815	0.97	0.745	0.709	0.979	0.451	0.949	0.994	0.775
nihtbx_cardsort	0.327	0.354	0.342	0.188	0.867	0.058	0.72	0.976	0.24
nihtbx_pattern	0.002	0.003	0.023	0.415	0.902	0.12	0.158	0.319	0.081
nihtbx_picture	0.003	0.002	0.056	0.598	0.832	0.459	0.895	0.989	0.74
nihtbx_reading	0	0	0	0.024	0.324	0.003	0.657	0.962	0.331
wiscv_matrix	0	0.898	0	0	1	0	0.272	0.985	0.033
cbcl_internal	0.321	0.137	0.797	0.547	0.544	0.721	0.491	0.514	0.467
cbcl_external	0.086	0.097	0.253	0.082	0.409	0.037	0.315	0.887	0.109
cbcl_total	0.013	0	0.537	0.265	0.267	0.252	0.368	0.704	0.143
cbcl_anxdep	0.16	0.058	0.436	0.1	0.11	0.147	0.372	0.381	0.379
cbcl_withdep	0.001	0.001	0.119	0.429	0.39	0.61	0.202	0.281	0.187
cbcl_somatic	0.976	0.874	0.999	0.753	0.646	0.841	0.621	0.556	0.631
cbcl_social	0.098	0.021	0.44	0.081	0.231	0.078	0.589	0.85	0.315
cbcl_thought	0.201	0.042	0.738	0.412	0.709	0.217	0.237	0.569	0.081
cbcl_attention	0.76	0.385	0.975	0.74	0.973	0.323	0.527	0.964	0.107
cbcl_rulebreak	0	0.06	0.001	0	0.247	0	0.138	0.805	0.027
cbcl_aggressive	0.113	0.115	0.167	0.062	0.439	0.021	0.108	0.799	0.019
pps_yes_num	0.001	0.001	0.993	0.033	0.107	0.552	0	0	0.388
pps_severity	0.15	0.035	0.973	0.371	0.552	0.475	0.009	0.016	0.212
bis	0.51	0.19	0.904	0.815	0.678	0.84	0.445	0.481	0.416
bas_rr	0.861	0.876	0.791	0.442	0.952	0.129	0.201	0.677	0.063
bas_drive	0.02	0.682	0.009	0.088	0.739	0.032	0.249	0.421	0.289
bas_fs	0.042	0.067	0.047	0.253	0.545	0.087	0.25	0.426	0.168
upps_perserverance	0.205	0.051	0.636	0.416	0.372	0.454	0.957	0.982	0.874
upps_planning	0.668	0.321	0.955	0.11	0.08	0.145	0.377	0.45	0.269
upps_seeking	0.247	0.225	0.518	0.316	0.559	0.219	0.871	0.983	0.681
upps_pos_urgency	0.381	0.775	0.321	0.167	0.802	0.127	0.827	0.9	0.719
upps_neg_urgency	0.053	0.01	0.74	0.046	0.089	0.089	0.555	0.688	0.525

Table S2: Significance (p-values) of motion impact score 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. Omnibus p-values are given for overestimation scores, underestimation scores, and impact from both/either type of score. The omnibus p-values account for multiple comparisons across edges, but they are not corrected for the multiple comparisons across the 45 different traits shown.

Variable	SHAMAN Motion Score								
	No Censoring			FD < 0.2 mm			FD < 0.1 mm		
	Impact	Over	Under	Impact	Over	Under	Impact	Over	Under
male	108.0	38.0	47.9	48.8	-3.2	43.3	13.8	0.0	13.4
age_months	31.7	47.6	-12.7	4.1	4.7	2.1	3.0	4.0	1.7
height	24.3	29.9	1.5	17.9	24.8	-2.6	10.1	14.1	1.6
weight	100.2	56.9	45.5	53.9	7.8	37.6	15.5	4.9	11.7
bmi	147.3	91.9	53.7	64.7	-5.5	54.9	9.4	-3.5	11.5
home_roster	9.3	19.0	-6.7	-4.2	-0.4	-7.7	3.4	9.4	-3.4
parental_income	128.0	-69.9	145.7	106.6	-75.4	127.4	56.3	-21.7	62.7
parents_married	90.5	-36.6	98.5	72.9	-31.3	76.0	35.4	2.7	26.7
parental_education	97.6	-55.7	115.4	73.4	-49.9	92.0	33.1	-6.1	36.9
residential_deprivation	84.4	-19.6	86.3	65.7	-44.2	84.8	39.7	-23.8	54.2
mri_runs	27.2	23.4	15.5	15.4	11.1	9.9	10.8	-1.5	16.4
sleepy	7.9	13.4	-1.0	5.3	0.7	10.0	-0.6	-1.8	2.6
nihtbx_total	26.2	17.5	22.3	4.5	-11.5	17.8	-5.2	-12.2	3.5
nihtbx_cryst	28.2	13.7	29.6	22.6	3.5	29.7	6.3	-3.4	13.3
nihtbx_fluid	16.0	9.4	15.0	3.6	-9.9	12.4	-2.9	-8.9	1.8
nihtbx_picvocab	13.1	-1.0	18.7	15.5	0.2	23.6	4.5	-2.7	10.2
nihtbx_flanker	10.7	8.0	9.3	6.8	-1.3	10.7	-5.1	-7.6	-1.9
nihtbx_list	-1.8	-5.4	-0.5	-1.0	-7.4	2.5	-6.3	-10.2	-1.9
nihtbx_cardsort	4.7	4.2	4.6	7.0	-3.7	11.9	-1.4	-8.0	5.9
nihtbx_pattern	18.3	18.5	13.2	3.2	-4.9	8.9	8.3	4.4	11.0
nihtbx_picture	18.9	20.1	12.0	0.2	-3.4	2.5	-5.2	-8.7	-1.7
nihtbx_reading	37.1	33.1	30.1	14.7	4.4	18.7	-0.4	-7.0	4.5
wiscv_matrix	45.8	-6.8	55.0	26.7	-22.9	44.2	5.5	-12.1	16.9
cbcl_internal	4.6	8.3	-1.7	1.2	1.3	-1.1	1.9	1.7	2.2
cbcl_external	9.7	8.7	6.1	9.9	3.0	12.4	4.5	-4.6	9.4
cbcl_total	14.8	22.0	2.0	5.3	5.3	5.4	3.6	-1.2	8.3
cbcl_anxdep	8.2	11.5	2.8	9.6	9.0	7.7	3.9	3.8	3.8
cbcl_withdep	25.2	29.8	9.3	2.5	3.2	0.2	7.1	5.5	7.7
cbcl_somatic	-7.1	-3.0	-9.9	-1.7	-0.2	-3.1	-0.1	0.9	-0.2
cbcl_social	11.0	17.1	2.9	9.7	6.0	10.4	0.4	-4.2	4.7
cbcl_thought	6.6	12.9	-0.9	2.9	-0.8	6.2	6.1	1.0	10.3
cbcl_attention	-1.0	3.9	-7.0	-1.4	-8.3	4.8	1.5	-8.2	10.0
cbcl_rulebreak	25.9	11.6	29.0	26.2	6.7	31.3	8.7	-3.8	15.8
cbcl_aggressive	9.6	8.8	7.8	11.2	2.6	15.2	10.4	-3.0	18.0
pps_yes_num	19.6	34.8	-18.1	13.6	13.4	0.3	30.8	34.0	3.9
pps_severity	8.5	19.5	-12.9	3.7	0.5	2.2	20.9	20.9	8.6
bis	2.2	6.8	-3.6	-2.8	-0.6	-3.0	2.5	2.2	3.0
bas_rr	-2.7	-3.1	-1.7	2.5	-6.6	8.9	6.9	-0.9	11.3
bas_drive	15.2	-0.9	21.6	10.2	-3.0	17.5	5.8	3.2	5.8
bas_fs	11.6	10.2	11.6	5.7	1.0	10.5	5.9	2.9	7.4
upps_perserverance	7.4	11.4	0.4	2.8	3.4	2.3	-6.7	-8.3	-4.1
upps_planning	0.3	4.5	-5.3	9.3	10.0	7.9	3.6	2.6	5.6
upps_seeking	5.8	6.0	2.0	4.4	1.1	5.9	-4.1	-8.1	-0.8
upps_pos_urgency	4.1	-2.7	5.1	7.1	-3.7	10.1	-3.1	-6.2	-2.0
upps_neg_urgency	11.7	17.4	-0.7	11.6	9.8	10.1	0.7	-1.2	1.2

Table S3: Motion impact scores (omnibus Stouffer's Z) 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. Higher scores correspond to a greater impact of motion. Motion overestimation scores, underestimation scores, and overall impact (both/either over/underestimation) are shown separately.

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.

	No Motion Censoring											
Variable	Impact (Over or Under)				Overestimation				Underestimation			
	p-Value		Motion Score		p-Value		Motion Score		p-Value		Motion Score	
Number of Frames	All	600	All	600	All	600	All	600	All	600	All	600
male	0	0	108	109.13	0	0	38	46.581	0	0.0313	47.9	38.213
age_months	0	0.008	31.7	36.469	0	0	47.6	54.339	0.99	1	-12.7	-14.04
height	0.002	0.039	24.3	30.613	0.001	0	29.9	32.815	0.574	0.9297	1.5	7.3372
weight	0	0	100.2	123.43	0	0	56.9	71.582	0	0	45.5	58.992
bmi	0	0	147.3	170.56	0	0	91.9	111.48	0	0	53.7	62.503
home_roster	0.121	0.594	9.3	15.562	0.016	0.2188	19	22.867	0.946	1	-6.7	0.701
parental_income	0	0	128	128.83	1	1	-69.9	-73.48	0	0	145.7	151.69
parents_married	0	0	90.5	96.469	1	1	-36.6	-35.19	0	0	98.5	107.14
parental_education	0	0	97.6	106.33	1	1	-55.7	-61.43	0	0	115.4	128.84
residential_deprivation	0	0	84.4	86.392	0.999	1	-19.6	-24.25	0	0	86.3	91.359
mri_runs	0.004	0	27.2	58.324	0.007	0.0078	23.4	39.126	0.063	0	15.5	44.473
sleepy	0.271	0.047	7.9	34.309	0.14	0	13.4	45.127	0.62	0.8203	-1	10.334
nihtbx_total	0	0.219	26.2	22.017	0.013	0.5547	17.5	15.496	0.001	0.3828	22.3	19.248
nihtbx_cryst	0	0.063	28.2	27.588	0.037	0.6172	13.7	15.224	0	0.0625	29.6	28.868
nihtbx_fluid	0.012	0.602	16	15.575	0.1	0.9688	9.4	7.6907	0.014	0.6797	15	14.328
nihtbx_picvocab	0.039	0.82	13.1	11.685	0.71	1	-1	1.3712	0.013	0.5156	18.7	16.774
nihtbx_flanker	0.068	0.688	10.7	15.111	0.149	0.8516	8	11.575	0.106	0.7031	9.3	14.245
nihtbx_list	0.815	1	-1.8	-1.353	0.97	1	-5.4	-5.946	0.745	1	-0.5	0.7405
nihtbx_cardsort	0.327	1	4.7	2.5511	0.354	0.9922	4.2	2.6785	0.342	1	4.6	0.9357
nihtbx_pattern	0.002	0.531	18.3	16.705	0.003	0.375	18.5	19.169	0.023	0.9609	13.2	8.5012
nihtbx_picture	0.003	0.844	18.9	11.538	0.002	0.8125	20.1	11.811	0.056	0.9375	12	7.4315
nihtbx_reading	0	0.016	37.1	35.877	0	0.0313	33.1	30.726	0	0.0625	30.1	29.305
wiscv_matrix	0	0	45.8	52.419	0.898	1	-6.8	-3.442	0	0	55	60.913
cbcl_internal	0.321	1	4.6	-1.28	0.137	0.9922	8.3	3.9331	0.797	1	-1.7	-9.307
cbcl_external	0.086	0.953	9.7	9.0023	0.097	0.9297	8.7	9.7443	0.253	1	6.1	4.3582
cbcl_total	0.013	0.898	14.8	11.031	0	0.3672	22	19.372	0.537	1	2	-2.834
cbcl_anxdep	0.16	0.914	8.2	8.1403	0.058	0.7266	11.5	13.189	0.436	1	2.8	0.7301
cbcl_withdep	0.001	0.289	25.2	21.419	0.001	0.0781	29.8	28.301	0.119	0.9609	9.3	5.5037
cbcl_somatic	0.976	1	-7.1	-10.36	0.874	1	-3	-4.277	0.999	1	-9.9	-15.4
cbcl_social	0.098	0.406	11	18.235	0.021	0.1563	17.1	25.154	0.44	0.9141	2.9	8.0428
cbcl_thought	0.201	0.688	6.6	14.571	0.042	0.3828	12.9	19.494	0.738	0.9844	-0.9	6.0049
cbcl_attention	0.76	0.656	-1	14.89	0.385	0.5938	3.9	15.823	0.975	0.875	-7	9.7579
cbcl_rulebreak	0	0.023	25.9	30.934	0.06	0.5938	11.6	15.945	0.001	0.0078	29	33.641
cbcl_aggressive	0.113	0.906	9.6	10.038	0.115	0.8516	8.8	11.488	0.167	0.9688	7.8	6.5077
pps_yes_num	0.001	0.133	19.6	24.755	0.001	0.0313	34.8	38.941	0.993	1	-18.1	-12.84
pps_severity	0.15	0.375	8.5	18.898	0.035	0.1328	19.5	26.885	0.973	0.9922	-12.9	-4.167
bis	0.51	1	2.2	-0.712	0.19	1	6.8	5.1135	0.904	1	-3.6	-6.548
bas_rr	0.861	1	-2.7	-3.386	0.876	1	-3.1	-3.34	0.791	1	-1.7	-3.479
bas_drive	0.02	0.555	15.2	16.863	0.682	1	-0.9	-3.937	0.009	0.0547	21.6	27.429
bas_fs	0.042	0.844	11.6	11.255	0.067	0.9141	10.2	9.9637	0.047	0.8594	11.6	10.819
upps_perserverance	0.205	0.984	7.4	5.0833	0.051	0.9453	11.4	7.4026	0.636	1	0.4	2.0256
upps_planning	0.668	0.992	0.3	5.9279	0.321	0.8984	4.5	9.5865	0.955	1	-5.3	0.9064
upps_seeking	0.247	1	5.8	-2.688	0.225	1	6	0.2879	0.518	1	2	-6.989
upps_pos_urgency	0.381	0.898	4.1	9.8227	0.775	0.9219	-2.7	7.2436	0.321	0.9844	5.1	5.2049
upps_neg_urgency	0.053	0.977	11.7	7.6357	0.01	0.7422	17.4	14.585	0.74	1	-0.7	-5.375

Table S5: Comparison between using all fMRI frames and clamping to 600 frames per participant. The 600 frames were chosen randomly without replacement. There was no motion censoring. Motion impact scores that were significant ($p < 0.05$) using all frames but became not-significant ($p > 0.05$) with clamping to 600 frames are highlighted in red. Motion impact scores that were significant only with clamping to 600 frames are highlighted in blue.

No Motion Censoring

Variable	Impact (Over or Under)				Overestimation				Underestimation			
	p-Value		Motion Score		p-Value		Motion Score		p-Value		Motion Score	
	Default	Block	Default	Block	Default	Block	Default	Block	Default	Block	Default	Block
male	0	0	93.9	108	0	0	43.2	38	0.016	0	24.1	47.9
age_months	0	0	32.5	31.7	0	0	57.2	47.6	1	0.99	-30.1	-12.7
height	0.001	0.002	22.3	24.3	0	0.001	30.9	29.9	0.779	0.574	-3.8	1.5
weight	0	0	88	100.2	0	0	57.7	56.9	0.001	0	28.3	45.5
bmi	0	0	118.8	147.3	0	0	91.2	91.9	0.038	0	16.6	53.7
home_roster	0.127	0.121	6.8	9.3	0.006	0.016	17.9	19	0.955	0.946	-9.2	-6.7
parental_income	0	0	108.5	128	1	1	-51.3	-69.9	0	0	119.3	145.7
parents_married	0	0	81.3	90.5	0.995	1	-28.2	-36.6	0	0	84.7	98.5
parental_education	0	0	82.3	97.6	1	1	-38.6	-55.7	0	0	94.4	115.4
residential_deprivation	0	0	78	84.4	0.802	0.999	-7.7	-19.6	0	0	70.4	86.3
mri_runs	0	0.004	30.2	27.2	0.014	0.007	21.9	23.4	0.049	0.063	16	15.5
sleepy	0.007	0.271	16.8	7.9	0.002	0.14	22.9	13.4	0.295	0.62	3.6	-1
nihtbx_total	0.001	0	20.6	26.2	0.006	0.013	21.6	17.5	0.416	0.001	2.4	22.3
nihtbx_cryst	0	0	21.3	28.2	0.017	0.037	15.5	13.7	0.021	0	16.1	29.6
nihtbx_fluid	0.009	0.012	16.4	16	0.041	0.1	13.9	9.4	0.228	0.014	5.7	15
nihtbx_picvocab	0.072	0.039	8.9	13.1	0.751	0.71	-3.3	-1	0.033	0.013	14.3	18.7
nihtbx_flanker	0.049	0.068	10.2	10.7	0.196	0.149	4.6	8	0.052	0.106	9.9	9.3
nihtbx_list	0.709	0.815	-2.1	-1.8	0.838	0.97	-4.2	-5.4	0.62	0.745	-1	-0.5
nihtbx_cardsort	0.29	0.327	3.7	4.7	0.554	0.354	-0.7	4.2	0.199	0.342	6	4.6
nihtbx_pattern	0	0.002	20	18.3	0	0.003	24.1	18.5	0.24	0.023	5.2	13.2
nihtbx_picture	0.006	0.003	14.9	18.9	0.002	0.002	23.3	20.1	0.736	0.056	-3	12
nihtbx_reading	0	0	27.9	37.1	0	0	32.3	33.1	0.057	0	10.3	30.1
wiscv_matrix	0	0	45.8	45.8	0.813	0.898	-8	-6.8	0	0	54.5	55
cbcl_internal	0.228	0.321	4.7	4.6	0.059	0.137	8.7	8.3	0.728	0.797	-2.4	-1.7
cbcl_external	0.012	0.086	14.8	9.7	0.026	0.097	13.4	8.7	0.082	0.253	9.6	6.1
cbcl_total	0.004	0.013	17.3	14.8	0	0	30.2	22	0.884	0.537	-6.9	2
cbcl_anxdep	0.099	0.16	7.8	8.2	0.037	0.058	11	11.5	0.34	0.436	2.8	2.8
cbcl_withdep	0	0.001	29.1	25.2	0	0.001	37.3	29.8	0.382	0.119	2.3	9.3
cbcl_somatic	0.807	0.976	-3.8	-7.1	0.549	0.874	-0.1	-3	0.945	0.999	-6.9	-9.9
cbcl_social	0	0.098	28	11	0	0.021	30.8	17.1	0.046	0.44	14	2.9
cbcl_thought	0.036	0.201	11.3	6.6	0.003	0.042	20.5	12.9	0.62	0.738	-1.6	-0.9
cbcl_attention	0.008	0.76	14.7	-1	0.041	0.385	16.7	3.9	0.345	0.975	3.9	-7
cbcl_rulebreak	0	0	36.2	25.9	0.12	0.06	9.7	11.6	0	0.001	41.7	29
cbcl_aggressive	0.007	0.113	16.7	9.6	0.031	0.115	13.8	8.8	0.038	0.167	12.9	7.8
pps_yes_num	0	0.001	23.9	19.6	0	0.001	37.6	34.8	0.963	0.993	-15.4	-18.1
pps_severity	0.021	0.15	12.4	8.5	0.021	0.035	20.7	19.5	0.863	0.973	-9.4	-12.9
bis	0.299	0.51	3.2	2.2	0.086	0.19	7.8	6.8	0.739	0.904	-2.5	-3.6
bas_rr	0.57	0.861	-0.4	-2.7	0.833	0.876	-4.3	-3.1	0.241	0.791	4.3	-1.7
bas_drive	0.002	0.02	21.7	15.2	0.915	0.682	-10.3	-0.9	0.001	0.009	36.8	21.6
bas_fs	0.004	0.042	16.3	11.6	0.142	0.067	7.2	10.2	0	0.047	20.4	11.6
upps_perserverance	0.07	0.205	8.7	7.4	0.011	0.051	17.2	11.4	0.833	0.636	-4.9	0.4
upps_planning	0.198	0.668	4.9	0.3	0.027	0.321	12.4	4.5	0.873	0.955	-5.4	-5.3
upps_seeking	0.142	0.247	6.2	5.8	0.14	0.225	6.2	6	0.316	0.518	2.9	2
upps_pos_urgency	0.048	0.381	9.9	4.1	0.44	0.775	1.6	-2.7	0.167	0.321	9.2	5.1
upps_neg_urgency	0.001	0.053	17.7	11.7	0	0.01	27	17.4	0.685	0.74	-2.6	-0.7

Table S6: Comparison between motion blocks and unconstrained permutation with no motion censoring. Motion impact scores that were significant ($p < 0.05$) with unconstrained permutation but became not-significant ($p > 0.05$) with motion blocks permutation are highlighted in red. Motion impact scores that were significant only with motion blocks permutation are highlighted in blue.

FD < 0.2

Motion Censoring, FD < 0.2 mm

Variable	Impact (Over or Under)				Overestimation				Underestimation			
	p-Value		Motion Score		p-Value		Motion Score		p-Value		Motion Score	
	Default	Block	Default	Block	Default	Block	Default	Block	Default	Block	Default	Block
male	0	0	45.2	48.8	0.679	0.731	-3.2	-3.2	0	0	38.1	43.3
age_months	0.096	0.348	8.3	4.1	0.487	0.315	0.4	4.7	0.067	0.491	11.5	2.1
height	0.001	0.012	17.8	17.9	0	0	24.1	24.8	0.678	0.768	-1.9	-2.6
weight	0	0	48.5	53.9	0.259	0.243	5.9	7.8	0	0	42.2	37.6
bmi	0	0	56.7	64.7	0.736	0.828	-4.8	-5.5	0	0	56.6	54.9
home_roster	0.847	0.873	-4.6	-4.2	0.582	0.649	-0.8	-0.4	0.919	0.955	-6.9	-7.7
parental_income	0	0	81.9	106.6	1	1	-57.5	-75.4	0	0	100.9	127.4
parents_married	0	0	58.2	72.9	0.999	1	-26.8	-31.3	0	0	62.2	76
parental_education	0	0	56	73.4	1	1	-38.2	-49.9	0	0	73.1	92
residential_deprivation	0	0	56.5	65.7	1	1	-37.1	-44.2	0	0	73	84.8
mri_runs	0.05	0.014	10.9	15.4	0.165	0.129	8.4	11.1	0.187	0.151	6.6	9.9
sleepy	0.234	0.265	4.5	5.3	0.536	0.584	0.2	0.7	0.08	0.086	8.9	10
nihtbx_total	0.228	0.311	4.5	4.5	0.994	1	-12	-11.5	0.004	0.011	18.1	17.8
nihtbx_cryst	0	0.004	22.3	22.6	0.233	0.389	5	3.5	0	0	28.3	29.7
nihtbx_fluid	0.093	0.376	8.1	3.6	0.997	0.997	-14.6	-9.9	0.002	0.032	20.7	12.4
nihtbx_picvocab	0.006	0.024	16.5	15.5	0.429	0.591	1.7	0.2	0	0.006	24	23.6
nihtbx_flanker	0.071	0.203	9.4	6.8	0.657	0.745	-1.8	-1.3	0.007	0.072	14.3	10.7
nihtbx_list	0.623	0.709	-1.1	-1	0.959	0.979	-7.6	-7.4	0.324	0.451	2.8	2.5
nihtbx_cardsort	0.063	0.188	9.8	7	0.911	0.867	-6.2	-3.7	0.006	0.058	17.3	11.9
nihtbx_pattern	0.213	0.415	4.8	3.2	0.947	0.902	-7.2	-4.9	0.019	0.12	13.3	8.9
nihtbx_picture	0.442	0.598	1.2	0.2	0.857	0.832	-4.7	-3.4	0.195	0.459	5.5	2.5
nihtbx_reading	0.017	0.024	13.8	14.7	0.207	0.324	4.9	4.4	0.001	0.003	17.5	18.7
wiscv_matrix	0	0	23.1	26.7	0.999	1	-21	-22.9	0	0	40.8	44.2
cbcl_internal	0.489	0.547	1.1	1.2	0.463	0.544	1.2	1.3	0.621	0.721	-1	-1.1
cbcl_external	0.045	0.082	10.4	9.9	0.437	0.409	1.2	3	0.008	0.037	15.1	12.4
cbcl_total	0.106	0.265	7.8	5.3	0.233	0.267	4.4	5.3	0.054	0.252	10.3	5.4
cbcl_anxdep	0.097	0.1	8.4	9.6	0.093	0.11	7.9	9	0.153	0.147	6.6	7.7
cbcl_withdep	0.224	0.429	4.9	2.5	0.252	0.39	4.4	3.2	0.321	0.61	3.2	0.2
cbcl_somatic	0.686	0.753	-2.1	-1.7	0.582	0.646	-0.5	-0.2	0.756	0.841	-3.2	-3.1
cbcl_social	0.036	0.081	11.8	9.7	0.301	0.231	3.9	6	0.027	0.078	14.8	10.4
cbcl_thought	0.216	0.412	5.1	2.9	0.646	0.709	-1.7	-0.8	0.064	0.217	10.8	6.2
cbcl_attention	0.326	0.74	3.2	-1.4	0.973	0.973	-10.7	-8.3	0.037	0.323	14	4.8
cbcl_rulebreak	0	0	25.4	26.2	0.235	0.247	5.1	6.7	0	0	31.5	31.3
cbcl_aggressive	0.03	0.062	12.1	11.2	0.523	0.439	-0.1	2.6	0.005	0.021	18.4	15.2
pps_yes_num	0.028	0.033	11.7	13.6	0.117	0.107	11.8	13.4	0.55	0.552	-0.8	0.3
pps_severity	0.492	0.371	0.8	3.7	0.581	0.552	-1.2	0.5	0.491	0.475	0.6	2.2
bis	0.872	0.815	-5	-2.8	0.738	0.678	-2.6	-0.6	0.881	0.84	-5	-3
bas_rr	0.331	0.442	2.8	2.5	0.929	0.952	-7.3	-6.6	0.077	0.129	9.9	8.9
bas_drive	0.052	0.088	9.7	10.2	0.815	0.739	-6.1	-3	0.017	0.032	19	17.5
bas_fs	0.2	0.253	5.3	5.7	0.525	0.545	-0.2	1	0.063	0.087	10.7	10.5
upps_perserverance	0.289	0.416	3.5	2.8	0.311	0.372	3.1	3.4	0.282	0.454	3.6	2.3
upps_planning	0.072	0.11	9	9.3	0.058	0.08	9.7	10	0.1	0.145	7.6	7.9
upps_seeking	0.383	0.316	1.9	4.4	0.601	0.559	-1	1.1	0.294	0.219	3.3	5.9
upps_pos_urgency	0.147	0.167	6.5	7.1	0.782	0.802	-5.2	-3.7	0.116	0.127	10.1	10.1
upps_neg_urgency	0.034	0.046	11.7	11.6	0.089	0.089	8.9	9.8	0.057	0.089	10.3	10.1

Table S7: Comparison between motion blocks and unconstrained permutation with censoring at FD < 0.2 mm. Motion impact scores that were significant ($p < 0.05$) with unconstrained permutation but became not-significant ($p > 0.05$) with motion blocks permutation are highlighted in red. Motion impact scores that were significant only with motion blocks permutation are highlighted in blue.

FD < 0.1

Motion Censoring, FD < 0.1 mm

Variable	Impact (Over or Under)				Overestimation				Underestimation			
	p-Value		Motion Score		p-Value		Motion Score		p-Value		Motion Score	
	Default	Block	Default	Block	Default	Block	Default	Block	Default	Block	Default	Block
male	0.035	0.041	12.5	13.8	0.531	0.577	0	0	0.077	0.087	12.1	13.4
age_months	0.338	0.401	3	3	0.363	0.339	2.5	4	0.314	0.492	3.4	1.7
height	0.037	0.085	11.2	10.1	0.014	0.03	14.7	14.1	0.353	0.514	2.7	1.6
weight	0.019	0.022	14.1	15.5	0.238	0.35	6.1	4.9	0.091	0.116	10.8	11.7
bmi	0.145	0.108	6.6	9.4	0.682	0.769	-3	-3.5	0.137	0.108	9.3	11.5
home_roster	0.411	0.396	1.8	3.4	0.132	0.148	8	9.4	0.818	0.8	-5.1	-3.4
parental_income	0	0	48.7	56.3	0.958	0.99	-16	-21.7	0	0	53	62.7
parents_married	0	0	30.9	35.4	0.418	0.461	3.2	2.7	0.031	0.02	22.2	26.7
parental_education	0	0	27.9	33.1	0.675	0.779	-3.8	-6.1	0.003	0.001	30.6	36.9
residential_deprivation	0	0	35.2	39.7	0.995	0.996	-21.5	-23.8	0	0	48.4	54.2
mri_runs	0.054	0.083	11	10.8	0.545	0.679	-0.4	-1.5	0.009	0.029	17.1	16.4
sleepy	0.604	0.659	-0.8	-0.6	0.679	0.733	-1.9	-1.8	0.367	0.423	2.4	2.6
nihtbx_total	0.853	0.9	-5.1	-5.2	0.997	0.999	-12.3	-12.2	0.306	0.39	3.6	3.5
nihtbx_cryst	0.165	0.228	6.5	6.3	0.76	0.837	-3.1	-3.4	0.029	0.04	13.3	13.3
nihtbx_fluid	0.606	0.81	-0.9	-2.9	0.983	0.989	-9	-8.9	0.224	0.52	4.7	1.8
nihtbx_picvocab	0.252	0.328	4.2	4.5	0.714	0.799	-2.7	-2.7	0.07	0.096	10.2	10.2
nihtbx_flanker	0.785	0.901	-3.8	-5.1	0.917	0.977	-6.5	-7.6	0.545	0.756	-0.3	-1.9
nihtbx_list	0.885	0.949	-5.2	-6.3	0.994	0.994	-9.5	-10.2	0.596	0.775	-0.9	-1.9
nihtbx_cardsort	0.464	0.72	1.1	-1.4	0.904	0.976	-6.2	-8	0.114	0.24	8	5.9
nihtbx_pattern	0.126	0.158	7.6	8.3	0.317	0.319	3	4.4	0.038	0.081	10.8	11
nihtbx_picture	0.808	0.895	-4.3	-5.2	0.977	0.989	-9.1	-8.7	0.523	0.74	0.2	-1.7
nihtbx_reading	0.603	0.657	-1.1	-0.4	0.94	0.962	-7.2	-7	0.298	0.331	3.7	4.5
wiscv_matrix	0.235	0.272	4.6	5.5	0.971	0.985	-11.2	-12.1	0.034	0.033	14.9	16.9
cbcl_internal	0.474	0.491	0.8	1.9	0.478	0.514	0.9	1.7	0.436	0.467	1.2	2.2
cbcl_external	0.288	0.315	3.6	4.5	0.864	0.887	-5.2	-4.6	0.078	0.109	8.5	9.4
cbcl_total	0.322	0.368	3.1	3.6	0.668	0.704	-1.9	-1.2	0.095	0.143	8	8.3
cbcl_anxdep	0.312	0.372	3.4	3.9	0.313	0.381	3.2	3.8	0.309	0.379	3.4	3.8
cbcl_withdep	0.135	0.202	7.1	7.1	0.221	0.281	5.2	5.5	0.122	0.187	7.7	7.7
cbcl_somatic	0.517	0.621	0.3	-0.1	0.424	0.556	1.5	0.9	0.535	0.631	0	-0.2
cbcl_social	0.532	0.589	-0.1	0.4	0.846	0.85	-5.2	-4.2	0.251	0.315	4.8	4.7
cbcl_thought	0.226	0.237	4.8	6.1	0.594	0.569	-0.7	1	0.07	0.081	9.7	10.3
cbcl_attention	0.401	0.527	2	1.5	0.937	0.964	-8.5	-8.2	0.056	0.107	11	10
cbcl_rulebreak	0.146	0.138	7.6	8.7	0.786	0.805	-4.8	-3.8	0.028	0.027	14.8	15.8
cbcl_aggressive	0.076	0.108	10.1	10.4	0.737	0.799	-3.6	-3	0.015	0.019	17.9	18
pps_yes_num	0	0	27.8	30.8	0	0	31.7	34	0.415	0.388	2	3.9
pps_severity	0.005	0.009	19.7	20.9	0.007	0.016	20.8	20.9	0.218	0.212	6.9	8.6
bis	0.438	0.445	1.3	2.5	0.455	0.481	1.1	2.2	0.372	0.416	2	3
bas_rr	0.16	0.201	6.7	6.9	0.567	0.677	-0.4	-0.9	0.044	0.063	10.8	11.3
bas_drive	0.202	0.249	5.1	5.8	0.405	0.421	2.2	3.2	0.234	0.289	5.7	5.8
bas_fs	0.215	0.25	5.1	5.9	0.366	0.426	2.3	2.9	0.137	0.168	6.5	7.4
upps_perserverance	0.913	0.957	-6.6	-6.7	0.957	0.982	-8.1	-8.3	0.799	0.874	-3.9	-4.1
upps_planning	0.337	0.377	2.7	3.6	0.413	0.45	1.6	2.6	0.214	0.269	4.7	5.6
upps_seeking	0.805	0.871	-3.9	-4.1	0.952	0.983	-7.6	-8.1	0.586	0.681	-0.6	-0.8
upps_pos_urgency	0.734	0.827	-2.9	-3.1	0.836	0.9	-5.9	-6.2	0.641	0.719	-2.3	-2
upps_neg_urgency	0.566	0.555	-0.2	0.7	0.671	0.688	-1.8	-1.2	0.5	0.525	0.5	1.2

Table S8: Comparison between motion blocks and unconstrained permutation with censoring at FD < 0.1 mm. Motion impact scores that were significant ($p < 0.05$) with unconstrained permutation but became not-significant ($p > 0.05$) with motion blocks permutation are highlighted in red. Motion impact scores that were significant only with motion blocks permutation are highlighted in blue

Supplemental Figures

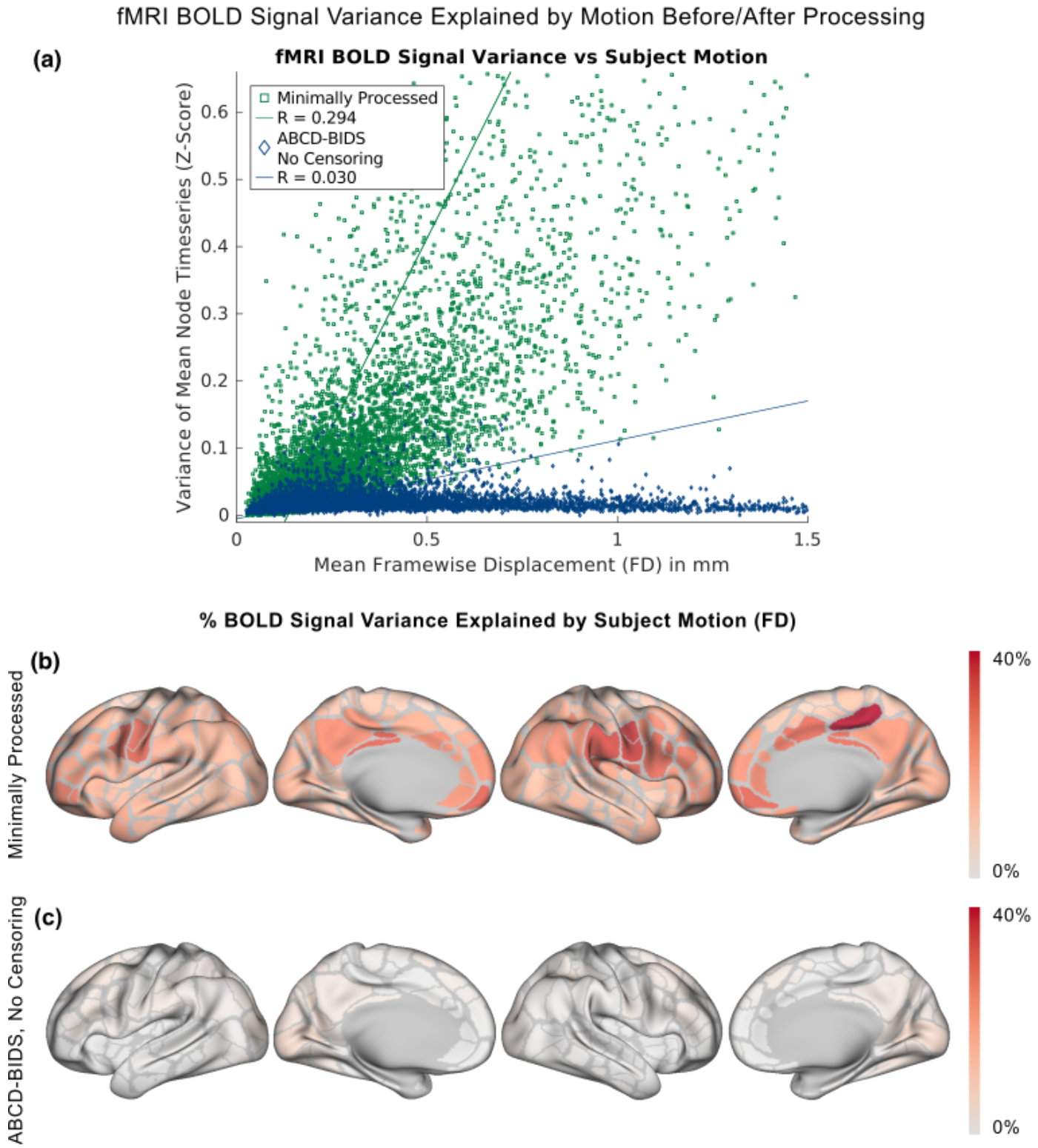

Figure S1: Proportion of fMRI BOLD signal variance explained by head motion before and after processing with the ABCD-BIDS (DCAN Labs) motion reduction algorithm. The comparison is made prior to any framewise motion censoring. **(a)** Scatterplot of variance (Z-score) of the mean node timeseries *vs.* 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.

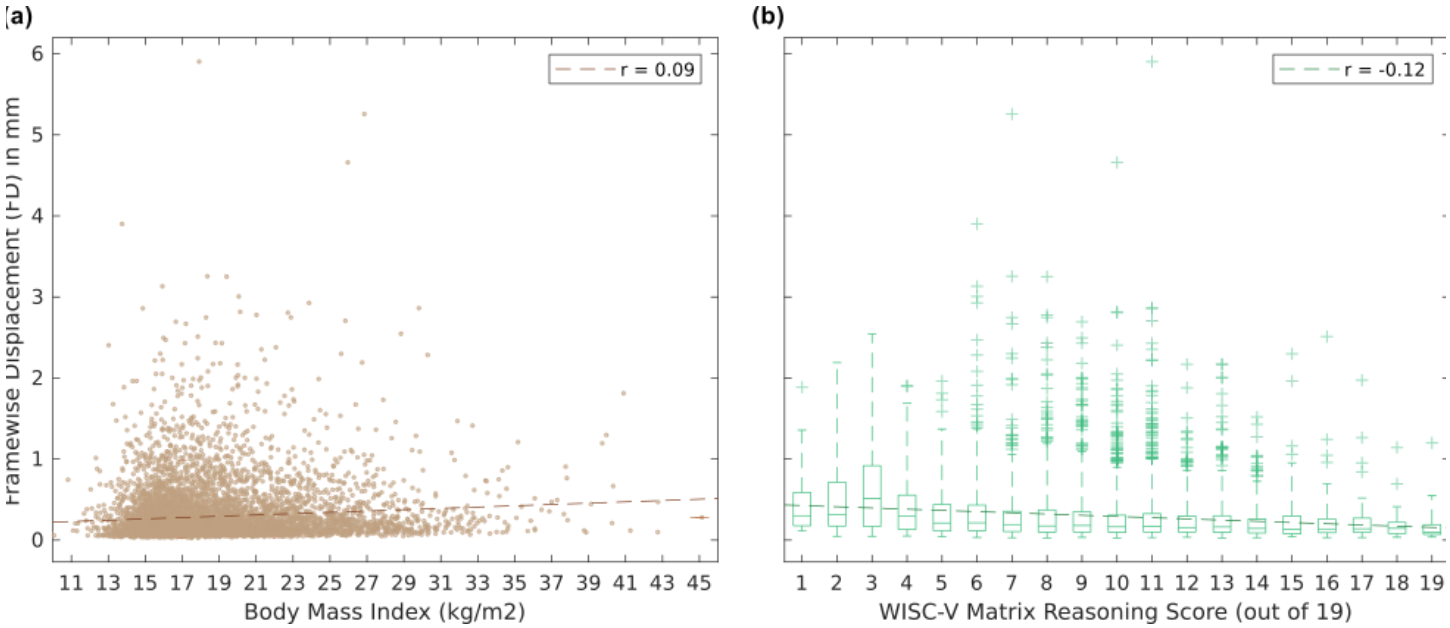

Figure S2: Correlation between trait variables and in-scanner head motion. (a) Body mass index (BMI, kg/m^2, mean 16.8, standard deviation 3.1), Pearson r = 0.09, Spearman r = 0.13. **(b)** Wechsler Intelligence Scale for Children 5th Edition (WISC-V)[51] matrix reasoning score (mean 10, standard deviation 3, maximum 19), Pearson r = -0.12, Spearman 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.

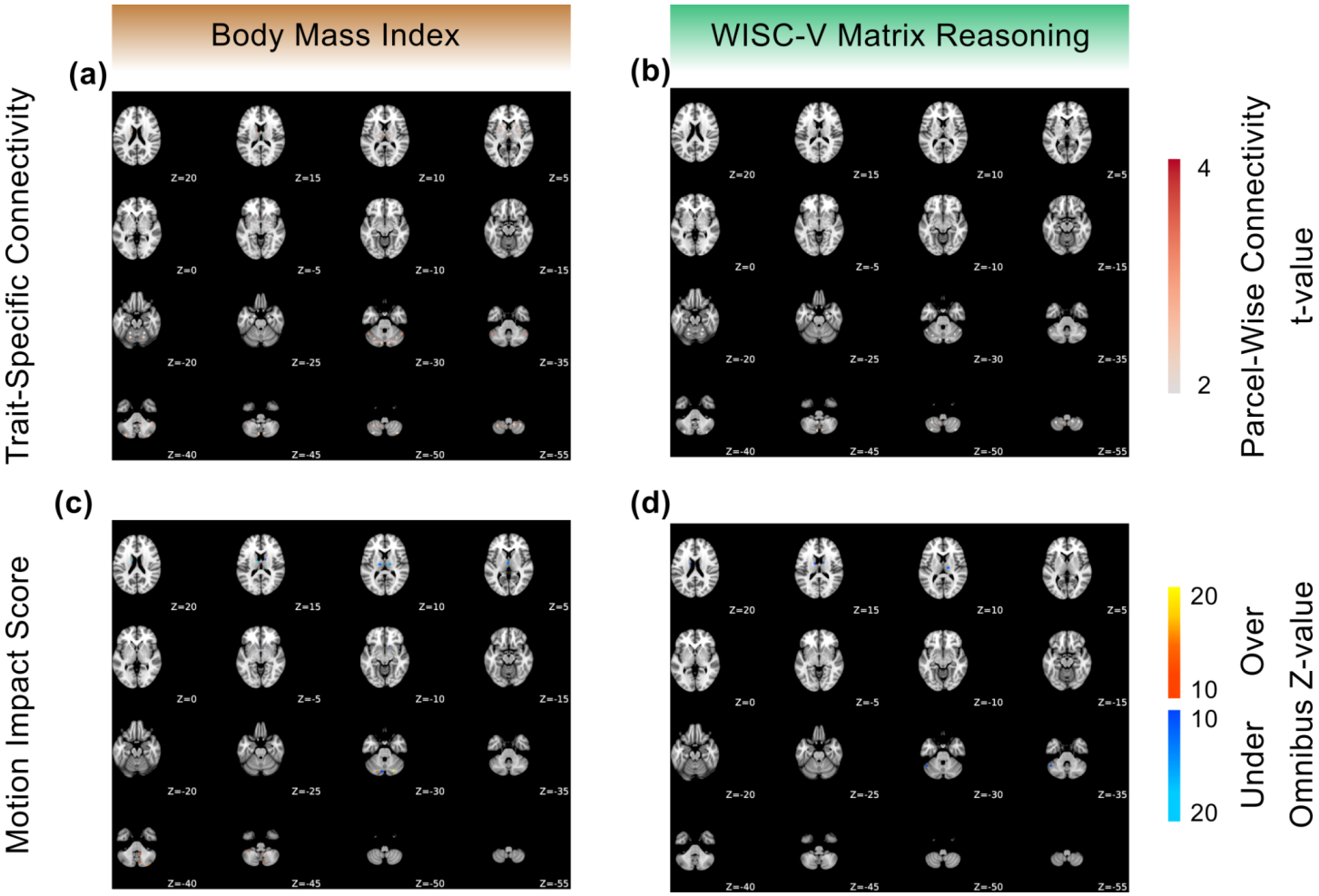

Figure S3: Trait-specific impact of motion on functional connectivity (FC). This companion to Figure 4 renders results on the 61 Seitzman [66] subcortical regions of interest.**Top:** Parcel-wise FC, computed as the root mean square (RMS) of connectivity values for each parcel/node in the trait-FC effect matrix for**(a)** body mass index (BMI) and **(b)** WISC-V matrix reasoning score. **Bottom:** Motion impact score for **(c)** BMI and **(d)** WISC-V. Motion overestimation scores are labeled “Over” in orange and motion underestimation scores are labeled “Under” in blue.

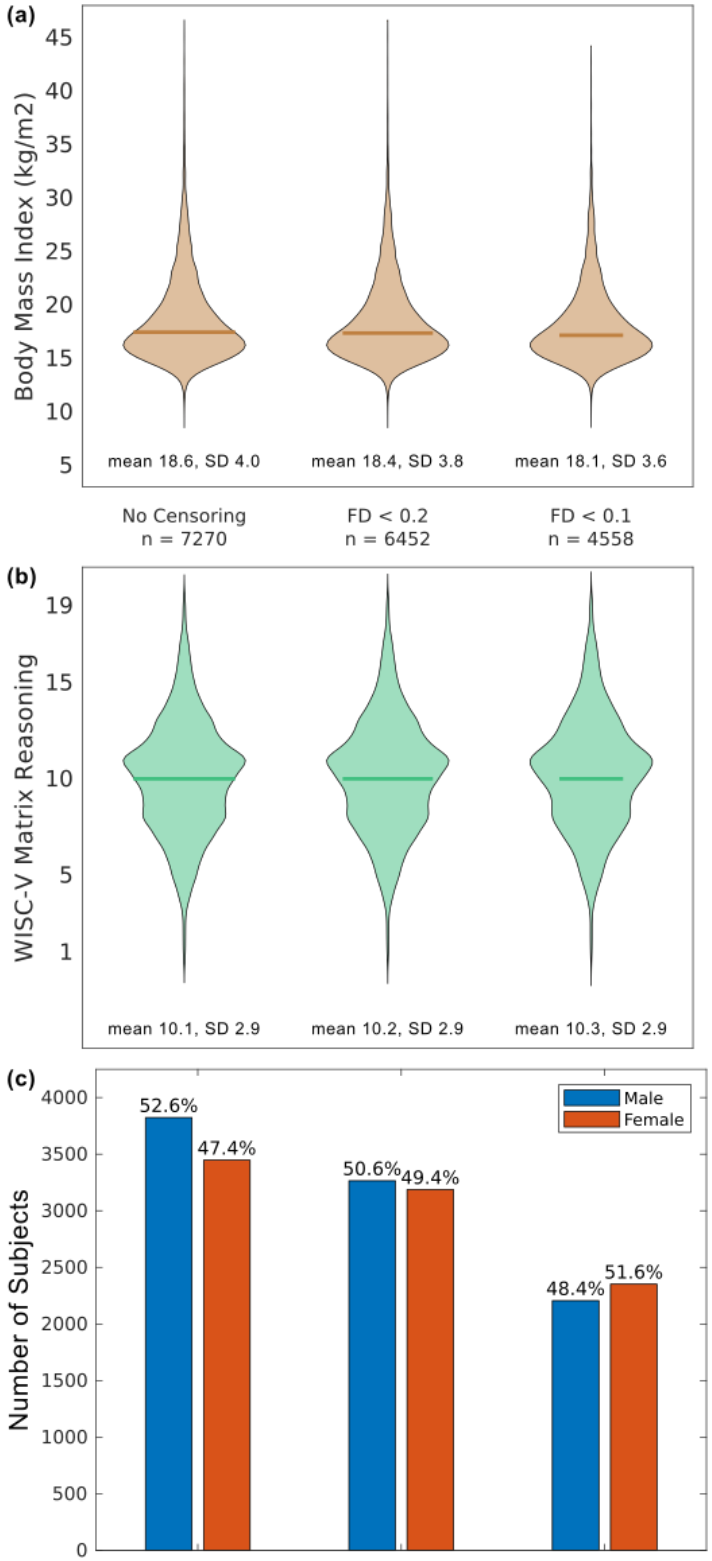

Figure S4: Sampling bias at different levels of framewise censoring for selected variables. Sample distributions are shown for **(a)** body mass index (BMI), **(b)** WISC-V matrix reasoning score, and **(c)** sex assigned at birth. Mean (solid line) and standard deviation (SD) for (a) BMI and (b) matrix 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.

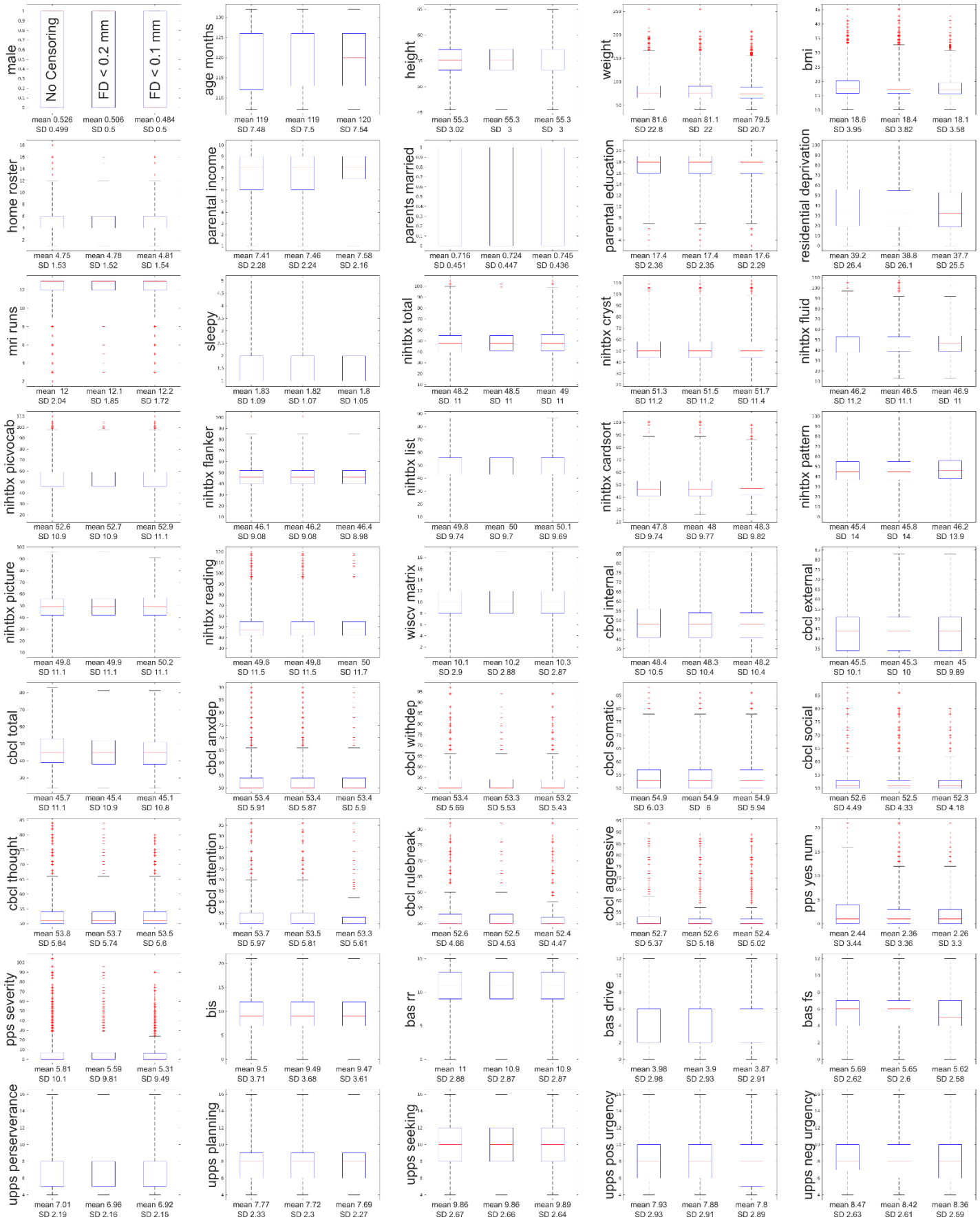
No Censoring
FD < 0.2 mm
FD < 0.1 mm
male
mean 0.526 SD 0.499
mean 0.506 SD 0.5
mean 0.484 SD 0.5
age months
mean 119 SD 7.48
mean 119 SD 7.5
mean 120 SD 7.54
height
mean 55.3 SD 3.02
mean 55.3 SD 3
mean 55.3 SD 3
weight
mean 81.6 SD 22.8
mean 81.1 SD 22
mean 79.5 SD 20.7
bmi
mean 18.6 SD 3.95
mean 18.4 SD 3.82
mean 18.1 SD 3.58
home roster
mean 4.75 SD 1.53
mean 4.78 SD 1.52
mean 4.81 SD 1.54
parental income
mean 7.41 SD 2.28
mean 7.46 SD 2.24
mean 7.58 SD 2.16
parents married
mean 0.716 SD 0.451
mean 0.724 SD 0.447
mean 0.745 SD 0.436
parental education
mean 17.4 SD 2.36
mean 17.4 SD 2.35
mean 17.6 SD 2.29
residential deprivation
mean 39.2 SD 26.4
mean 38.8 SD 26.1
mean 37.7 SD 25.5
mri runs
mean 12 SD 2.04
mean 12.1 SD 1.85
mean 12.2 SD 1.72
sleepy
mean 1.83 SD 1.09
mean 1.82 SD 1.07
mean 1.8 SD 1.05
nihtbx total
mean 48.2 SD 11
mean 48.5 SD 11
mean 49 SD 11
nihtbx cryst
mean 51.3 SD 11.2
mean 51.5 SD 11.2
mean 51.7 SD 11.4
nihtbx fluid
mean 46.2 SD 11.2
mean 46.5 SD 11.1
mean 46.9 SD 11
nihtbx picvocab
mean 52.6 SD 10.9
mean 52.7 SD 10.9
mean 52.9 SD 11.1
nihtbx flanker
mean 46.1 SD 9.08
mean 46.2 SD 9.08
mean 46.4 SD 8.98
nihtbx list
mean 49.8 SD 9.74
mean 50 SD 9.7
mean 50.1 SD 9.69
nihtbx cardsort
mean 47.8 SD 9.74
mean 48 SD 9.77
mean 48.3 SD 9.82
nihtbx pattern
mean 45.4 SD 14
mean 45.8 SD 14
mean 46.2 SD 13.9
nihtbx picture
mean 49.8 SD 11.1
mean 49.9 SD 11.1
mean 50.2 SD 11.1
nihtbx reading
mean 49.6 SD 11.5
mean 49.8 SD 11.5
mean 50 SD 11.7
wiscv matrix
mean 10.1 SD 2.9
mean 10.2 SD 2.88
mean 10.3 SD 2.87
cbcl internal
mean 48.4 SD 10.5
mean 48.3 SD 10.4
mean 48.2 SD 10.4
cbcl external
mean 45.5 SD 10.1
mean 45.3 SD 10
mean 45 SD 9.89
cbcl total
mean 45.7 SD 11.1
mean 45.4 SD 10.9
mean 45.1 SD 10.8
cbcl anxdep
mean 53.4 SD 5.91
mean 53.4 SD 5.87
mean 53.4 SD 5.9
cbcl withdep
mean 53.4 SD 5.69
mean 53.3 SD 5.53
mean 53.2 SD 5.43
cbcl somatic
mean 54.9 SD 6.03
mean 54.9 SD 6
mean 54.9 SD 5.94
cbcl social
mean 52.6 SD 4.49
mean 52.5 SD 4.33
mean 52.3 SD 4.18
cbcl thought
mean 53.8 SD 5.84
mean 53.7 SD 5.74
mean 53.5 SD 5.6
cbcl attention
mean 53.7 SD 5.97
mean 53.5 SD 5.81
mean 53.3 SD 5.61
cbcl rulebreak
mean 52.6 SD 4.66
mean 52.5 SD 4.53
mean 52.4 SD 4.47
cbcl aggressive
mean 52.7 SD 5.37
mean 52.6 SD 5.18
mean 52.4 SD 5.02
pps yes num
mean 2.44 SD 3.44
mean 2.36 SD 3.36
mean 2.26 SD 3.3
pps severity
mean 5.81 SD 10.1
mean 5.59 SD 9.81
mean 5.31 SD 9.49
bis
mean 9.5 SD 3.71
mean 9.49 SD 3.68
mean 9.47 SD 3.61
bas rr
mean 11 SD 2.88
mean 10.9 SD 2.87
mean 10.9 SD 2.87
bas drive
mean 3.98 SD 2.98
mean 3.9 SD 2.93
mean 3.87 SD 2.91
bas fs
mean 5.69 SD 2.62
mean 5.65 SD 2.6
mean 5.62 SD 2.58
upps perserverance
mean 7.01 SD 2.19
mean 6.96 SD 2.16
mean 6.92 SD 2.15
upps planning
mean 7.77 SD 2.33
mean 7.72 SD 2.3
mean 7.69 SD 2.27
upps seeking
mean 9.86 SD 2.67
mean 9.86 SD 2.66
mean 9.89 SD 2.64
upps pos urgency
mean 7.93 SD 2.93
mean 7.88 SD 2.91
mean 7.8 SD 2.89
upps neg urgency
mean 8.47 SD 2.63
mean 8.42 SD 2.61
mean 8.36 SD 2.59

Figure S5: Sampling bias at different levels of framewise censoring. Box and whisker plots are shown for all 45 traits examined in the ABCD study at different levels of motion censoring. From left to right, no censoring, FD < 0.2 mm, and FD < 0.1 mm.

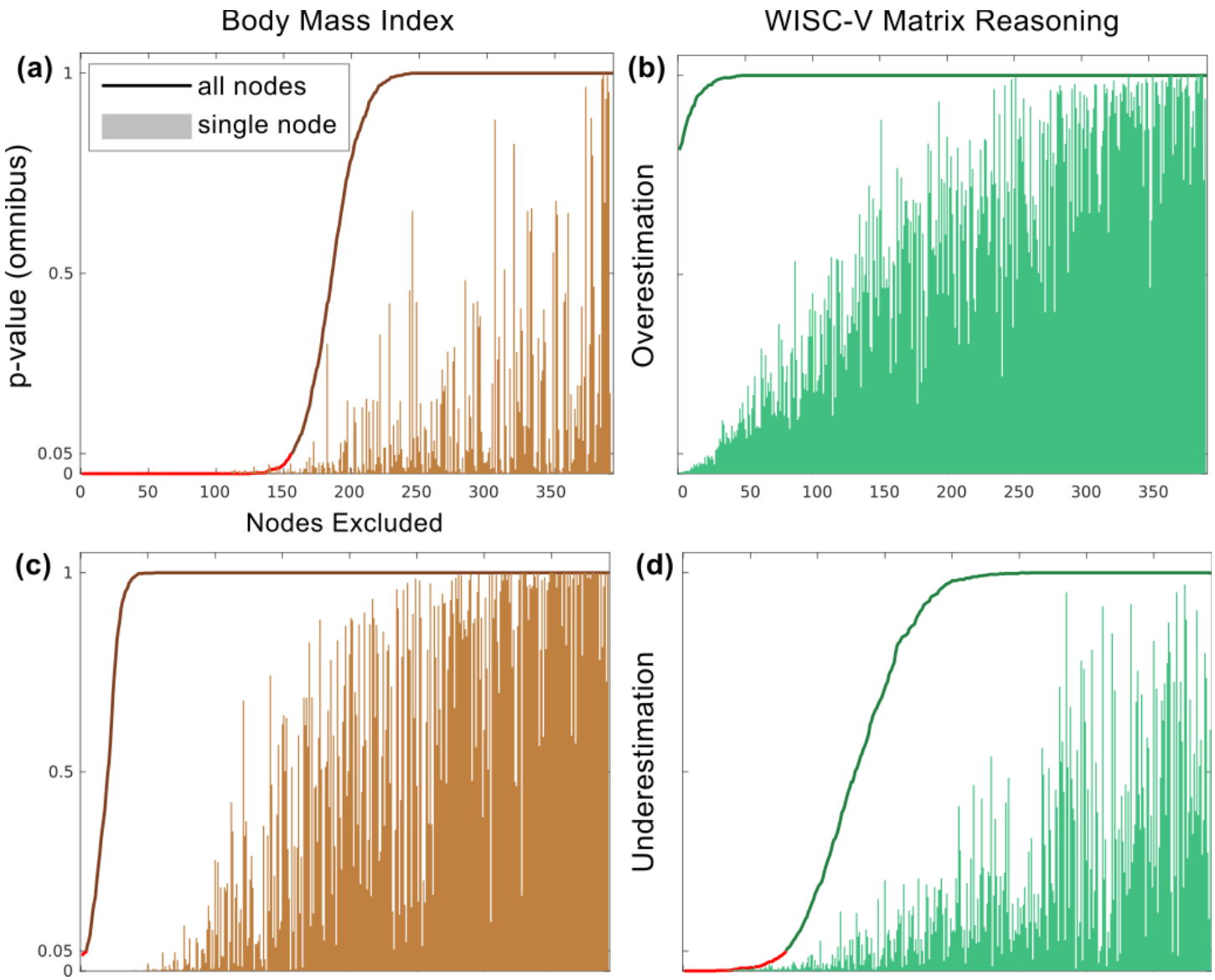

Figure S6: Number of nodes contributing to significance of motion impact. Top: Nodes driving motion overestimation score for **(a)** body mass index (BMI) and **(b)** WISC-V matrix reasoning score. **Bottom:** Nodes driving motion underestimation score for **(c)** BMI and **(d)** matrix reasoning. Reading 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.

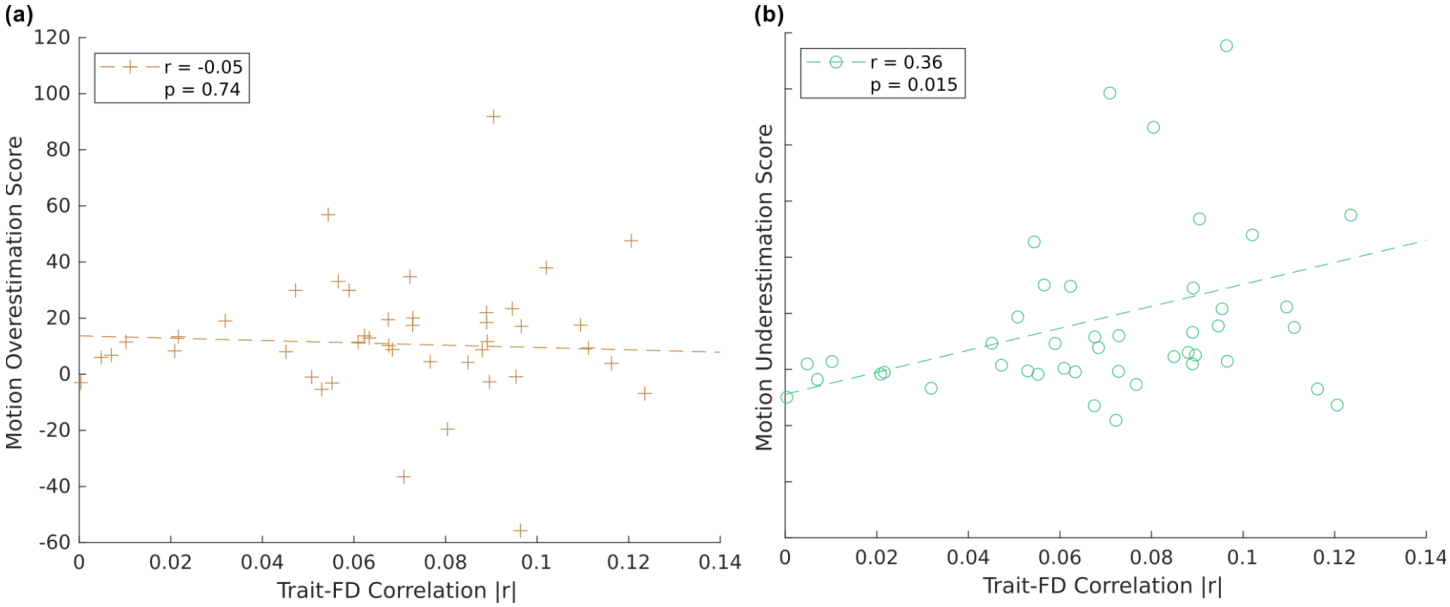

Figure S7: Correlation between SHAMAN motion impact score and in-scanner head motion. (a) Motion overestimation and **(b)** motion underestimation scores (omnibus Stouffer's Z-score) for selected variables from the ABCD study are plotted vs the correlation (absolute value) of those variables with average participant head motion (framewise displacement, FD). Results reflect fMRI data after motion-reduction with ABCD-BIDS and without additional motion censoring. See Table S1.

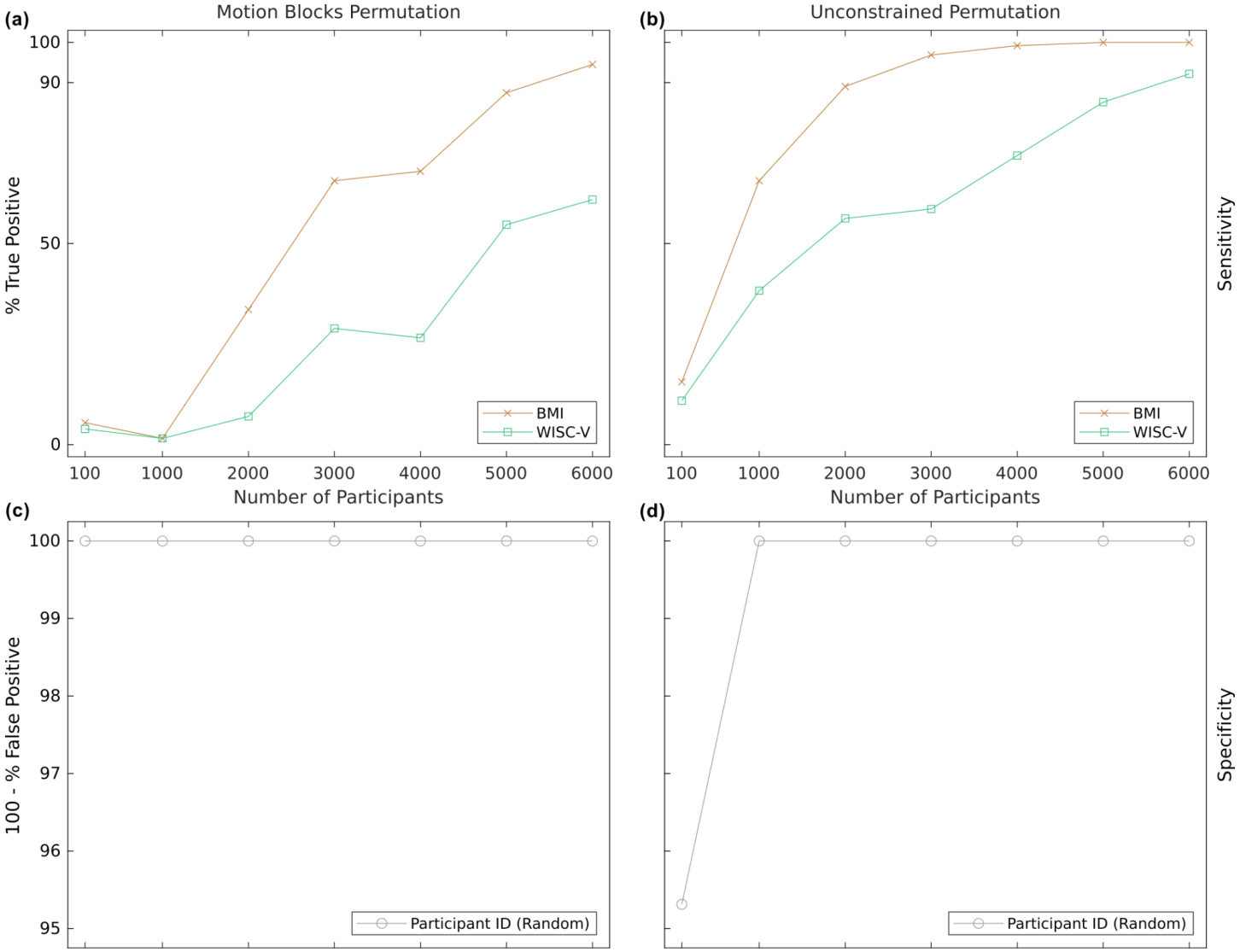

Figure S8: Statistical Power of SHAMAN Method. (a,b) True positive detection rate (sensitivity, power) of SHAMAN computed using bootstrapped subsamples of different sizes (numbers of participants) with body mass index as an exemplar variable for significant motion overestimation score and WISC-V matrix reasoning score as an exemplar for motion underestimation score. **(c,d)** Specificity of motion impact score using a random variable generated from the unique participant identifier. Panels **(a,c)** use the conservative, motion blocks permutation scheme, which accounts for weak temporal autocorrelation and is used throughout this study. Panels **(b,d)** show the theoretical upper limits on power using the less-conservative unconstrained permutation scheme. All data use 128 bootstraps and 128 permutations at each sample size.

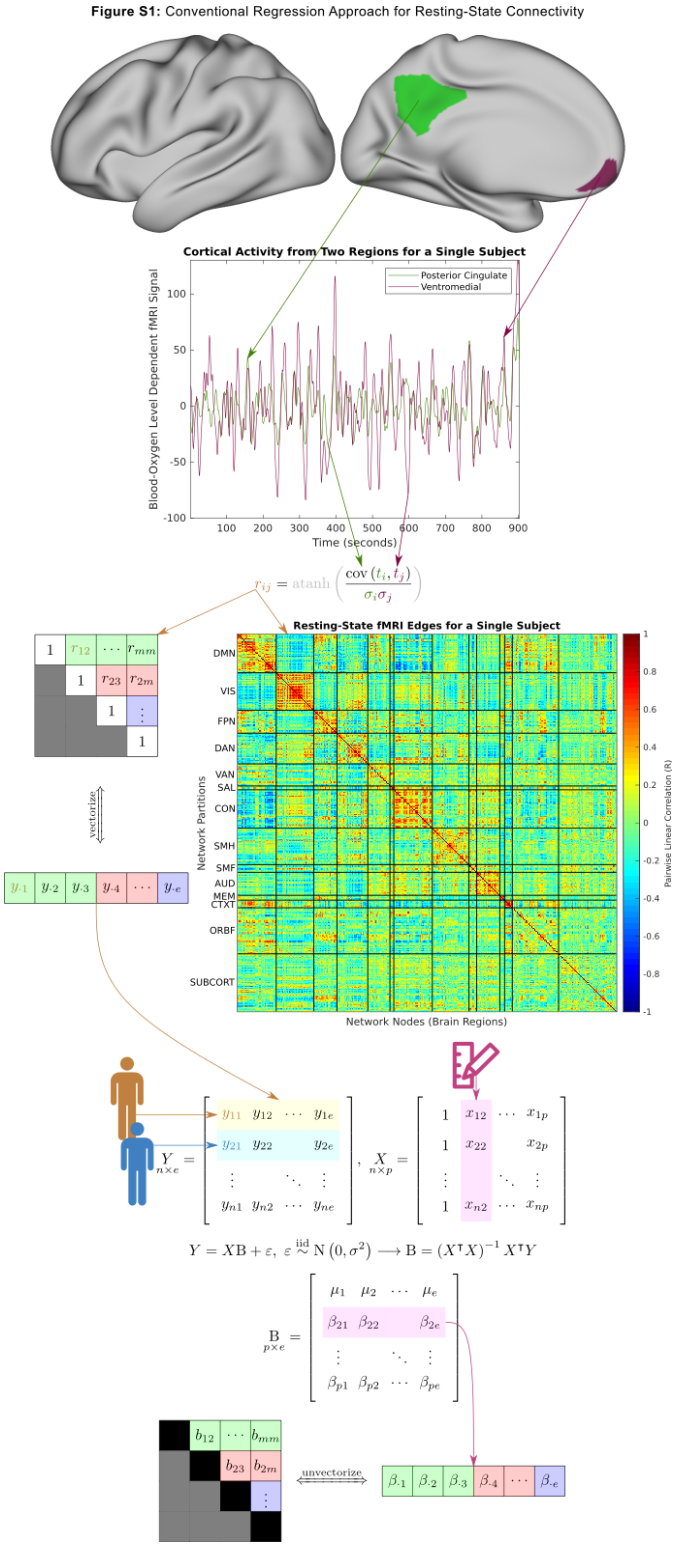

Figure S9: Conventional regression model for resting-state functional connectivity.

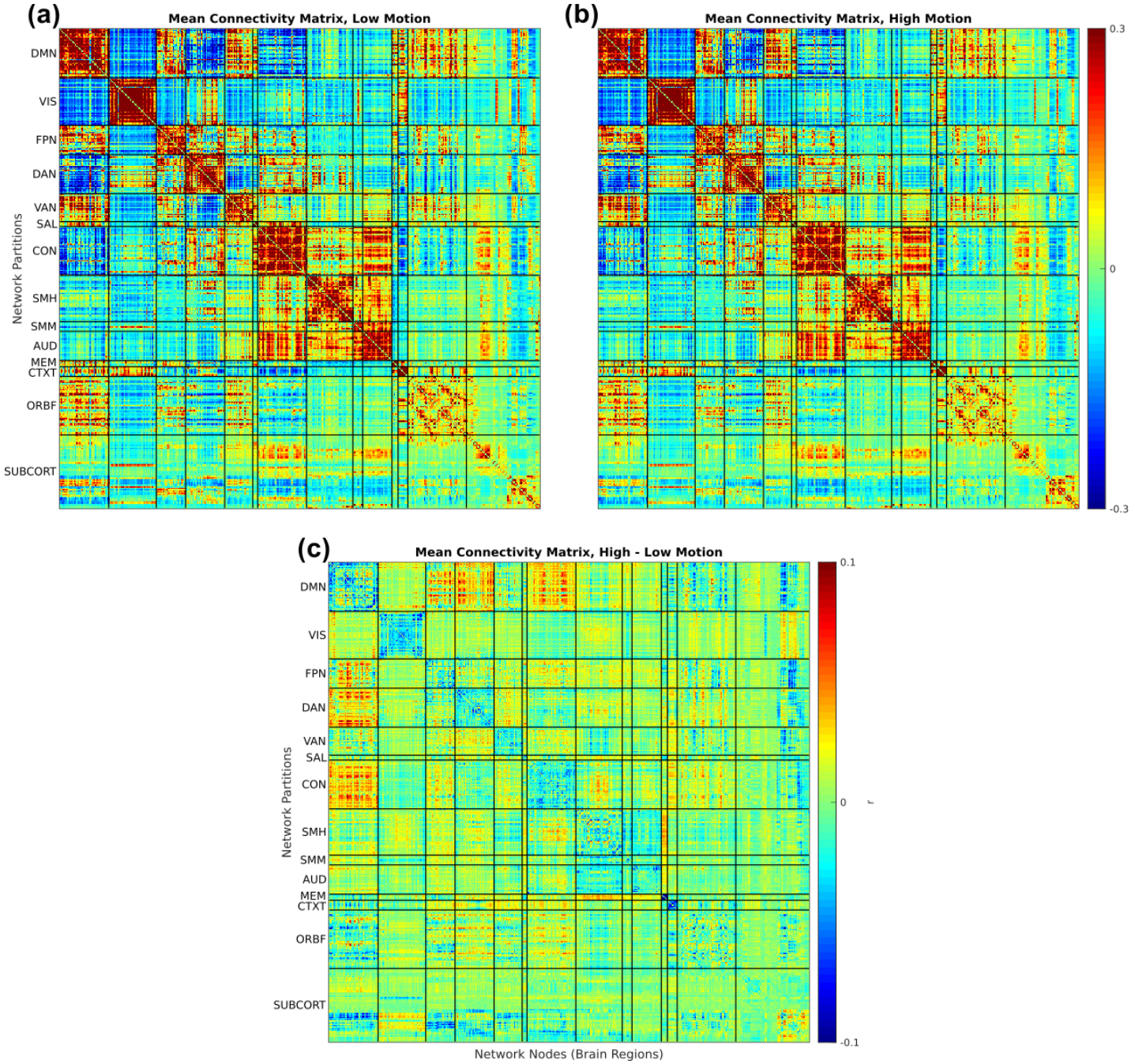

Figure S10: Mean (average) connectivity matrix computed from the low-motion **(a)** and high-motion **(b)** halves of each participant's resting-state fMRI data. The difference (rescaled to show visual detail) is shown in **(c)**. Figures are generated from data after denoising with ABCD-BIDS but with*out* motion censoring.

Figure S2: SHAMAN Algorithm (Permuted Null Model)

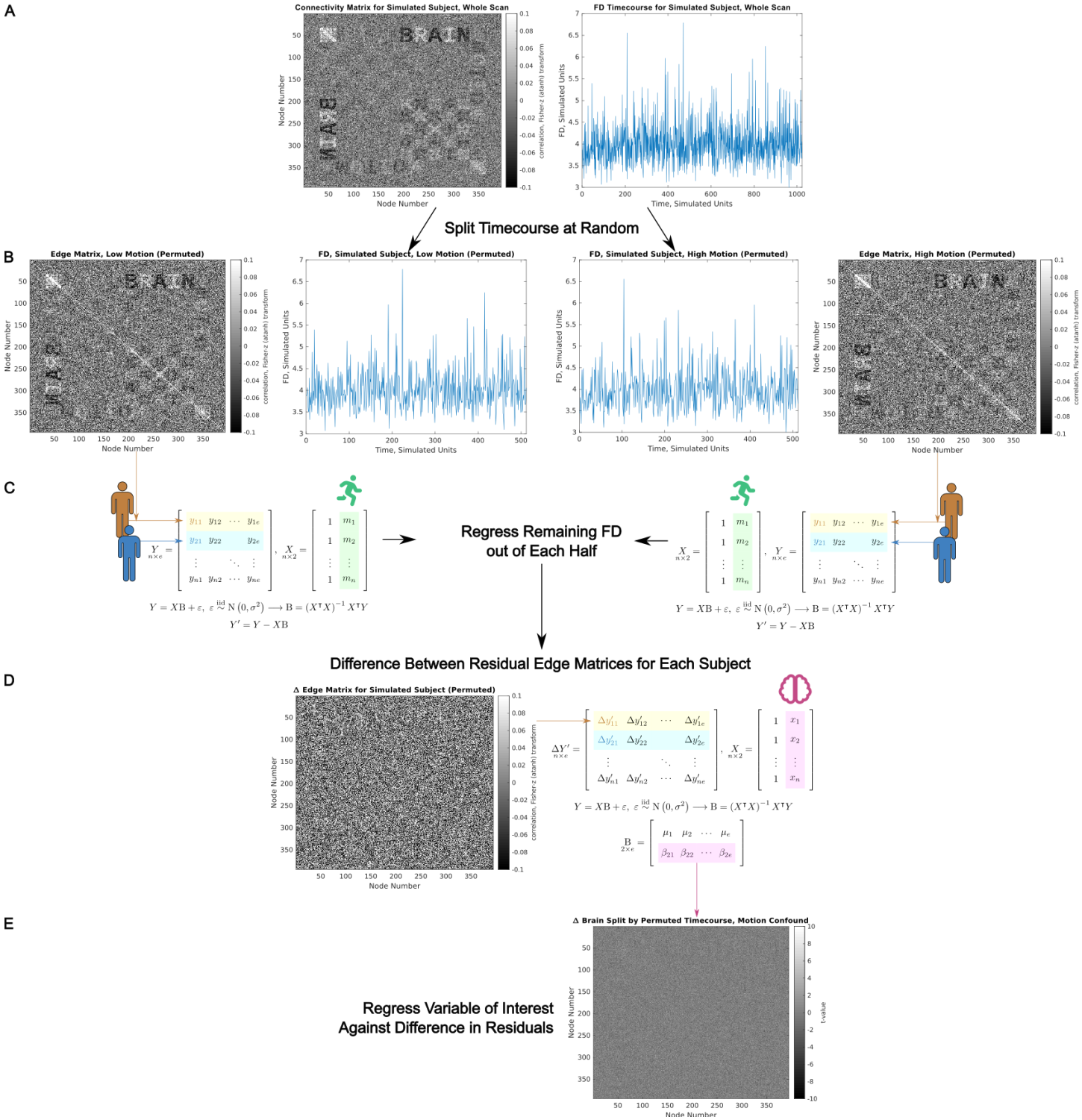

Figure S11: Null model permutation scheme for SHAMAN algorithm. In **(B)** the data is split in half at random, without respect to the FD timeseries. The rest of the algorithm proceeds as before. The algorithm is repeated many times for each random split (permutation) to generate a null distribution for the motion-associated connectivity matrix in **(E)**.

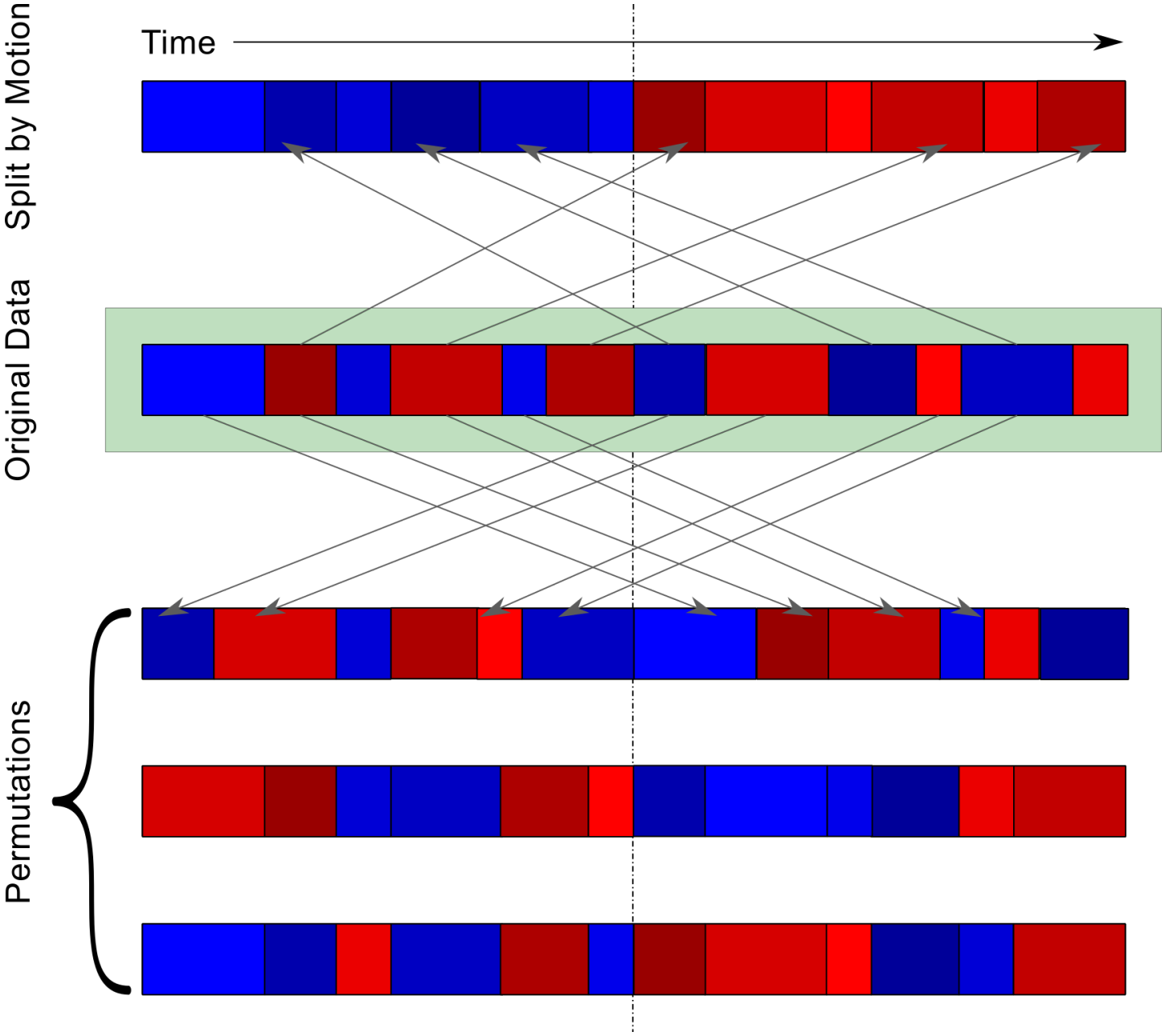

Figure S12: Motion blocks permutation method. A single participant's fMRI timeseries data is visualized as a rectangle with time along the horizontal axis. A dashed vertical line represents division of the data into two halves. Data with associated motion (e.g. framewise displacement, FD) below the median is labeled blue, and data with motion above the median is labeled red. Consecutive timepoints labeled the same color form temporally consecutive "blocks" of data. Different shades of blue and red are used to highlight the boundaries of different motion-derived blocks. When split by motion **(top)**, all of the low-motion blocks and all of the high-motion blocks are grouped together in the same split half of the data. During permutation **(bottom)**, blocks are sorted randomly into each split half. Preserving temporally consecutive timepoints within blocks during permutation ensures the same amount of temporal autocorrelation in permuted and un-permuted divisions of the data.

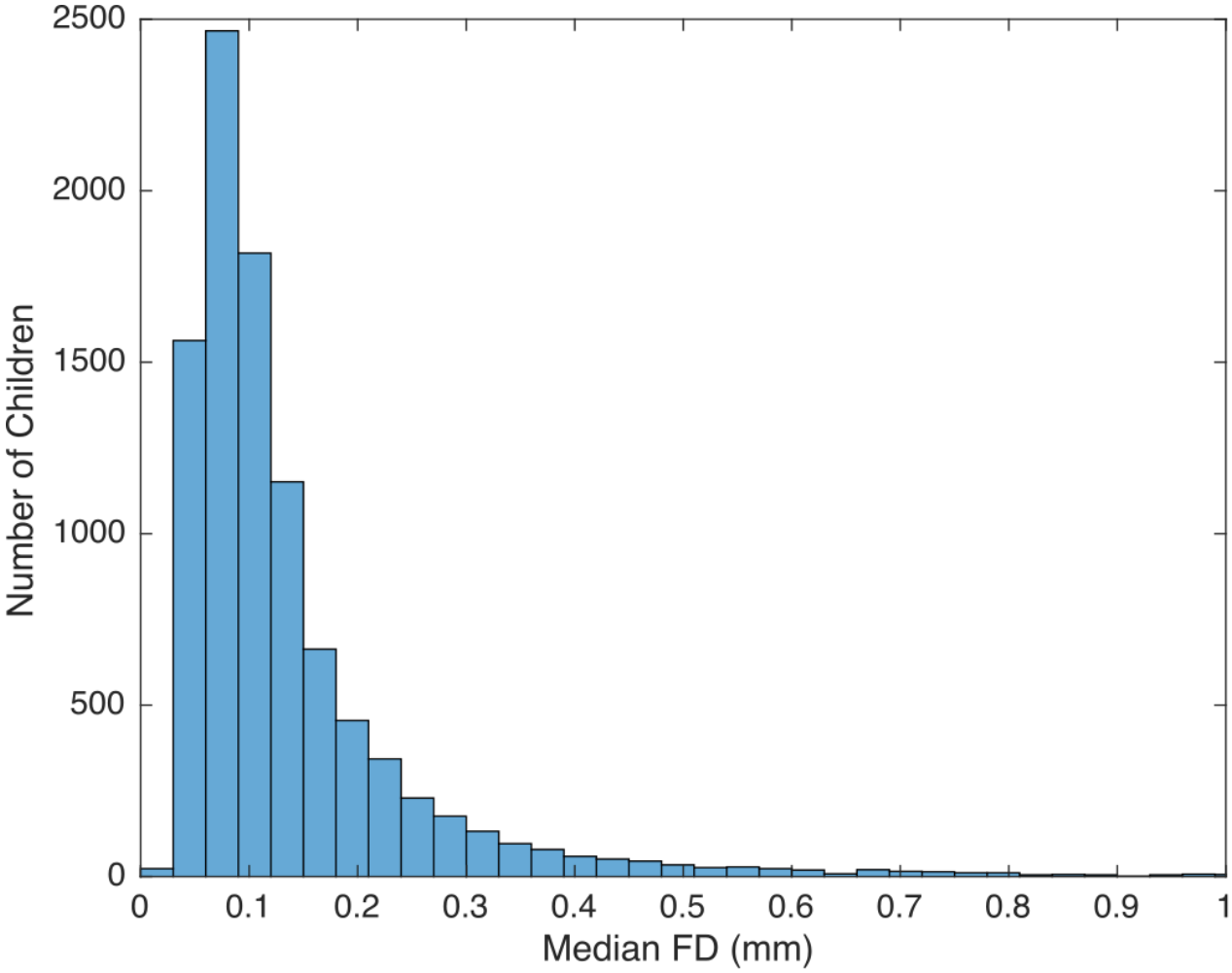

Figure S13: Distribution of median FD. Data are shown for n = 9,625 children in the ABCD study with at least 8 minutes of rs-fMRI data before motion censoring. The histogram is truncated at 1 mm for readability; the maximum median FD was 2.9 mm.

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 [71], 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 S9) 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 [65,66] 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.

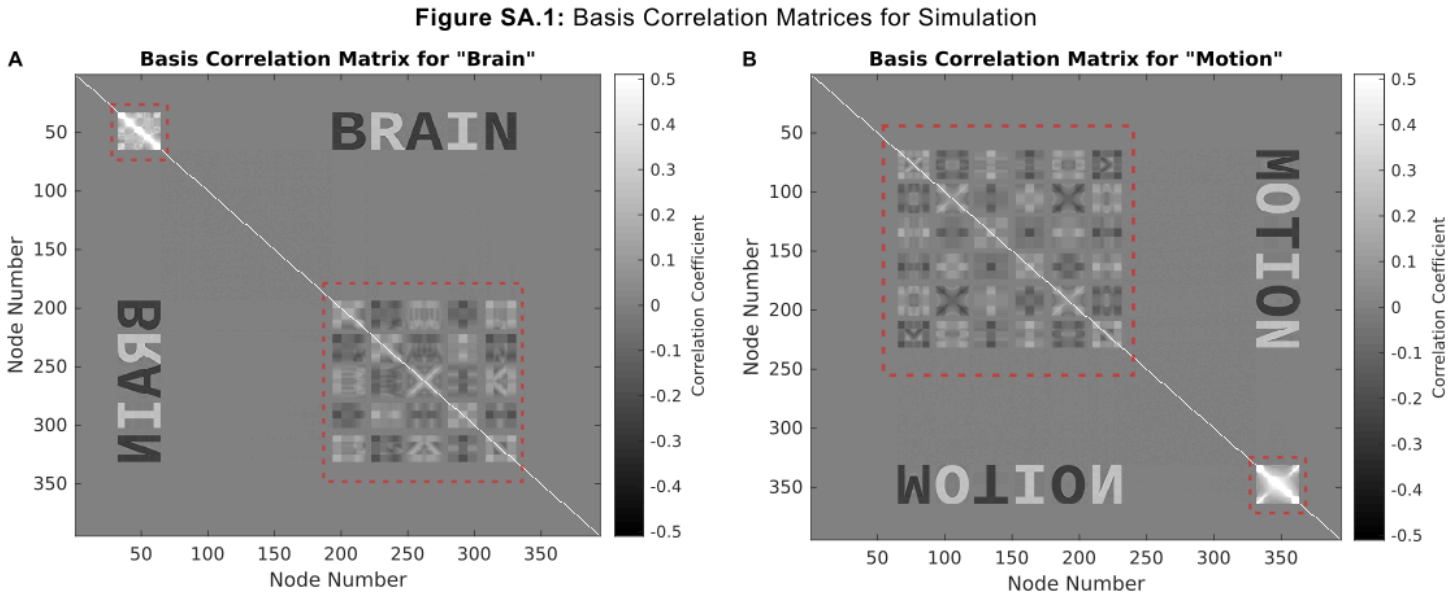

Figure SA.1: Basis correlation matrices. (a) “Brain” signal. **(b)** “Motion” artifact. The dimensions of each matrix are 394 x 394 nodes. The portions of the matrices outlined in the red dashed boxes were algorithmically added to make each matrix positive semi-definite.

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 S9.

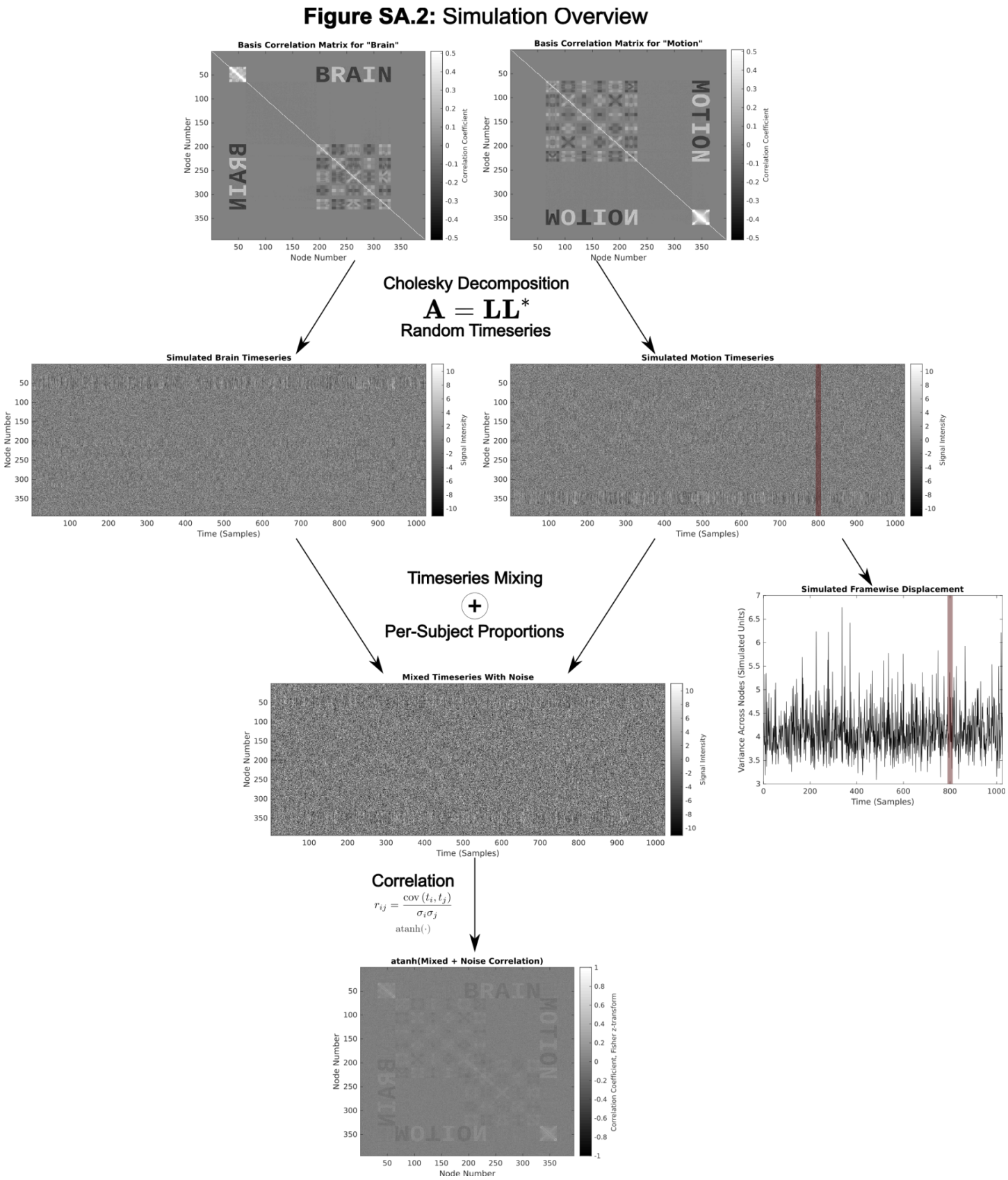

Figure SA.2: Method for generating simulated fMRI timeseries and connectivity matrices. 394 x 394 basis correlation matrices for “brain” and “motion” were Cholesky-decomposed to generate random 394 x 1,024 timeseries for each of 1,024 participants. The simulated FD (framewise displacement) timeseries 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

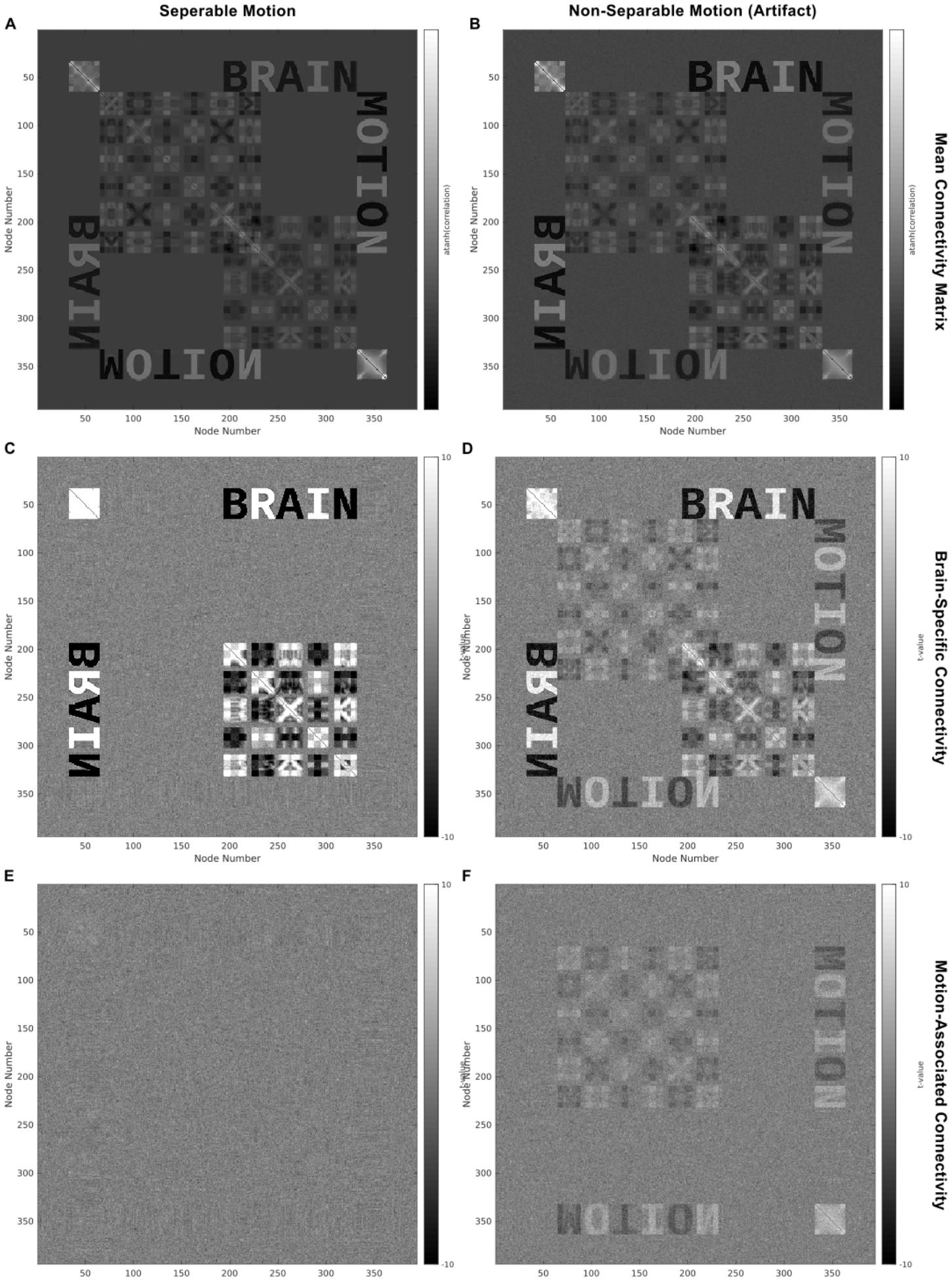

Figure SA.3: Simulation results when motion is linearly separable (left) and when motion is not linearly separable (right). In both cases **(a, b)** motion is present in the mean connectivity matrix. The linear regression model for "brain" with motion (participants' mean FD) as a covariate successfully 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 S9, 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

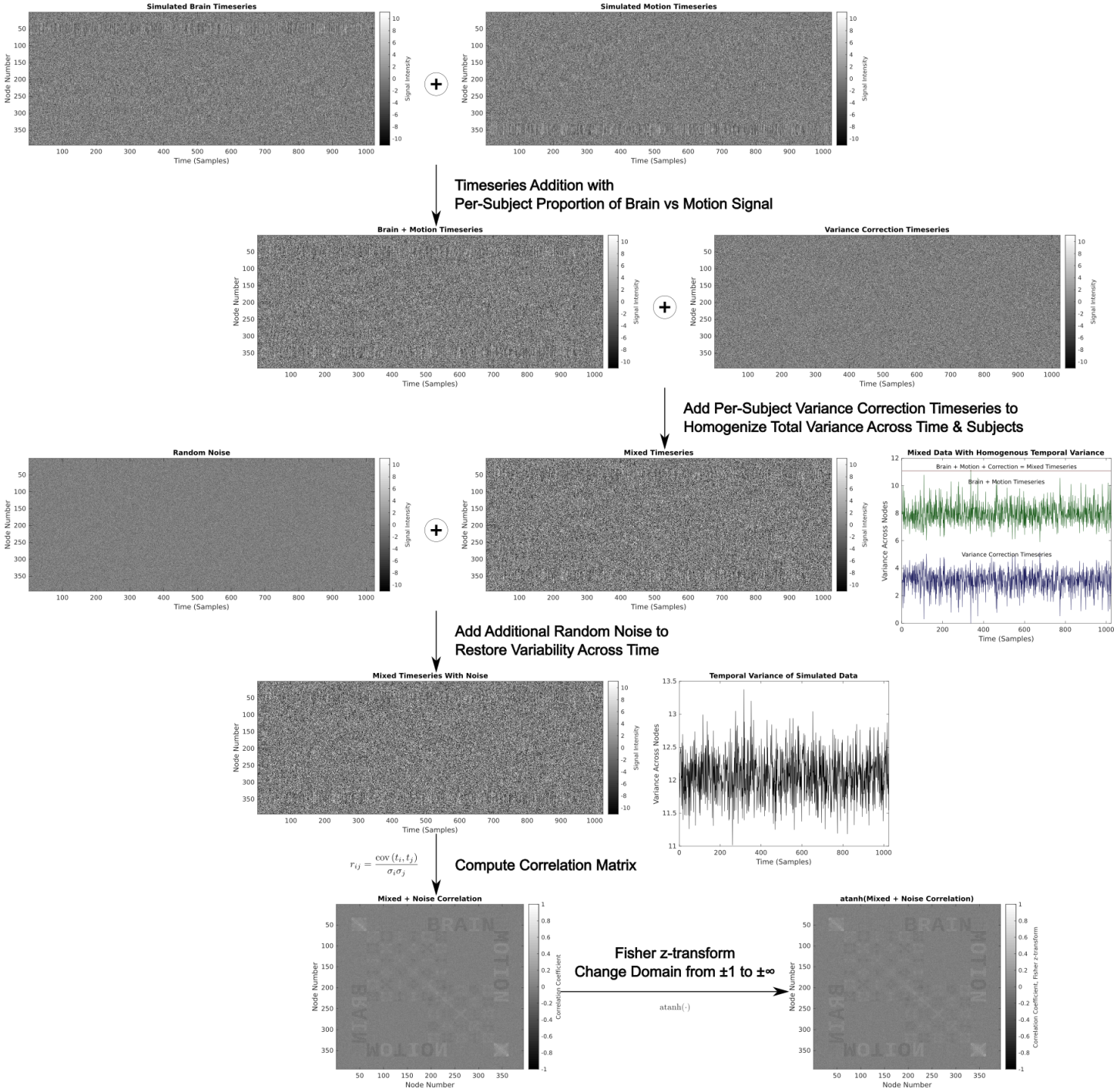

Figure SA.4: Method for proper timeseries mixing of brain and motion signals in controlled proportions. A provisional mixing of the brain and motion timeseries is obtained by addition in the time domain. A random variance-correction timeseries is then added such that, at each time point, the across-node 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.

Data Comparison Supplement

Rationale

Innumerable methods exist for quantifying and mitigating head motion artifact within many large neuroimaging data sets. DVARS, an alternative to framewise displacement (FD) for quantifying in-scanner head motion, was first proposed by Smyer et al[72] and later popularized by Power[14] and Afyouni & Nichols[73]. More recently, Pham et al[41] demonstrated a method for selecting an optimal DVARS motion censoring threshold. DVARS is an acronym for, "D referring to temporal derivative of timecourses, VARS referring to RMS variance over voxels."[14] It is calculated as the root mean square difference in voxel intensity (within a brain mask) between one fMRI timepoint/volume and the next. A critical conceptual difference between FD and DVARS is that FD measures motion specifically, being itself derived from the six rigid-body motion co-registration parameters, whereas DVARS is sensitive to any change in fMRI signal intensity, whether due to motion artifact or another source.

In this supplementary analysis we apply SHAMAN to data processed with different exemplar methods and from different data sets. Specifically, we apply SHAMAN to ABCD data using DVARS in place of FD, and to Human Connectome Project (HCP) data, which is processed using ICA-FIX, using both FD and DVARS. The purpose of this supplement is not to exhaustively catalog all possible methodological variations, nor to make specific prescriptions for data processing. Rather, we show that the general trends seen in our main analysis are replicated in different data processed in different ways.

Methods

<u>Calculation of DVARS in ABCD</u>
The ABCD-BIDS pipeline [25,37] computes DVARS as one of its quality control measures. ABCD-BIDS computes DVARS at three different stages of processing: before motion co-registration, after co-registration and before further denoising, and after both co-registration and de-noising. We used DVARS after co-registration and before further denoising. Crucially, this version of DVARS was computed *before* global signal regression. We substituted this measure of DVARS for FD when running SHAMAN. For computational expediency, we only ran 128 permutations with DVARS and chose a single DVARS threshold of 200 for motion censoring rather than explore two different motion censoring thresholds. A DVARS threshold of 200 was selected because it resulted in the exclusion of 1,058 participants (retention of 6,212 participants), comparable to the number of participants excluded by FD censoring.

<u>HCP Data Acquisition and Processing</u>
The Human Connectome Project (HCP) includes fMRI data from 1200 young, healthy adult volunteers [49,50]. Resting-state fMRI data was acquired at 3T using gradient-echo EPI with TR = 720 ms, slice thickness 2.0 mm, and 72 slices for 2.0 mm isotropic voxels. There were 4 resting state runs with 1200 frames per run [74]. We used data after minimal processing [38] and ICA-FIX [7,16]. Part of our rationale for analyzing HCP data was to test SHAMAN on data processed with ICA-FIX, which automatically removes motion artifact using independent component analysis. As with the ABCD data, at each level of motion censoring, we excluded participants with fewer than 600 fMRI frames after censoring. This resulted in n = 1017 (out of 1200 participants recruited) with usable resting-state fMRI data before censoring. For computational expediency, we only ran 128 permutations of SHAMAN for each motion measure and censoring threshold in HCP data.

<u>Calculation of FD and DVARS in HCP</u>
FD and DVARS were computed using the methods described by Power et al [14] with the following modifications from ABCD. FD was computed without correction for respiratory motion artifact, which is part of the

ABCD-BIDS pipeline. The FD censoring thresholds were the same as those used in the ABCD data, 0.2 and 0.1 mm. For DVARS, we selected a threshold of 800.

Selection of Traits in HCP

The following traits of interest in HCP were selected to parallel the traits selected in ABCD. The meanings of variable names can be found in the HCP data dictionary:
https://wiki.humanconnectome.org/display/PublicData/HCP-YA+Data+Dictionary-+Updated+for+the+1200+Subject+Release

Demographic:

SSAGA_Income, SSAGA_Educ

Physical:

Age_in_Yrs, Height, Weight, BMI, Handedness, Hematocrit_1, ThyroidHormone, HbA1C, Dexterity_AgeAdj, Strength_AgeAdj, Noise_Comp, Odor_AgeAdj, Taste_AgeAdj

Personality:

PSQI_Score, DDisc_AUC_200, DDisc_AUC_40K, ER40_CR, ER40_CRT, AngAffect_Unadj, AngHostil_Unadj, AngAggr_Unadj, FearAffect_Unadj, FearSomat_Unadj, Sadness_Unadj, LifeSatisf_Unadj, MeanPurp_Unadj, PosAffect_Unadj, Friendship_Unadj, Loneliness_Unadj, PercHostil_Unadj, PercReject_Unadj, EmotSupp_Unadj, InstruSupp_Unadj, PercStress_Unadj, SelfEff_Unadj, NEOFAC_A, NEOFAC_O, NEOFAC_C, NEOFAC_E, ASR_Anxd_Pct, ASR_Witd_T, ASR_Soma_T, ASR_Thot_T, ASR_Attn_T, ASR_Aggr_T, ASR_Rule_T, ASR_Intr_T, ASR_Intn_T, ASR_Extn_T, ASR_Totp_T, DSM_Depr_T, DSM_Anxi_T, DSM_Somp_T, DSM_Avoid_T, DSM_Adh_T, DSM_Antis_T

Cognition:

MMSE_Score, PicSeq_AgeAdj, CardSort_AgeAdj, Flanker_AgeAdj, PMAT24_A_CR, PMAT24_A_RTCR, ReadEng_AgeAdj, PicVocab_AgeAdj, ProcSpeed_AgeAdj, VSPLOT_TC, VSPLOT_CRTE, IWRD_TOT, IWRD_RTC, ListSort_AgeAdj, CogFluidComp_AgeAdj, CogTotalComp_AgeAdj, CogCrystalComp_AgeAdj

Results

Head Motion and Motion Impact Scores Were Lower in HCP

The ABCD and HCP data sets differed substantially in terms of sample size (ABCD n = 7270, HCP n = 1017), age (ABCD 9.9 ± 0.6 years, HCP 28.7 ± 3.7 years), and amount of in-scanner head motion. Mean FD was greater in ABCD, 0.29 ± 0.35 mm, compared to HCP, 0.16 ± 0.06 mm, see Figure SB.1. The distribution of mean FD was heavier-tailed in ABCD. The largest mean FD in the ABCD data was 6.0 mm.

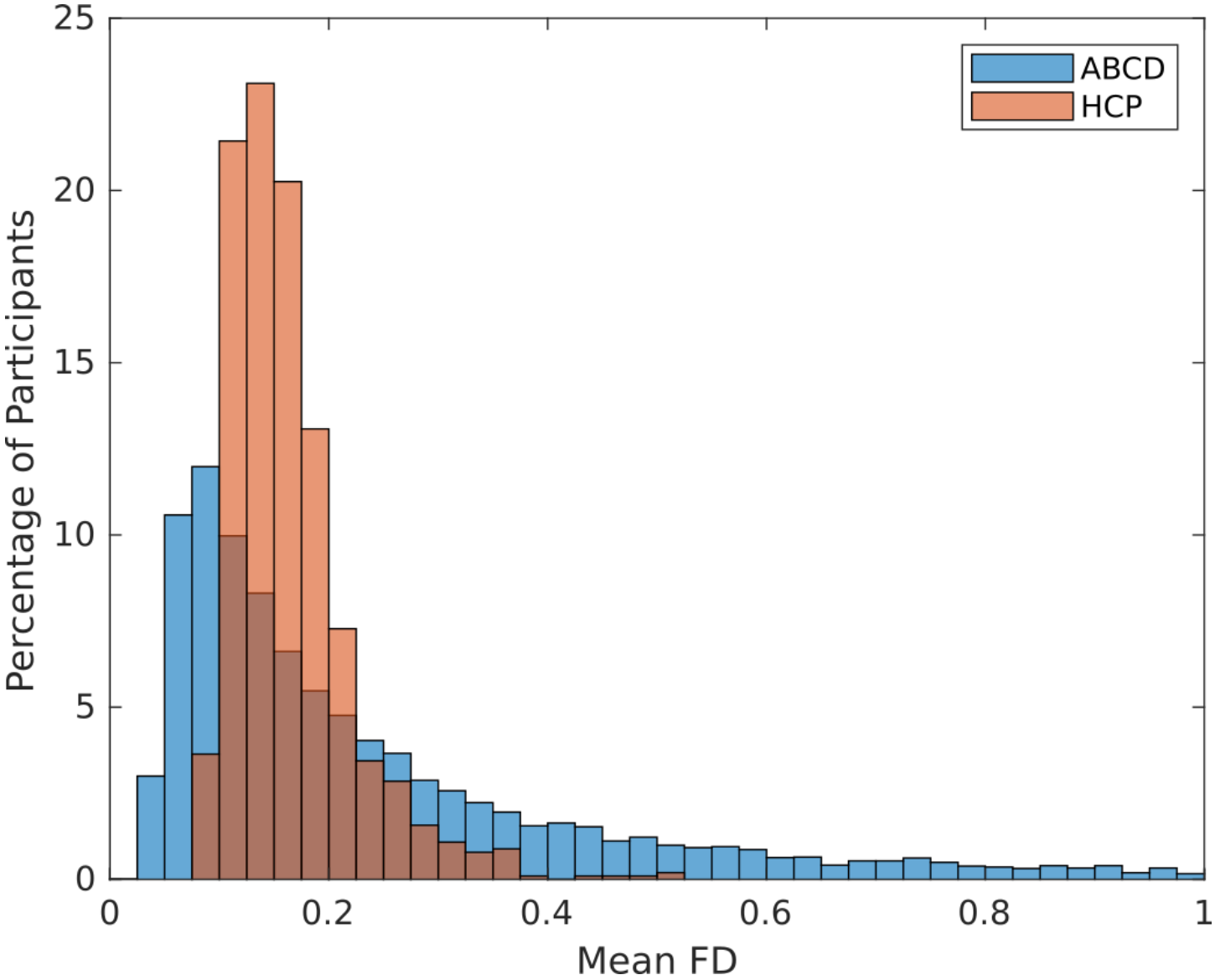

Figure SB.1: In-scanner head motion in ABCD and HCP. Head motion was quantified by mean framewise displacement (in mm). The histogram for ABCD is truncated at FD = 1.0 mm. The largest mean FD in the ABCD data is 6.0 mm. Data are shown without motion censoring.

WIthout motion censoring, 13% (10/76) of traits in HCP had a significant (omnibus $p < 0.05$) motion impact score (over- or under-estimation of the trait-FC effect) compared to 53% (24/45) traits in ABCD. 12% (9/76) of traits in HCP had a significant motion overestimation score and 14% (11/76) had a significant motion underestimation score. Our selection of personality traits in HCP covered a wider variety of tests than the personality variables we selected from ABCD. Personality traits were the most numerous (16%, 7/43) to have significant motion impact scores in HCP (either over- or under-estimation), but as in ABCD, physical traits in HCP were still the most likely to have a significant motion impact score (23%, 3/13).

<u>Motion Censoring Reduced Motion Impact Score in HCP</u>

As in ABCD, motion censoring reduced motion impact scores in HCP. Only two traits, age and NIH Toolbox anger-aggression subscore, had significant motion overestimation scores at FD < 0.2 mm. Only NIH Toolbox card sorting subscore had significant motion overestimation at FD < 0.1 mm. While the motion overestimation score for age remained highly significant at FD < 0.2 mm ($p < 0.001$), the motion impact scores of the toolbox measures may have been spuriously significant due to the low number of permutations performed ($p = 0.047$).

Motion censoring was more effective at reducing motion underestimation scores in HCP. Besides age, no traits in HCP had a significant motion underestimation score in HCP after censoring at FD < 0.2 mm.

Unlike ABCD, in which motion censoring at FD < 0.2 mm excluded 818 participants due to insufficient remaining data, motion censoring HCP at FD < 0.2 mm did not result in the exclusion of any participants. Motion censoring HCP at FD < 0.1 mm resulted in the exclusion of 89 participants (8.8%), with 928 participants remaining.

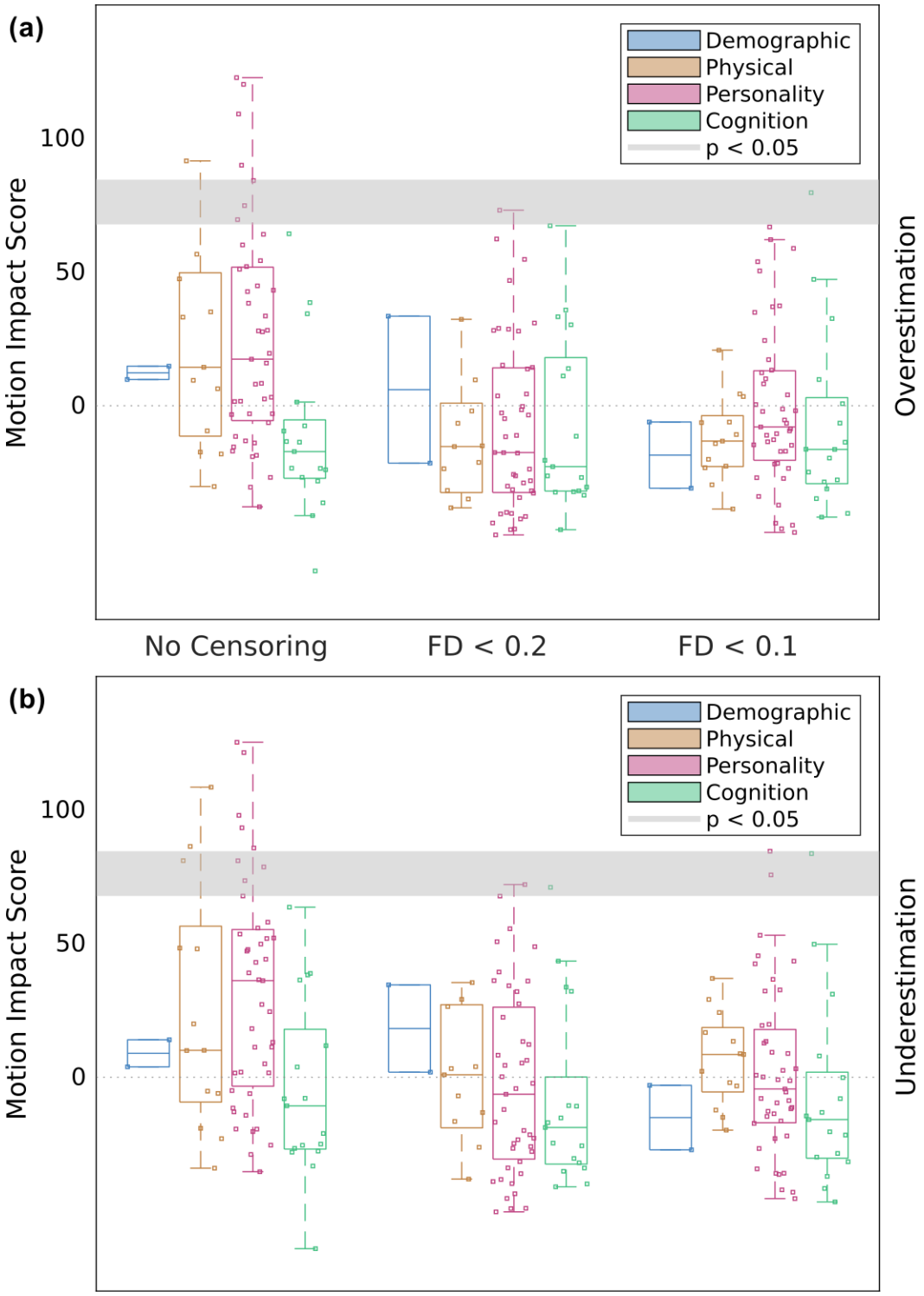

Figure SB.2: Effects of frame censoring on motion impact score in HCP data using FD. Motion impact score (omnibus Stouffer's Z, higher = more motion) for traits of a given category at different levels of motion censoring. The cutoff score for significance at $p < 0.05$ is different for each trait. The range of significance thresholds is indicated by a gray band. **(a)** Motion overestimation score. **(b)** Motion underestimation score.

Variable	Correlation w/FD		SHAMAN p-Value			Shaman Motion Score		
	Pearson	Spearman	Impact	Over	Under	Impact	Over	Under
Age_in_Yrs	0.116	0.102	0.102	0.102	0.102	51.47	47.47	48.37
Height	-0.004	-0.026	0.023	0.188	0.008	91.75	33.15	81.07
Weight	0.501	0.466	0.055	0.008	0.578	73.25	91.64	10.02
BMI	0.623	0.612	0	0	0.016	386.83	416.40	86.40
Handedness	-0.031	0.007	0.469	0.563	0.367	15.23	9.52	19.99
SSAGA_Income	-0.059	-0.074	0.586	0.57	0.648	8.24	9.84	3.92
SSAGA_Educ	-0.215	-0.171	0.391	0.438	0.422	15.79	14.80	14.04
Hematocrit_1	0.007	0.001	0.055	0.063	0.094	64.24	56.72	48.06
ThyroidHormone	-0.057	-0.077	0.906	0.891	0.914	-18.25	-17.34	-19.08
HbA1C	0.19	0.154	0.438	0.414	0.477	12.30	14.42	10.10
MMSE_Score	-0.066	-0.029	0.844	0.844	0.828	-9.43	-9.46	-8.04
PSQI_Score	0.145	0.095	0.82	0.813	0.828	-4.75	-3.28	-4.97
PicSeq_AgeAdj	-0.096	-0.081	0.914	0.938	0.914	-13.01	-13.38	-10.67
CardSort_AgeAdj	-0.133	-0.15	0.055	0.055	0.055	67.20	64.36	63.66
Flanker_AgeAdj	-0.045	-0.101	0.984	0.984	0.992	-28.11	-23.36	-28.04
PMAT24_A_CR	-0.161	-0.176	0.992	0.984	1	-23.24	-17.14	-26.60
PMAT24_A_RTCR	-0.055	-0.072	0.664	0.695	0.648	2.45	1.38	3.85
ReadEng_AgeAdj	-0.165	-0.185	0.414	0.898	0.211	17.76	-13.62	36.41
PicVocab_AgeAdj	-0.164	-0.18	1	1	1	-33.72	-26.79	-25.43
ProcSpeed_AgeAdj	-0.039	-0.056	0.867	0.883	0.898	-7.08	-7.52	-7.86
DDisc_AUC_200	-0.136	-0.156	0.922	0.945	0.922	-15.43	-16.98	-11.50
DDisc_AUC_40K	-0.154	-0.159	0.969	1	0.969	-16.29	-15.42	-12.96
VSPLOT_TC	-0.11	-0.138	0.133	0.172	0.148	40.93	34.42	38.29
VSPLOT_CRTE	-0.047	-0.028	0.148	0.156	0.156	40.92	38.62	38.88
IWRD_TOT	-0.101	-0.066	1	1	1	-46.27	-41.11	-33.21
IWRD_RTC	0.035	-0.001	1	1	1	-64.45	-61.82	-64.13
ListSort_AgeAdj	-0.055	-0.094	1	1	1	-28.80	-28.19	-27.65
CogFluidComp_AgeAdj	-0.113	-0.131	1	1	1	-24.99	-23.44	-24.99
CogTotalComp_AgeAdj	-0.178	-0.21	1	1	0.992	-33.20	-36.33	-21.03
ogCrystalComp_AgeAdj	-0.181	-0.199	0.789	0.984	0.508	-5.88	-23.95	11.83
ER40_CR	-0.096	-0.082	0.68	0.703	0.703	3.34	1.58	1.60
ER40_CRT	0.048	0.078	0.875	0.859	0.93	-16.81	-11.50	-19.35
AngAffect_Unadj	0.08	0.018	0.008	0.008	0.008	127.47	122.78	125.38
AngHostil_Unadj	0.071	0.05	0.008	0.016	0.008	79.51	69.66	81.03
AngAggr_Unadj	0.114	0.096	0.008	0.008	0.008	114.74	109.19	98.02
FearAffect_Unadj	0.021	-0.017	0.094	0.109	0.086	55.18	51.10	53.59
FearSomat_Unadj	0.082	0.049	0.641	0.656	0.633	1.34	1.71	1.94
Sadness_Unadj	0.009	-0.028	0.008	0.008	0.008	96.15	90.02	93.36
LifeSatisf_Unadj	-0.048	-0.015	0.063	0.07	0.039	66.30	60.17	67.86
MeanPurp_Unadj	0.034	0.085	0.008	0.008	0.008	126.89	120.28	121.52
PosAffect_Unadj	-0.004	0.03	0.039	0.039	0.039	77.16	74.90	73.59
Friendship_Unadj	-0.019	0.018	0.961	0.93	0.961	-14.93	-13.13	-14.22
Loneliness_Unadj	-0.046	-0.034	0.102	0.102	0.109	52.99	52.09	47.22
PercHostil_Unadj	0.08	0.081	0.133	0.141	0.125	46.67	42.73	47.80
PercReject_Unadj	0.048	0.031	0.195	0.188	0.18	39.88	38.36	39.09
EmotSupp_Unadj	-0.061	-0.013	0.836	0.789	0.852	-6.14	-3.00	-6.10
InstruSupp_Unadj	-0.02	-0.024	1	1	1	-30.66	-30.47	-28.88
PercStress_Unadj	0.041	0.013	0.43	0.453	0.438	18.64	17.43	18.22
SelfEff_Unadj	0.062	0.068	0.953	0.953	0.961	-19.87	-19.04	-20.31
Dexterity_AgeAdj	-0.102	-0.113	0.883	0.945	0.859	-8.95	-9.39	-5.18
Strength_AgeAdj	0.099	0.069	0.008	0.172	0	124.94	35.13	108.56
NEOFAC_A	-0.161	-0.112	0.008	0	0	89.61	84.28	85.81
NEOFAC_O	-0.117	-0.146	0.719	0.938	0.555	0.26	-13.93	11.27
NEOFAC_C	-0.044	-0.012	0.273	0.578	0.18	30.26	8.07	43.03
NEOFAC_N	0.013	-0.006	0.281	0.305	0.375	25.86	24.11	18.98
NEOFAC_E	0.037	0.019	0.977	0.977	0.977	-19.64	-18.55	-19.38
Noise_Comp	0.09	0.111	1	1	1	-34.30	-30.21	-34.01
Odor_AgeAdj	-0.01	0.004	0.672	0.594	0.781	1.44	6.37	-6.02
Taste_AgeAdj	0.017	0.031	0.953	0.953	0.977	-21.19	-18.05	-22.96
ASR_Anxd_Pct	-0.033	-0.047	0.141	0.18	0.117	53.55	44.87	55.85
ASR_Witd_T	0.004	-0.021	0.211	0.25	0.203	34.04	28.02	36.51
ASR_Soma_T	0.106	0.055	1	1	1	-39.71	-37.82	-35.35
ASR_Thot_T	0.045	0.005	0.133	0.117	0.148	55.55	54.30	49.86
ASR_Attn_T	0.064	0.047	0.359	0.563	0.297	20.88	8.38	27.26
ASR_Aggr_T	0.16	0.086	0.203	0.273	0.195	34.18	27.67	36.18
ASR_Rule_T	0.127	0.167	0.078	0.094	0.063	76.52	64.16	78.72
ASR_Intr_T	0.16	0.112	0.711	0.688	0.68	2.22	2.54	5.04
ASR_Intn_T	0.021	-0.017	0.203	0.25	0.133	40.96	33.59	44.20
ASR_Extn_T	0.185	0.146	0.148	0.414	0.086	41.29	15.97	51.92
ASR_Totp_T	0.096	0.049	0.102	0.258	0.063	51.31	28.24	58.04
DSM_Depr_T	0.018	0.005	0.734	0.836	0.711	-1.82	-6.33	1.65
DSM_Anxi_T	0.028	-0.005	0.367	0.398	0.352	22.62	19.62	24.56
DSM_Somp_T	0.11	0.072	0.992	0.992	0.992	-28.52	-26.82	-25.40
DSM_Avoid_T	-0.062	-0.116	0.492	0.578	0.438	7.67	3.19	11.31
DSM_Adh_T	0.104	0.071	0.539	0.719	0.438	6.94	-2.99	13.09
DSM_Antis_T	0.172	0.18	0.148	0.195	0.141	50.33	43.24	52.12

Table SB.1: Relationship between selected variables in the HCP 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 overestimation scores,

underestimation scores, and impact from both/either type of score. Results in this table reflect data after motion-reduction and without additional motion censoring.

Variable	SHAMAN p-Value								
	No Censoring			FD < 0.2 mm			FD < 0.1 mm		
	Impact	Over	Under	Impact	Over	Under	Impact	Over	Under
Age_in_Yrs	0.102	0.102	0.102	0	0	0	0.688	0.805	0.648
Height	0.023	0.188	0.008	0.938	0.984	0.742	0.766	1	0.414
Weight	0.055	0.008	0.578	0.734	1	0.289	0.5	0.977	0.289
BMI	0	0	0.016	0.945	1	0.633	0.445	0.992	0.234
Handedness	0.469	0.563	0.367	0.859	0.844	0.867	0.844	0.867	0.789
SSAGA_Income	0.586	0.57	0.648	0.242	0.258	0.227	0.742	0.773	0.703
SSAGA_Educ	0.391	0.438	0.422	0.859	0.945	0.625	0.984	0.992	0.977
Hematocrit_1	0.055	0.063	0.094	0.781	0.797	0.836	0.398	0.414	0.391
ThyroidHormone	0.906	0.891	0.914	0.336	0.328	0.352	0.781	0.773	0.797
HbA1C	0.438	0.414	0.477	1	1	1	0.781	0.789	0.773
MMSE_Score	0.844	0.844	0.828	0.977	0.969	0.977	0.82	0.844	0.828
PSQI_Score	0.82	0.813	0.828	1	1	1	0.859	0.852	0.875
PicSeq_AgeAdj	0.914	0.938	0.914	0.953	0.992	0.93	0.914	0.977	0.875
CardSort_AgeAdj	0.055	0.055	0.055	0.078	0.086	0.055	0.039	0.047	0.039
Flanker_AgeAdj	0.984	0.984	0.992	0.992	0.984	0.992	0.148	0.164	0.141
PMAT24_A_CR	0.992	0.984	1	1	1	1	0.977	0.992	0.969
PMAT24_A_RTCR	0.664	0.695	0.648	0.156	0.25	0.133	0.508	0.492	0.531
ReadEng_AgeAdj	0.414	0.898	0.211	1	1	0.906	1	1	0.867
PicVocab_AgeAdj	1	1	1	0.875	0.469	1	1	0.992	1
ProcSpeed_AgeAdj	0.867	0.883	0.898	0.195	0.203	0.211	1	0.984	1
DDisc_AUC_200	0.922	0.945	0.922	0.219	0.258	0.195	0.188	0.219	0.18
DDisc_AUC_40K	0.969	1	0.969	0.922	0.953	0.93	0.57	0.609	0.594
VSPLOT_TC	0.133	0.172	0.148	0.633	0.469	0.883	0.984	0.992	0.969
VSPLOT_CRTE	0.148	0.156	0.156	0.258	0.305	0.273	0.25	0.25	0.242
IWRD_TOT	1	1	1	1	1	0.992	1	0.953	1
IWRD_RTC	1	1	1	0.805	0.828	0.805	0.93	0.922	0.945
ListSort_AgeAdj	1	1	1	1	1	1	0.797	0.789	0.82
CogFluidComp_AgeAdj	1	1	1	0.992	1	0.977	0.664	0.656	0.664
CogTotalComp_AgeAdj	1	1	0.992	1	1	1	0.977	0.898	0.984
ogCrystalComp_AgeAdj	0.789	0.984	0.508	1	1	1	1	1	1
ER40_CR	0.68	0.703	0.703	1	1	1	0.984	0.977	0.992
ER40_CRT	0.875	0.859	0.93	0.133	0.117	0.156	0.172	0.141	0.195
AngAffect_Unadj	0.008	0.008	0.008	0.547	0.57	0.516	0.992	1	0.992
AngHostil_Unadj	0.008	0.016	0.008	0.211	0.281	0.172	0.148	0.148	0.148
AngAggr_Unadj	0.008	0.008	0.008	0.07	0.047	0.078	0.852	0.859	0.852
FearAffect_Unadj	0.094	0.109	0.086	1	1	1	0.398	0.453	0.359
FearSomat_Unadj	0.641	0.656	0.633	0.703	0.727	0.68	0.656	0.688	0.641
Sadness_Unadj	0.008	0.008	0.008	0.43	0.508	0.375	0.492	0.523	0.469
LifeSatisf_Unadj	0.063	0.07	0.039	0.867	0.938	0.836	0.281	0.297	0.273
MeanPurp_Unadj	0.008	0.008	0.008	0.68	0.727	0.656	0.477	0.516	0.469
PosAffect_Unadj	0.039	0.039	0.039	0.82	0.875	0.805	0.773	0.758	0.836
Friendship_Unadj	0.961	0.93	0.961	1	1	1	0.875	0.875	0.867
Loneliness_Unadj	0.102	0.102	0.109	1	1	1	0.359	0.406	0.359
PercHostil_Unadj	0.133	0.141	0.125	0.219	0.266	0.211	0.063	0.109	0.055
PercReject_Unadj	0.195	0.188	0.18	0.133	0.148	0.117	0.078	0.086	0.07
EmotSupp_Unadj	0.836	0.789	0.852	1	1	1	0.477	0.438	0.477
InstruSupp_Unadj	1	1	1	1	1	1	0.219	0.219	0.227
PercStress_Unadj	0.43	0.453	0.438	0.992	1	0.992	0.859	0.852	0.867
SelfEff_Unadj	0.953	0.953	0.961	0.992	0.992	0.992	0.938	0.93	0.938
Dexterity_AgeAdj	0.883	0.945	0.859	1	1	1	0.773	0.789	0.742
Strength_AgeAdj	0.008	0.172	0	0.219	0.805	0.195	0.898	1	0.453
NEOFAC_A	0.008	0	0	1	1	1	1	1	1
NEOFAC_O	0.719	0.938	0.555	0.898	0.969	0.672	0.836	0.828	0.805
NEOFAC_C	0.273	0.578	0.18	0.484	0.867	0.289	0.828	0.945	0.688
NEOFAC_N	0.281	0.305	0.375	0.82	0.883	0.836	0.859	0.844	0.883
NEOFAC_E	0.977	0.977	0.977	0.945	0.922	0.969	1	1	1
Noise_Comp	1	1	1	0.531	0.555	0.711	0.742	0.82	0.727
Odor_AgeAdj	0.672	0.594	0.781	0.961	0.961	0.992	0.578	0.594	0.5
Taste_AgeAdj	0.953	0.953	0.977	0.898	0.93	0.898	0.594	0.641	0.547
ASR_Anxd_Pct	0.141	0.18	0.117	0.266	0.258	0.289	0.266	0.234	0.328
ASR_Witd_T	0.211	0.25	0.203	0.992	0.984	0.984	0.82	0.836	0.813
ASR_Soma_T	1	1	1	1	1	0.992	1	1	0.992
ASR_Thot_T	0.133	0.117	0.148	0.641	0.68	0.594	0.602	0.633	0.594
ASR_Attn_T	0.359	0.563	0.297	0.547	0.688	0.445	0.961	0.953	0.984
ASR_Aggr_T	0.203	0.273	0.195	0.883	0.93	0.844	0.672	0.711	0.672
ASR_Rule_T	0.078	0.094	0.063	0.102	0.109	0.102	0.82	0.914	0.734
ASR_Intr_T	0.711	0.688	0.68	1	1	1	0.82	0.75	0.906
ASR_Intn_T	0.203	0.25	0.133	0.586	0.516	0.547	0.516	0.57	0.477
ASR_Extn_T	0.148	0.414	0.086	0.266	0.438	0.195	0.898	0.977	0.758
ASR_Totp_T	0.102	0.258	0.063	0.578	0.75	0.508	0.703	0.75	0.672
DSM_Depr_T	0.734	0.836	0.711	0.914	0.93	0.914	0.758	0.734	0.797
DSM_Anxi_T	0.367	0.398	0.352	0.969	0.969	0.945	0.836	0.852	0.828
DSM_Somp_T	0.992	0.992	0.992	0.969	1	0.977	0.992	1	0.992
DSM_Avoid_T	0.492	0.578	0.438	0.531	0.453	0.656	0.18	0.195	0.227
DSM_Adh_T	0.539	0.719	0.438	0.977	1	0.969	1	1	1
DSM_Antis_T	0.148	0.195	0.141	0.172	0.258	0.133	0.563	0.664	0.539

Table SB.2: Significance (p-values) of motion impact score for selected variables in the HCP study after motion-reduction and with (varying from left to right) no motion censoring, censoring at framewise displacement (FD) < 0.2 mm, and FD < 0.1. Omnibus p-values are given for overestimation scores, underestimation scores, and impact from both/either type of score. The omnibus p-values account for multiple comparisons across edges, but they are not corrected for the multiple comparisons across the 76 different traits shown.

Variable	SHAMAN Motion Score								
	No Censoring			FD < 0.2 mm			FD < 0.1 mm		
	Impact	Over	Under	Impact	Over	Under	Impact	Over	Under
Age_in_Yrs	51.5	47.5	48.4	355.3	313.4	338.9	-2.1	-6.3	2.2
Height	91.7	33.2	81.1	-20.3	-23.6	0.9	-5.7	-23.2	16.8
Weight	73.3	91.6	10.0	1.1	-31.7	26.5	12.2	-20.0	29.1
BMI	386.8	416.4	86.4	-21.4	-38.2	3.3	11.9	-29.6	37.0
Handedness	15.2	9.5	20.0	-17.5	-15.3	-16.5	-15.7	-14.2	-12.3
SSAGA_Income	8.2	9.8	3.9	36.1	33.5	34.6	-4.4	-6.1	-3.0
SSAGA_Educ	15.8	14.8	14.0	-11.0	-21.5	1.9	-33.0	-30.8	-27.1
Hematocrit_1	64.2	56.7	48.1	-8.2	-6.6	-7.0	23.7	20.8	24.2
ThyroidHormone	-18.3	-17.3	-19.1	31.9	32.3	29.2	-14.9	-13.2	-15.0
HbA1C	12.3	14.4	10.1	-107.1	-102.2	-103.4	-22.0	-22.6	-19.8
MMSE_Score	-9.4	-9.5	-8.0	-20.8	-20.4	-18.8	-15.7	-16.3	-14.5
PSQI_Score	-4.8	-3.3	-5.0	-45.1	-43.9	-39.0	-18.6	-14.6	-17.2
PicSeq_AgeAdj	-13.0	-13.4	-10.7	-23.0	-26.2	-17.0	-20.7	-24.8	-15.8
CardSort_AgeAdj	67.2	64.4	63.7	69.9	67.4	71.1	84.5	79.7	83.7
Flanker_AgeAdj	-28.1	-23.4	-28.0	-27.2	-22.8	-24.6	48.8	47.3	49.8
PMAT24_A_CR	-23.2	-17.1	-26.6	-40.8	-32.3	-41.0	-35.2	-34.7	-29.9
PMAT24_A_RTCR	2.4	1.4	3.9	41.8	33.3	43.5	9.6	9.8	8.0
ReadEng_AgeAdj	17.8	-13.6	36.4	-40.2	-46.4	-15.2	-35.2	-41.7	-13.1
PicVocab_AgeAdj	-33.7	-26.8	-25.4	-12.6	11.1	-35.2	-43.1	-28.5	-41.6
ProcSpeed_AgeAdj	-7.1	-7.5	-7.9	38.2	35.8	33.7	-35.9	-31.1	-37.1
DDisc_AUC_200	-15.4	-17.0	-11.5	35.3	28.2	36.2	42.9	35.0	42.5
DDisc_AUC_40K	-16.3	-15.4	-13.0	-18.8	-17.5	-16.8	3.1	0.4	0.8
VSPLOT_TC	40.9	34.4	38.3	0.4	13.9	-10.6	-25.2	-19.6	-20.4
VSPLOT_CRTE	40.9	38.6	38.9	32.9	30.3	32.2	33.0	32.6	31.1
IWRD_TOT	-46.3	-41.1	-33.2	-43.9	-32.2	-30.4	-37.5	-16.3	-46.6
IWRD_RTC	-64.4	-61.8	-64.1	-11.2	-11.4	-10.8	-29.0	-27.7	-28.5
ListSort_AgeAdj	-28.8	-28.2	-27.6	-33.5	-31.9	-32.0	-7.8	-6.5	-8.0
CogFluidComp_AgeAdj	-25.0	-23.4	-25.0	-28.4	-26.9	-25.6	-0.7	0.8	-0.1
CogTotalComp_AgeAdj	-33.2	-36.3	-21.0	-40.3	-33.5	-33.9	-23.7	-13.7	-21.6
ogCrystalComp_AgeAdj	-5.9	-24.0	11.8	-48.3	-30.4	-39.9	-48.9	-40.3	-31.6
ER40_CR	3.3	1.6	1.6	-51.2	-48.3	-50.4	-33.9	-26.9	-34.3
ER40_CRT	-16.8	-11.5	-19.3	59.4	62.4	50.8	53.4	53.9	45.5
AngAffect_Unadj	127.5	122.8	125.4	5.0	3.7	8.3	-31.4	-33.9	-26.7
AngHostil_Unadj	79.5	69.7	81.0	37.2	29.0	39.4	53.4	50.5	53.2
AngAggr_Unadj	114.7	109.2	98.0	74.1	73.2	67.8	-8.6	-7.9	-7.9
FearAffect_Unadj	55.2	51.1	53.6	-41.7	-40.5	-38.2	17.2	12.3	19.4
FearSomat_Unadj	1.3	1.7	1.9	-1.5	-2.7	0.2	-0.6	-2.1	0.1
Sadness_Unadj	96.2	90.0	93.4	19.6	15.2	22.5	11.5	8.2	12.8
LifeSatisf_Unadj	66.3	60.2	67.9	-14.9	-17.5	-12.1	29.8	24.4	32.3
MeanPurp_Unadj	126.9	120.3	121.5	1.6	-4.8	4.5	11.7	10.1	13.4
PosAffect_Unadj	77.2	74.9	73.6	-9.0	-11.6	-6.3	-12.9	-11.0	-14.6
Friendship_Unadj	-14.9	-13.1	-14.2	-45.8	-39.9	-45.3	-14.2	-13.5	-12.7
Loneliness_Unadj	53.0	52.1	47.2	-34.2	-30.1	-33.9	20.1	17.3	19.8
PercHostil_Unadj	46.7	42.7	47.8	32.6	28.7	34.3	82.1	66.9	84.6
PercReject_Unadj	39.9	38.4	39.1	52.9	46.9	55.6	71.7	62.2	75.7
EmotSupp_Unadj	-6.1	-3.0	-6.1	-51.1	-46.3	-49.1	11.3	13.3	9.4
InstruSupp_Unadj	-30.7	-30.5	-28.9	-41.6	-40.3	-39.8	37.4	37.0	36.7
PercStress_Unadj	18.6	17.4	18.2	-31.0	-31.4	-26.6	-13.8	-12.7	-13.6
SelfEff_Unadj	-19.9	-19.0	-20.3	-25.7	-25.7	-24.8	-23.7	-21.8	-22.9
Dexterity_AgeAdj	-8.9	-9.4	-5.2	-41.9	-34.9	-38.1	-6.5	-6.1	-2.0
Strength_AgeAdj	124.9	35.1	108.6	35.5	-1.9	35.4	-19.5	-38.6	13.5
NEOFAC_A	89.6	84.3	85.8	-49.9	-46.1	-43.6	-42.2	-44.0	-36.0
NEOFAC_O	0.3	-13.9	11.3	-17.3	-26.2	-4.2	-12.8	-10.5	-9.6
NEOFAC_C	30.3	8.1	43.0	13.5	-11.1	32.1	-13.9	-21.5	-0.9
NEOFAC_N	25.9	24.1	19.0	-13.0	-13.8	-11.9	-17.5	-15.6	-17.4
NEOFAC_E	-19.6	-18.6	-19.4	-21.1	-17.7	-23.4	-38.4	-37.2	-36.3
Noise_Comp	-34.3	-30.2	-34.0	10.2	9.7	4.0	-5.8	-10.8	-3.2
Odor_AgeAdj	1.4	6.4	-6.0	-24.5	-21.2	-26.2	4.4	4.4	8.8
Taste_AgeAdj	-21.2	-18.0	-23.0	-14.4	-15.0	-13.2	6.1	3.5	8.6
ASR_Anxd_Pct	53.6	44.9	55.8	29.1	27.9	27.5	37.5	37.4	32.7
ASR_Witd_T	34.0	28.0	36.5	-35.3	-34.4	-31.6	-18.5	-17.1	-16.4
ASR_Soma_T	-39.7	-37.8	-35.3	-41.6	-42.3	-36.1	-45.2	-46.0	-42.1
ASR_Thot_T	55.6	54.3	49.9	2.5	-1.5	5.3	0.9	-1.2	2.6
ASR_Attn_T	20.9	8.4	27.3	6.4	-0.3	13.4	-32.7	-27.2	-35.9
ASR_Aggr_T	34.2	27.7	36.2	-25.0	-28.9	-19.9	-5.0	-5.3	-4.4
ASR_Rule_T	76.5	64.2	78.7	70.6	54.8	72.1	-11.7	-17.0	-5.5
ASR_Intr_T	2.2	2.5	5.0	-50.0	-41.4	-49.0	-15.7	-6.6	-22.0
ASR_Intn_T	41.0	33.6	44.2	3.7	4.4	6.4	7.3	4.1	8.9
ASR_Extn_T	41.3	16.0	51.9	29.6	13.8	36.0	-21.7	-23.4	-8.7
ASR_Totp_T	51.3	28.2	58.0	4.5	-3.5	12.3	-6.4	-9.4	-1.3
DSM_Depr_T	-1.8	-6.3	1.7	-23.4	-23.7	-21.5	-9.9	-8.6	-11.8
DSM_Anxi_T	22.6	19.6	24.6	-28.3	-28.1	-25.9	-13.5	-14.8	-11.3
DSM_Somp_T	-28.5	-26.8	-25.4	-31.5	-31.8	-27.8	-45.4	-44.7	-43.0
DSM_Avoid_T	7.7	3.2	11.3	6.7	14.3	-2.2	60.5	58.9	43.5
DSM_Adh_T	6.9	-3.0	13.1	-30.0	-32.7	-22.8	-48.5	-47.4	-45.4
DSM_Antis_T	50.3	43.2	52.1	45.1	31.0	48.9	1.3	-1.9	3.2

Table SB.3: Motion impact scores (omnibus Stouffer's Z) for selected variables in the HCP study after motion-reduction and with (varying from left to right) no motion censoring, censoring at framewise displacement (FD) < 0.2 mm, and FD < 0.1. Higher scores correspond to a greater impact of motion. Motion overestimation scores, underestimation scores, and overall impact (both/either over/underestimation) are shown separately.

Motion Impact Scores Were Similar for FD and DVARS in ABCD

Motion impact scores, and the effect of motion censoring, were comparable for FD and DVARS in ABCD data. Using DVARS, 82% (37/45) of traits had a significant motion overestimation score and 13% (6/45) had a significant motion underestimation score without motion censoring. After motion censoring at DVARS < 200, no traits had a significant motion overestimation score and 18% (8/45) of traits had a significant motion underestimation score.

Whereas ABCD + DVARS and HCP + FD yielded similar results to our main analysis with ABCD + FD, HCP + DVARS failed to detect any significant motion impact scores, either with or without censoring (Figure SB.4, Table, SB.7-9). As with a threshold of FD < 0.2 mm, the DVARS < 800 threshold did not exclude any participants in HCP.

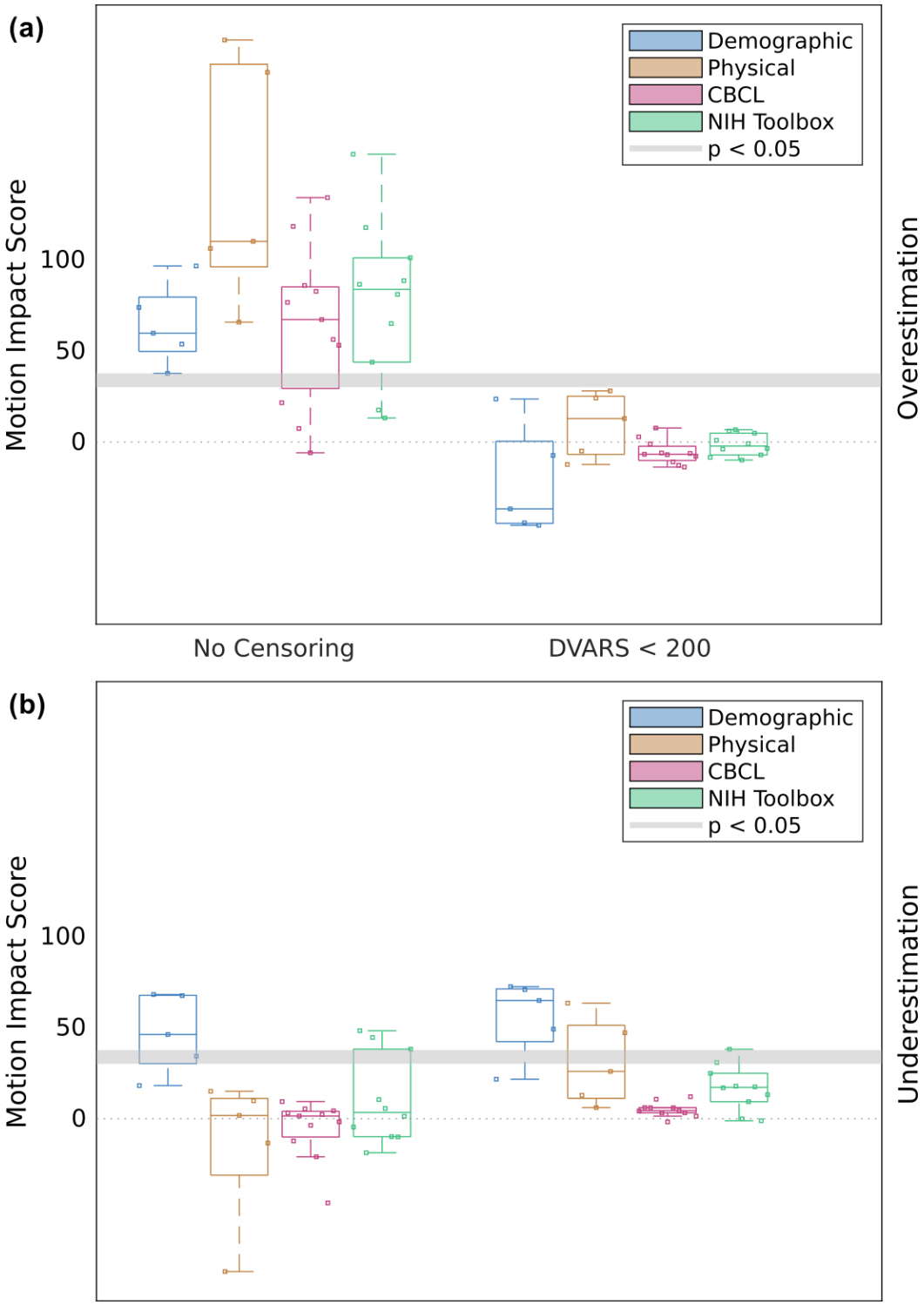

Figure SB.3: Effects of frame censoring on motion impact score in ABCD data using DVARS. Motion impact score (omnibus Stouffer's Z, higher = more motion) for traits of a given category at different levels of motion censoring. The cutoff score for significance at $p < 0.05$ is different for each trait. The range of significance thresholds is indicated by a gray band. **(a)** Motion overestimation score. **(b)** Motion underestimation score.

Variable	Correlation w/DVARS		SHAMAN p-Value			Shaman Motion Score		
	Pearson	Spearman	Impact	Over	Under	Impact	Over	Under
male	0.016	0.089	0	0	0.57	171.10	106.32	15.07
age_months	-0.019	-0.153	0	0	1	178.07	220.76	-83.90
height	-0.017	-0.06	0	0	1	54.03	65.84	1.79
weight	0.006	0.061	0	0	0.836	125.69	110.19	9.88
bmi	0.017	0.123	0	0	1	221.09	203.04	-13.35
home_roster	-0.008	-0.052	0.023	0	0.359	61.90	73.94	18.22
parental_income	-0.029	-0.123	0	0	0	175.87	59.82	68.23
parents_married	-0.005	-0.109	0	0.063	0.008	110.84	37.73	46.25
parental_education	-0.034	-0.119	0	0	0	158.91	53.78	67.61
residential_deprivation	0	0.097	0	0	0.07	163.32	96.67	34.31
mri_runs	-0.009	0.001	0	0	1	82.21	100.09	-14.79
sleepy	-0.014	0.032	0.125	0.016	0.977	24.40	34.23	2.60
nihtbx_total	-0.01	-0.128	0	0	1	141.73	158.07	-4.54
nihtbx_cryst	0.002	-0.075	0	0	0	92.71	86.61	48.34
nihtbx_fluid	-0.017	-0.124	0	0	1	96.39	117.78	-18.65
nihtbx_picvocab	0	-0.061	0	0.008	0.008	57.38	43.88	44.64
nihtbx_flanker	-0.006	-0.051	0.484	0.477	0.906	17.66	17.63	10.55
nihtbx_list	-0.014	-0.055	0.805	0.758	1	12.54	13.24	5.66
nihtbx_cardsort	-0.003	-0.093	0	0	1	47.23	65.07	-9.82
nihtbx_pattern	-0.01	-0.104	0	0	1	61.67	81.09	-9.95
nihtbx_picture	-0.021	-0.089	0	0	1	71.82	88.62	1.33
nihtbx_reading	0.001	-0.072	0	0	0.008	95.79	101.18	38.22
wiscv_matrix	-0.021	-0.127	0	0	0.914	127.00	114.82	4.10
cbcl_internal	0.018	0.003	0.43	0.227	0.938	18.26	21.60	9.45
cbcl_external	0.035	0.078	0	0	0.984	62.81	76.72	3.15
cbcl_total	0.032	0.085	0	0	1	93.35	118.42	-12.17
cbcl_anxdep	0.023	-0.01	0.992	0.984	1	4.94	7.49	1.51
cbcl_withdep	0.02	0.041	0	0	0.891	69.27	85.99	5.44
cbcl_somatic	0.015	-0.023	1	1	1	-5.72	-5.88	-3.60
cbcl_social	0.034	0.09	0.008	0	1	57.07	82.79	-20.87
cbcl_thought	0.02	0.051	0	0	0.984	50.68	67.27	2.31
cbcl_attention	0.044	0.118	0	0	1	99.54	134.21	-46.20
cbcl_rulebreak	0.033	0.098	0	0.008	0.945	47.48	56.34	4.39
cbcl_aggressive	0.024	0.068	0	0	1	37.87	53.15	-1.68
pps_yes_num	0.013	0.123	0	0	1	107.73	137.88	-50.74
pps_severity	0.007	0.123	0	0	1	92.95	116.11	-32.78
bis	-0.005	0.017	1	1	1	-3.00	2.52	-9.32
bas_rr	0.02	0.068	0.148	0.016	1	25.23	40.07	-7.46
bas_drive	0.012	0.079	0	0	1	60.83	79.69	-11.64
bas_fs	0.011	0.064	0.141	0	1	22.93	37.43	-4.71
upps_perserverance	0.008	0.05	0	0	1	58.65	75.91	1.42
upps_planning	0.019	0.048	0	0	1	64.54	82.70	4.07
upps_seeking	0.009	0.005	0.742	0.734	0.922	13.17	13.22	9.52
upps_pos_urgency	0.016	0.091	0	0	1	57.00	83.78	-34.41
upps_neg_urgency	0.017	0.077	0	0	1	57.04	77.36	-12.79

Table SB.4: Relationship between selected variables in the ABCD study with head motion (DVARS) 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 overestimation scores, underestimation scores, and impact from both/either type of score. Results in this table reflect data after motion-reduction with ABCD-BIDS and without additional motion censoring.

Variable	SHAMAN p-Value					
	No Censoring			DVARS < 200		
	Impact	Over	Under	Impact	Over	Under
male	0	0	0.57	0	1	0
age_months	0	0	1	0.992	1	0.75
height	0	0	1	0.336	0.172	0.977
weight	0	0	0.836	0	0.109	0.18
bmi	0	0	1	0	0.719	0
home_roster	0.023	0	0.359	0.156	0.172	0.266
parental_income	0	0	0	0	1	0
parents_married	0	0.063	0.008	0	1	0
parental_education	0	0	0	0.008	1	0
residential_deprivation	0	0	0.07	0	1	0
mri_runs	0	0	1	0.094	0.945	0.094
sleepy	0.125	0.016	0.977	1	1	1
nihtbx_total	0	0	1	0.875	1	0.156
nihtbx_cryst	0	0	0	0.266	1	0.031
nihtbx_fluid	0	0	1	0.875	1	0.508
nihtbx_picvocab	0	0.008	0.008	0.039	0.922	0.016
nihtbx_flanker	0.484	0.477	0.906	0.656	0.977	0.453
nihtbx_list	0.805	0.758	1	1	1	1
nihtbx_cardsort	0	0	1	0.93	1	0.906
nihtbx_pattern	0	0	1	0.773	0.984	0.453
nihtbx_picture	0	0	1	1	1	1
nihtbx_reading	0	0	0.008	0.977	1	0.758
wiscv_matrix	0	0	0.914	0.961	1	0.406
cbcl_internal	0.43	0.227	0.938	0.992	1	0.992
cbcl_external	0	0	0.984	1	1	0.953
cbcl_total	0	0	1	1	1	0.977
cbcl_anxdep	0.992	0.984	1	0.891	0.93	0.875
cbcl_withdep	0	0	0.891	1	1	1
cbcl_somatic	1	1	1	1	1	1
cbcl_social	0.008	0	1	1	1	0.984
cbcl_thought	0	0	0.984	1	1	0.992
cbcl_attention	0	0	1	1	1	1
cbcl_rulebreak	0	0.008	0.945	0.984	1	0.781
cbcl_aggressive	0	0	1	1	1	1
pps_yes_num	0	0	1	0.961	0.953	0.945
pps_severity	0	0	1	1	1	0.82
bis	1	1	1	1	1	1
bas_rr	0.148	0.016	1	0.664	0.352	1
bas_drive	0	0	1	1	1	0.711
bas_fs	0.141	0	1	1	1	1
upps_perserverance	0	0	1	0.563	0.727	0.57
upps_planning	0	0	1	0.977	0.961	0.984
upps_seeking	0.742	0.734	0.922	0.977	0.992	0.969
upps_pos_urgency	0	0	1	0.992	1	0.859
upps_neg_urgency	0	0	1	0.938	1	0.891

Table SB.5: Significance (p-values) of motion impact score for selected variables in the ABCD study after motion-reduction with ABCD-BIDS and with (varying from left to right) no motion censoring and censoring at DVARS < 200. Omnibus p-values are given for overestimation scores, underestimation scores, and impact from both/either type of score. The omnibus p-values account for multiple comparisons across edges, but they are not corrected for the multiple comparisons across the 45 different traits shown.

Variable	SHAMAN Motion Score					
	No Censoring			DVARS < 200		
	Impact	Over	Under	Impact	Over	Under
male	171.1	106.3	15.1	64.2	-12.3	63.5
age_months	178.1	220.8	-83.9	4.5	-4.9	12.9
height	54.0	65.8	1.8	19.6	24.2	6.2
weight	125.7	110.2	9.9	53.4	28.1	26.0
bmi	221.1	203.0	-13.3	64.2	12.9	47.2
home_roster	61.9	73.9	18.2	24.3	23.7	21.7
parental_income	175.9	59.8	68.2	53.1	-36.7	72.5
parents_married	110.8	37.7	46.3	41.0	-44.3	70.9
parental_education	158.9	53.8	67.6	40.1	-45.7	64.9
residential_deprivation	163.3	96.7	34.3	52.5	-7.3	49.2
mri_runs	82.2	100.1	-14.8	26.9	7.1	31.3
sleepy	24.4	34.2	2.6	-2.6	-0.6	-2.9
nihtbx_total	141.7	158.1	-4.5	11.1	-8.4	25.0
nihtbx_cryst	92.7	86.6	48.3	22.2	1.0	30.9
nihtbx_fluid	96.4	117.8	-18.6	9.9	-3.9	17.0
nihtbx_picvocab	57.4	43.9	44.6	30.2	6.0	38.1
nihtbx_flanker	17.7	17.6	10.5	15.0	6.8	17.9
nihtbx_list	12.5	13.2	5.7	-5.2	-9.9	0.0
nihtbx_cardsort	47.2	65.1	-9.8	7.8	-0.8	9.4
nihtbx_pattern	61.7	81.1	-10.0	13.4	4.9	17.5
nihtbx_picture	71.8	88.6	1.3	-5.4	-7.0	-1.1
nihtbx_reading	95.8	101.2	38.2	6.0	-3.5	13.2
wiscv_matrix	127.0	114.8	4.1	6.3	-14.9	18.4
cbcl_internal	18.3	21.6	9.4	3.2	2.9	4.2
cbcl_external	62.8	76.7	3.2	0.7	-6.7	6.2
cbcl_total	93.4	118.4	-12.2	2.7	-1.0	6.1
cbcl_anxdep	4.9	7.5	1.5	9.9	7.7	10.7
cbcl_withdep	69.3	86.0	5.4	-0.5	-6.0	3.3
cbcl_somatic	-5.7	-5.9	-3.6	-4.8	-6.9	-1.7
cbcl_social	57.1	82.8	-20.9	-2.6	-10.9	6.1
cbcl_thought	50.7	67.3	2.3	-4.6	-12.7	4.5
cbcl_attention	99.5	134.2	-46.2	-4.8	-13.7	3.3
cbcl_rulebreak	47.5	56.3	4.4	5.4	-6.2	12.1
cbcl_aggressive	37.9	53.2	-1.7	-4.2	-7.8	1.4
pps_yes_num	107.7	137.9	-50.7	8.4	3.8	3.9
pps_severity	93.0	116.1	-32.8	2.2	-7.0	9.6
bis	-3.0	2.5	-9.3	0.0	0.2	-0.4
bas_rr	25.2	40.1	-7.5	14.0	20.3	1.1
bas_drive	60.8	79.7	-11.6	-0.2	-13.5	11.6
bas_fs	22.9	37.4	-4.7	-0.2	-2.3	2.4
upps_perserverance	58.7	75.9	1.4	16.8	14.2	16.5
upps_planning	64.5	82.7	4.1	7.9	8.3	4.8
upps_seeking	13.2	13.2	9.5	4.5	3.8	5.3
upps_pos_urgency	57.0	83.8	-34.4	5.7	-3.3	10.1
upps_neg_urgency	57.0	77.4	-12.8	8.7	4.3	10.3

Table SB.6: Motion impact scores (omnibus Stouffer's Z) for selected variables in the ABCD study after motion-reduction with ABCD-BIDS and with (varying from left to right) no motion censoring and censoring at DVARS < 200. Higher scores correspond to a greater impact of motion. Motion overestimation scores, underestimation scores, and overall impact (both/either over/underestimation) are shown separately.

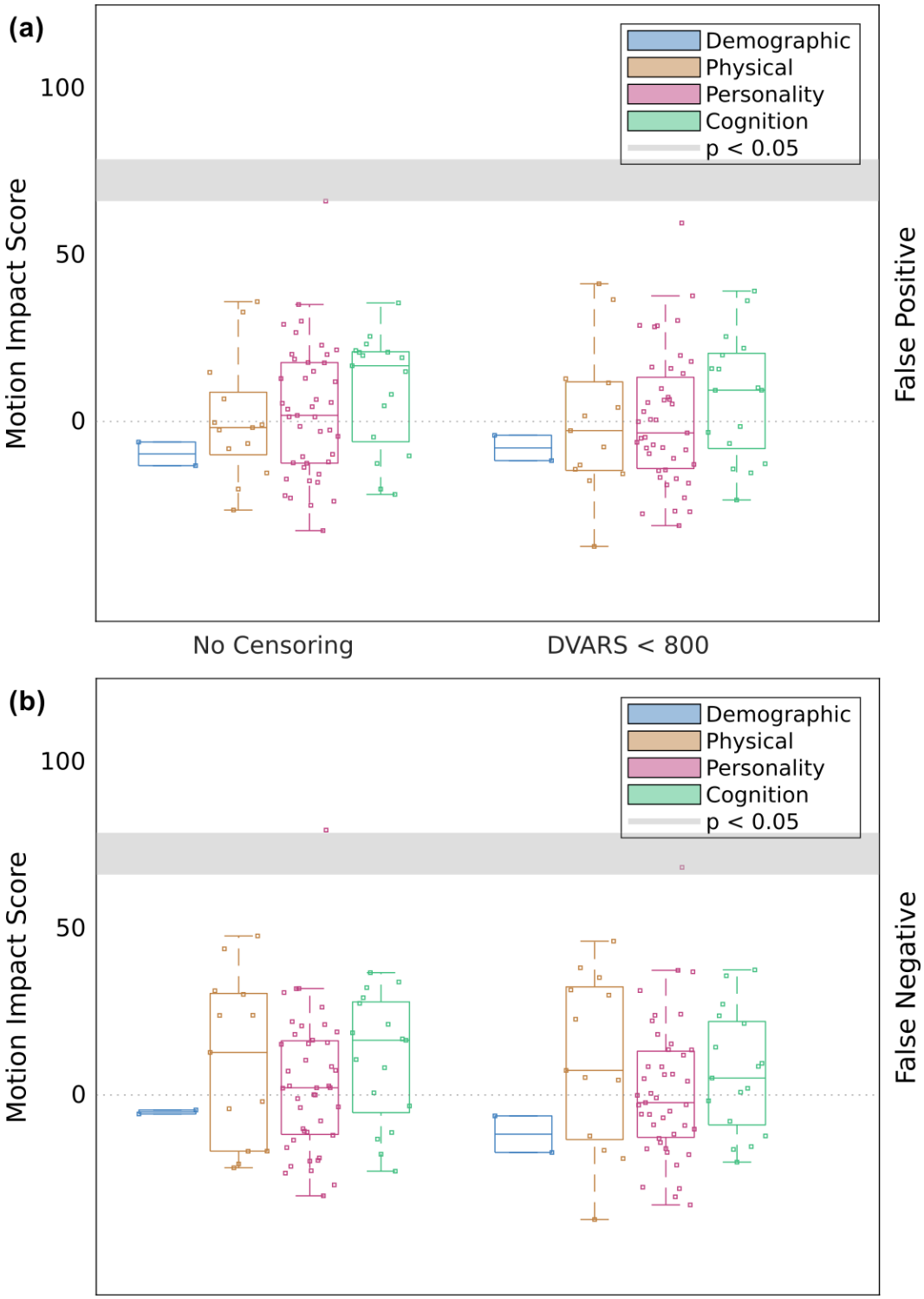

Figure SB.4: Effects of frame censoring on motion impact score in HCP data using DVARS. Motion impact score (omnibus Stouffer's Z, higher = more motion) for traits of a given category at different levels of motion censoring. The cutoff score for significance at $p < 0.05$ is different for each trait. The range of significance thresholds is indicated by a gray band. **(a)** Motion overestimation score. **(b)** Motion underestimation score.

Variable	Correlation w/DVARS		SHAMAN p-Value			Shaman Motion Score		
	Pearson	Spearman	Impact	Over	Under	Impact	Over	Under
Age_in_Yrs	-0.03	-0.021	0.477	0.469	0.5	13.66	14.74	12.78
Height	0.309	0.343	0.25	0.844	0.281	33.01	-0.32	31.28
Weight	0.476	0.502	0.492	0.883	0.367	15.48	-2.54	23.88
BMI	0.392	0.392	0.203	0.664	0.125	36.53	6.77	43.87
Handedness	-0.054	-0.118	0.789	0.828	0.758	-7.19	-8.17	-4.13
SSAGA_Income	-0.057	-0.065	0.789	0.828	0.797	-5.27	-6.19	-5.68
SSAGA_Educ	-0.073	-0.074	0.852	0.945	0.813	-8.97	-13.27	-4.49
Hematocrit_1	0.205	0.212	0.984	0.992	0.977	-27.86	-26.58	-21.85
ThyroidHormone	-0.06	-0.031	0.938	0.938	0.93	-21.72	-20.29	-20.70
HbA1C	0.079	0.043	0.297	0.305	0.305	32.70	32.83	30.21
MMSE_Score	-0.052	-0.026	0.383	0.43	0.406	19.16	16.72	18.67
PSQI_Score	0.077	0.074	0.453	0.477	0.477	14.95	12.87	15.24
PicSeq_AgeAdj	-0.186	-0.19	0.375	0.32	0.484	16.51	21.28	10.64
CardSort_AgeAdj	-0.062	-0.04	0.281	0.344	0.266	25.78	20.74	27.52
Flanker_AgeAdj	0.009	0.021	0.359	0.391	0.266	25.46	19.79	29.20
PMAT24_A_CR	-0.011	-0.021	0.258	0.352	0.25	31.05	23.24	32.22
PMAT24_A_RTCR	-0.014	0.002	0.25	0.328	0.195	32.91	25.53	36.72
ReadEng_AgeAdj	-0.028	-0.017	0.797	0.844	0.75	-3.92	-4.70	0.65
PicVocab_AgeAdj	0.04	0.048	0.938	0.906	0.945	-15.35	-12.62	-13.23
ProcSpeed_AgeAdj	0.005	0.002	0.984	0.984	0.984	-20.36	-20.30	-17.76
DDisc_AUC_200	0.027	-0.005	0.594	0.609	0.641	4.43	5.45	2.05
DDisc_AUC_40K	0.006	0.009	0.219	0.266	0.242	34.08	29.14	30.75
VSPLOT_TC	0.086	0.077	0.664	0.695	0.633	5.75	4.69	8.19
VSPLOT_CRTE	-0.099	-0.076	0.391	0.398	0.391	21.61	20.79	21.23
IWRD_TOT	-0.14	-0.125	0.781	0.633	0.906	-2.46	8.12	-11.26
IWRD_RTC	-0.006	0.009	0.945	0.945	0.945	-23.12	-21.90	-22.87
ListSort_AgeAdj	-0.037	-0.037	0.18	0.219	0.195	37.52	35.56	33.92
CogFluidComp_AgeAdj	-0.087	-0.08	0.391	0.398	0.445	18.13	19.14	16.81
CogTotalComp_AgeAdj	-0.052	-0.051	0.414	0.516	0.453	18.53	14.96	16.43
ogCrystalComp_AgeAdj	0	0.009	0.82	0.906	0.758	-8.53	-10.34	-3.29
ER40_CR	-0.103	-0.081	0.961	0.961	0.961	-24.37	-22.25	-23.45
ER40_CRT	0.044	0.059	0.828	0.844	0.828	-18.55	-17.29	-15.79
AngAffect_Unadj	-0.002	-0.01	0.648	0.68	0.617	4.96	3.69	7.16
AngHostil_Unadj	0.074	0.061	0.734	0.773	0.742	2.43	1.39	2.71
AngAggr_Unadj	0.136	0.135	0.969	0.977	0.961	-23.07	-22.99	-21.41
FearAffect_Unadj	-0.063	-0.079	0.383	0.406	0.398	22.71	20.16	22.03
FearSomat_Unadj	-0.014	-0.057	0.891	0.883	0.891	-14.48	-12.42	-13.54
Sadness_Unadj	-0.049	-0.066	0.422	0.414	0.445	18.77	18.71	18.19
LifeSatisf_Unadj	-0.057	-0.064	0.25	0.281	0.242	31.24	26.70	31.91
MeanPurp_Unadj	-0.017	-0.029	0.711	0.703	0.797	0.61	1.81	-1.07
PosAffect_Unadj	-0.003	-0.028	0.203	0.203	0.234	35.20	35.09	31.96
Friendship_Unadj	0.052	0.047	0.797	0.797	0.805	-3.66	-1.49	-3.79
Loneliness_Unadj	-0.042	-0.044	0.234	0.18	0.313	27.34	30.10	20.74
PercHostil_Unadj	0.072	0.047	0.844	0.844	0.844	-10.58	-10.57	-10.14
PercReject_Unadj	-0.016	-0.002	0.945	0.961	0.906	-13.65	-13.78	-10.87
EmotSupp_Unadj	-0.043	-0.055	0.5	0.516	0.555	13.70	12.96	10.46
InstruSupp_Unadj	0.017	0.017	0.93	0.945	0.922	-12.15	-12.54	-11.15
PercStress_Unadj	-0.03	-0.042	0.422	0.422	0.477	18.25	17.69	15.39
SelfEff_Unadj	0.091	0.095	0.977	0.977	0.977	-19.58	-17.83	-19.83
Dexterity_AgeAdj	-0.15	-0.155	0.938	0.898	0.977	-13.07	-6.64	-16.88
Strength_AgeAdj	0.386	0.415	0.391	0.82	0.344	19.98	-1.85	23.93
NEOFAC_A	-0.164	-0.153	0.992	1	0.992	-26.82	-25.18	-22.74
NEOFAC_O	0.044	0.038	0.484	0.664	0.398	11.25	4.42	16.46
NEOFAC_C	-0.102	-0.087	0.578	0.445	0.75	9.00	15.04	0.11
NEOFAC_N	-0.022	-0.026	0.711	0.695	0.734	-0.93	0.63	-4.40
NEOFAC_E	0.048	0.026	0.758	0.758	0.773	0.53	1.35	0.02
Noise_Comp	0.094	0.119	0.156	0.195	0.141	45.55	35.95	47.74
Odor_AgeAdj	-0.102	-0.084	0.852	0.805	0.844	-2.26	-1.00	-1.95
Taste_AgeAdj	0.032	0.021	0.93	0.922	0.945	-18.35	-15.46	-16.85
ASR_Anxd_Pct	0.023	0.001	0.508	0.484	0.57	4.82	6.59	2.16
ASR_Witd_T	0.063	0.066	0.867	0.859	0.883	-19.38	-18.27	-19.61
ASR_Soma_T	0.059	0.091	0.875	0.852	0.898	-19.28	-15.83	-18.88
ASR_Thot_T	0.079	0.073	0.609	0.602	0.656	-6.15	-2.97	-7.76
ASR_Attn_T	0.095	0.065	0.273	0.367	0.266	28.48	22.89	26.35
ASR_Aggr_T	0.099	0.071	0.875	0.883	0.875	-31.60	-32.76	-30.25
ASR_Rule_T	0.101	0.089	0.422	0.461	0.398	20.07	17.60	21.15
ASR_Intr_T	0.12	0.12	0.023	0.039	0.023	78.14	66.13	79.50
ASR_Intn_T	0.036	0.013	0.398	0.383	0.445	19.24	20.08	15.75
ASR_Extn_T	0.141	0.135	0.789	0.867	0.688	-5.16	-12.16	2.69
ASR_Totp_T	0.092	0.066	0.633	0.711	0.609	0.64	-2.58	2.19
DSM_Depr_T	0.011	-0.012	0.516	0.531	0.516	7.98	5.70	8.53
DSM_Anxi_T	0.026	-0.005	0.664	0.672	0.68	-11.24	-9.86	-12.07
DSM_Somp_T	0.079	0.072	0.938	0.914	0.945	-27.18	-23.94	-26.99
DSM_Avoid_T	0.005	-0.035	0.563	0.531	0.617	10.61	11.92	7.38
DSM_Adh_T	0.117	0.098	0.352	0.352	0.383	21.62	21.48	18.93
DSM_Antis_T	0.151	0.123	0.648	0.641	0.641	-4.70	-4.47	-3.57

Table SB.7: Relationship between selected variables in the HCP study with head motion (DVARS) 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 overestimation scores, underestimation scores, and impact from both/either type of score. Results in this table reflect data after motion-reduction and without additional motion censoring.

Variable	SHAMAN p-Value					
	No Censoring			DVARS < 800		
	Impact	Over	Under	Impact	Over	Under
Age_in_Yrs	0.477	0.469	0.5	0.523	0.469	0.586
Height	0.25	0.844	0.281	0.281	0.875	0.211
Weight	0.492	0.883	0.367	0.625	0.953	0.367
BMI	0.203	0.664	0.125	0.383	0.969	0.156
Handedness	0.789	0.828	0.758	0.609	0.648	0.578
SSAGA_Income	0.789	0.828	0.797	0.852	0.859	0.867
SSAGA_Educ	0.852	0.945	0.813	0.922	0.914	0.961
Hematocrit_1	0.984	0.992	0.977	0.891	0.938	0.891
ThyroidHormone	0.938	0.938	0.93	0.914	0.914	0.922
HbA1C	0.297	0.305	0.305	0.234	0.211	0.242
MMSE_Score	0.383	0.43	0.406	0.727	0.766	0.742
PSQI_Score	0.453	0.477	0.477	0.758	0.781	0.75
PicSeq_AgeAdj	0.375	0.32	0.484	0.461	0.414	0.57
CardSort_AgeAdj	0.281	0.344	0.266	0.477	0.531	0.461
Flanker_AgeAdj	0.359	0.391	0.266	0.383	0.422	0.344
PMAT24_A_CR	0.258	0.352	0.25	0.289	0.383	0.281
PMAT24_A_RTCR	0.25	0.328	0.195	0.258	0.328	0.211
ReadEng_AgeAdj	0.797	0.844	0.75	0.922	0.883	0.898
PicVocab_AgeAdj	0.938	0.906	0.945	0.953	0.961	0.984
ProcSpeed_AgeAdj	0.984	0.984	0.984	0.953	0.961	0.961
DDisc_AUC_200	0.594	0.609	0.641	0.719	0.719	0.742
DDisc_AUC_40K	0.219	0.266	0.242	0.211	0.266	0.242
VSPLOT_TC	0.664	0.695	0.633	0.773	0.852	0.758
VSPLOT_CRTE	0.391	0.398	0.391	0.359	0.359	0.367
IWRD_TOT	0.781	0.633	0.906	0.313	0.188	0.688
IWRD_RTC	0.945	0.945	0.945	0.898	0.898	0.891
ListSort_AgeAdj	0.18	0.219	0.195	0.195	0.195	0.203
CogFluidComp_AgeAdj	0.391	0.398	0.445	0.563	0.555	0.563
CogTotalComp_AgeAdj	0.414	0.516	0.453	0.5	0.539	0.563
ogCrystalComp_AgeAdj	0.82	0.906	0.758	0.953	0.945	0.961
ER40_CR	0.961	0.961	0.961	0.828	0.828	0.844
ER40_CRT	0.828	0.844	0.828	0.938	0.938	0.914
AngAffect_Unadj	0.648	0.68	0.617	0.68	0.703	0.641
AngHostil_Unadj	0.734	0.773	0.742	0.758	0.789	0.773
AngAggr_Unadj	0.969	0.977	0.961	0.914	0.906	0.969
FearAffect_Unadj	0.383	0.406	0.398	0.539	0.594	0.555
FearSomat_Unadj	0.891	0.883	0.891	0.766	0.805	0.742
Sadness_Unadj	0.422	0.414	0.445	0.695	0.703	0.711
LifeSatisf_Unadj	0.25	0.281	0.242	0.391	0.484	0.375
MeanPurp_Unadj	0.711	0.703	0.797	0.813	0.813	0.82
PosAffect_Unadj	0.203	0.203	0.234	0.281	0.281	0.336
Friendship_Unadj	0.797	0.797	0.805	0.719	0.703	0.742
Loneliness_Unadj	0.234	0.18	0.313	0.305	0.258	0.414
PercHostil_Unadj	0.844	0.844	0.844	0.867	0.883	0.867
PercReject_Unadj	0.945	0.961	0.906	0.945	0.953	0.93
EmotSupp_Unadj	0.5	0.516	0.555	0.492	0.531	0.57
InstruSupp_Unadj	0.93	0.945	0.922	0.883	0.883	0.883
PercStress_Unadj	0.422	0.422	0.477	0.625	0.633	0.633
SelfEff_Unadj	0.977	0.977	0.977	0.961	0.945	0.961
Dexterity_AgeAdj	0.938	0.898	0.977	0.906	0.867	0.953
Strength_AgeAdj	0.391	0.82	0.344	0.203	0.539	0.195
NEOFAC_A	0.992	1	0.992	0.945	0.945	0.945
NEOFAC_O	0.484	0.664	0.398	0.422	0.516	0.375
NEOFAC_C	0.578	0.445	0.75	0.68	0.602	0.781
NEOFAC_N	0.711	0.695	0.734	0.68	0.672	0.734
NEOFAC_E	0.758	0.758	0.773	0.508	0.508	0.508
Noise_Comp	0.156	0.195	0.141	0.109	0.164	0.094
Odor_AgeAdj	0.852	0.805	0.844	0.703	0.688	0.688
Taste_AgeAdj	0.93	0.922	0.945	0.961	0.953	0.969
ASR_Anxd_Pct	0.508	0.484	0.57	0.555	0.578	0.547
ASR_Witd_T	0.867	0.859	0.883	0.906	0.906	0.914
ASR_Soma_T	0.875	0.852	0.898	0.977	0.938	0.977
ASR_Thot_T	0.609	0.602	0.656	0.813	0.805	0.852
ASR_Attn_T	0.273	0.367	0.266	0.195	0.273	0.188
ASR_Aggr_T	0.875	0.883	0.875	0.922	0.938	0.922
ASR_Rule_T	0.422	0.461	0.398	0.414	0.453	0.383
ASR_Intr_T	0.023	0.039	0.023	0.047	0.055	0.055
ASR_Intn_T	0.398	0.383	0.445	0.484	0.477	0.508
ASR_Extn_T	0.789	0.867	0.688	0.891	0.984	0.875
ASR_Totp_T	0.633	0.711	0.609	0.773	0.828	0.766
DSM_Depr_T	0.516	0.531	0.516	0.664	0.711	0.609
DSM_Anxi_T	0.664	0.672	0.68	0.836	0.836	0.828
DSM_Somp_T	0.938	0.914	0.945	0.977	0.969	0.984
DSM_Avoid_T	0.563	0.531	0.617	0.438	0.438	0.484
DSM_Adh_T	0.352	0.352	0.383	0.227	0.234	0.227
DSM_Antis_T	0.648	0.641	0.641	0.789	0.805	0.773

Table SB.8: Significance (p-values) of motion impact score for selected variables in the HCP study after motion-reduction and with (varying from left to right) no motion censoring and censoring at DVARS < 0.07. Omnibus p-values are given for overestimation scores, underestimation scores, and impact from both/either type of score. The omnibus p-values account for multiple comparisons across edges, but they are not corrected for the multiple comparisons across the76 different traits shown.

Variable	SHAMAN Motion Score					
	No Censoring			DVARS < 800		
	Impact	Over	Under	Impact	Over	Under
Age_in_Yrs	13.7	14.7	12.8	9.7	12.8	7.4
Height	33.0	-0.3	31.3	29.2	-2.7	31.5
Weight	15.5	-2.5	23.9	6.7	-14.4	22.8
BMI	36.5	6.8	43.9	21.8	-13.1	38.2
Handedness	-7.2	-8.2	-4.1	3.3	1.7	5.3
SSAGA_Income	-5.3	-6.2	-5.7	-4.4	-4.2	-6.3
SSAGA_Educ	-9.0	-13.3	-4.5	-15.4	-11.8	-17.2
Hematocrit_1	-27.9	-26.6	-21.8	-17.0	-17.8	-12.3
ThyroidHormone	-21.7	-20.3	-20.7	-38.9	-37.5	-37.4
HbA1C	32.7	32.8	30.2	40.8	41.3	35.2
MMSE_Score	19.2	16.7	18.7	-2.0	-3.3	-1.7
PSQI_Score	14.9	12.9	15.2	-3.3	-6.2	-0.1
PicSeq_AgeAdj	16.5	21.3	10.6	10.4	15.9	5.1
CardSort_AgeAdj	25.8	20.7	27.5	12.5	9.4	14.3
Flanker_AgeAdj	25.5	19.8	29.2	19.9	15.7	23.8
PMAT24_A_CR	31.1	23.2	32.2	26.9	19.9	27.3
PMAT24_A_RTCR	32.9	25.5	36.7	32.3	25.5	35.8
ReadEng_AgeAdj	-3.9	-4.7	0.6	-8.8	-6.6	-7.9
PicVocab_AgeAdj	-15.4	-12.6	-13.2	-17.1	-14.3	-16.3
ProcSpeed_AgeAdj	-20.4	-20.3	-17.8	-23.5	-23.6	-20.1
DDisc_AUC_200	4.4	5.5	2.1	-0.8	-0.1	-3.0
DDisc_AUC_40K	34.1	29.1	30.8	34.4	28.8	31.3
VSPLOT_TC	5.8	4.7	8.2	-1.4	-1.5	0.8
VSPLOT_CRTE	21.6	20.8	21.2	22.1	22.0	21.5
IWRD_TOT	-2.5	8.1	-11.3	22.6	36.2	2.0
IWRD_RTC	-23.1	-21.9	-22.9	-16.2	-15.4	-15.5
ListSort_AgeAdj	37.5	35.6	33.9	41.7	39.1	37.5
CogFluidComp_AgeAdj	18.1	19.1	16.8	9.2	10.1	8.6
CogTotalComp_AgeAdj	18.5	15.0	16.4	11.8	9.4	9.5
ogCrystalComp_AgeAdj	-8.5	-10.3	-3.3	-14.1	-12.7	-12.3
ER40_CR	-24.4	-22.3	-23.5	-5.6	-5.0	-5.7
ER40_CRT	-18.6	-17.3	-15.8	-30.9	-27.7	-27.7
AngAffect_Unadj	5.0	3.7	7.2	3.0	3.0	4.9
AngHostil_Unadj	2.4	1.4	2.7	-3.8	-4.8	-2.3
AngAggr_Unadj	-23.1	-23.0	-21.4	-11.6	-8.0	-16.2
FearAffect_Unadj	22.7	20.2	22.0	8.6	5.7	8.5
FearSomat_Unadj	-14.5	-12.4	-13.5	-8.3	-9.7	-5.8
Sadness_Unadj	18.8	18.7	18.2	1.1	0.6	0.5
LifeSatisf_Unadj	31.2	26.7	31.9	21.0	16.3	22.3
MeanPurp_Unadj	0.6	1.8	-1.1	-8.1	-7.0	-8.9
PosAffect_Unadj	35.2	35.1	32.0	27.8	28.4	23.9
Friendship_Unadj	-3.7	-1.5	-3.8	-1.1	0.5	-0.9
Loneliness_Unadj	27.3	30.1	20.7	25.7	28.7	18.2
PercHostil_Unadj	-10.6	-10.6	-10.1	-14.0	-14.8	-13.0
PercReject_Unadj	-13.6	-13.8	-10.9	-16.8	-16.8	-14.3
EmotSupp_Unadj	13.7	13.0	10.5	10.9	9.9	8.5
InstruSupp_Unadj	-12.2	-12.5	-11.1	-7.6	-7.9	-6.8
PercStress_Unadj	18.2	17.7	15.4	7.6	6.4	6.1
SelfEff_Unadj	-19.6	-17.8	-19.8	-16.2	-14.5	-16.1
Dexterity_AgeAdj	-13.1	-6.6	-16.9	-13.7	-7.7	-16.6
Strength_AgeAdj	20.0	-1.8	23.9	32.7	11.6	30.0
NEOFAC_A	-26.8	-25.2	-22.7	-20.6	-19.0	-17.2
NEOFAC_O	11.2	4.4	16.5	10.6	7.3	13.6
NEOFAC_C	9.0	15.0	0.1	1.2	6.6	-4.9
NEOFAC_N	-0.9	0.6	-4.4	6.3	8.2	2.6
NEOFAC_E	0.5	1.4	0.0	15.7	15.9	15.3
Noise_Comp	45.5	36.0	47.7	44.9	36.6	46.2
Odor_AgeAdj	-2.3	-1.0	-2.0	4.1	4.2	4.5
Taste_AgeAdj	-18.3	-15.5	-16.8	-19.8	-15.7	-19.1
ASR_Anxd_Pct	4.8	6.6	2.2	6.5	5.2	6.3
ASR_Witd_T	-19.4	-18.3	-19.6	-11.3	-11.1	-11.8
ASR_Soma_T	-19.3	-15.8	-18.9	-31.2	-26.9	-30.6
ASR_Thot_T	-6.2	-3.0	-7.8	-20.1	-17.2	-21.0
ASR_Attn_T	28.5	22.9	26.3	38.9	30.3	37.4
ASR_Aggr_T	-31.6	-32.8	-30.3	-29.8	-31.3	-28.1
ASR_Rule_T	20.1	17.6	21.1	22.5	19.8	24.3
ASR_Intr_T	78.1	66.1	79.5	67.9	59.5	68.3
ASR_Intn_T	19.2	20.1	15.7	14.4	14.4	12.0
ASR_Extn_T	-5.2	-12.2	2.7	-17.1	-22.9	-9.1
ASR_Totp_T	0.6	-2.6	2.2	-5.3	-8.5	-2.9
DSM_Depr_T	8.0	5.7	8.5	1.3	-3.4	4.1
DSM_Anxi_T	-11.2	-9.9	-12.1	-18.9	-18.5	-17.9
DSM_Somp_T	-27.2	-23.9	-27.0	-32.0	-27.1	-33.0
DSM_Avoid_T	10.6	11.9	7.4	17.6	18.0	13.5
DSM_Adh_T	21.6	21.5	18.9	39.4	37.7	37.0
DSM_Antis_T	-4.7	-4.5	-3.6	-12.2	-12.9	-10.2

Table SB.9: Motion impact scores (omnibus Stouffer's Z) for selected variables in the HCP study after motion-reduction and with (varying from left to right) no motion censoring and censoring at DVAR < 0.07. Higher scores correspond to a greater impact of motion. Motion overestimation scores, underestimation scores, and overall impact (both/either over/underestimation) are shown separately.

Discussion

Our supplementary analyses, using imaging data from the Human Connectome Project (HCP) in addition to ABCD and using DVARS as an alternative to framewise displacement for quantifying head motion, demonstrate a broadly consistent pattern of results applying SHAMAN to compute motion impact score. One limitation of our analysis was intrinsic differences in data processing, sample size, sample population, and amount of head motion in ABCD and HCP. The HCP study was processed using ICA-FIX, it had fewer participants, the participants were adults, and the participants had considerably less in-scanner head motion than the children in ABCD, consistent with prior observations of children as a high-motion cohort [19,22,29]. This made it difficult to disentangle whether lower motion impact scores in HCP were due to superiority of ICA-FIX as a method for mitigating head motion artifact, limitations in SHAMAN's sensitivity at smaller sample sizes, or lower head motion in the adult cohort. Despite these differences, clear parallels emerged in both data sets. Multiple traits, especially physical traits such as BMI, had significant motion overestimation scores in both ABCD and HCP. Age had a greater motion overestimation score in HCP, consistent with the wider distribution of ages in the HCP cohort. Motion censoring using FD < 0.2 mm reduced to near-zero the number of traits with significant motion overestimation scores in both ABCD and HCP. (Remarkably, in HCP, this reduction was possible without the exclusion of any participants at FD < 0.2 mm.) Thus, data processed with ICA-FIX are not immune to residual motion artifact, and it is reasonable to suspect that residual motion artifact is an issue in most resting-state fMRI data regardless of processing approach.

The supplementary analysis comparing FD to DVARS revealed advantages and disadvantages of each method for quantifying in-scanner head motion. DVARS is sensitive to sources of transient fMRI signal change due to any cause, not just head motion. It is therefore unsurprising that SHAMAN was more sensitive to motion overestimation in ABCD using DVARS compared to FD in the uncensored data. Combined with methods to select an optimal DVARS censoring threshold [41], DVARS has the potential to minimize both the risk of spurious trait-FC inference and sample bias/attrition due to censoring and exclusion of interesting participants.

On the other hand, divergence of the results for HCP-DVARS compared to HCP-FD or ABCD-DVARS highlights the difficulty of standardizing DVARS measurement. While a full exploration of methods for DVARS calculation is beyond the scope of this study, we observe that the calculation of FD using the six rigid-body motion co-registration parameters is standardized across datasets, whereas the calculation of DVARS depends both on how the data are normalized and at what point in the data processing pipeline the calculation is made. Afyouni & Nichols [73] describe a method for harmonizing DVARS calculation in data with different normalizations, but we are not aware of any method to render the calculation of DVARS equivalent in data processed using entirely different pipelines (e.g. ABCD-BIDS vs ICA-FIX). Thus, while DVARS may have distinct advantages for motion mitigation under certain circumstances, FD may still be useful for comparing motion between data processed using different methods.