



Children's head motion during fMRI tasks is heritable and stable over time



Laura E. Engelhardt^{a,*}, Mary Abbe Roe^a, Jenifer Juranek^b, Dana DeMaster^b,
K. Paige Harden^{a,c}, Elliot M. Tucker-Drob^{a,c}, Jessica A. Church^{a,d}

^a Department of Psychology, The University of Texas at Austin, United States

^b The Children's Learning Institute, The University of Texas Health Science Center at Houston, United States

^c Population Research Center, The University of Texas at Austin, United States

^d Imaging Research Center, The University of Texas at Austin, United States

ARTICLE INFO

Article history:

Received 27 September 2016

Received in revised form

21 November 2016

Accepted 30 January 2017

Available online 1 February 2017

Keywords:

Head motion

fMRI

Twin study

Repeated measures

Individual differences

ABSTRACT

Head motion during fMRI scans negatively impacts data quality, and as post-acquisition techniques for addressing motion become increasingly stringent, data retention decreases. Studies conducted with adult participants suggest that movement acts as a relatively stable, heritable phenotype that serves as a marker for other genetically influenced phenotypes. Whether these patterns extend downward to childhood has critical implications for the interpretation and generalizability of fMRI data acquired from children. We examined factors affecting scanner motion in two samples: a population-based twin sample of 73 participants (ages 7–12 years) and a case-control sample of 32 non-struggling and 78 struggling readers (ages 8–11 years), 30 of whom were scanned multiple times. Age, but not ADHD symptoms, was significantly related to scanner movement. Movement also varied as a function of task type, run length, and session length. Twin pair concordance for head motion was high for monozygotic twins and moderate for dizygotic twins. Cross-session test-retest reliability was high. Together, these findings suggest that children's head motion is a genetically influenced trait that has the potential to systematically affect individual differences in BOLD changes within and across groups. We discuss recommendations for future work and best practices for pediatric neuroimaging.

© 2017 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

1.1. In-scanner movement negatively impacts fMRI data

Among the many potential confounds complicating functional magnetic resonance imaging (fMRI) research, head motion is perhaps the most intractable. Head movement causes misalignment of spatial units from one time point to the next, resulting in difficulty accurately localizing blood oxygenation level dependent (BOLD) activity. More problematically, movement-induced spatial misalignment leads to changes in the BOLD signal and structural estimates that aren't necessarily attributable to true effects, potentially obscuring or distorting investigations of correlates of brain function and structure (Friston et al., 1996; Siegel et al., 2016).

Scanner movement and its negative effects on data quality are especially pronounced in pediatric samples. Age-related differences in scanner movement are hardly surprising; relative to adults, children have reduced inhibitory control (Bedard et al., 2002; Williams et al., 1999) and may find it more challenging to monitor their movement, especially in the face of scanner distractions or attention-consuming tasks (Greene et al., 2016; Olesen et al., 2007). Individual differences in children's head movement are readily apparent, both in terms of absolute head displacement and movement fluctuations across a scan session (Van Dijk et al., 2012; Yan et al., 2013). On average, however, children and adolescents exhibit significantly more head motion than adults (e.g., Satterthwaite et al., 2012; Kelly et al., 2008), resulting in potentially exaggerated differences in brain activity and structure between age groups (Bullmore et al., 1999; Power et al., 2012; Power et al., 2014; Satterthwaite et al., 2012). Thus, differences in the BOLD response that are commonly attributed to age differences in cognitive processes are likely confounded by age differences in scanner movement, a problem recognized by the developmental neuroimaging community (Church et al., 2010; Greene et al., 2016).

* Corresponding author at: 108 E. Dean Keeton, Stop A8000, Austin, Texas 78712, United States.

E-mail address: lauraeng@utexas.edu (L.E. Engelhardt).

For instance, in the context of resting-state functional-connectivity (RSFC) analyses of synchronization in the spontaneous activity of brain regions across time, head movement dampens estimates of long-range connectivity (i.e., signal correlations between spatially distant brain regions) while increasing estimates of short-range connectivity (Power et al., 2012; Satterthwaite et al., 2012; Van Dijk et al., 2012; Yan et al., 2013; Zeng et al., 2014). Previously reported age group differences in RSFC may be the product of motion differences rather than neural differences (Satterthwaite et al., 2012; Van Dijk et al., 2012; Yan et al., 2013).

1.2. Approaches to handling movement

Given the potential for movement to bias fMRI results and artificially inflate group differences, the field has developed a variety of post-acquisition methods to remove motion components from the BOLD signal. A critical step in the processing of fMRI data involves spatially transforming images to match the location of a previously acquired reference image (Friston et al., 1996). Because the resulting transformation estimates directly reflect spatial changes across time, movement parameters (linear and degree movement in x, y, and z planes) may be added as covariates in the general linear model as a secondary technique for reducing motion bias (Poldrack et al., 2011). Despite recent gains in understanding and controlling for the effects of motion artifact, movement in the scanner remains problematic. Appropriately modeling movement parameters cannot remove all of the bias resulting from head motion (Power et al., 2012, 2014; Satterthwaite et al., 2012, 2013; Van Dijk et al., 2012; Yan et al., 2013). A complementary method for reducing motion bias entails excluding time points for which movement exceeds a certain threshold. The criteria for such motion scrubbing practices have become increasingly stringent as the effects of movement on data quality have reached greater recognition (Siegel et al., 2014; Power et al., 2014; Laumann et al., 2016). Although motion scrubbing significantly reduces motion-related noise in the BOLD signal, it necessarily leads to lower data retention and the omission of information that may otherwise be valuable if it wasn't contaminated. If the missing values or, alternatively, the movements themselves, are related to the correlates under study, scientific inferences may be distorted (Rubin, 1976; Siegel et al., 2016).

1.3. Individual differences in head movement may be systematic

A growing body of research, based largely on adult samples, indicates that individual differences in head movement during MRI scans are systematic (Covy-Duchesne et al., 2014; Van Dijk et al., 2012; Zeng et al., 2014). For instance, movement is relatively stable across independent scanning sessions, with cross-session correlations falling in the 0.5–0.7 range (Van Dijk et al., 2012; Zeng et al., 2014). Covy-Duchesne et al. (2014) investigated multiple measures of head movement in a sample of 231 adult twin pairs and found that heritability in motion ranged from 35% to 54%, depending on the metric; these estimates are consistent with heritabilities reported for other involuntary movements in adults (Anokhin et al., 2003) and children (Fisher et al., 2010). Clinically relevant phenotypes such as impulsivity and externalizing behaviors are of particular interest to the study of interindividual variation in movement. Participants with ADHD, which is characterized by phenotypes like impulsivity and externalizing behaviors, have been reported to be more prone to scanner movement than age-matched controls (e.g., Durston et al., 2003; Epstein et al., 2007). Importantly, common genetic factors contribute to individual differences in head motion and ADHD in adults (Covy-Duchesne et al., 2016), suggesting that head motion serves as a marker for other cognitive and behavioral traits that are commonly studied in neuroimaging research.

Despite recent gains in our understanding of the sources and correlates of head motion in adults, the extent to which children's head motion follows these patterns remains unknown. Identifying predictors of scanner movement in pediatric samples has the potential to inform future study designs and interpretations of neuroimaging data. Creating studies that maximize data retention is a major goal in developmental cognitive imaging, not only because data loss reduces overall power, but also because the removal of data on the basis of movement may disproportionately affect certain populations (e.g., children with externalizing disorders) in ways not previously addressed in the literature. As researchers seek to become more representative in their neuroimaging samples, it is important to better understand how movement systematically differs across individuals.

1.4. The current study approach

The present study explores the consistency and correlates of head movement during fMRI scans, specifically by estimating the reliability and familial similarity of in-scanner motion in two pediatric samples. First, we investigate movement patterns within a single scan session to determine how movement differs as a function of age, ADHD symptoms, scan run and session duration, and task type. To further assess the extent to which motion is a stable property of the individual, we examine familial resemblance by comparing movement within and across twin pairs, and by measuring movement within children across repeated scan sessions over a period of several months. To the extent that characteristics of the individual or of the scan influence movement, we also determine whether these properties continue to associate with movement after omission of high-motion frames.

2. Materials and methods

2.1. Participants

2.1.1. Twin sample

Individuals in the twin sample were 3rd- through 8th-grade twins and triplets recruited through the Texas Twin Project in the summer and fall of 2015 (Harden et al., 2013). Participants that completed an in-laboratory assessment of cognitive, academic, and executive function abilities were invited to return for a one-time MRI scan. Participants were excluded from participating in the MRI scan if they presented any contraindication for MRI or if parents reported a history of major developmental delay. We report data for 73 of the 79 children that returned for the MRI scan. Two participants were not scanned due to the presence of MRI contraindications at the time of MRI visit; one participant declined to enter the scanner; two participants declined to continue the scan before functional data were collected; and a technological problem prevented us from collecting one participants' functional data. The final sample of 73 individuals ranged between 7 and 12 years of age ($M = 10.11$, $SD = 1.23$) and included 36 boys. The sample was racially and socioeconomically diverse, with 18% of participants identified as Hispanic, 45% as white, 5% as black, 3% as Asian, 3% as another race, and 26% as multiracial. Thirty-eight percent of families reported having received need-based public assistance at some point since the twins' birth.

The sample consisted of 32 twin pairs, 2 individuals from a triplet set (henceforth referred to as a twin pair for convenience), and 7 children whose twins were not scanned. Within the sample, opposite-sex twins were automatically classified as dizygotic (DZ). Zygosity of same-sex twins was determined by entering parent and experimenter ratings of the twins' physical similarity into a latent class analyses. This procedure has been found to be more than 99%

Table 1
Sample characteristics.

Group	Group N's		Mean age (yrs) at session			Mean time (mth) between sessions	
	Total	Intervention: Control	Session 1	Session 2	Session 3	Sessions 1 & 2	Sessions 2 & 3
Nonstruggling readers	32 (16 male)	–	9.79 [0.89]	–	–	–	–
Struggling readers: All	78 (44 male)	44:36	10.14 [0.70]	10.73 [0.70]	11.73 [0.62]	7.52 [3.33]	12.99 [1.32]
Struggling readers: 1 Scan	48 (31 male)	21:27	10.24 [0.64]	–	–	–	–
Struggling readers: 2 Scans	23 (10 male)	16:7	10.09 [0.47]	10.76 [0.48]	–	7.95 [3.66]	–
Struggling readers: 3 Scans	7 (3 male)	5:2	10.14 [0.63]	10.65 [.57]	11.73 [.62]	6.09 [1.12]	12.99 [1.32]
MZ individuals	24 (10 male)	–	10.61 [1.00]	–	–	–	–
DZ individuals	49 (26 male)	–	9.87 [1.27]	–	–	–	–

Note. Sample numbers, ages, and time between scan sessions for the reading and twin samples. Standard deviations are in brackets. MZ = monozygotic, DZ = dizygotic.

accurate, as validated by genotyping (Heath et al., 2003). Twelve pairs (36%) were classified as monozygotic (MZ), 10 were classified as same-sex dizygotic (DZ), and the remaining 11 pairs were classified as opposite-sex DZ (Table 1).

2.1.2. Reading sample

Individuals in the reading sample were 4th- and 5th-grade children recruited for a reading comprehension intervention as part of the Learning Disabilities Research Center of Texas, centered in Houston and Austin, Texas. The Meadows Center for Preventing Educational Risk at the University of Texas at Austin administered the intervention for the Austin site. The intervention team identified children as struggling readers based on an in-school screening assessment using a standardized reading test (e.g., the Test of Sentence Reading Efficiency and Comprehension; Wagner et al., 2010). All children identified as struggling readers (standard score <90) were invited to participate in the imaging study. Nonstruggling readers in 4th and 5th grade were recruited from the greater Austin community. For participants recruited from the community, reading status was confirmed by scores on the standardized Sight Word Efficiency subtest of the Test of Word Reading Efficiency-2 (Torgesen et al., 2012). Participants were excluded from the MRI scan if they presented any contraindication for MRI, a history of developmental delay, or use of stimulant medication (control participants only). We report data for 110 of the 118 children that participated in the MRI study at the Austin site. The data from three participants were excluded due to irregularities in brain structure; one participant did not fall within the age range set by the intervention; one participant manifested a developmental delay; one control participant reported an ADHD diagnosis, which was only permissible for the struggling reader group; and three other participants' data were unusable due to scanner problems.

We report in Table 1 age, sex, and additional descriptive information for the 110 participants in reading sample. Thirty-two non-struggling readers provided usable data collected from a single session. We also analyzed data from 48 struggling readers who completed a single MRI session, 23 who completed two sessions, and 7 who completed three sessions. Mean participant age at the first session, regardless of intervention status or study wave, was 10.07 (SD = 0.71). Of the participants that returned for a second scan, most did so at the end of the intervention period, approximately six months after the first scan. Participants that returned for a third scan did so one year after the end of the first intervention period, approximately 13 months after the second scan. The sample was racially diverse: 44% of participants were Hispanic, 42% were non-

Hispanic white, 9% were black, 1% were Native American, and 4% were multiracial.

2.2. Assessment of ADHD symptoms

In the twin sample, 69 individuals' ADHD symptoms were assessed using parent ratings on 18 items from the Conners' Rating Scales (Conners, 1997). The study used nine items from the Inattentive subscale (e.g., "Twin 1 is forgetful in daily activities") and nine items from the Hyperactive-Impulsive subscales (e.g., "Twin 1 has difficulty waiting for his/her turn"). Parents used a four-point scale to rate how true each item was for each twin, with ratings ranging from 0 = Not True at All to 3 = Very True. We report the sums of individual subscales and all 18 items, with higher scores corresponding to worse behavior.

In the reading sample, teachers or parents of 76 participants reported ADHD behaviors and symptoms using the Strengths and Weaknesses of ADHD Symptoms and Normal Behavior rating scale (SWAN; Swanson, 2011). Adults were asked to rate the target child relative to his/her peers on nine Inattentive (e.g., "Sustains attention on tasks or play activities") and nine Hyperactive-Impulsive (e.g., "Modulates verbal activity [control excess talking]") items. The seven-point scale ranged from -3 = Positive Behavior Far Above Average to 3 = Positive Behavior Far Below Average. We report the sums of individual subscales and all 18 items, with more positive scores corresponding to worse behavior. We also report total symptom count, which corresponds to the number of items for which a participant was rating as being below average. Adults completed the assessment multiple times for 18 participants. Table 2 shows mean SWAN ratings collected closest to the first scan session. We report ADHD rating-movement relations for 11 nonstruggling and 41 struggling readers with SWAN ratings collected closest to their first scan, 23 struggling readers with SWAN ratings collected closest to their second scan, and 1 struggling reader with SWAN ratings collected closest to his/her third scan. Parents reported a diagnosis of AD(H)D and/or use of neurostimulant medication for one participant from the twin sample and eleven participants from the reading sample.

2.3. fMRI sessions

On the day of the scan, parents provided informed consent for their child(ren)'s participation, and children provided informed assent. The current paper does not report the behavioral or functional imaging results of the tasks listed below.

Table 2
Descriptive statistics for the first scan session.

Group	Mean frames analyzed	Mean FD (mm)	FD skew	Mean frames analyzed after scrubbing	Mean FD (mm) after scrubbing	FD skew after scrubbing	Proportion with ADHD data	Mean ADHD severity rating
Twins	931.05 [215.30]	0.63 [0.63]	2.48	717.70 [243.54]	0.20 [0.08]	0.52	0.95	11.70 ^a [9.97]
Nonstruggling readers	1298.66 [131.52]	0.50 [0.42]	1.92	1051.32 [216.62]	0.19 [0.06]	0.08	0.34	–15.00 ^b [12.16]
Struggling readers	1207.59 [273.53]	0.62 [0.56]	2.03	939.12 [295.47]	0.20 [0.07]	0.53	0.82	.58 ^b [17.20]

Note. Descriptive statistics for number of frames analyzed, session-wide framewise displacement (FD) in millimeters before and after motion scrubbing, and attention deficit hyperactivity disorder (ADHD) ratings. Standard deviations are in brackets.

^a ADHD ratings correspond to the sum of inattentive and hyperactive-impulsive item scores from the Conners' rating scales.

^b ADHD ratings correspond to the sum of inattentive and hyperactive-impulsive item scores and SWAN rating scales.

All twin participants were invited to get inside of a mock scanner at the end of the initial behavioral visit. Inside the mock scanner, participants practiced remaining still for approximately 1 min, viewing objects through an overhead mirror, and listening to the scanner noise. When they returned for the MRI scan, twins were scanned back-to-back. The fMRI sessions for twins consisted of two resting state runs (180 vol each), two runs of a flexible switching task (161 vol each), two runs of a working memory task (106 vol each), and one run of a response inhibition task (180 vol). In the vast majority of visits, the session proceeded as follows: resting state, switching task, working memory task, response inhibition task, switching task, working memory task, resting state. Run order was adjusted occasionally to maximize collection from each participant. Sessions carried out to completion lasted approximately 75 min. Unlike the twin sample, few participants from the reading sample were exposed to the mock scanner, primarily due to the distance they traveled, which limited collection to one day. fMRI sessions for participants in the reading sample consisted of two resting-state runs (180 vol each), two runs of a response inhibition task (180 vol each), and three runs of a sentence comprehension task (212 vol). In the majority of visits, runs proceeded as follows: resting state, sentence comprehension task, response inhibition task, sentence comprehension task, resting state, response inhibition task, sentence comprehension task. Sessions carried out to completion lasted approximately 90 min.

Across both samples, data were acquired on the same Siemens Skyra 3T scanner with a Siemens 32-channel head coil. Participants' head motion was restricted with tight foam pads and the flat headphones participants wore. The research staff provided verbal feedback about movement between scan runs. We acquired T₁-weighted structural images with an MPRAGE sequence and T₂-weighted structural images with a turbo-spin echo sequence. Functional images for both samples were collected using a multi-band echo-planar sequence (TR = 2000 ms, TE = 30 ms, flip angle = 60°, MB factor = 2, 48 axial slices, 2 × 2 × 2 mm voxels).

2.4. fMRI processing

Image processing was carried out in FMRIB Software Library (FSL) version 5.0. Following standard motion correction procedures, images were realigned to an inter-run reference image via trilinear interpolation. Six rigid body transformations were estimated for each time point. We selected framewise displacement (FD; Power et al., 2012) as our movement metric. FD represents the sum of movements across the six rigid body motion parameters. Prior to summing, rotations were converted to displacements about a 50 mm sphere. FD is computed for each volume in relation to displacement since the previous volume. FD is a marker of relative movement across a scan, rather than absolute movement since the start of a scan.

2.5. Analyses

The first frame of each run served as the initial reference point for movement (i.e., FD for this frame was always 0 mm); these frames were removed prior to analysis. Analyses were conducted in R version 3.2.3 (R Core Team, 2015). Raw movement estimates were highly skewed to the right, so subsequent analyses used log transformed FD estimates. When appropriate, in order to account for the nonindependence of data from twins nested within families and from multiple observations nested within individuals (e.g., multiple runs within a scan session), we used the *nlme* package (Pinheiro et al., 2016) to estimate regressions as linear mixed models with random intercepts. We investigated the contributions of person-level characteristics (age, sex, ADHD symptoms) and scan characteristics (run and session length, task type) to movement averaged across the entire scan session or to movement averaged within runs. Note that we do not control for the time of day in which scans occurred, though the vast majority were collected midday on weekends. Next, we evaluated cross-session stability in scanner movement for struggling readers that returned for additional scan sessions. We also evaluated familial resemblance in movement by estimating co-twin correlations in movement across the scan session (MZ vs. DZ relationships in movement). Finally, we omitted high-movement frames per current scrubbing recommendations (Power et al., 2014) and reran key analyses to determine whether the characteristics identified as important to movement continued to impact a narrower band of movement. We did not log transform FD estimates of scrubbed data, as those data more closely adhered to a normal distribution.

3. Results

3.1. Descriptive statistics

3.1.1. Frames collected and session-wide movement

Table 2 presents descriptive statistics for frames collected, raw FD, and ADHD scores at or closest to participants' first scan session. In the twin sample, the number of frames collected was not significantly related to sex ($b = -28.64$, $SE = 48.77$, $p = 0.56$), age ($b = 24.36$, $SE = 22.26$, $p = 0.28$), or zygosity ($b = -74.53$, $SE = 64.99$, $p = 0.26$). In the reading sample, the number of frames collected did not vary by age sample ($b = 10.56$, $SE = 33.04$, $p = 0.75$) or by reading status (struggling vs nonstruggling; $b = -91.07$, $SE = 50.69$, $p = 0.075$). Compared to males, females completed significantly more frames ($b = -141.46$, $SE = 44.91$, $p < 0.01$). A one-way, independent samples ANOVA indicated that struggling readers' movement at the first scan session was not significantly associated with intervention cohort ($F(1, 2) = 0.96$, $p = 0.39$). The point during the intervention at which the first scan occurred (pre-intervention, immediately post-intervention, one year post-intervention) also did not significantly

relate to movement ($F(1, 2) = 1.83, p = 0.18$). Thus, we do not control for cohort or intervention period in subsequent analyses.

3.1.2. ADHD scores

In the twin sample, Conners' ADHD scores ranged from 0 to 43, out of a possible range of 0 (complete disagreement that problematic behaviors apply to child) to 54 (complete agreement that problematic behaviors apply to child). In the reading sample, SWAN scores collected closest to the first scan session ranged from -37 to 38, out of a possible range of -54 (child's positive behaviors far above average) to 54 (child's positive behaviors far below average). The range for struggling readers was -34 to 38, and that of nonstruggling readers was -37 to 0. Ratings on the Inattentive and Hyperactive-Impulsive subscales were significantly correlated ($r_{\text{twin}} = 0.56, p < 0.001$; $r_{\text{reading}} = 0.74, p < 0.001$).

In the twin sample, neither age ($b = -0.26, SE = 0.97, p = 0.79$) nor sex ($b = -0.35, SE = 2.42, p = 0.89$) was associated with severity of ADHD symptoms. In the reading sample, age did not significantly relate to symptom severity ($b = 0.94, SE = 3.25, p = 0.77$) or symptom count ($b = 0.22, SE = 1.09, p = 0.84$). In the reading sample, sex was significantly associated with symptom severity ($b = 9.95, SE = 3.89, p < 0.05$) but not symptom count ($b = -1.65, SE = 1.35, p = 0.22$). Specifically, ADHD severity was significantly lower (i.e., positive behaviors were more likely to be reported) for girls than boys. Compared to nonstruggling readers, struggling readers were reported to have greater ADHD severity ($b = 15.58, SE = 5.42, p < 0.01$) and total symptoms ($b = -4.16, SE = 1.84, p < 0.01$). These differences persisted even after removing participants with known diagnoses or neurostimulant use. ADHD ratings in the reading sample differed by reporter (parent or teacher), with teachers reporting significantly greater symptom severity ($b = 10.16, SE = 3.90, p < 0.05$) and parents tending to report a greater number of symptoms ($b = -2.27, SE = 1.33, p = 0.092$).

3.2. Individual characteristics

We first evaluated the extent to which properties of the individual related to movement during participants' first scan session (see Fig. 1). FD averaged across all volumes in the first scan session was significantly associated with age in the twin sample ($b = -0.43, SE = 0.07, p < 0.001$), such that younger participants moved more in the scanner. Age was not significantly related to FD in the reading sample ($b = -0.084, SE = 0.11, p = 0.44$). Mean session FD did not differ significantly across sexes in the twin sample ($b = -0.07, SE = 0.20, p = 0.74$) or the reading sample ($b = 0.24, SE = 0.15, p = 0.12$). Although mean FD was lower for nonstruggling readers (.50 mm) than for struggling readers (.62 mm), this difference was not significant ($b = 0.19, SE = 0.17, p = 0.25$).

Next, we investigated the effect of ADHD symptomology on movement at the first scan session. Across both samples, mean session FD was not significantly related to Inattentive scores ($b_{\text{twin}} = 0.021, SE = 0.021, p = 0.32$; $b_{\text{reading}} = 0.00076, SE = 0.0098, p = 0.94$), Hyperactive-Impulsive scores ($b_{\text{twin}} = 0.018, SE = 0.014, p = 0.21$; $b_{\text{reading}} = 0.0017, SE = 0.010, p = 0.87$), total ADHD symptom severity (sum of the two subscales; $b_{\text{twin}} = 0.012, SE = 0.00094, p = 0.20$; $b_{\text{reading}} = 0.00068, SE = 0.0054, p = 0.90$), or symptom count ($b_{\text{reading}} = -0.0038, SE = 0.016, p = 0.82$).

3.3. Scan characteristics

We next examined movement as a function of components of the scan session. We computed Cronbach's alpha on movement averaged within each run; within-person movement was highly reliable throughout the scan session ($\alpha_{\text{twin}} = 0.91, 95\% \text{ CI } [.87, 0.94]$; $\alpha_{\text{reading}} = 0.91, 95\% \text{ CI } [.84, 0.97]$). Random-intercepts linear regressions of movement onto run number revealed a significant increase

in mean FD across scan runs in both the twin sample ($b = 0.080, SE = 0.015, p < 0.001$) and the reading sample ($b = 0.094, SE = 0.025, p < 0.001$).

To determine the effect of run length and total session length on movement, we estimated (mixed) linear regressions of movement onto the number of frames collected within and across runs. Total number of frames collected was negatively and significantly associated with FD in both samples ($b_{\text{twin}} = -0.0010, SE = 0.00045, p < 0.05$; $b_{\text{reading}} = -0.00067, SE = 0.00031, p < 0.05$), indicating that individuals who completed more frames overall moved less. Run length and run-specific FD were negatively associated in the twin sample ($b = -0.0032, SE = 0.00094, p < 0.001$) and positively associated in the reading sample ($b = 0.012, SE = 0.0013, p < 0.001$). We conducted Kaplan-Meier survival analyses in R (Therneau, 2015) to identify points during the scan at which participants are likely to discontinue their participation. As depicted in Fig. 2, the proportion of individuals continuing with the scan dropped below 90% after the fourth run in the twin sample and after third run in the reading sample. This dropout point corresponds to approximately 40 min of scanning in both samples.

FD was lowest, on average, for the rapid response inhibition task in both samples (see Fig. 3). One-way ANOVAs estimated as mixed linear models with random intercepts revealed that FD varied significantly by task in both samples ($F_{\text{twin}}(1,3) = 2.91, p < 0.05$; $F_{\text{reading}}(1,2) = 34.01, p < 0.001$). Capitalizing on the relatively wider age range in this sample, we looked at task-specific movement effects across three age groups: 16 participants ages 7–8 years, 37 participants ages 9–10 years, and 20 participants ages 11–13 years. As depicted in Fig. 3c, the youngest group exhibited the most movement during rest and the least movement during the inhibition task. Among the 9- to 10-year-olds, movement was highest during the working memory task and lowest during the inhibition task. In the oldest age group, movement was highest during the working memory task and lowest during the switching task. Given the small sizes of these bins, we did not formally test for within-group, between-task differences or interaction effects. In the reading sample, FD was significantly greater during the reading task relative to both rest and the inhibition tasks. Reading status and task did not significantly interact to predict FD ($F(1,2) = 1.44, p = 0.24$). These patterns held in both samples when we included only participants providing data for all possible tasks.

3.4. Cross-session stability

Movement was highly consistent across scan sessions for repeat participants from the reading sample. Group-level FD did not change significantly over the course of repeat sessions ($b = -0.17, SE = 0.087, p = 0.062$), despite participants having previous experience in the scan environment. Mean FD at the first session correlated with mean FD at the second session (~8 months apart) at $r = 0.62 (p < 0.001)$. The correlation between mean FD at the second and third (~13 months apart) sessions was $r = 0.69 (p = 0.085)$. We were interested in whether the observed stability of FD was due simply to sex and age heterogeneity of the return participants. We therefore computed sex- and age- partialled correlations. The correlation between mean FD at the first and second sessions, partialling out age and sex, was $r = 0.57 (p < 0.01)$; for the second and third sessions, it was $r = 0.78 (p = 0.22)$. As a further sensitivity analysis correlated each individual's first scan movement with an age-matched and sex-matched person's movement at their first scan session, and they were unrelated ($r = -0.07, p = 0.72$).

3.5. Familial resemblance

The correlation between co-twins for mean session FD was positive and significant ($r = 0.59, p < 0.001$). Zygosity-specific anal-

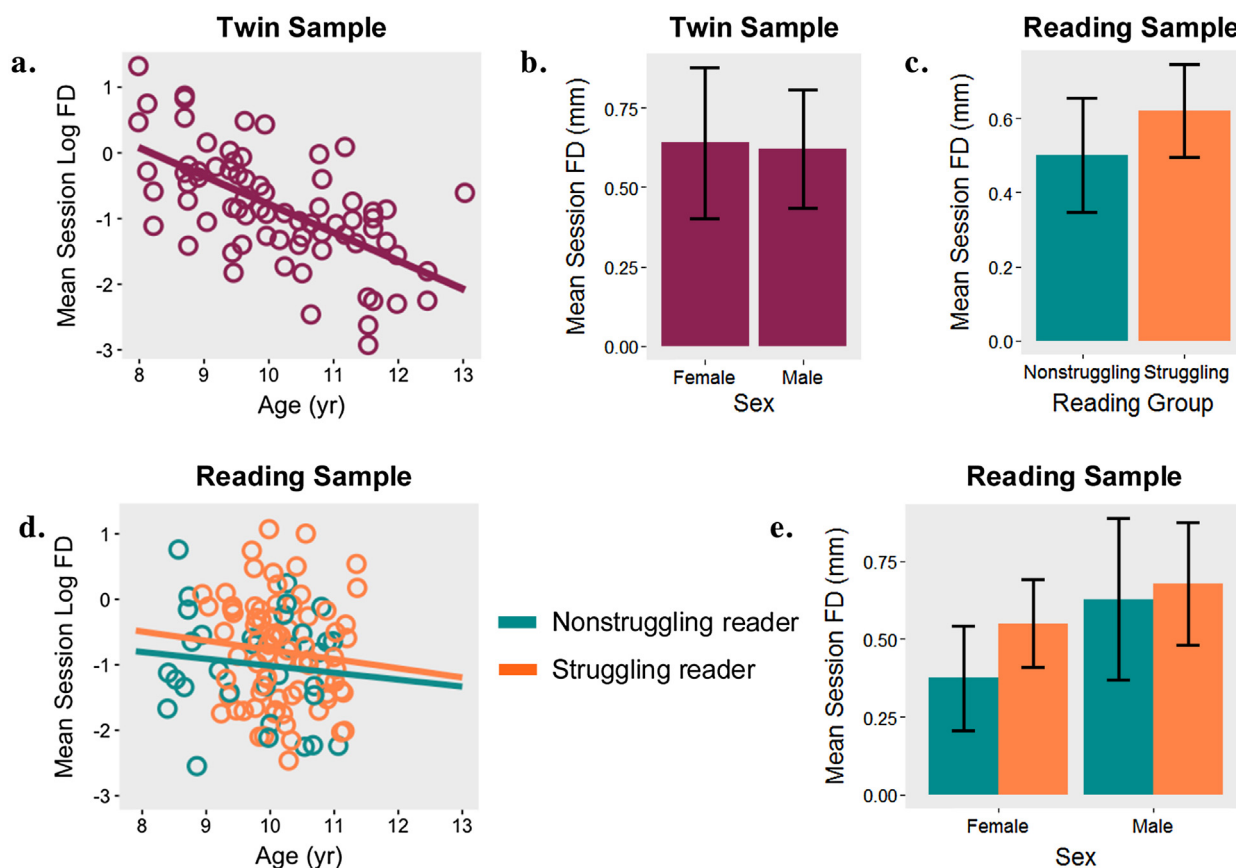


Fig. 1. Session-wide movement as a function of age, sex, and reading status.

Note. Mean framewise displacement (FD) across a scan session and its relation to age (1a) and sex (1b) in a sample of 73 twins. Relationship between FD, reading status (1c), age (1d), and sex (1e) in a sample of 78 struggling and 32 nonstruggling readers.

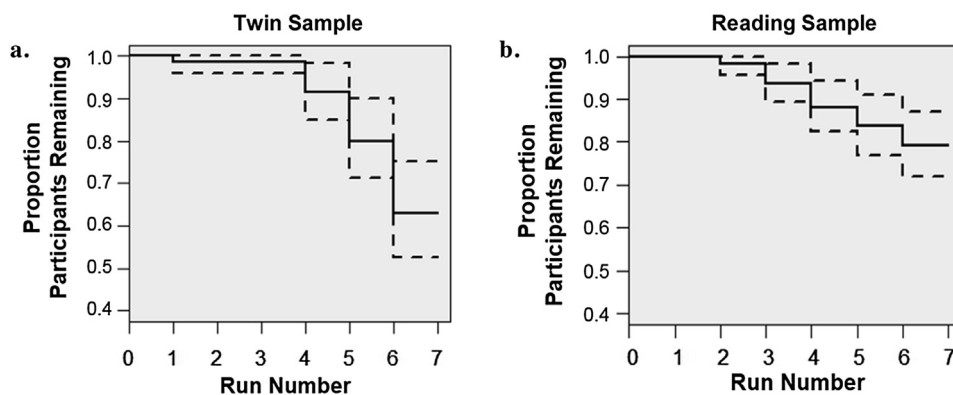


Fig. 2. Survival curves for continued participation across the first scan session.

Note. Proportion of participants providing data for each run in the twin (2a) and reading (2b) samples at the first visit. Dotted lines represent 95% confidence intervals.

yses indicated that the correlation among MZ pairs ($r=0.79$, $p<0.01$) was numerically larger than the correlation within DZ pairs ($r=0.45$, $p<0.05$), suggesting at least a partial genetic contribution to scanner movement (see Fig. 4a). As might be expected given the sample sizes, the difference between the MZ and DZ correlation coefficients was not significant ($z=1.43$, $p=0.15$; tested using *cocor* R package, which transforms coefficients to Fisher's z-scores and assesses differences between them; Diedenhofen and Musch, 2015). We were interested in whether the observed familial similarity in FD was due to participants' common age and, in many cases, common sex. We therefore computed age- and sex-partialled correlations. These were $r=0.29$ ($p=0.11$) for all twins,

$r=0.52$ ($p=0.16$) for MZ twins, and $r=0.20$ ($p=0.43$) for DZ twins. Thus, while partialling age and sex resulted in somewhat higher p -values, the patterns of familiarity remained quite similar. As an additional sensitivity analysis, we correlated participants' mean FD with that of an age- and sex-matched control (mean age difference = 3.34 months). The resulting correlation ($r=0.22$, $p=0.65$) was smaller than that obtained for the twin pairs ($r=0.59$), thus indicating that the observed co-twin similarity in FD was not simply the result of common age and sex.

In order to probe the effects of familial factors on the consistency of movement across the fMRI session, we investigated cross-twin correlations across first and second halves of the scan session. We

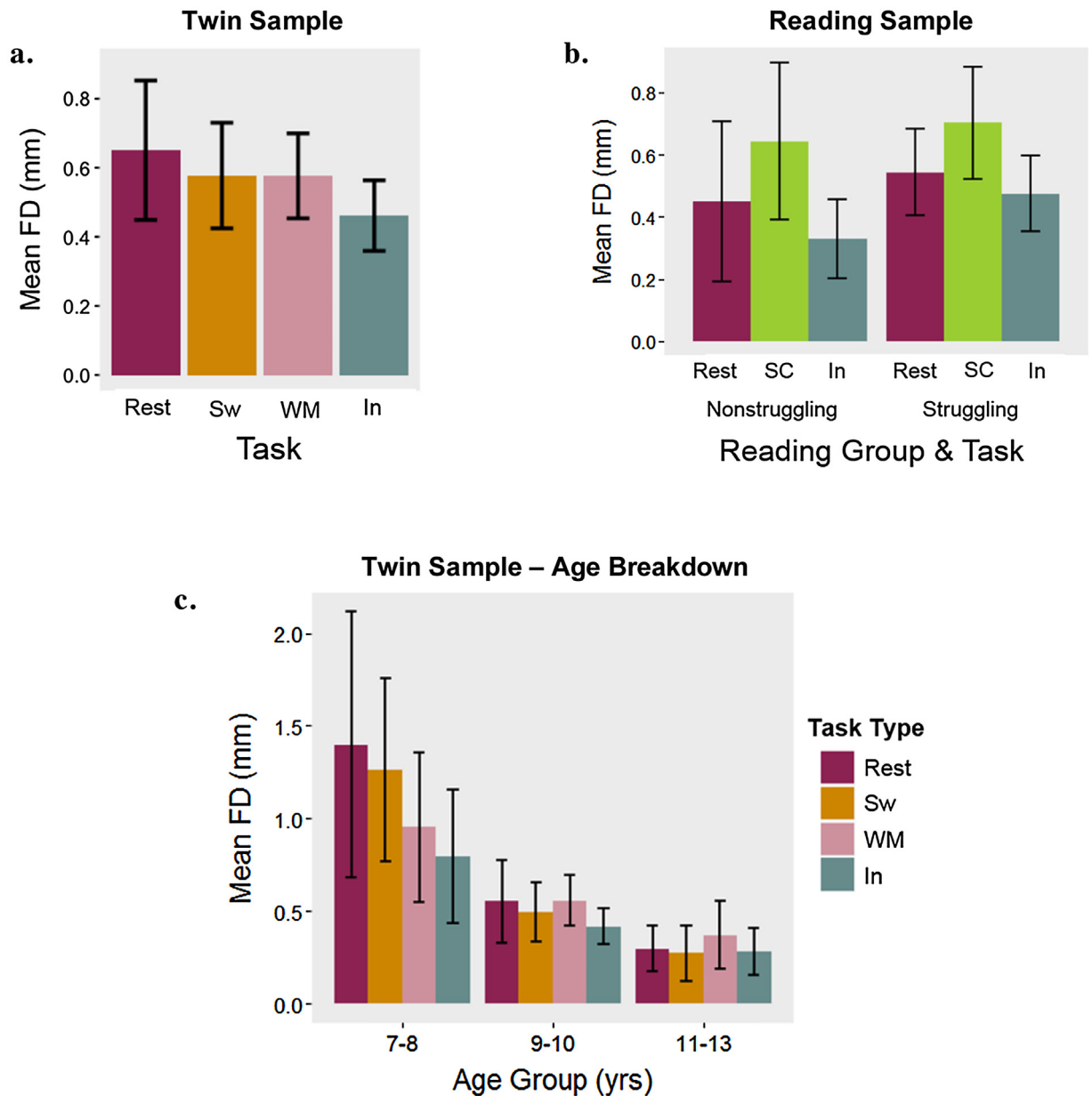


Fig. 3. Movement as a function of task type.

Note. Mean framewise displacement (FD) for each task in the twin sample (3a) and reading sample (3b). Task- and age-group specific means in the twin sample (3c). Sw = Switching, WM = Working Memory, In = Inhibition, SC = Sentence Comprehension.

Table 3
Cross-twin, split-half correlations for in-scanner movement.

	Twin 1 1 st Half	Twin 1 2 nd Half	Twin 2 1 st Half	Twin 2 2 nd Half
Twin 1 1 st Half	1	<i>r</i> = 0.71 <i>p</i> < 0.001	<i>r</i> = 0.47 <i>p</i> < 0.05	<i>r</i> = 0.53 <i>p</i> < 0.05
Twin 1 2 nd Half	<i>r</i> = 0.84 <i>p</i> < 0.001	1	<i>r</i> = 0.27 <i>p</i> = 0.24	<i>r</i> = 0.32 <i>p</i> = 0.16
Twin 2 1 st Half	<i>r</i> = 0.87 <i>p</i> < 0.001	<i>r</i> = 0.69 <i>p</i> < 0.05	1	<i>r</i> = 0.79 <i>p</i> < 0.001
Twin 2 2 nd Half	<i>r</i> = 0.64 <i>p</i> = 0.023	<i>r</i> = 0.71 <i>p</i> < 0.01	<i>r</i> = 0.73 <i>p</i> < 0.01	1

Note. Zero-order Pearson correlation coefficients. MZ correlations are provided below the diagonal; DZ correlations are provided above the diagonal.

computed mean FD for the first 50% of frames collected for an individual, then compared it to one's own mean FD for the second 50% of frames collected, as well as to the mean half-session FD of

the associated co-twin (see Table 3). Average within-twin, cross-time consistency was estimated at 0.77. By comparison, average cross-twin, cross-time consistency was estimated at 0.67 for MZ

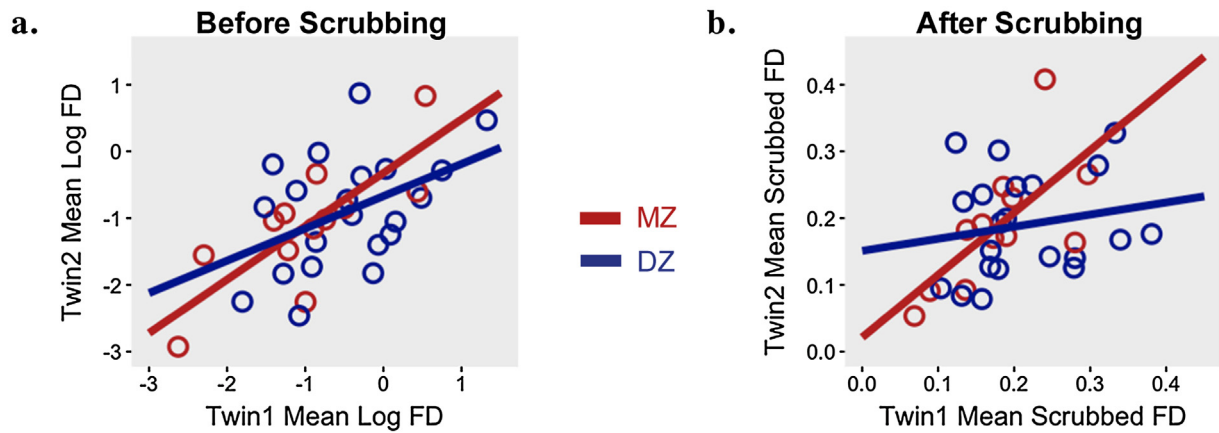


Fig. 4. Familial resemblance in movement in the twin sample, split by zygosity.

Note. Mean framewise displacement (FD) across a scan session for monozygotic (MZ) and dizygotic (DZ) twins before (4a) and after (4b) removing frames with movement above 0.2mm for rest runs and 0.9mm for task runs.

twins and 0.40 for DZ twins. This suggests that familial, partially genetic factors contribute to the consistency of movement across individuals over the course of the scan session.

3.6. Scrubbing

We sought to determine whether the key movement relationships within and across individuals observed above persisted after removing high-motion frames from the data. Following conventional motion scrubbing recommendations, we omitted frames with movement over 0.9 mm FD for task runs (Siegel et al., 2014) and over 0.2 mm FD for rest runs (Power et al., 2015). In the twin sample, an average of 58% of rest frames and 86% of task frames were retained. The mean number of motion spikes (i.e., number of frames for which movement exceeded the given threshold) was as follows: 76.13 during the rest task (42% of 180 total frames), 22.68 during the switching task (14% of 161 total frames), 15.80 during the working memory task (15% of 106 total frames), and 19.67 during the inhibition task (11% of 180 total frames). In the reading sample, an average of 63% of rest frames and 86% of task frames were retained. The mean number of motion spikes for each task in the reading sample was 72.51 for the rest task (40% of 180 total frames), 37.94 for the sentence comprehension task (18% of 212 total frames), and 17.11 for the inhibition task (10% of 180 total frames). In the twin sample, there was an inverse relationship between age and the number of motion spikes in rest runs ($b = -0.10$, $SE = 0.033$, $p < 0.01$) and task runs ($b = -0.068$, $SE = 0.013$, $p < 0.001$). No significant relationships between sex and motion spikes or ADHD and motion spikes emerged (p 's > 0.05). The number of motion spikes did not vary significantly by reading status, age, sex, or ADHD symptoms in the reading sample (p 's > 0.05), regardless of whether we assessed rest runs, task runs, or all runs.

After scrubbing, mean session FD continued to be associated with age ($b = -0.032$, $SE = 0.0070$, $p < 0.001$), but not sex ($b = 0.013$, $SE = 0.019$, $p = 0.49$), in the twin sample. In the reading sample, FD at the first scan session was not significantly related to age ($b = -0.0061$, $SE = 0.0088$, $p = 0.49$), sex ($b = 0.015$, $SE = 0.012$, $p = 0.22$), or reading status ($b = 0.011$, $SE = 0.014$, $p = 0.44$). Scrubbed movement did not significantly related to ADHD severity or symptom count in either sample (p 's > 0.05). This was true regardless of acquisition type: rest only, task only, or all runs.

Consistent with the unscrubbed results, within-person movement was highly reliable throughout the scan session (Cronbach's

$\alpha_{\text{twin}} = 0.88$; Cronbach's $\alpha_{\text{reading}} = 0.85$). Mixed linear regressions of movement onto run number with random intercepts revealed a significant increase in mean FD across runs in both samples ($b_{\text{twin}} = 0.012$, $SE = 0.0025$, $p < 0.001$; $b_{\text{reading}} = 0.0073$, $SE = 0.0018$, $p < 0.001$). The results of one-way ANOVAs estimated as mixed linear models indicated significant differences in movement as a function of task in both samples ($F_{\text{twin}}(1,2) = 7.69$, $p < 0.001$; $F_{\text{reading}}(1,1) = 87.62$, $p < 0.001$). Post-hoc tests indicated that scrubbed movement during the inhibition response task was significantly lower than movement during the other non-rest tasks.

Among repeat participants from the reading sample, scrubbed movement at the first session significantly correlated with scrubbed movement at the second session ($r = 0.54$, $p < 0.01$); this pattern was consistent when comparing rest-only ($r = 0.52$, $p < 0.01$) and task-only ($r = 0.71$, $p < 0.001$) runs across scan sessions. After scrubbing, the twin pair correlation for mean session FD remained significant ($r = 0.36$, $p < 0.05$). Interestingly, even within the narrow band of acceptable movement, zygosity differences were present (see Fig. 4b): The correlation within MZ pairs ($r = 0.69$, $p < 0.05$) was higher than the correlation within DZ pairs ($r = 0.19$, $p = 0.42$). As with the unscrubbed results ($r_{\text{MZ}} = 0.79$, $r_{\text{DZ}} = 0.45$), the difference between the correlation coefficients was not significant ($z = 1.60$, $p = 0.11$).

4. Discussion

Head movement during fMRI acquisition has emerged as a critical methodological issue, not only because of its deleterious effects on data quality, but also because methods that exclude data from participants with excessive head motion have the potential to alter the characteristics of study samples and bias results. These issues may be especially pronounced in developmental samples, as evidenced by greater mean movement among children and adolescents relative to adults (Siegel et al., 2014). When motion-confounded signals are subject to group comparison, it is challenging to tease apart the effects of group differences in motion from group differences in the task or cognitive processes under study. Motivated by recent findings that adults' head movement is highly traitlike (e.g., Van Dijk et al., 2012; Zeng et al., 2014), we examined children's in-scanner movement and its relationship to person- and scan-related characteristics in two developmental samples.

4.1. Stability in scanner movement across time and within families

The magnitude of session-wide motion reported for the current samples (0.50–0.63 mm) is consistent with that reported for other pediatric samples: For example, Kelly et al. (2008) reported mean translational (X, Y, Z) movement of 0.50 mm for a sample of 8–12-year-olds; Cantlon and Li (2013) reported mean translational movement of 1.26 mm for a sample of 4–10-year-olds, with even lower movement (0.39 mm) when participants watched a video. We found that children's in-scanner movement was highly stable within ($\alpha = 0.91$) and across ($r = 0.79$) scan sessions, consistent with stability estimates for adults ($r_s = 0.54 - 0.57$; Van Dijk et al., 2012; Zeng et al., 2014). These data suggest that children's movement is not simply a random event, but rather a lasting characteristic of the individual. Stability across repeated assessment suggests that head motion may not only confound activation differences between groups, but also has the potential to systematically affect inferences regarding individual differences in BOLD changes within a group.

Our twin analyses indicated substantial familial similarity in mean session FD. The small size of the twin sample rendered statistical comparisons of twin correlations by zygosity underpowered. Nevertheless, twin correlations differed quantitatively by zygosity, with MZ pairs exhibiting greater similarity in motion than DZ pairs. This suggests a role of genetics in individual differences in head movement (Couvry-Duchesne et al., 2014). This finding is especially relevant to studies that use family characteristics (e.g., socioeconomic status) as predictors of children's fMRI activation, as the relationship under study could be an artifact of familial differences in head movement.

4.2. Predictors of individual differences in movement

One characteristic related to the individual that emerged as a predictor of session-wide movement was age (although only in the twin sample, likely due to a more restricted age range in the reading sample). Contrary to our hypothesis, movement in the twin and reading samples did not covary with Inattentive scores, Hyperactive-Impulsivity scores, total ADHD scores, or symptom count as rated by parents or teachers. This adds to a growing literature documenting both positive (Rauch, 2005) and null (Costa Dias et al., 2013) relationships between ADHD and scanner movement. From a data retention standpoint, it is promising that the children most affected by ADHD symptoms did not move significantly more in the scanner. Nevertheless, it is possible that more severe levels of ADHD symptomology than were observed in the current study have more consistent effects on movement.

Characteristics related to the scan session that significantly predicted movement were session length, scan run length, and task type. Despite group-level increases in mean run movement across ~7 runs, overall session length, measured by the total number of frames collected, negatively correlated with movement in both samples. This relationship was likely due to discontinuing the scan sessions of high-movement individuals earlier than the scan sessions of lower-movement individuals. Movement positively and significantly correlated with run length in the reading sample, which had runs lasting 6–7 min. Both samples exhibited the lowest amount of movement during the response inhibition task, despite it being the longest task run in the twin sample. The low movement observed in both samples for this task could be because the response inhibition task (a version of Stop Signal) was engaging, fast-paced, and required frequent responses.

4.3. Persistence of motion effects after scrubbing

Many of the relationships we found between characteristics of individuals and scan sessions with movement persisted even after the removal of high-motion frames. For example, age and movement were still associated in the twin sample, within-person reliability across the session remained high, and zygoty differences in co-twin similarity persisted. Thus, even within a narrow band of “tolerable” movement, traitlike individual differences in motion were still apparent. While movement control is clearly a necessary step for isolating brain signals more likely attributable to the cognitive process under study, new approaches and techniques that are more robust to movement artifact, but that can study brain activity, are clearly needed. Because we find that head movement is a systematic property of the individual and not simply random noise, failures to deal with head movement in a sophisticated way will not only result in lower power or more conservative effects, but may also lead to biased results. Increasing sample size, while effective in bolstering power, will continue to undersample individuals from populations whose defining characteristics correlate with movement, thus providing a potentially unrepresentative picture of brain activity across the range of development.

4.4. Recommendations for experimenters

The current study underscores the need to continue investigating potential sources or correlates of children's scanner movement, which will allow us to better understand the factors that buffer or exacerbate movement during data acquisition. As we seek to become more diverse and inclusive in our neuroimaging samples, we must consider ways to combat movement prior to the application of traditional post-acquisition motion controls. Toward this end, careful characterization of study samples is critical, as it will allow us to identify predictors of dropped frames and determine best practices for screening.

In more practical terms, the current results suggest that developmental neuroimagers may benefit from using fast-paced, high-response tasks, especially for tasks requiring longer acquisition periods (exceeding 5 min). Given that movement significantly increased as a function of run time – but not aggregate session time – in the reading sample, it may be productive to keep task runs under six or seven minutes when possible. This has the additional benefit of allowing for more mixing of different task runs at the beginning of the scan session, prior to points at which participants become more likely to discontinue the scan session. Survival curves indicated that, in both samples, participation dropped below 90% approximately 40 min into the scan session. This indicates that children as young as seven years old tolerate scan sessions under an hour, with dropout becoming more substantial with increasing time. As such, frontloading all tasks may increase the likelihood that children will provide usable data for each task, even in the absence of repeated task exposure. Previous publications supply many other recommendations relevant for scanning children (Church et al., 2010; Greene et al., 2016), including the use of mock scanners to familiarize participants with the scanning environment and presentation of movies during anatomical scans. Recent studies investigating network connectivity during movie watching suggest that showing movies during functional scans can significantly reduce motion and increase data retention (e.g., Cantlon and Li, 2013; Emerson et al., 2015; Vanderwal et al., 2015). Future studies should investigate the impact of different types of exposure or mock scanner training on participant anxiety, engagement, and movement. Further, we find face-to-face training on scanner tasks, real-time monitoring of movement and image quality, and frequent communication with child participants to be highly useful in producing good quality data, although we have not manipulated

these factors directly. New methods for addressing motion artifact in real-time (e.g., Frame-wise Real-time Integrated MRI Motion Monitoring, FIRMM, developed by Nico Dosenbach and others at Washington University) offer new hope to improve collection from higher moving participants.

4.5. Limitations

The current study did not examine how movement or scrubbing affected image quality or neural patterns; these questions should be addressed in future work in order to fully understand the impact of systematic differences in children's head motion. Although the current study had relatively small samples of repeat participants and twin pairs, it benefited from the use of two datasets uniquely suited to the investigation of individual differences in scanner movement during childhood and early adolescence. Nevertheless, there remains much work to be done in characterizing the predictors and outcomes of movement in younger or older samples, as well as in those with greater clinical burden.

Conflict of interest

None.

Acknowledgements

This research project was supported by National Institutes of Health grants P50 HD052117 (JAC and JJ), R21HD081437 (JAC and ETD), and R01HD083613 (ETD). The Population Research Center at The University of Texas at Austin is supported by National Institutes of Health grant R24HD042849. Additional support came from the University of Texas Imaging Research Center Pilot Grant 20141031a. L. E. Engelhardt was supported by a National Science Foundation Graduate Research Fellowship. We wish to acknowledge the entire Texas Learning Disabilities Research Center team, especially Jack Fletcher and Sharon Vaughn. We also wish to acknowledge the Core for Advanced MRI (CAMRI) at Baylor College of Medicine. Finally, we thank our participating families for their time and effort.

References

- Anokhin, A.P., Heath, A.C., Myers, E., Ralano, A., Wood, S., 2003. Genetic influences on prepulse inhibition of startle reflex in humans? *Neurosci. Lett.* 353 (1), 45–48.
- Bedard, A.-C., Nichols, S., Barbosa, J.A., Schachar, R., Logan, G.D., Tannock, R., 2002. The development of selective inhibitory control across the life span. *Dev. Neuropsychol.* 21 (1), 93–111, http://dx.doi.org/10.1207/S15326942DN2101_5.
- Bullmore, E.T., Brammer, M.J., Rabe-Hesketh, S., Curtis, V.A., Morris, R.G., Williams, S.C.R., et al., 1999. Methods for diagnosis and treatment of stimulus-correlated motion in generic brain activation studies using fMRI. *Hum. Brain Mapp.* 7 (1), 38–48.
- Cantlon, J.F., Li, R., 2013. Neural activity during natural viewing of sesame street statistically predicts test scores in early childhood. *PLoS Biol.* 11 (1), e1001462, <http://dx.doi.org/10.1371/journal.pbio.1001462>.
- Church, J.A., Petersen, S.E., Schlaggar, B.L., 2010. The Task B problem and other considerations in developmental functional neuroimaging. *Hum. Brain Mapp.* 31 (6), 852–862, <http://dx.doi.org/10.1002/hbm.21036>.
- Conners, C.K., 1997. *Conners' Parent Rating Scales-Revised: Technical Manual*. Multi-Health Systems Inc., Toronto, Canada.
- Costa Dias, T.G., Wilson, V.B., Bathula, D.R., Iyer, S.P., Mills, K.L., Thurlow, B.L., et al., 2013. Reward circuit connectivity relates to delay discounting in children with attention-deficit/hyperactivity disorder. *Eur. Neuropsychopharmacol.* 23 (1), 33–45, <http://dx.doi.org/10.1016/j.euroneuro.2012.10.015>.
- Couvry-Duchesne, B., Blokland, G.A.M., Hickie, I.B., Thompson, P.M., Martin, N.G., de Zubicaray, G.I., et al., 2014. Heritability of head motion during resting state functional MRI in 462 healthy twins. *Neuroimage* 102 (Part 2), 424–434, <http://dx.doi.org/10.1016/j.neuroimage.2014.08.010>.
- Couvry-Duchesne, B., Ebejer, J.L., Gillespie, N.A., Duffy, D.L., Hickie, I.B., Thompson, P.M., et al., 2016. Head motion and inattention/hyperactivity share common genetic influences: implications for fMRI studies of ADHD. *PLoS One* 11 (1), e0146271 <http://dx.doi.org/10.1371/journal.pone.0146271>.
- Diedenhofen, B., Musch, J., 2015. cocor: a comprehensive solution for the statistical comparison of correlations. *PLoS One* 10 (4), <http://dx.doi.org/10.1371/journal.pone.0121945> (e0121945).
- Durston, S., Tottenham, N.T., Thomas, K.M., Davidson, M.C., Eigsti, I.-M., Yang, Y., et al., 2003. Differential patterns of striatal activation in young children with and without ADHD. *Biol. Psychiatry* 53 (10), 871–878.
- Emerson, R.W., Short, S.J., Lin, W., Gilmore, J.H., Gao, W., 2015. Network-level connectivity dynamics of movie watching in 6-year-old children. *Front. Hum. Neurosci.* 9 (631), <http://dx.doi.org/10.3389/fnhum.2015.00631>.
- Epstein, J.N., Casey, B.J., Tonev, S.T., Davidson, M., Reiss, A.L., Garrett, A., et al., 2007. Assessment and prevention of head motion during imaging of patients with attention deficit hyperactivity disorder. *Psychiatry Res.: Neuroimag.* 155 (1), 75–82, <http://dx.doi.org/10.1016/j.pscychresns.2006.12.009>.
- Fisher, A., van Jaarsveld, C.H.M., Llewellyn, C.H., Wardle, J., 2010. Environmental influences on children's physical activity: quantitative estimates using a twin design. *PLoS One* 5 (4), e10110, <http://dx.doi.org/10.1371/journal.pone.0010110>.
- Friston, K.J., Williams, S., Howard, R., Frackowiak, R.S.J., Turner, R., 1996. Movement-related effects in fMRI time-series? *Magn. Reson. Med.* 35 (3), 346–355.
- Greene, D.J., Black, K.J., Schlaggar, B.L., 2016. Considerations for MRI study design and implementation in pediatric and clinical populations. *Dev. Cognit. Neurosci.* 18, 101–112, <http://dx.doi.org/10.1016/j.dcn.2015.12.005>.
- Harden, K.P., Tucker-Drob, E.M., Tackett, J.L., 2013. The Texas Twin Project. *Twin Res. Hum. Genet.* 16 (1), 385–390, <http://dx.doi.org/10.1017/thg.2012.97>.
- Heath, A.C., Nyholt, D.R., Neuman, R., Madden, P.A., Bucholz, K.K., Todd, R.D., et al., 2003. Zygosity diagnosis in the absence of genotypic data: an approach using latent class analysis. *Twin Res.* 6 (1), 22–26.
- Kelly, A.M.C., Di Martino, A., Uddin, L.Q., Shehzad, Z., Gee, D.G., Reiss, P.T., et al., 2008. Development of anterior cingulate functional connectivity from late childhood to early adulthood. *Cereb. Cortex* 19 (3), 640–657, <http://dx.doi.org/10.1093/cercor/bhn117>.
- Laumann, T.O., Snyder, A.Z., Mitra, A., Gordon, E.M., Gratton, C., Adeyemo, B., et al., 2016. On the stability of BOLD fMRI correlations. *Cereb. Cortex*, <http://dx.doi.org/10.1093/cercor/bhw265>, in press.
- Olesen, P.J., Macoveanu, J., Tegner, J., Klingberg, T., 2007. Brain activity related to working memory and distraction in children and adults. *Cereb. Cortex* 17 (5), 1047–1054, <http://dx.doi.org/10.1093/cercor/bhl014>.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team, 2016. nlme: Linear and Nonlinear Mixed Effects Models, R Package Version 3., pp. 1–128 (Retrieved from) <http://CRAN.R-project.org/package=nlme>.
- Poldrack, R.A., Mumford, J.A., Nichols, T.E., 2011. *Handbook of Functional MRI Data Analysis*. University Press, New York: Cambridge.
- Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2012. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* 59 (3), 2142–2154, <http://dx.doi.org/10.1016/j.neuroimage.2011.10.018>.
- Power, J.D., Mitra, A., Laumann, T.O., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2014. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* 84 (0), 320–341, <http://dx.doi.org/10.1016/j.neuroimage.2013.08.048>.
- Power, J.D., Schlaggar, B.L., Petersen, S.E., 2015. Recent progress and outstanding issues in motion correction in resting state fMRI. *Neuroimage* 105, 536–551, <http://dx.doi.org/10.1016/j.neuroimage.2014.10.044>.
- R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (Retrieved from) <http://www.R-project.org/>.
- Rauch, S.L., 2005. Neuroimaging and attention-deficit/hyperactivity disorder in the 21st century: what to consider and how to proceed. *Biol. Psychiatry* 57 (11), 1261–1262, <http://dx.doi.org/10.1016/j.biopsych.2005.02.014>.
- Rubin, D.B., 1976. Inference and missing data. *Biometrika* 63 (3), 581–592, <http://dx.doi.org/10.1093/biomet/63.3.581>.
- Satterthwaite, T.D., Wolf, D.H., Loughhead, J., Ruparel, K., Elliott, M.A., Hakonarson, H., et al., 2012. Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. *Neuroimage* 60 (1), 623–632, <http://dx.doi.org/10.1016/j.neuroimage.2012.08.052>.
- Satterthwaite, T.D., Elliott, M.A., Gerraty, R.T., Ruparel, K., Loughhead, J., Calkins, M.E., et al., 2013. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage* 64, 240–256, <http://dx.doi.org/10.1016/j.neuroimage.2012.08.052>.
- Siegel, J.S., Power, J.D., Dubis, J.W., Vogel, A.C., Church, J.A., Schlaggar, B.L., Petersen, S.E., 2014. Statistical improvements in functional magnetic resonance imaging analyses produced by censoring high-motion data points. *Hum. Brain Mapp.* 35 (5), 1981–1996, <http://dx.doi.org/10.1002/hbm.22307>.
- Siegel, J.S., Mitra, A., Laumann, T.O., Seitzman, B.A., Raichle, M., Corbetta, M., Snyder, A.Z., 2016. Data quality influences observed links between functional connectivity and behavior. *Cereb. Cortex*, <http://dx.doi.org/10.1093/cercor/bhw253>, in press.
- Swanson, J., 2011. *Strengths and Weaknesses of ADHD Symptoms and Normal Behavior Rating Scale (SWAN)*.
- Therneau, T., 2015. A Package for Survival Analysis in S.R. Package, Version 2.38, Retrieved from <http://CRAN.R-project.org/package=survival>.
- Torgesen, J.K., Wagner, R.K., Rashotte, C.A., 2012. *Test of Word Reading Efficiency-2. TX: Pro-Ed, Austin*.

- Van Dijk, K.R.A., Sabuncu, M.R., Buckner, R.L., 2012. The influence of head motion on intrinsic functional connectivity MRI. *Neuroimage* 59 (1), 431–438, <http://dx.doi.org/10.1016/j.neuroimage.2011.07.044>.
- Vanderwal, T., Kelly, C., Eilbott, J., May, L.C., Castellanos, F.X., 2015. *Inscapes*: A movie paradigm to improve compliance in functional magnetic resonance imaging. *Neuroimage* 122, 222–232, <http://dx.doi.org/10.1016/j.neuroimage.2015.07.069>.
- Wagner, R.K., Torgesen, J., Rashotte, C.A., Pearson, N., 2010. *Test of Sentence Reading Efficiency and Comprehension*. TX: Pro-Ed, Austin.
- Williams, B.R., Ponesse, J.S., Schachar, R.J., Logan, G.D., Tannock, R., 1999. Development of inhibitory control across the life span. *Dev. Psychol.* 35 (1), 205–213.
- Yan, C.-G., Cheung, B., Kelly, C., Colcombe, S., Craddock, R.C., Di Martino, A., et al., 2013. A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *Neuroimage* 76, 183–201, <http://dx.doi.org/10.1016/j.neuroimage.2013.03.004>.
- Zeng, L.-L., Wang, D., Fox, M.D., Sabuncu, M., Hu, D., Ge, M., et al., 2014. Neurobiological basis of head motion in brain imaging. *Proc. Natl. Acad. Sci.* 111 (16), 6058–6062, <http://dx.doi.org/10.1073/pnas.1317424111>.