NeuroImage 88 (2014) 79–90

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Spurious group differences due to head motion in a diffusion MRI study

Anastasia Yendiki [a,*], Kami Koldewyn [b], Sita Kakunoori [a], Nancy Kanwisher [b], Bruce Fischl [a,c]

[a] Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
[b] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA
[c] Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA

ARTICLE INFO

Article history:
Accepted 14 November 2013
Available online 21 November 2013

Keywords:
Diffusion MRI
Tractography
Motion
Autism

ABSTRACT

Diffusion-weighted MRI (DW-MRI) has become a popular imaging modality for probing the microstructural properties of white matter and comparing them between populations in vivo. However, the contrast in DW-MRI arises from the microscopic random motion of water molecules in brain tissues, which makes it particularly sensitive to macroscopic head motion. Although this has been known since the introduction of DW-MRI, most studies that use this modality for group comparisons do not report measures of head motion for each group and rely on registration-based correction methods that cannot eliminate the full effects of head motion on the DW-MRI contrast. In this work we use data from children with autism and typically developing children to investigate the effects of head motion on differences in anisotropy and diffusivity measures between groups. We show that group differences in head motion can induce group differences in DW-MRI measures, and that this is the case even when comparing groups that include control subjects only, where no anisotropy or diffusivity differences are expected. We also show that such effects can be more prominent in some white-matter pathways than others, and that they can be ameliorated by including motion as a nuisance regressor in the analyses. Our results demonstrate the importance of taking head motion into account in any population study where one group might exhibit more head motion than the other.

© 2013 Elsevier Inc. All rights reserved.

Introduction

Diffusion-weighted MRI (DW-MRI) encodes information on the direction and speed of the diffusion of water molecules in the intensity values of the acquired images. In neuroimaging this has become a tool for inferring the local orientation of white-matter (WM) pathways at every voxel in the brain, as well as deriving measures of diffusivity and anisotropy that are thought to reflect the local structure and integrity of those pathways. These measures have been used to follow progressive changes in the brain across the lifespan (Yoshida et al., 2013; Salat, in press) and to study the effects of a variety of conditions, including Alzheimer's disease (Stebbins and Murphy, 2009), Huntington's disease (Bohanna et al., 2008), Parkinson's disease (Cochrane and Ebmeier, 2013), multiple sclerosis (Inglese and Bester, 2010), schizophrenia (Kubicki et al., 2007), and autism (Travers et al., 2012).

However, the populations compared in such studies may differ not only in terms of WM structure, but also in how likely they are to exhibit head motion during the scan. Remaining still in the scanner may be more challenging for some age groups than others. It may also be more challenging for subjects with one of the aforementioned disorders than control subjects. This can make group comparisons of measures derived from DW-MRI scans problematic. Subject motion during the acquisition of a DW-MRI series will not only result in misalignment between the images in the series; but can also alter the intensity values in the images, because motion during the diffusion-encoding gradient pulses leads to attenuation of the image intensity. That is, the very phenomenon that gives rise to the DW-MRI contrast is also what makes it particularly sensitive to subject motion. Signal attenuation due to macroscopic head motion can confound the measurement of interest, which is signal attenuation due to microscopic random motion of water molecules in tissues. If a subject moves only during the application of one diffusion-encoding gradient, this can give the appearance of preferential diffusion in the direction of that gradient and lead to an overestimation of diffusion anisotropy. If a subject moves randomly throughout the scan, this can reduce the contrast between diffusion directions and lead to an underestimation of diffusion anisotropy.

The deleterious effects of head motion on DW-MRI have been known since the early days of its application to neuroimaging (Anderson and Gore, 1994). However, the issue has received surprisingly little attention in the numerous DW-MRI studies of clinical populations that have been published since then. It is common to realign the images in a DW-MRI series to each other (Andersson and Skare, 2002; Rohde et al., 2004). This will mitigate motion artifacts but not remove them completely, and most studies do not report the levels of detected motion by group. For example, 48 studies of autism spectrum disorders (ASD) that use DW-MRI are reviewed in Travers et al. (2012). Almost all of these studies report significant differences in diffusion measures between subjects with ASD and

* Corresponding author.
E-mail address: ayendiki@nmr.mgh.harvard.edu (A. Yendiki).

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.11.027

control subjects. However, only five of the studies evaluate some measure related to head motion for each group and report that it is comparable between groups.

In this work, we use data collected from children with ASD and typically developing (TD) children to investigate the effects of head motion on measures of anisotropy and diffusivity derived from DW-MRI and tractography. We show that group differences in such measures can increase substantially for small increases in the difference in head motion between groups. This is the case not only when the groups being compared are children with ASD vs. TD children, but also when both groups include TD children only. We also show that DW-MRI findings may be more sensitive to head motion for some WM pathways than others. Our results have implications not only for autism studies but for a wide range of neurological and psychiatric applications where the population under study and the control population are likely to exhibit different levels of head motion.

Materials and methods

Data acquisition

All MRI data was collected at the Massachusetts Institute of Technology, using a Siemens 3 T *Magnetom Tim Trio* scanner (Siemens, Erlangen, Germany) with a custom-made 32-channel pediatric head coil (Keil et al., 2011). All sessions included DW images and T_1-weighted images.

The DW images were acquired using a conventional 2D spin-echo echo-planar imaging (EPI) sequence. The series included 30 images acquired with diffusion weighting along non-colinear directions ($b = 700$ s $\cdot$ mm^{-2}), and 10 images acquired without diffusion weighting ($b = 0$). The acquisition parameters were: 2 mm isotropic resolution, matrix size 128×128, number of slices ranging from 52 to 74 and chosen for full brain coverage, no inter-slice gap, TE = 84ms, TR ranging from 8.04 s to 14.18 s depending on the number of slices, BW = 1395 Hz/px, and GRAPPA acceleration factor 2.

The T_1-weighted images were acquired using a 3D multi-echo magnetization-prepared gradient echo (MP-RAGE) sequence with prospective motion correction (van der Kouwe et al., 2008; Tisdall et al., 2012). The acquisition parameters were: 1 mm isotropic resolution, $192 \times 192 \times 176$ image matrix, 12 echos with minimum TE = 1.64 ms and maximum TE = 27 ms, TR = 2.53 s, BW = 651 Hz/pixel, flip angle 7°, and GRAPPA acceleration factor 3.

Participants

The data described above was collected for 112 children, 50 in the ASD group and 62 in the TD group. Several of the subjects were scanned twice, leading to a total of 165 scans. The subjects' ages were 5–12 years. Their non-verbal IQ was evaluated using the Kaufman brief intelligence test II (Kaufman and Kaufman, 2004). All subjects included in the study had non-verbal IQ of at least 80; no history of birth or brain trauma; and normal or corrected-to-normal vision. The children in the ASD group were evaluated using the diagnostic criteria in the DSM-IV, as well as the autism diagnostic observation schedule (ADOS) (Lord et al., 2000). In addition, all children were evaluated on the social responsiveness scale (SRS) (Constantino et al., 2007). For more information on the standardized tests administered as part of this study, see Koldewyn et al. (2013).

Image analysis

For each DW-MRI scan, we aligned all images in the series to the first non-diffusion-weighted image using affine registration (Jenkinson et al., 2002) and reoriented the corresponding diffusion-weighting gradient vectors accordingly (Rohde et al., 2004; Leemans and Jones, 2009). Affine registration between volumes is a processing step that is commonly applied to DW-MRI data to reduce misalignment between the images due to head motion and eddy currents. To quantify head motion in each scan, we derived volume-by-volume translation and rotation from this affine registration, as well as slice-by-slice signal drop-out measures that are specific to DW-MRI (Benner et al., 2011). The registration-based measures are better at capturing slower, between-volume motion, whereas the intensity-based measures are better at capturing more rapid, within-volume motion. In more detail the motion measures were:

1. *Average volume-by-volume translation*: We used the translation component of the affine registration from each volume to the first volume to compute the translation vector between each pair of consecutive volumes. We averaged the magnitude of these translation vectors over all volumes in the scan.
2. *Average volume-by-volume rotation*: We used the rotation component of the affine registration from each volume to the first volume to compute the rotation angles between each pair of consecutive volumes. We averaged the sum of the absolute values of these rotation angles over all volumes in the scan.
3. *Percentage of slices with signal drop-out*: We computed the signal drop-out score proposed in Benner et al. (2011) for each slice in each volume. Slices with a score greater than 1 are considered to have suspect signal drop-out. We computed the percentage of slices in the entire scan that had a score greater than 1.
4. *Signal drop-out severity*: We computed the average signal drop-out score over all slices in the scan that had a score greater than 1.

We used TRActs Constrained by UnderLying Anatomy (TRACULA) to delineate 18 major WM fascicles in each scan (Yendiki et al., 2011). This is an algorithm for automated global probabilistic tractography that estimates the posterior probability of each of the 18 pathways given the DW-MRI data. The posterior probability is decomposed into a data likelihood term, which uses the "ball-and-stick" model of diffusion (Behrens et al., 2007), and a pathway prior term, which incorporates prior anatomical knowledge on the pathways from a set of training subjects. The information extracted from the training subjects is the probability of each pathway passing through (or next to) each anatomical segmentation label. This probability is calculated separately for every point along the trajectory of the pathway. Thus there is no assumption that the pathways have the same shape in the study subjects and training subjects, only that the pathways traverse the same regions relative to the surrounding anatomy. In other words, TRACULA does not rely on perfect alignment between the study subjects and training subjects. The anatomical segmentation labels required by TRACULA were obtained by processing the T_1-weighted images of the study subjects with the automated cortical parcellation and subcortical segmentation tools in FreeSurfer (Fischl et al., 2002, 2004a,b). More details on the tractography method, as well as an evaluation of its accuracy on healthy subjects and schizophrenia patients, can be found in Yendiki et al. (2011).

The pathways reconstructed by TRACULA are: corticospinal tract (CST), uncinate fasciculus (UNC), inferior longitudinal fasciculus (ILF), anterior thalamic radiations (ATR), cingulum–cingulate gyrus bundle (CCG), cingulum–angular bundle (CAB), superior longitudinal fasciculus-parietal terminations (SLFP), superior longitudinal fasciculus-temporal terminations (SLFT), corpus callosum–forceps major (FMAJ), and corpus callosum–forceps minor (FMIN). Other than the corpus callosum, all other pathways are reconstructed for the left (L) and right (R) hemisphere. Fig. 1 shows an example reconstruction, where an isosurface of the probability distribution of each pathway is displayed.

We obtained mean values of the fractional anisotropy (FA), mean diffusivity (MD), radial diffusivity (RD), and axial diffusivity (AD) in each of the 18 WM pathways reconstructed by TRACULA for each subject. To compute these mean values, the pathway distributions were thresholded at 20% of their maximum value, and the FA, MD, RD, and AD values at each voxel were weighted by the pathway probability at that voxel. We also computed the average FA, MD, RD, and AD in the

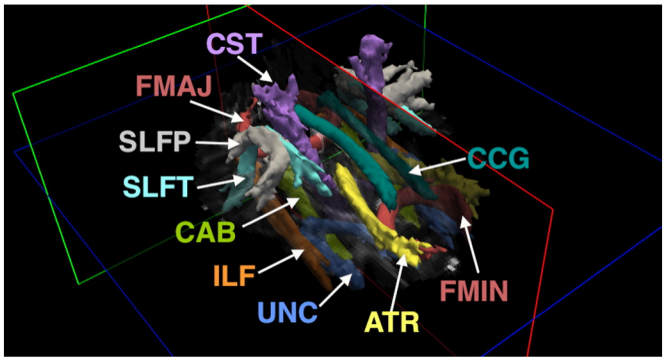

Fig. 1. WM pathways reconstructed by TRACULA. ATR: anterior thalamic radiations; CAB: cingulum–angular bundle; CCG: cingulum–cingulate gyrus bundle; CST: corticospinal tract; FMAJ: corpus callosum–forceps major; FMIN: corpus callosum–forceps minor; ILF: inferior longitudinal fasciculus; SLFP: superior longitudinal fasciculus-parietal terminations; SLFT: superior longitudinal fasciculus-temporal terminations; UNC: uncinate fasciculus.

entire WM for each subject. For this purpose we generated a WM mask from the subject's anatomical segmentation and mapped it from the space of the T_1-weighted image to the space of the DWIs. Note that the tensor model was fit to the data only to extract these anisotropy and diffusivity measures, and not to perform the tractography in TRACULA, which relies on the ball-and-stick model of diffusion instead.

Our quality assurance procedure involved careful visual inspection of all the DW images, FA maps, and tractography reconstructions, in conjunction with the motion measures described above. As will be discussed further in Discussion section, we excluded from any further analysis 17 scans (12 of them from children with ASD) that were deemed to have excessive motion. Thus all analyses presented in the following include the remaining 148 scans only.

Group comparisons

Our goal was to examine whether more group differences in DW-MRI measures would be detected between groups with different levels of head motion than between groups with similar levels of head motion. We had at our disposal subjects with different amounts of motion and, for some of those subjects, two scans with different amounts of motion. Thus we were able to generate multiple combinations of scans to include in each group and, for each combination, quantify the differences in motion parameters and the differences in DW-MRI measures between groups. We first considered the case where a group of children with ASD was compared to an age-matched group of TD children. We then considered the case where both groups consisted of age-matched TD children only, thus no group differences in DW-MRI measures were expected. Finally, using the TD subjects that had two scans, we examined differences between scans from the same children but with different levels of motion. All the aforementioned analyses where performed on mean values of the DW-MRI measures over the entire pathways. To further investigate how the effects were distributed spatially, we also performed a voxel-wise analysis of the association between FA and motion.

ASD vs. TD

First we investigated how motion affected the differences between subjects with autism and control subjects. We generated 50,000 random combinations of 30 children with ASD and 30 age-matched TD children. For subjects that had two scans, we chose one of the two scans at random each time. For each of the 18 pathways and for the entire WM, we computed *T*-tests on the difference in the mean FA, MD, RD, and AD between the ASD and TD group, using age as a nuisance regressor. For each of the 50,000 scan combinations, we recorded how many of the 18 pathways exhibited significant differences in each DW-MRI measure at the $p < 0.05$ level between the ASD and TD group. We also computed the difference in the average motion measures between the ASD and TD group.

Introducing nuisance regressors is a common ad hoc approach to accounting for confounds in neuroimaging studies. We examined whether the use of a motion score as a nuisance regressor would reduce findings of statistically significant differences in DW-MRI measures between groups. We define here the following *total motion index* (TMI) for the *i*-th subject:

$$\mathrm{TMI}_i \triangleq \sum_{j=1}^{4} \frac{x_{ij} - M_j}{Q_j - q_j},$$

where $j = 1,\ldots,4$ indexes the four motion measures described in Image analysis section, x_{ij} is the value of the j-th motion measure for the i-th subject, and M_j, Q_j, and q_j are, respectively, the median, upper quartile, and lower quartile of the j-th motion measure over all subjects included in a group comparison. Note that the mean and standard deviation are not good measures of central tendency and dispersion for the four motion parameters, as their distributions are skewed (see Results section). We repeated the group comparisons of the mean FA, MD, RD, and AD for each of the 50,000 scan combinations, using TMI as a nuisance regressor (in addition to age).

TD vs. TD

To confirm that motion-induced differences were not specific to autism, we repeated the previous experiment using only control subjects. This time we generated random combinations of 60 TD children that could be split into two age-matched groups of 30. We adopted the following procedure for generating groups that had subtle differences in head motion. For subjects with a single scan, that scan could be drawn either for group 1 or for group 2. For subjects with two scans, we used the lower-motion scan when the subject was drawn for group 1 and the higher-motion scan when the subject was drawn for group 2. Comparisons between groups 1 and 2 were carried out for each of the 18 pathways and for the entire WM, as described in the previous section. The frequency with which significant group differences in DW-MRI measures were detected was now the false positive rate, as no differences are expected between random combinations of TD children.

Test vs. retest

To demonstrate definitively that motion can generate false positives where no true differences exist, we used data from the subjects that had two scans. In this experiment we included only TD children whose motion parameters were below a rather stringent threshold (the median plus 1.5 times the interquartile range of the cohort) for both of their scans. This left us with 25 subjects. We used paired *T*-tests to test for differences in DW-MRI measures in each pathway between the lower- and higher-motion scans of these 25 children.

Voxel-based analysis

The tractography-based approach that we followed in this work was to compute mean values of DW-MRI measures over each pathway in each subject's native space, and compare these mean values across subjects. A popular alternative is the voxel-based approach, where the images of the subjects are aligned in a common (template) space, and the values of DW-MRI measures at individual voxels are compared across subjects in this template space. To further investigate the effects of motion with the voxel-based approach, we used Tract-Based Spatial Statistics (TBSS), a popular method for aligning FA maps across subjects in a template space and performing voxel-wise statistics on FA values on the interior skeleton of the WM (Smith et al., 2006). In this analysis we included data from all subjects. We used group, age, and motion parameters as regressors. We tested for voxels with a statistically significant association of FA with motion, using non-parametric permutation testing

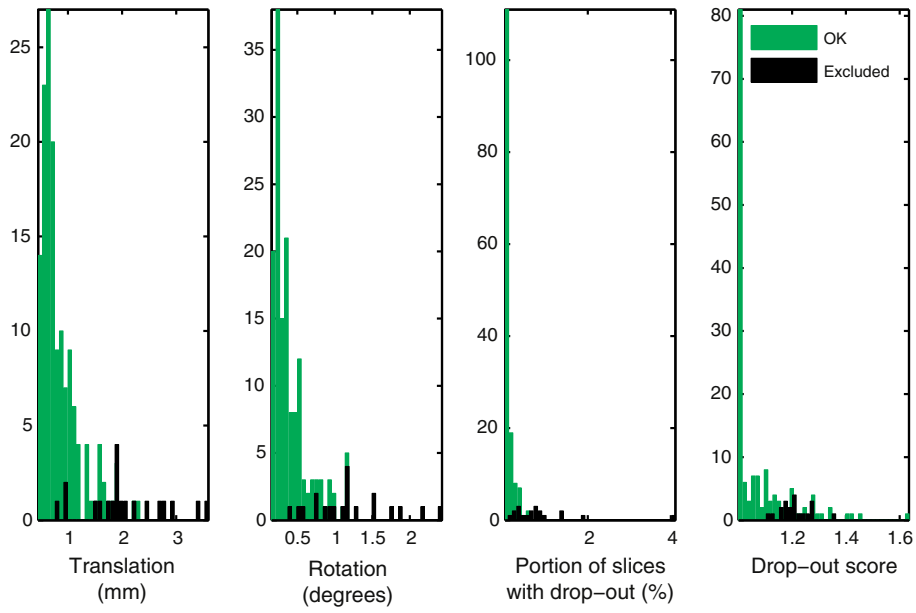

Fig. 2. Overview of motion measures. Histograms of the four motion measures are shown for the 148 scans that were included in our analyses (green) and 17 scans that were excluded due to excessive motion (black).

(Nichols and Holmes, 2002) and threshold-free cluster enhancement (Smith and Nichols, 2009).

Results

Overview of motion in the data

Fig. 2 shows histograms of the four measures of motion described in Image analysis section for the 148 data sets that were included in the analyses, and for the 17 data sets that were excluded by visual inspection due to egregious motion artifacts. Fig. 3 shows box-and-whisker plots of these measures by group for the 148 scans that were deemed acceptable. As seen in these plots, the median of all four motion measures was higher in the scans of children with ASD than those of TD children. A Wilcoxon rank-sum test showed statistically significant group differences (translation: $p = 0.009$; rotation: $p = 0.0006$; portion of slices with signal drop-out: $p = 0.01$; signal drop-out score: $p = 0.02$). For children that had two scans, the median time between scans was 29 days and the interquartile range was 40 days. There was no general tendency for more or less motion in the earlier scan compared to the later scan (translation: $p = 0.48$; rotation: $p = 0.57$; portion of slices with signal drop-out: $p = 0.89$; signal drop-out score: $p = 0.83$). There was no difference in the time between scans between groups ($p = 0.97$).

Table 1 contains demographic information on the subjects whose scans were included in the analyses, including age, IQ, SRS score, and ADOS score. Pearson correlation coefficients of each of these variables to translational and rotational motion are also shown. There were no significant correlations of age, IQ, or SRS scores with motion measures. However, the ADOS symptom severity scores of the autistic children were positively correlated with rotational motion ($p = 0.04$).

ASD vs. TD

In Fig. 4 we have grouped the 50,000 random combinations of 30 children with ASD and 30 age-matched TD children based on how many of the 18 pathways reconstructed by TRACULA were found to have significant FA differences between the ASD and TD group at the $p < 0.05$ level. The plots show the average difference in motion measures between the groups (ASD-TD), plotted against the number of pathways that exhibited significant FA group differences. As seen in the figure, the trials with significantly different FA between the ASD and TD groups in a greater number of pathways were also, on average, the trials with a greater difference in motion measures between the two groups. The most frequent outcome was only one pathway with a significant group difference (10,998 trials with no findings, 21,867 with one, 12,131 with two, 3719 with three, 946 with four, and 339 with five or more). The one-pathway outcome was also associated with the

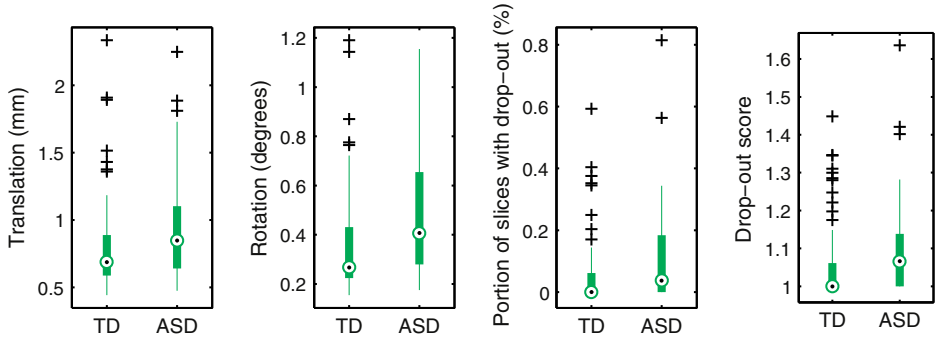

Fig. 3. Motion measures by group. The four motion measures are plotted for the 148 scans that were included in the analyses. Median motion measures (black dots) were higher for children with ASD than for TD children.

Table 1
Demographic information for the children whose 148 scans were included in the analyses. For each group the number of subjects (n_{subj}) and total number of scans (n_{scan}) are shown. For the age, IQ, SRS and ADOS scores of each group, the columns show average value (μ), standard deviation (σ), correlation with translational motion (r_T) and correlation with rotational motion (r_R). Pearson correlation coefficients are provided with the respective p-values in parentheses.

	ASD ($n_{subj} = 45$, $n_{scan} = 57$)				TD ($n_{subj} = 61$, $n_{scan} = 91$)			
	μ	σ	r_T	r_R	μ	σ	r_T	r_R
Age	8.7	1.7	−0.14 (0.31)	−0.22 (0.11)	8.4	1.9	−0.02 (0.86)	−0.04 (0.71)
IQ	108.0	16.9	−0.15 (0.35)	−0.12 (0.43)	115.2	14.0	0.01 (0.92)	−0.00 (0.99)
SRS	77.5	9.2	−0.08 (0.66)	−0.08 (0.63)	47.0	8.9	−0.20 (0.18)	−0.20 (0.18)
ADOS	7.0	1.9	0.24 (0.12)	0.31 (0.04)	–	–	–	–

lowest difference in motion between the ASD and TD groups, for three out of the four motion measures.

In general, as the difference in motion between the children with ASD and the TD children increased, differences in DW-MRI measures between the groups increased, and this was more pronounced for some pathways than others. To illustrate this, we show results from two sets of trials: the 500 trials with the lowest group differences in rotational motion and the 500 trials with the highest group differences in rotational motion. In the former set, the groups had average differences (ASD-TD) in translation: 0.040 ± 0.034 mm; rotation: 0.054 ± 0.012°; portion of slices with drop-out: 0.040 ± 0.018%; and drop-out score: 0.022 ± 0.017. In the latter set, the groups had average differences in translation: 0.358 ± 0.031 mm; rotation: 0.276 ± 0.011°; portion of slices with drop-out: 0.113 ± 0.015%; drop-out score: 0.075 ± 0.016. Fig. 5 shows group differences in FA, MD, RD, and AD, averaged over the 500 trials with low or high differences in motion, for each pathway and for the entire WM. Fig. 6 shows the frequency (fraction of the 500 trials) with which these differences reached statistical significance at the $p < 0.05$ level.

As seen in Fig. 6, the right ILF is the pathway that exhibits significant group differences in DW-MRI measures the most consistently, for both lower and higher group differences in motion. Other pathways show significant differences in DW-MRI measures mostly when there are higher differences in motion between groups. As seen in Fig. 5, when there is more motion in the ASD group relative to the TD group, the FA of the ASD group tends to decrease and its RD tends to increase relative to the TD group. Note that for some of the pathways the FA is somewhat higher in the ASD than the TD group (although these differences may not be significant) when the differences in motion are small. For those pathways, the FA differences decrease and then change sign as the motion differences become greater. This is, perhaps, why in Fig. 4 the trials where FA differences do not reach statistical significance in any pathways have somewhat more motion than the trials where there is a significant difference in one pathway.

Using TMI as a nuisance regressor

Fig. 7 shows the frequency of significant differences in DW-MRI measures at the $p < 0.05$ level, for the same trials as the ones shown in Fig. 6, when the TMI of the subjects is used as a nuisance regressor. The plots show that, with the introduction of the motion regressor in the analysis, the results become very similar between the trials with low group differences in motion (Fig. 7a) and the ones with high group differences in motion (Fig. 7b). A comparison of these plots to the respective plots in Fig. 6 shows that the frequency of significant findings is decreased when TMI is used as a regressor, and that this decrease is much more substantial for the trials with high group differences in motion than for the trials with low group differences in motion.

TD vs. TD

Figs. 8, 9, and 10 show results from comparisons of random combinations of TD children, i.e., groups of 30 TD children with less head motion vs. 30 age-matched TD children with more head motion. As in the previous section, sets of 500 trials with lower differences in head motion between the two groups and 500 trials with higher

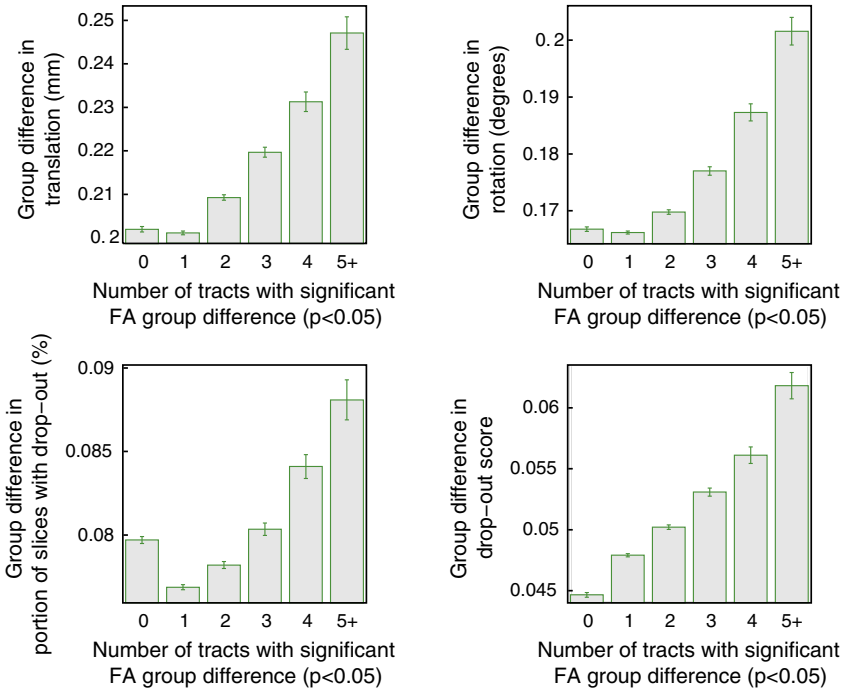

Fig. 4. Subjects with autism vs. control subjects. The difference in motion measures between groups of 30 children with ASD and 30 age-matched TD children, plotted against the number of pathways that exhibited significant FA differences between the ASD and TD group at the $p < 0.05$ level. Averages and standard error bars are shown for each of the four motion measures, over a total of 50,000 randomly drawn subject combinations. On average, the greater the difference in motion between the ASD and TD group, the greater the number of pathways with significant FA differences.

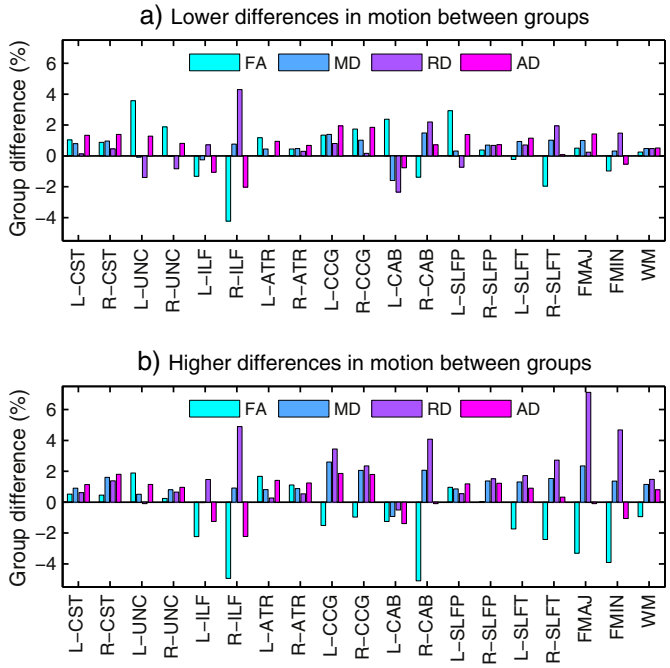

Fig. 5. Subjects with autism vs. control subjects. Group differences in FA, MD, RD, and AD, for each of the 18 pathways reconstructed by TRACULA and for the entire WM, averaged over 500 trials with low differences in motion (a) and 500 trials with high differences in motion between groups (b). Differences in DW-MRI measures are expressed as $100 \cdot (x_{ASD} - x_{TD})/x_{TD}$, where x_{ASD} and x_{TD} are the measures for the ASD and TD group, respectively. There were greater group differences in DW-MRI measures for some pathways when the group differences in motion were higher.

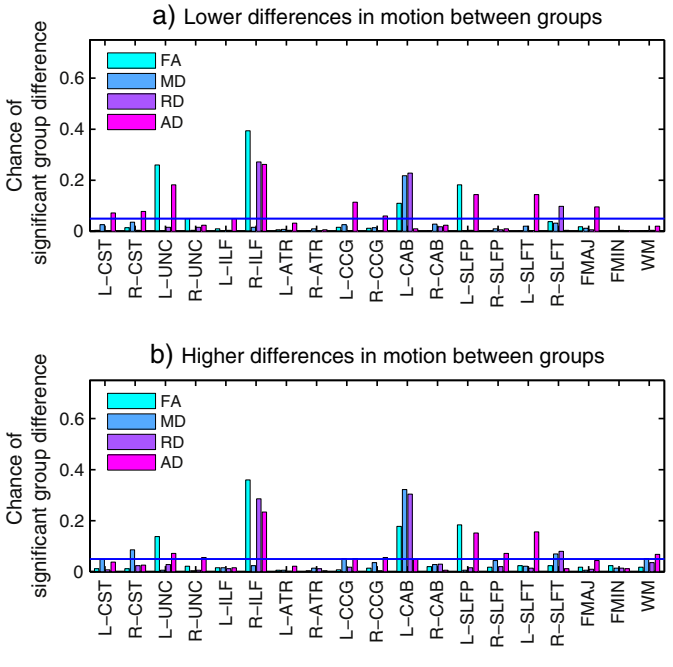

Fig. 7. Subjects with autism vs. control subjects, regressing motion. Frequency of significant group differences in FA, MD, RD, and AD at the $p < 0.05$ level, for each of the 18 pathways reconstructed by TRACULA and for the entire WM. Results are shown for 500 trials with low differences in motion (a) and 500 trials with high differences in motion between groups (b). The horizontal blue line indicates the type-I error rate of 0.05. Introducing the motion regressor led to similar results between trials with low and high group differences in motion.

differences in head motion between the two groups were identified. In the former set, the groups had average differences in translation: 0.041 ± 0.031 mm; rotation: 0.0003 ± 0.0002°; portion of slices with

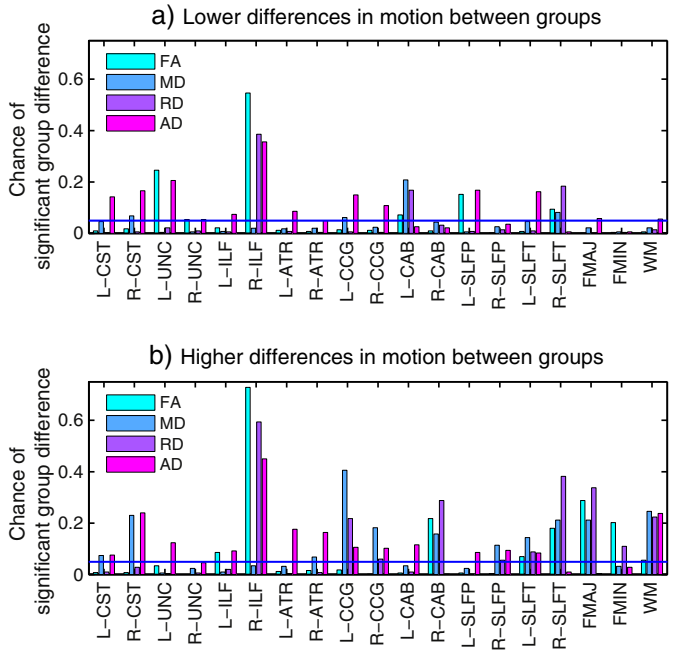

Fig. 6. Subjects with autism vs. control subjects. Frequency of significant group differences in FA, MD, RD, and AD at the $p < 0.05$ level, for each of the 18 pathways reconstructed by TRACULA and for the entire WM. Results are shown for 500 trials with low differences in motion (a) and 500 trials with high differences in motion between groups (b). The horizontal blue line indicates the type-I error rate of 0.05. Some pathways showed significant group differences in DW-MRI measures more frequently when the group differences in motion were higher.

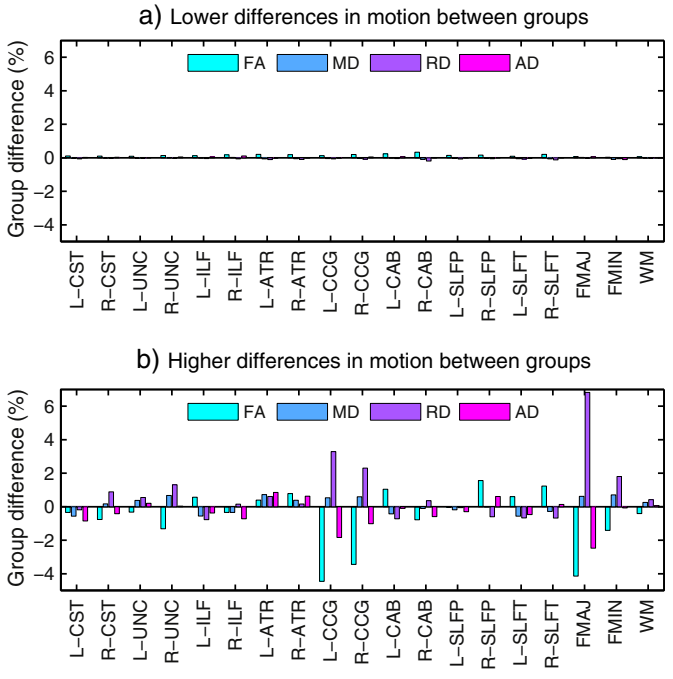

Fig. 8. Control subjects only. Group differences in FA, MD, RD, and AD, for each of the 18 pathways reconstructed by TRACULA and for the entire WM, averaged over 500 trials with low differences in motion (a) and 500 trials with high differences in motion between groups (b). Differences in DW-MRI measures are expressed as $100 \cdot (x_{TD2} - x_{TD1})/x_{TD1}$, where x_{TD1} and x_{TD2} are the measures for the TD group with less and more motion, respectively. There were group differences in DW-MRI measures for some pathways, only when the group differences in motion were higher.

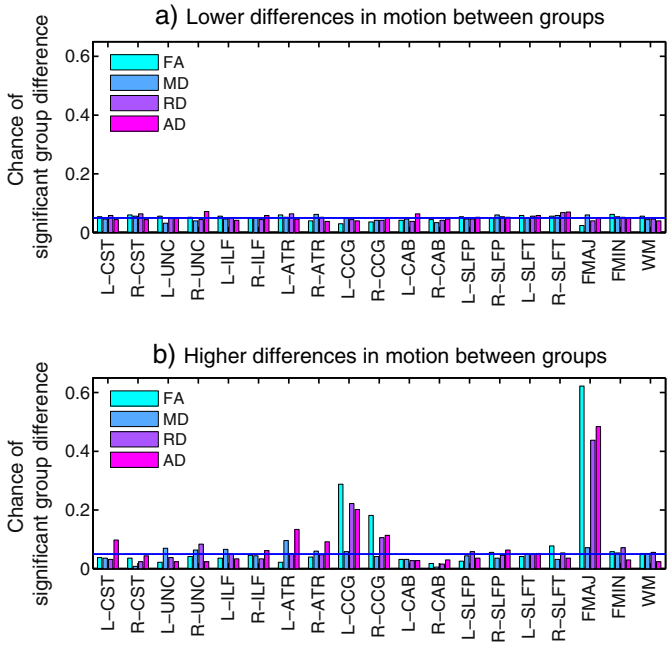

Fig. 9. Control subjects only. False positive rates for FA, MD, RD, and AD at the $p < 0.05$ level, for each of the 18 pathways reconstructed by TRACULA and for the entire WM. Results are shown for 500 trials with low differences in motion (a) and 500 trials with high differences in motion between groups (b). The horizontal blue line indicates the type-I error rate of 0.05. False positive rates increased for some pathways when the group differences in motion increased.

drop-out: 0.013 ± 0.010%; and drop-out score: 0.016 ± 0.012. In the latter set, the groups had average differences in translation: 0.276 ± 0.047 mm; rotation: 0.204 ± 0.011°; portion of slices with drop-out: 0.071 ± 0.013%; and drop-out score: 0.065 ± 0.016.

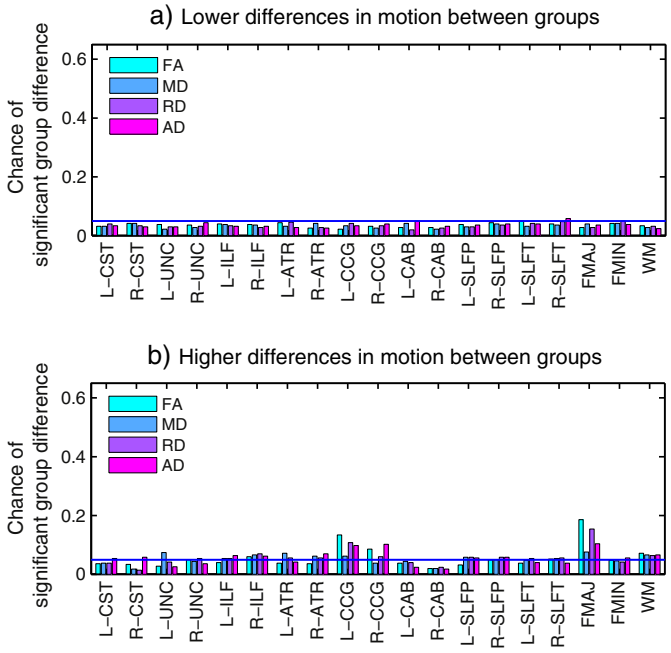

Fig. 10. Control subjects only, regressing motion. False positive rates for FA, MD, RD, and AD at the $p < 0.05$ level, for each of the 18 pathways reconstructed by TRACULA and for the entire WM. Results are shown for 500 trials with low differences in motion (a) and 500 trials with high differences in motion between groups (b). The horizontal blue line indicates the type-I error rate of 0.05. Introducing the motion regressor led to similar false positive rates between trials with low and high group differences in motion.

Note that these differences between lower-motion and higher motion scans from TD children are more subtle than the respective differences between scans from children with ASD and TD children, which were reported in the previous section.

Fig. 8 shows group differences in FA, MD, RD, and AD, averaged over the 500 trials with low or high differences in motion, for each pathway and for the entire WM. Fig. 9 shows the frequency (fraction of the 500 trials) with which these differences reached statistical significance at the $p < 0.05$ level. This frequency is now a false positive rate, as no differences in DW-MRI measures are expected between random combinations of TD children. As seen in the plots, when the motion differences between the groups of TD children were low, there were no group differences in DW-MRI measures (Fig. 8a) and the false positive rate was around 5% for all pathways (Fig. 9a). This was consistent with the chosen threshold ($p < 0.05$) for the probability of detecting a difference between groups under the null hypothesis. When the motion difference between groups was higher, the false positive rates increased (Fig. 9b), particularly for the forceps major of the corpus callosum and the cingulum bundle. For these pathways, there was an increase in RD and a (smaller) decrease in AD for the group with more motion relative to the group with less motion (Fig. 8b). As a result, there was a decrease in FA but a much smaller increase in MD in the presence of motion.

Using TMI as a nuisance regressor

Fig. 10 shows the frequency of significant differences in DW-MRI measures at the $p < 0.05$ level, for the same trials as the ones shown in Fig. 9, when the TMI of the subjects is used as a nuisance regressor. The introduction of the motion regressor in the analysis reduced the false positives substantially in the trials with high group differences in motion (Fig. 10b), although it did not bring them quite to the same level as the trials with low group differences in motion (Fig. 10a). The false positive rate was decreased, when motion was used as a regressor, even for the trials with low group differences in motion, where it fell slightly below 5%. This may be an indication that the motion regressor introduces some noise in the analysis, as the motion parameters are themselves noisy estimates derived from image data. Thus it is plausible that the introduction of motion parameters in the analysis reduces the bias due to motion at the cost of a small increase in variance due to noise.

Test vs. retest

Here we used data from 25 TD children only. These children had test–retest scans that did not exceed a stringent motion threshold, so that the maximum translational and rotational motion in any of the scans included in this analysis was, respectively, 1.17 mm and 0.58°. The median number of days between test–retest scans was 22 and the interquartile range was 37. Paired *T*-tests between the earlier and later scans showed no significant change in FA; thus there were no significant longitudinal within-subject changes between these scans. The higher- and lower-motion scans of these 25 children had average differences in translation: 0.157 ± 0.032 mm; rotation: 0.126 ± 0.020°; portion of slices with drop-out: 0.025 ± 0.014%; and drop-out score: 0.037 ± 0.017.

Fig. 11 shows comparisons of FA, MD, RD, and AD, for each of the 18 pathways reconstructed by TRACULA and for the entire WM, between the subjects' lower-motion and higher-motion scans. The corpus callosum and cingulum bundle showed significant differences in FA between the lower-motion and higher-motion scans of these 25 TD children. The differences followed a similar pattern as the one seen in the previous section: higher motion led to decreased FA but largely unchanged MD, due to an increase in RD and decrease in AD. The sizes of these differences are shown in Table 2.

We also investigated whether the differences in DW-MRI measures could be caused by the inclusion of more gray matter voxels in the posterior probability distributions of the pathways in the presence of more motion. Due to the fact that the distributions were thresholded at 20% of their maximum value, and the voxels that remained were weighted by

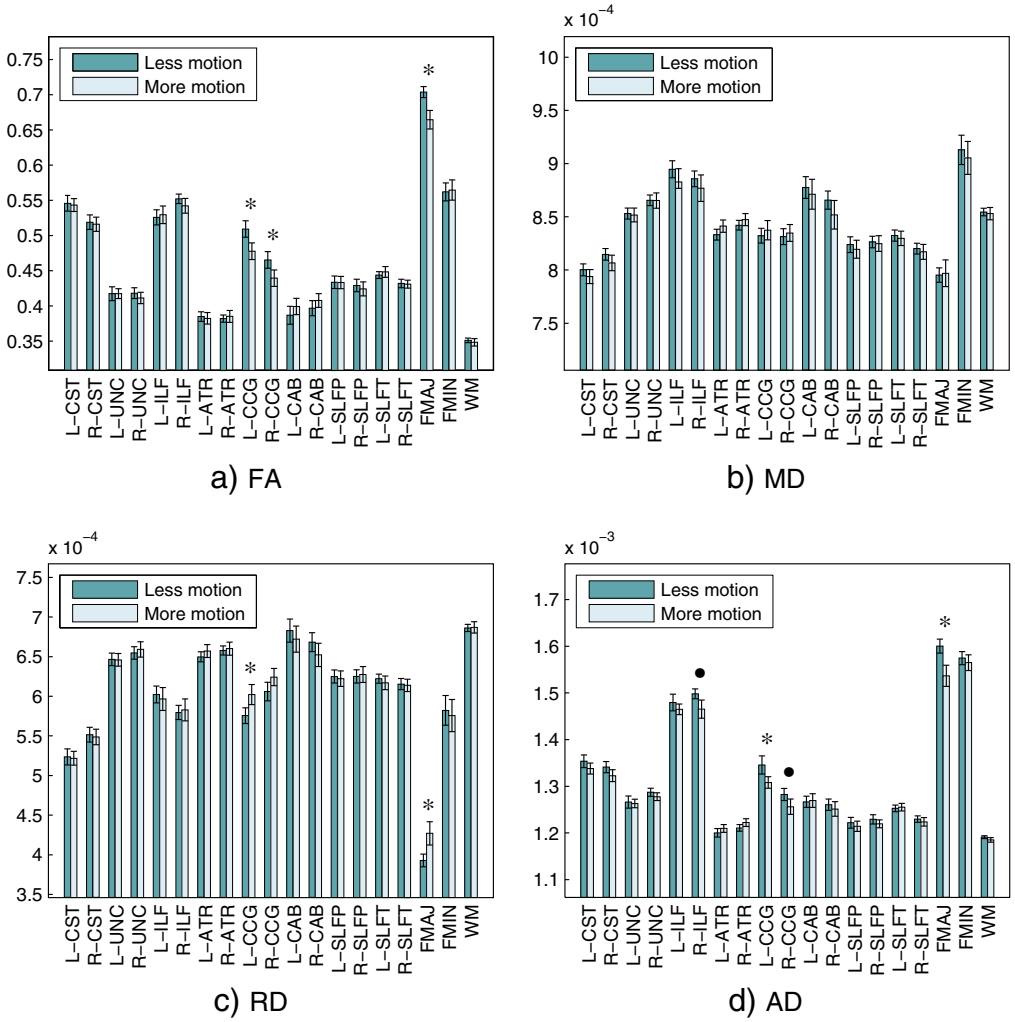

Fig. 11. Subjects with test–retest scans. Mean FA, MD, RD, and AD for each of the 18 pathways reconstructed by TRACULA and for the entire WM are plotted for the lower- and higher-motion scans of the same 25 children. Group averages and standard error bars are shown. An asterisk indicates a significant difference in FA between groups ($p < 0.05$) and a disk indicates a trend towards significance ($p < 0.1$) based on a paired T-test. Significant group differences were found between test and retest scans *of the same children*, particularly in the corpus callosum and cingulum bundle.

their probabilities, gray matter voxels contributed very little to the computation of the mean values of DW-MRI measures over each pathway. For example, for the forceps major of the corpus callosum, which showed the greatest sensitivity to motion, the sum of the probabilities of all included gray matter voxels was on average 0.01 ± 0.003 for the lower-motion scans, 0.02 ± 0.005 for the higher-motion scans, and their differences did not reach statistical significance based on a paired T-test ($p = 0.062$).

Table 2
Percent differences in FA, MD, RD, and AD between the test–retest scans of 25 TD children. Differences are expressed as $100 \cdot (x_{TD2} - x_{TD1})/x_{TD1}$, where x_{TD1} and x_{TD2} are the measures for the scans with less and more motion, respectively. Results are shown for the three pathways that showed statistically significant changes in FA (FMAJ, L-CCG, R-CCG). The columns show average (μ), standard error (ϵ), and p-value from a paired T-test on the difference between the lower-motion and higher-motion scans.

	FMAJ		L-CCG		R-CCG	
	$\mu \pm \epsilon$	p	$\mu \pm \epsilon$	p	$\mu \pm \epsilon$	p
FA	−4.9 ± 0.4	0.01	−5.5 ± 0.5	0.02	−5.1 ± 0.5	0.03
MD	0.1 ± 0.3	0.99	0.9 ± 0.2	0.38	1.1 ± 0.2	0.41
RD	8.1 ± 0.7	0.03	5.3 ± 0.4	0.03	4.6 ± 0.5	0.09
AD	−3.6 ± 0.7	0.03	−2.4 ± 0.2	0.03	−1.7 ± 0.2	0.10

Voxel-based analysis

Statistical significance maps from the voxel-wise statistical analysis of the association of FA and rotational motion in the full set of scans are shown in the supplementary figure. Increased motion is associated with decreased FA, and the corpus callosum shows the strongest association. This result replicates, in our cohort of children, what was shown for a group of middle-aged and older adults by Salat (in press), (Fig. 10). Using translational instead of rotational motion did not alter this result.

Discussion

Anisotropy and diffusivity measures derived from DW-MRI are sensitive to several confounding factors, including head motion, partial volume, and fiber crossing effects (Jones and Cercignani, 2010; Metzler-Baddeley et al., 2012; Jones et al., 2013). Although we have focused exclusively on head motion in this work, care must be taken by researchers to ensure that group differences in their studies are not caused by any of the above factors.

Effects of head motion

Our results illustrate that group differences in motion can have a non-negligible effect on group differences in DW-MRI measures. In

our comparison of subjects with autism to control subjects, the former always exhibited greater motion than the latter. Small increases in this difference in motion between the two groups were accompanied by increases in the number of pathways that had significantly different FA between the groups (Fig. 4). A comparison of results from trials with smaller motion differences between groups (Figs. 5a, 6a) to those from trials with larger motion differences between groups (Figs. 5b, 6b) showed that the latter were characterized by greater group differences in DW-MRI measures. In particular, more motion in the ASD group relative to the TD group was accompanied by an apparent increase in RD and decrease in FA in the ASD group relative to the TD group. This pattern of group differences in anisotropy and diffusivity has often been reported in the autism literature (Travers et al., 2012).

However, in the absence of ground truth on the structure of the autistic brain, it is difficult to determine if these findings are false positives. As we found head motion to be correlated with ADOS symptom severity scores in the children with ASD, in principle one could not eliminate the possibility of a true biological underpinning for the greater differences in DW-MRI measures for ASD groups with more head motion. Furthermore, group comparisons of measures derived from the tensor model of diffusion can be confounded by other factors. For example, subjects from different populations could differ not only in terms of head motion, but also in terms of partial-volume or fiber-crossing effects. To make sure that we can isolate the effects of head motion from other possible group differences, we went on to compare groups consisting of control subjects only.

When we compared groups of randomly drawn, age-matched TD children with similar levels of head motion, differences in DW-MRI measures of anisotropy and diffusivity were close to zero and the false positive rate was uniformly at 5% for all 18 pathways (Figs. 8a, 9a). However, the magnitude of the differences increased substantially when one group of TD children exhibited more head motion than the other (Figs. 8b, 9b). The motion differences between these groups were even subtler than the differences between ASD and TD groups in our previous set of experiments. We were also able to detect within-subject differences between the lower-motion and higher-motion scans of a set of TD children who had received test–retest scans, even after removing the subjects with the most head motion (Fig. 11). The pattern that emerged from our analyses of data from TD children was that, as motion increased, there was an overestimation of RD, accompanied by an underestimation of AD. This led to a significant underestimation of FA but only a very small (non-significant) overestimation of MD.

The pathways that exhibited the most substantial motion-induced group differences in our data were the corpus callosum and the cingulum bundle. Perhaps this is related to the proximity of non-brain voxels (such as the ventricles) to a sizeable portion of those pathways. Furthermore, in our voxel-based analysis of the association of FA and motion, deeper brain areas appear to be more affected than more superficial ones (see Supplementary figure). Thus distance from the head coils may also be a factor. Interestingly, anisotropy and diffusivity measures for the corpus callosum and cingulum bundle have often been reported in the literature to differ between a variety of clinical populations and healthy subjects (Salat, in press; Stebbins and Murphy, 2009; Bohanna et al., 2008; Cochrane and Ebmeier, 2013; Inglese and Bester, 2010; Kubicki et al., 2007; Travers et al., 2012). This raises questions about the extent to which differences in head motion between the clinical and control population may be a confound in such studies, particularly given that reporting levels of head motion by group is not commonplace in the literature.

Strategies for group comparisons

There are several approaches to comparing DW-MRI measures between groups. In this work, we have focused on mean values of such measures over major WM pathways, as obtained from tractography. Motion will affect these mean values if it affects a large enough portion of the pathway. As can be seen in the statistical maps produced by a voxel-wise analysis (Supplementary figure), the effects of motion may be more significant in certain parts of a given pathway than others. Thus, one would expect that if tractography were used to compare DW-MRI measures along the trajectory of a pathway, the effects of motion might not be homogeneous over the entire trajectory. In general, we would not expect results from voxel-wise statistical analysis, which is performed in a common template space and thus relies on good spatial alignment of subjects in that space, to agree perfectly with tract-wise statistical analysis, where measures are extracted in the native space of each subject.

An alternative approach to defining WM pathways, instead of tractography, is to use regions of interest from an atlas (e.g., Faria et al. (2011); Keihaninejad et al. (2013)). This would rely on accurate registration of the individual to the atlas but it would not rely on the DW-MRI data beyond that. Our approach is an intermediate one, where prior information from an atlas is used to constrain the tractography solutions in areas where there is uncertainty in the DW-MRI data. Finally, if a study were to examine more than one of the pathways included here, the *p*-values would typically be corrected for multiple comparisons. If we applied Bonferroni correction to the results of Fig. 9, then both the blue line indicating the threshold and the bars indicating the frequency of significant findings would be lower than they are now. However, for the trials with greater motion differences between groups, the frequency of significant findings would still be above the blue line, and thus greater than what would be expected by random chance alone, for the same pathways.

Our results are based on data from children aged 5–12, a population that is particularly challenging to scan regardless of diagnosis. However, our voxel-based analysis (Supplementary figure) replicates a result on the sensitivity of the corpus callosum to motion that has been shown elsewhere for a group of middle-aged and older adults (Salat, in press, Fig. 10). Importantly, even if the average adult moved less that the subjects in our study, studies of adult subjects could still be confounded by different amounts of head motion between groups, such as younger vs. older or clinical vs. control populations.

A potential concern when analyzing image data from children is that analysis methods that have been developed for adult subjects may not be appropriate. Our tractography method relies on an automated surface reconstruction and segmentation of our subjects' T_1-weighted images. The validity of this particular anatomical data processing stream and its lack of age-related bias have been shown previously for children of the same age range (Ghosh et al., 2010). Our tractography algorithm does not rely on exact spatial alignment of our subjects to an atlas, as it uses information only about which anatomical labels each pathway passes through or next to, and not about the exact spatial coordinates or shape of the pathways (Yendiki et al., 2011).

Strategies for motion compensation

Having illustrated the impact that head motion can have on DW-MRI group studies, our results underline the importance of accounting for motion in any such study where one population is more prone to motion in the scanner than the other. In general, methods for motion compensation in imaging data are either retrospective, i.e., they are performed as a post-processing step after the images are acquired, or prospective, i.e., they are built into the image acquisition. In the following we discuss these potential remedies briefly.

The most popular retrospective approach to motion correction in DW-MRI relies on image registration to align the DW images to a baseline image (Andersson and Skare, 2002; Rohde et al., 2004). However, the effects of head motion on DW-MRI are two-fold; it can cause misalignment between different volumes in the DW-MRI series but it can also alter the intensity values of a specific volume (or, more appropriately for a 2D acquisition, a specific slice), if head motion occurs during the diffusion-encoding gradient pulse. Although registration-based

correction will address the misalignment between volumes, the drop-out in intensity values will persist. All the results shown in the present study were obtained after performing registration-based correction.

A common strategy for controlling confounding factors in neuroimaging group studies is to introduce these factors as linear regressors in statistical analyses. In this work we used TMI, a composite index defined from four measures of motion, as a nuisance regressor in the statistical analysis of the mean values of FA, MD, RD, and AD in each pathway. The use of this regressor allowed us to reduce the frequency of significant findings among trials with high motion difference between groups, making it very similar to the respective frequency among trials with low motion difference between groups. This was the case in comparisons between children with ASD and TD children (Fig. 7), as well as comparisons between TD children with lower and higher motion (Fig. 10). We found TMI to be more effective in this regard than the individual motion measures (results not shown). Of course, any such linear regression approach is ad hoc, as anisotropy and diffusivity measures are not linear with respect to global motion parameters. Thus regressors cannot be expected to eliminate false positives completely. A true model-based approach would involve knowledge of the trajectory of head motion during the acquisition of each slice. Furthermore, measures of motion may be linearly dependent with respect to other regressors that are often included in DW-MRI group analyses, most notably age.

Another retrospective approach to addressing the effects of motion is outlier rejection. This can range from a procedure as simple as discarding an entire scan that exceeds acceptable levels of motion, either by visual inspection or based on a hard threshold on motion parameters, to statistical methods for detecting and discarding subsets of the data in each scan as outliers (Chang et al., 2005; Zwiers, 2010). When discarding subsets of the data in a scan, i.e., volumes or slices, the amount of redundancy in each data set is an important consideration. It has been shown that the minimum number of distinct gradient directions that are necessary for robust estimation of FA values is 30 (Jones, 2004). Thus it is desirable to have more than 30 gradient directions per data set, so that outlier volumes can be discarded and scans with intermittent head motion can still be used without introducing bias in the analyses. However, the data in the present study included only 30 gradient directions. In the absence of redundancy, we chose to discard egregiously poor scans in their entirety but maintain the full range of data quality in the scans that were included in our analyses.

Specifically, we inspected all scans and discarded only the ones where head motion led to visibly poor quality of the FA maps. This was the case for 17 scans, or roughly 10% of the full set of data. As seen in Fig. 2, the scans that we discarded based on visual inspection included most if not all of the scans that would be deemed outliers based on one of the four measures of motion. (Note that, if a scan had visible signal drop-out in a slice at the very top or base of the brain that would not affect the major WM pathways, we did not discard it. Hence a few of the scans included in the analyses may score high in terms of signal drop-out but none scores high in terms of the percentage of affected slices.) Some of the 17 discarded scans were outliers based on only one of the four motion measures, some were outliers based on more than one measure, and some were not outliers based on any of these measures. This illustrates the difficulty in setting a hard threshold on motion measures for excluding scans from a study and the importance of taking into account all such measures, as well as inspecting the data visually. Hard thresholds that have been used in other autism studies are 2 mm of translational motion (Knaus et al., 2010; Shukla et al., 2011) and 2° of rotational motion (Knaus et al., 2010). As seen in Fig. 3, none of the data sets included in our analyses exceeded the 2° threshold. Only two data sets exceeded the 2 mm threshold, and excluding them from the analyses did not change our results.

The approach that we followed here, applying one of the most widely available methods for registration-based correction (Jenkinson et al., 2002) and discarding poor-quality scans based on visual inspection, is certainly not the only possible route but it is one that we believe reflects current common practices. No matter which combination of registration-based and outlier-based corrections is used, however, a concern with all such retrospective methods is that data with different levels of motion will also be subjected to different levels of processing. In a group study where subjects from one group tend to move more, scans from this group will have more volumes smoothed due to the interpolation performed by a registration-based motion correction method, and more data points removed by a method for rejecting outlier slices or volumes.

A potential remedy for the residual effects of motion that cannot be removed by retrospective correction methods, as well as the effects of treating one group more than the other with such methods, is to make sure that the scans in the two groups are matched with respect to some summary measures of motion. Our results indicate that, when the groups do not differ in terms of the motion measures that we considered in this work, the chance of false positive findings is reduced substantially. However, even if matching groups for motion measures can decrease false positives, it will not address the potential for false negatives due to motion. That is, the contamination of the image data with motion, even if that motion is comparable between groups, may occlude true but subtle differences in WM microstructure. For accurate measures of this microstructure, one needs not only equal motion artifacts between groups, but ideally no motion artifacts at all.

An additional drawback of motion correction methods that rely on registration of DW images is that their performance is dependent on the b-value. Images acquired with very high b-values do not contain enough anatomical features to be registered accurately. For such data, it is not possible to use registration-based approaches, either to correct translational and rotational motion, or to quantify it and match it across populations. This makes motion correction particularly problematic for high-angular resolution DW-MRI scans that require the acquisition of data with higher b-values than routine scans. Such high-angular resolution acquisitions are at once more sensitive to head motion due to higher diffusion contrast at high b-values and more likely to include head motion due to the longer scan time needed to acquire more diffusion-encoding gradient directions.

Prospective methods for motion correction have been proposed to overcome some of the limitations of retrospective methods. Several motion-compensated sequences for DW-MRI have been introduced recently, using volume registration (Benner et al., 2011; Sarlls et al., 2012), external optical tracking systems (Aksoy et al., 2011), free-induction decay navigators (Kober et al., 2012), or volumetric navigators (Alhamud et al., 2012; Bhat et al., 2012). In Aksoy et al. (2011) and Alhamud et al. (2012) the authors compare their methods to standard retrospective methods for motion correction and show that the prospective approach leads to improved performance. A particular benefit of prospective motion correction is that it avoids the interpolation of image intensities performed by registration-based retrospective methods. In addition to motion-compensated sequences, a promising development for DW-MRI is accelerated acquisition methods (Feinberg and Setsompop, 2013), as reducing the duration of a scan will also make it less susceptible to subject motion.

Further analyses

The results presented here show the effects of motion on group comparisons of anisotropy and diffusivity. However, there are other types of analyses that we have not investigated in this work but that may be confounded by head motion:

- *Differences in lateralization of diffusion measures between groups.* It is not entirely clear how differences in head motion might affect differences in lateralization. If subjects in one group moved more, thus reducing the apparent anisotropy in both hemispheres, power to detect differences between hemispheres in that group could be reduced. Whether

these differences would still be detectable would also depend on the level of noise in the data.

- *Correlation of diffusion measures with other study parameters.* The amount of head motion that subjects exhibit in the scanner could be correlated with several demographic and behavioral attributes of the subjects, including age, IQ, cognitive performance metrics, or symptom severity for various conditions. If such a correlation existed, motion could induce spurious associations between demographic or behavioral parameters and the measures of anisotropy and diffusivity extracted from DW-MRI data.

Implications for autism studies and beyond

In the autism literature, which has been the motivation for this work, findings of all the types listed above have been reported (Travers et al., 2012). Among the 48 studies reviewed in Travers et al. (2012), the following strategies for mitigating motion artifacts are mentioned: retrospective registration-based correction in 30 studies; ad hoc outlier rejection (discarding entire scans either by visual inspection or by a hard threshold on a motion parameter) in 13 studies; and selective sedation (only for the most challenging subjects) in 10 studies. However, despite the limitations of the above approaches, very few of the 48 studies report measures of head motion for subjects in each group.

Specifically, Thakkar et al. report that there was no significant difference in translational motion between subjects with ASD and control subjects in their study (although those motion measures are estimated from a functional MRI scan). They find significant FA differences between groups in the subcortical WM underlying a number of cortical regions of interest (Thakkar et al., 2008). Knaus et al. report that there were no significant group differences in motion parameters in the DW-MRI scans of their subjects, based on a multivariate analysis of variance. They find no significant group differences in the FA of the arcuate fasciculus (Knaus et al., 2010). Shukla et al. report that there were no significant group differences in translational or rotational motion in their DW-MRI data. They find significant group differences in FA in a region-of-interest analysis (Shukla et al., 2010), as well as a voxel-based analysis (Shukla et al., 2011). Groen follows a statistical approach to outlier removal as part of the tensor estimation step. The author uses the mean diffusivity values in cerebrospinal fluid to quantify motion artifacts and reports that there was no significant difference between groups in that respect. Groen finds no significant group differences in FA after correcting for age and IQ, but does find significant differences in mean diffusivity. In addition, that work reports a significantly higher number of voxels classified as outliers in subjects with ASD than control subjects (Groen, 2011). Weinstein et al. use sedation for all of the subjects in their study. They find significant FA differences between groups in a voxel-based analysis (Weinstein et al., 2010). In a more recent study, Walker et al. compare sedated children with ASD to sleeping TD children. They find significant but small differences in FA, on the order of 1%–2%, in voxel-based analyses. Importantly, they also study the spatial distribution of artifacts, as identified by their method for detecting outlier voxels, and they report differences between the two groups. The authors conclude that these differences may be partly due to more head motion in the unsedated TD subjects than the sedated subjects with ASD, and that this could have affected certain aspects of their DW-MRI findings (Walker et al., 2012).

Taken together, the studies discussed above suggest that differences in measures derived from DW-MRI between subjects with autism and control subjects could not be explained entirely by head motion, and that true effects are likely to exist. However, as we have shown here, it is worth revisiting studies that have not reported measures of motion and examining whether group differences in motion exist and, if so, whether controlling for this and using analysis methods that are robust to outliers increase the specificity of the findings. In our own data, we find significant but not wide-spread differences between children with ASD and TD children, after matching the groups for motion (Koldewyn et al., in review). Specifically, our follow-up analyses confirm that differences between the two groups in the right ILF persist even after head motion is accounted for.

The small portion of studies that report measures of motion for each group is not characteristic only of the autism literature. It is representative of the wide range of applications that use DW-MRI to infer group differences in WM microstructure. Our results show the impact that differences in head motion between groups can have on DW-MRI group studies and thus underline the importance of reporting measures of head motion. If differences in the amount of motion are found between the groups, a combination of the retrospective correction methods described above, i.e., between-volume registration, outlier rejection, and using motion measures as nuisance regressors, will mitigate the effects of motion. However, it is important to note the limitations of these approaches and not to assume that they will eliminate all such effects completely. Ultimately, our results demonstrate the significance of developing motion-compensated acquisition methods for DW-MRI and incorporating them into the common practice of neuroimaging studies.

Conclusions

We found that small differences in the amount of head motion between two groups of subjects were sufficient to yield false positive findings of differences in anisotropy and diffusivity between the groups, and that some WM pathways were more sensitive to this than others. The popular post-processing approach to motion correction by registration of DW images to a baseline image did not eliminate the problem. The introduction of a motion index as a regressor in the analysis reduced the false positives substantially. Our results have implications for any diffusion MRI study where one population is less likely to remain still in the scanner than the other. Specifically, these results highlight the importance of (i) ensuring that there are no group differences in motion and reporting motion measures by group in any study that reports group differences in the DW-MRI measures that we studied here, and (ii) using motion-compensated acquisition methods for DW-MRI in future studies.

Acknowledgments

The authors would like to thank Drs. Dorit Kliemann of Freie Universität Berlin and Dylan Tisdall of Massachusetts General Hospital for helpful discussions on issues related to this paper.

Support for this research was provided in part by The Autism & Dyslexia Project funded by the Ellison Medical Foundation, the National Institute for Biomedical Imaging and Bioengineering (*Pathway to Independence* award K99/R00-EB008129, R01-EB006758), the National Center for Research Resources (P41-RR14075, U24-RR021382), the National Institute on Aging (AG022381, 5R01-AG008122-22), the National Center for Alternative Medicine (RC1-AT005728-01), and the National Institute for Neurological Disorders and Stroke (R01-NS052585-01, 1R21-NS072652-01, 1R01-NS070963), and was made possible by the resources provided by Shared Instrumentation Grants 1S10RR023401, 1S10RR019307, and 1S10RR023043. Additional support was provided by the National Institutes of Health Blueprint for Neuroscience Research (5U01-MH093765), part of the multi-institutional Human Connectome Project.

Conflict of interest

B.F. has a financial interest in CorticoMetrics, a company whose medical pursuits focus on brain imaging and measurement technologies. B.F.'s interests were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2013.11.027.

References

Aksoy, M.M., Forman, C.C., Straka, M.M., Skare, S.S., Holdsworth, S.S., Hornegger, J.J., Bammer, R.R., 2011. Real-time optical motion correction for diffusion tensor imaging. Magn. Reson. Med. 66 (2), 366–378 (Aug.).

Alhamud, A., Tisdall, M.D., Hess, A.T., Hasan, K.M., Meintjes, E.M., van der Kouwe, A.J.W., 2012. Volumetric navigators for real-time motion correction in diffusion tensor imaging. Magn. Reson. Med. 68 (4), 1097–1108 (Oct.).

Anderson, A.W., Gore, J.C., 1994. Analysis and correction of motion artifacts in diffusion weighted imaging. Magn. Reson. Med. 32 (3), 379–387 (Sep.).

Andersson, J.L.R., Skare, S., 2002. A model-based method for retrospective correction of geometric distortions in diffusion-weighted EPI. NeuroImage 16 (1), 177–199 (May).

Behrens, T.E.J., Berg, H.J., Jbabdi, S., Rushworth, M.F.S., Woolrich, M.W., 2007. Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? NeuroImage 34 (1), 144–155 (Jan).

Benner, T., van der Kouwe, A.J.W., Sorensen, A.G., 2011. Diffusion imaging with prospective motion correction and reacquisition. Magn. Reson. Med. 66 (1), 154–167 (Jul.).

Bhat, H., Tisdall, M.D., van der Kouwe, A.J.W., Feiweier, T., Heberlein, K., 2012. Epi navigator based prospective motion correction technique for diffusion neuroimaging. Proc. Int. Soc. Magn. Reson. Med. 20, 113.

Bohanna, I., Georgiou-Karistianis, N., Hannan, A., Egan, G., 2008. Magnetic resonance imaging as an approach towards identifying neuropathological biomarkers for Huntington's disease. Brain Res. Rev. 58 (1), 209.

Chang, L.-C., Jones, D.K., Pierpaoli, C., 2005. RESTORE: robust estimation of tensors by outlier rejection. Magn. Reson. Med. 53 (5), 1088–1095.

Cochrane, C.J.C., Ebmeier, K.P.K., 2013. Diffusion tensor imaging in parkinsonian syndromes: a systematic review and meta-analysis. Neurology 80 (9), 857–864 (Feb.).

Constantino, J., Gruber, C., (Firm), W.P.S., 2007. Social Responsiveness Scale (SRS). Western Psychological Services.

Faria, A.V., Hoon, A., Stashinko, E., Li, X., Jiang, H., Mashayekh, A., Akhter, K., Hsu, J., Oishi, K., Zhang, J., Miller, M.I., van Zijl, P.C.M., Mori, S., 2011. Quantitative analysis of brain pathology based on MRI and brain atlases—applications for cerebral palsy. NeuroImage 54 (3), 1854–1861 (Feb.).

Feinberg, D.A., Setsompop, K., 2013. Ultra-fast MRI of the human brain with simultaneous multi-slice imaging. J. Magn. Reson. 229, 90–100 (San Diego, Calif.: 1997, Apr.).

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33 (3), 341–355 (Jan).

Fischl, B., Salat, D.H., van der Kouwe, A.J.W., Makris, N., Ségonne, F., Quinn, B.T., Dale, A.M., 2004a. Sequence-independent segmentation of magnetic resonance images. NeuroImage 23 (Suppl. 1), S69–S84 (Jan).

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A.M., 2004b. Automatically parcellating the human cerebral cortex. Cereb. Cortex 14 (1), 11–22 (Jan).

Ghosh, S.S., Kakunoori, S., Augustinack, J., Nieto-Castanon, A., Kovelman, I., Gaab, N., Christodoulou, J.A., Triantafyllou, C., Gabrieli, J.D.E., Fischl, B., 2010. Evaluating the validity of volume-based and surface-based brain image registration for developmental cognitive neuroscience studies in children 4 to 11 years of age. NeuroImage 53 (1), 85–93 (Oct.).

Groen, W., 2011. Pervasive microstructural abnormalities in autism: a DTI study. J. Psychiatry Neurosci. 36 (1), 32–40 (Jan.).

Inglese, M., Bester, M., 2010. Diffusion imaging in multiple sclerosis: research and clinical implications. NMR Biomed. 23 (7), 865–872 (Aug.).

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage 17 (2), 825–841 (Oct).

Jones, D.K., 2004. The effect of gradient sampling schemes on measures derived from diffusion tensor MRI: a Monte Carlo study. Magn. Reson. Med. 51 (4), 807–815 (Apr).

Jones, D.K., Cercignani, M., 2010. Twenty-five pitfalls in the analysis of diffusion MRI data. NMR Biomed. 23 (7), 803–820 (Aug).

Jones, D.K., Knösche, T.R., Turner, R., 2013. White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion MRI. NeuroImage 73, 239–254 (Jun.).

Kaufman, A.S., Kaufman, N.L., 2004. Kaufman Brief Intelligence Test, 2nd ed. NCS, Pearson, Inc., Minneapolis, MN.

Keihaninejad, S., Zhang, H., Ryan, N.S., Malone, I.B., Modat, M., Cardoso, M.J., Cash, D.M., Fox, N.C., Ourselin, S., 2013. An unbiased longitudinal analysis framework for tracking white matter changes using diffusion tensor imaging with application to Alzheimer's disease. NeuroImage 72, 153–163 (May).

Keil, B.B., Alagappan, V.V., Mareyam, A.A., McNab, J.A.J., Fujimoto, K.K., Tountcheva, V.V., Triantafyllou, C.C., Dilks, D.D.D., Kanwisher, N.N., Lin, W.W., Grant, P.E.P., Wald, L.L.L., 2011. Size-optimized 32-channel brain arrays for 3 T pediatric imaging. Magn. Reson. Med. 66 (6), 1777–1787 (Dec.).

Knaus, T.A.T., Silver, A.M.A., Kennedy, M.M., Lindgren, K.A.K., Dominick, K.C.K., Siegel, J.J., Tager-Flusberg, H.H., 2010. Language laterality in autism spectrum disorder and typical controls: a functional, volumetric, and diffusion tensor MRI study. Brain Lang. 112 (2), 8-8 (Feb.).

Kober, T., Gruetter, R., Krueger, G., 2012. Prospective and retrospective motion correction in diffusion magnetic resonance imaging of the human brain. NeuroImage 59 (1), 10-10 (Jan.).

Koldewyn, K., Jiang, Y.V., Weigelt, S., Kanwisher, N., 2013. Global/local processing in autism: not a disability, but a disinclination. J. Autism Dev. Disord. 43 (10), 2329–2340 (Oct.).

Koldewyn, K., Yendiki, A., Weigelt, S., Gweon, H., Julian, J., Richardson, H., Malloy, C., Saxe, R., Fischl, B., Kanwisher, N., 2013w. Reduced Integrity of the Right Inferior Longitudinal Fasciculus but no General Reduction in White Matter Integrity in ASD (in review).

Kubicki, M., McCarley, R., Westin, C.-F., Park, H.-J., Maier, S., Kikinis, R., Jolesz, F.A., Shenton, M.E., 2007. A review of diffusion tensor imaging studies in schizophrenia. J. Psychiatr. Res. 41 (1–2), 15–30 (Jan).

Leemans, A., Jones, D.K., 2009. The B-matrix must be rotated when correcting for subject motion in DTI data. Magn. Reson. Med. 61 (6), 1336–1349 (Jun).

Lord, C.C., Risi, S.S., Lambrecht, L.L., Cook, E.H.E., Leventhal, B.L.B., DiLavore, P.C.P., Pickles, A.A., Rutter, M.M., 2000. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. J. Autism Dev. Disord. 30 (3), 205–223 (Jun.).

Metzler-Baddeley, C., O'Sullivan, M.J., Bells, S., Pasternak, O., Jones, D.K., 2012. How and how not to correct for CSF-contamination in diffusion MRI. NeuroImage 59 (2), 1394–1403 (Jan.).

Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. 15 (1), 1–25 (Jan.).

Rohde, G.K.G., Barnett, A.S.A., Basser, P.J.P., Marenco, S.S., Pierpaoli, C.C., 2004. Comprehensive approach for correction of motion and distortion in diffusion-weighted MRI. Magn. Reson. Med. 51 (1), 103–114 (Jan.).

Salat, D.H., 2013. Chapter 12 — diffusion tensor imaging in the study of aging and age-associated neural disease, In: Johansen-Berg, H., Behrens, T.E. (Eds.), Diffusion MRI, 2nd ed. Academic Press, San Diego (in press).

Sarlls, Joelle E., Shaw, Philip, Adleman, N.E., Rooopchansingh, V., 2012. Straightforward method to improve sensitivity in diffusion imaging studies of subjects who move. Proc. Int. Soc. Magn. Reson. Med. 20, 3551.

Shukla, D.K., Keehn, B., Lincoln, A.J., Müller, R.-A., 2010. White matter compromise of callosal and subcortical fiber tracts in children with autism spectrum disorder: a diffusion tensor imaging study. J. Am. Acad. Child Adolesc. Psychiatry 49 (12) (1269–1278.e2, Dec.).

Shukla, D.K., Keehn, B., Müller, R.-A., 2011. Tract-specific analyses of diffusion tensor imaging show widespread white matter compromise in autism spectrum disorder. J. Child Psychol. Psychiatry 52 (3), 286–295 (Mar.).

Smith, S.M., Nichols, T.E., Jan. 2009. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. NeuroImage 44 (1), 83–98.

Smith, S.M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T.E., Mackay, C.E., Watkins, K.E., Ciccarelli, O., Cader, M.Z., Matthews, P.M., Behrens, T.E.J., 2006. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. NeuroImage 31 (4), 1487–1505 (Jul.).

Stebbins, G.T., Murphy, C.M., 2009. Diffusion tensor imaging in Alzheimer's disease and mild cognitive impairment. Behav. Neurol. 21 (1), 39–49.

Thakkar, K.N.K., Polli, F.E.F., Joseph, R.M.R., Tuch, D.S.D., Hadjikhani, N.N., Barton, J.J.S.J., Manoach, D.S.D., 2008. Response monitoring, repetitive behaviour and anterior cingulate abnormalities in autism spectrum disorders (ASD). Brain 131 (Pt 9), 2464–2478 (Sep.).

Tisdall, M.D., Hess, A.T., Reuter, M., Meintjes, E.M., Fischl, B., van der Kouwe, A.J.W., 2012. Volumetric navigators for prospective motion correction and selective reacquisition in neuroanatomical MRI. Magn. Reson. Med. 68 (2), 389–399 (Aug.).

Travers, B.G., Adluru, N., Ennis, C., Tromp, D.P.M., Destiche, D., Doran, S., Bigler, E.D., Lange, N., Lainhart, J.E., Alexander, A.L., 2012. Diffusion tensor imaging in autism spectrum disorder: a review. Autism Res. 5 (5), 289–313 (Oct.).

van der Kouwe, A.J.W., Benner, T., Salat, D.H., Fischl, B., 2008. Brain morphometry with multiecho MPRAGE. NeuroImage 40 (2), 559–569 (Apr.).

Walker, L., Gozzi, M., Lenroot, R., Thurm, A., Behseta, B., Swedo, S., Pierpaoli, C., 2012. Diffusion tensor imaging in young children with autism: biological effects and potential confounds. Biol. Psychiatry 72 (12), 1043–1051 (Dec.).

Weinstein, M., Ben-Sira, L., Levy, Y., Zachor, D.A., Itzhak, E.B., Artzi, M., Tarrasch, R., Eksteine, P.M., Hendler, T., Bashat, D.B., 2010. Abnormal white matter integrity in young children with autism. Hum. Brain Mapp. 32 (4), 534–543 (May).

Yendiki, A., Panneck, P., Srinivasan, P., Stevens, A., Zöllei, L., Augustinack, J., Wang, R., Salat, D., Ehrlich, S., Behrens, T., Jbabdi, S., Gollub, R., Fischl, B., 2011. Automated probabilistic reconstruction of white-matter pathways in health and disease using an atlas of the underlying anatomy. Front. Neuroinformatics 5, 23.

Yoshida, S., Oishi, K., Faria, A.V., Mori, S., 2013. Diffusion tensor imaging of normal brain development. Pediatr. Radiol. 43 (1), 15–27 (Jan.).

Zwiers, M.P., 2010. Patching cardiac and head motion artefacts in diffusion-weighted images. NeuroImage 53 (2), 565–575 (Nov.).