A site-wise reliability analysis of the ABCD diffusion fractional anisotropy data, impact of the scanner and analytical pipeline

Yezhi Pan[1†], L. Elliot Hong[2], Ashley Acheson[3], Paul Thompson[4], Neda Jahanshad[4], Jiaao Yu[5], Chixiang Chen[6], Tianzhou Ma[7], Anthony Liu[8], Els Fieremans[9], Peter Kochunov[2*], Shuo Chen[1,6*]

[1]Maryland Psychiatric Research Center, Department of Psychiatry, School of Medicine, University of Maryland, Baltimore, United States
[2]Department of Psychiatry and Behavioral Science, University of Texas Health Science Center Houston, United States
[3]Department of Psychiatry, University of Arkansas for Medical Sciences, United States
[4]Imaging Genetics Center, Mark & Mary Stevens Institute for Neuroimaging & Informatics, Keck School of Medicine, University of Southern California, United States
[5]Department of Mathematics, University of Maryland, College Park, United States
[6]Division of Biostatistics and Bioinformatics, Department of Epidemiology and Public Health, School of Medicine, University of Maryland, Baltimore, United States
[7]Department of Epidemiology and Biostatistics, School of Public Health, University of Maryland, College Park, United States
[8]Department of Radiology, University of Texas MD Anderson Center, United States
[9]Center for Biomedical Imaging, Department of Radiology, New York University School of Medicine, United States

Abstract

The Adolescent Brain and Cognitive Development (ABCD) project is the largest longitudinal study of brain development that tracts 11,820 subjects from 21 sites using standardized protocols for multi-site data collection and analysis. Adolescence is a critical period of brain development associated with white matter myelination and requires reliable measures to detect these changes. We assessed confounding non-biological variances in diffusion tensor imaging (DTI) data that may be present due to technological variations, participant compliance and data analysis protocols. ABCD imaging data were collected biannually, and thus ongoing maturation may artificially introduce bias to classical test-retest approaches such as the interclass correlation coefficients (ICC). To address this, we developed a site-wise adaptive ICC (AICC) to systematically evaluate the quality of imaging-derived phenotypes while accounting for ongoing brain development. We measured the age-related brain development trajectory and estimated site-wise AICC iteratively, adjusting the weight for each site based on the ICC scores. We evaluated longitudinal reliability of diffusion fractional anisotropy (FA) data for each site, considering the impact of MRI scanner platform and standard ABCD versus ENIGMA-DTI data extraction pipelines, and comparing longitudinal stability of FA measurements to these of the cortical thickness. The ENIGMA structural and diffusion pipeline with QA/QC improved the average reliability for cortical FA to AICC=0.70±0.19, compared to 0.61±0.19 for the standard ABCD pipeline (Wilcoxon test $p<0.001$). Furthermore, we showed that the AICC for sites that used Siemens scanners significantly outperformed those using GE/Phillips scanners (AICC=0.78±0.11 vs 0.62±0.21, $p<0.001$). In conclusion, variations in data quality among study

sites and preprocessing pipelines underscore the necessity for meticulous data curation in subsequent association analyses.

1.Introduction

Adolescence is a crucial period for brain development that includes continual myelination of cerebral white matter facilitating more efficient neural communications and acquisition of higher cognitive functions (Gogtay, Giedd et al. 2004, Gogtay and Thompson 2009, Bartzokis, Lu et al. 2010, Kochunov, Williamson et al. 2012, Kochunov, Thompson et al. 2015). Adolescence is also associated with the onset of severe mental illnesses (SMI), such as schizophrenia, bipolar and major depressive disorders, substance use and others (Rapoport, Addington et al. 2005, Casey, Nigg et al. 2007, Kalia 2008). Adolescent Brain Cognitive Development (ABCD) is the largest longitudinal study of brain development and child health consisting of over N=11,820 subjects at baseline, ascertained at 21 sites across the US (Karcher and Barch 2021). The ABCD data collection covers this critical period, and the neuroimaging data collection and analysis were intended to be optimized for quantitative analysis of multi-site diffusion, structural and functional maturation changes (Casey, Cannonier et al. 2018, Hagler, Hatton et al. 2019) . However, despite rigorous efforts to implement standardized imaging protocols and preprocessing pipelines in the ABCD study (Casey, Cannonier et al. 2018, Hagler, Hatton et al. 2019), the multi-site and multi-scanner approach, when applied to a period of rapid brain development, imposes new challenges that existing data processing pipeline and quality assessment statistical methods may not be fully equipped to handle. As in most imaging data, non-biological variations in data quality are present due to differences in scanners, deviations from the protocol, imaging artifacts and participant motion (Nielson, Pereira et al. 2018, Elyounssi, Kunitoki et al. 2023). Pipelines to address these issues for homogenizing multi-site, multiple manufacturer scanner data have been developed (Jahanshad, Kochunov et al. 2013, Fortin, Cullen et al. 2018).The challenges in data from ABCD arise when these confounds may vary not just due to different manufactures' scanners and site-specific procedural differences, but when these confounds may interact with the actual neurobiological change signals during rapid brain development, which may also vary across sites. The site/pipeline-varying signal-to-noise ratio and measurement errors may further jeopardize the fidelity of longitudinal multimodal high dimensional neuroimaging data analysis by introducing additional parameter estimation bias, inflated false positive findings, and reduced sensitivity (i.e. power) (Hagler, Hatton et al. 2019). Therefore, data quality assessment and quality control may provide a potential remedy to this issue.

Here, we performed quality assessment of longitudinal regional measurements of fractional anisotropy (FA) of water diffusion extracted for major white matter tracts using the ABCD recommended pipeline and compared it with Enhancing Neuro Imaging Genetic Meta Analyses (ENIGMA) DTI pipelines. This study focused on measuring the impact of non-biological variance on adolescent longitudinal DTI-FA measurements of cerebral white matter (WM). FA is a sensitive biomarker for WM development and increases in FA values during adolescent development indicate ongoing myelination of WM tracts (Basser 1994, Ulug, Barker et al. 1995, Conturo, McKinstry et al. 1996, Pierpaoli and Basser 1996). FA values are sensitive to many parameters (Beaulieu 2002) but the longitudinal changes in regional FA values during normal maturation are thought to be predominantly due to changes in myelination (Song, Sun et al. 2003,

Song, Yoshino et al. 2005, Budde, Kim et al. 2007, Madler, Drabycz et al. 2008, Ryan, Sherman et al. 2017, Ryan, Sherman et al. 2018). The regional changes in cerebral FA values during adolescent development agree with the classical work performed by Flechsig who demonstrated that continued myelination of WM during adolescence and early adulthood is the basis for development of higher cognitive function (Flechsig 1901, Kochunov, Williamson et al. 2012). Disorders such as schizophrenia, major depressive disorder and others are associated with altered childhood-to-adolescent development and are often described in terms of altered cerebral connectivity(Friston and Frith 1995, Lillrank, Lipska et al. 1995, Weinberger 1996, Repovs, Csernansky et al. 2011). Adult subjects with SMI show significant deficits in regional FA values, especially in multi-modal area that develop during adolescence (Kochunov, Coyle et al. 2017, Kochunov, Thompson et al. 2019, Kochunov, Hong et al. 2020). Herein, we evaluated the ability to detect longitudinal changes in FA, compared with longitudinal changes in cortical gray matter thickness and speculated on potential causes of non-biological variances. We show that longitudinal FA values demonstrate significant site-and-scanner differences, however, the FA measurements extracted by ENIGMA-DTI pipeline show significantly improved fidelity versus values derived using the currently recommended ABCD platform.

One of the key evaluation parameters for pipelines is the reproducibility in neuroimaging data. There is no single established metric for evaluating reproducibility in neuroimaging studies. While there isn't a universally established metric for this purpose, many prior studies have employed the classic test-retest reliability measurement, interclass correlation coefficients (ICC), to demonstrate excellent reproducibility for metrics such as fractional anisotropy (FA), cortical thickness, and other measurements (Shrout and Fleiss 1979, Wijtenburg, Gaston et al. 2013, Zuo and Xing 2014, Acheson, Wijtenburg et al. 2017, Xue, Yuan et al. 2021). However, ICC or other approaches may not be directly applicable to the ABCD data because the neuroimaging measures are collected every two years in participants who are undergoing developmental changes including myelination of associative white matter tracts and pruning of cortical gray matter (Kochunov, Glahn et al. 2011, Kochunov, Glahn et al. 2011, Kochunov, Williamson et al. 2012). ICC assesses the stability of repeated measures performed under similar conditions, and the dynamic alterations in the developing adolescent brain will affect the reliability analysis (Barnea-Goraly, Menon et al. 2005, Barnhart, Haber et al. 2007, Casey, Jones et al. 2008, Konrad, Firk et al. 2013).

In the present work, we described an adaptive ICC (AICC) measure to evaluate the reliability of imaging-derived phenotypes acquired from the ABCD study while accounting for the normative age-related changes. We first estimate developmental trajectory using the complete dataset, for it is more robust and accurate than site-wise estimation. The iterative AICC estimation process also factors site-wise data reliability into the calculation of age-effect. The resulting site-wise AICC can be integrated into subsequent statistical analyses to reduce bias and enhance inference efficiency. We specifically, compared ABCD and ENIGMA-DTI pipelines because the latter was specifically developed for homogenization of multi-site data and showed good performance in longitudinal analysis of adolescent development (Acheson, Wijtenburg et al. 2017).

2.Materials and Methods

2.1 Study samples

This study used baseline, two-year, and four-year follow up data from the NIMH Data Archive ABCD Curated Data Release 5.0 (https://abcdstudy.org/). The cohort and study protocols can be found in Garavan et al. (Garavan, Bartsch et al. 2018). Overall, the ABCD release 5.0 included early longitudinal data on 11,820 demographically diverse subjects, including neuroimaging data and other phenotypic data. For inclusion in the analyses, subjects were required to have both imaging data and relevant imaging acquisition information available. Additionally, subjects meeting any of the following exclusion criteria were excluded for the evaluation of site-wise data reliability: 1) attendance at different study sites during follow-up visits; 2) absence of longitudinal information.

2.2 ABCD image acquisition

In the ABCD study, imaging data were acquired using Siemens (Prisma VE11B-C), Philips (Achieva dStream, Ingenia), and GE (MR750, DV25-26) 3-Tesla MRI scanners. Siemens scanners were equipped with either 32 or 64 channel head coils. Philips scanners used 32 channel head coil. GE protocol required the use of Nova Medical 32 channel coil. The scanner-specific sequences and sequence parameters were standardized across different scanner platform with exception of minor discrepancies due to the hardware and software constraints (Casey, Cannonier et al. 2018, Hagler, Hatton et al. 2019).

2.2.1 Siemens scanners

For the T1-weighted session, a matrix of 256□×□256, 176 slices, 1.0mm isotropic resolution, a repetition time (TR) of 2500ms, echo time (TE) of 2.88ms, and a field of view (FOV) of 256□×□256 were used. T2-weighted employed the same matrix, slices, and resolution, with a TR of 3200ms and a TE of 565ms. The dMRI session had a matrix of 140□×□140, 81 slices, 1.7mm isotropic resolution, TR of 4100ms, TE of 88ms, multi-band acceleration of 3 (Moeller et al., 2010), and a FOV of 240□×□240. The protocol used 6/8 partial Fourier acquisition in phase direction.

2.2.2 Philips scanners

Matrix, FOV, and resolution parameters were identical across Philips and Siemens scanners in sMRI sessions. The T1-weighted session employed 225 slices, TR of 6.31ms, and TE of 2.9ms; the T2-weighted session used 256 slices, TR of 2500ms, and TE of 251.6ms. dMRI sessions shared similar slices, FOV, resolution, and multi-band acceleration factors, with variations in matrix (140 × 141). Phillips protocol had longer TR (5300ms) and TE (89ms). The protocol used 0.6 partial Fourier acquisition in phase direction.

2.2.3 GE scanners

GE sMRI sessions mirrored Siemens sessions in matrix, FOV, resolution, and TR, with 208 slices and TE of 2ms for T1-weighted and TE of 60ms for T2-weighted. GE dMRI sessions slightly differed from Siemens dMRI sessions in TE (81.9ms). The protocol used 5.5/8 partial Fourier acquisition in phase direction.

2.3 Imaging data preprocessing

2.3.1 ABCD sMRI and dMRI preprocessing

The details of the preprocessing pipelines from ABCD Data Analysis and Informatics Core is described elsewhere(Hagler, Hatton et al. 2019). Briefly, for sMRI scans, this pipeline performs

gradient distortion correction(Wald, Schmitt et al. 2001, Jovicich, Czanner et al. 2006) using scanner-specific and non-linear transformations provided by each scanner manufacturer, and intensity inhomogeneity correction(Sled, Zijdenbos et al. 1998, Ashburner and Friston 2000). T2-weighted images are then registered to T1-weighted images by maximizing mutual information(Wells, Viola et al. 1996), and resampled with a 1.0mm isotropic resolution reference brain in standard space. Subsequently, cortical reconstruction was performed using FreeSurfer 5.3.0, and reconstructed cortical surfaces were then registered to the Desikan atlas(Desikan, Ségonne et al. 2006). Average cortical thickness (CT) within each fuzzy-cluster parcels was calculated using smoothed surface maps(Chen, Gutierrez et al. 2012), denoted as ABCD-CT in the following analysis.

dMRI scans underwent eddy current correction using a diffusion gradient and model-based approach(Zhuang, Hrabe et al. 2006). Dark slices affected by abrupt head motion were identified through robust tensor fitting, and frames exceeding a normalized residual error threshold were censored. Head motion correction was performed by rigid-body-registering each frame to the corresponding volume synthesized from the post-ECC censored tensor fit. Gradient nonlinearity distortions were also corrected for each frame. The dMRI images were registered to T1-weighted structural images using mutual information(Wells, Viola et al. 1996), resampled with 1.7mm isotropic resolution, and adjusted for head rotation to achieve consistent diffusion orientations across participants.

2.3.2 ABCD regional FA analysis

Regional FA values were extracted for major white matter fiber tracts using AtlasTrack approach that included fitting of the diffusion tensor model to the pre-processed diffusion images (Hagler, Ahmadi et al. 2008). Diffusion tensor imaging (DTI) measures, including fractional anisotropy (FA) were obtained using a standard linear estimation approach, with two different tensor model fits: one excluding frames with b>1000 s/mm² (DTI inner shell) and another including all gradient strengths/shells (Alexander, Lee et al. 2007). In this study, we analyzed ABCD FA data extracted the full shell white matter FA modality, denoted as ABCD-FA.

2.3.2 ENIGMA regional FA analysis

We used partially preprocessed ABCD dMRI data that included all gradient strengths/shells and ENIGMA-DTI protocol to extract whole-brain and tract-wise average FA values. These protocols are detailed elsewhere (Jahanshad, Kochunov et al. 2013) and are available online, at http://www.nitrc.org/projects/enigma_dti. Briefly, DTI images were processed using the diffusion tensor analysis toolkit (FDT) distributed with FSL (http://fsl.fmrib.ox.ac.uk/fsl/fdt/)(Behrens, Woolrich et al. 2003). FA images from subjects were non-linearly registered to the ENIGMA-DTI target FA image using FSL's FNIRT(Smith, Jenkinson et al. 2004). This target was created as a "minimal deformation target" based on images from the participating studies as described previously(Kochunov, Lancaster et al. 1999, Kochunov, Lancaster et al. 2001, Jahanshad, Kochunov et al. 2013). The data were then processed using FSL's tract-based spatial statistics (TBSS; http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/TBSS) analytic method (Smith, Jenkinson et al. 2006) modified to project individual FA values onto the ENIGMA-DTI skeleton. After extracting the skeletonized white matter and the projection of individual FA values, ENIGMA tract-wise

regions of interest (Desikan, Ségonne et al.), derived from the Johns Hopkins University (JHU) white matter parcellation atlas (S, K et al. 2008), were transferred to extract the mean FA across the full skeleton and average FA values for major white matter tracts (Table S1). The whole brain average FA values were calculated to include all voxels in the ENIGMA-DTI skeleton. The protocol, ENIGMA-DTI template FA image along with its skeleton and mask, source code and executables, are all publicly available (http://www.nitrc.org/projects/enigma_dti). The FA values derived from the ENIGMA-DTI pipeline is denoted as ENIGMA-FA in the subsequent analysis.

ENIGMA-DTI Quality Assurance/Quality Control (QA/QC)
Quality assurance and quality control QA/QC was conducted according to the ENIGMA-DTI protocol (http://enigma.usc.edu/protocols/dti-protocols/). QA/QC checks consisted of inspecting the average skeleton projection distance. This measure assesses the registration quality between individual images and the ENIGMA-DTI template by calculating the average projection distance for the group-wide skeleton. The skeleton projection step corrects for minor spatial mismatch between the centers of the track on the template versus the individual image. A higher projection distance may indicate problems with aligning individual brain to the template. The ENIGMA-DTI protocol currently specifies no critical cut-off for the projection distance, and average projection distance for ENIGMA studies of 3.6±0.2 mm(Acheson, Wijtenburg et al. 2017). However, in this study, we excluded subject scans for sessions with projection distances that exceeded 3.8 mm.

2.5 Reliability analysis
In reliability analysis of neuroimaging data, the commonly used test-retest reliability scores are often built on the contrast of intra-subject vs inter-subject variances. A commonly used rationale is that the reliability is higher when the proportion of intra-subject variance is lower. The underlying assumption for this rationale is that each subject is measured repeatedly in the same condition (or with very high similarity). However, this assumption may not be applied to neuroimaging data that are subjected to the rapid brain development of the adolescents, where intra-subject changes are assumed and the relatively lower or lack of intra-subject changes in specific tracts or individuals may even infer neurodevelopment abnormality. Therefore, the contribution of intra-subject variability to reliability estimate under typical ICC is inflated, and age correction is required for unbiased estimation of testing-retesting measures. Due to the similar enrollment age across sites, the population-level age-related brain (imaging measures) developmental trajectories is relatively invariant across the 21 sites, allowing us to integrate data from all sites to estimate the age trajectories. Specifically, we calculate the AICC as follows:

Step 1: Fit a weighted mixed model across all-sites to estimate the age trajectory parameters.
Step 2: For each site, perform age-correction using the estimated age trajectory parameters and calculate ICC. Update the weights based on the site-wise ICC values.
Step 3: Repeat step 1&2 until convergence and report the updated AICC values for all sites.

In step 1, the weighted mixed model is expressed as:

$$y_{i,t} = \alpha_0 + f(age_i|\boldsymbol{X}_i) + \boldsymbol{\beta}^T\boldsymbol{X}_i + b_i + \varepsilon_{i,t} \qquad (1)$$

where $y_{i,t}$ is the outcome (e.g., FA value on a white matter fiber tract, *i* denotes the subject ($i = 1, \ldots., n$) and *t* denotes the time point; $f(age_i|\boldsymbol{X}_i)$ is function of developmental trajectory

by aging and trajectory can be sex and ethnical group specific; $\boldsymbol{X}_i$ is the vector of demographic variables; b_i is the random effect. The random slopes are generally not used because they tend to give rise to the convergence issue due to singular fit (i.e., overfit) (José C. Pinheiro 2024). The weight is incorporated in the covariance matrix $\boldsymbol{V} = \boldsymbol{Z}^T\boldsymbol{G}\boldsymbol{Z} + \boldsymbol{W}^{-\frac{1}{2}}\boldsymbol{R}\boldsymbol{W}^{-\frac{1}{2}}$, where $\boldsymbol{Z}$ is the coefficients for the random effects and $\boldsymbol{G}$ is the covariance for these coefficients and $\boldsymbol{R}$ denotes the dependence between repeated measures. The weight $\{w_{i,t}\}$ is calculated in step 2 as the input.

In step 2, we first perform ICC for each site. The weights are calculated using min-max scaling: $w_{i,t} = (\frac{ICC_i - ICC_{min}}{ICC_{max} - ICC_{min}})^2$, ICC_i is the ICC for the site where participant *i* was collected.
By repeating steps 1 and 2, $f(age_i|\boldsymbol{X}_i)$ can be estimated based on all participants while not being affected by the measurement errors. In practice, AICC converges fast. Since AICC is built on the mixed model, it is robust to drop-outs. The simulation analysis shows that AICC can effectively reduce the bias of ICC than classical testing-testing measures that ignore the developmental trends.

3.Results

AICC reveals site differences related to scanner manufacturer

The ABCD study release 5.0 includes the baseline scans from 11,820 subjects at 21 sites. The follow up scans included: N=3360 participants underwent one assessment only; 5744 participants underwent two assessments; 2619 participants completed all three assessments. 97 participants who went to different study sites during the baseline visit and follow-up visits were excluded from the study. After implementing the QC and preprocessing pipelines, the ABCD study provided brain white matter FA data for 11,542 subjects and morphometric measures (i.e., CT) for 11,802 subjects. The ENIGMA-DTI pipeline retained 11,335 subjects. Subjects with only one assessment were also excluded, as longitudinal information is imperative for the evaluation of imaging data quality in our proposed method. The distribution of subjects with multiple assessments across the 21 study sites is summarized in Table S1 in the supplementary material. There are 7,889 subjects with ABCD-FA data, 6,728 with ENIGMA-FA data, and 8,326 with ABCD-CT data.

The site-wise reliability analysis indicated that, overall, the morphometric measure (i.e., ABCD-CT) has higher reliability than the white matter FAs. Specifically, the mean AICC was estimated to be 0.61 (standard deviation [SD] 0.19) for ABCD-FA, 0.70 (SD 0.19) for ENIGMA-FA, and 0.76 (SD 0.12) for CT across all brain regions. Figure 1 visualizes the mean site-wise AICC for each of the three measures. The variance in AICCs across sites was predominantly influenced by the type of scanners employed, with Siemens scanners demonstrating notably superior reliability (AICC=0.78±0.11 vs 0.62±0.21). This distinction was substantiated by a Wilcoxon test, which revealed a statistically significant difference in AICC values between Siemens scanners and those of other manufacturers ($p < 0.001$).

Figure 3 shows the significant and positive correlation of the reliability among all three imaging outcomes. The reliability of ABCD-CT was highly correlated with that of ABCD-FA ($r = 0.86$, $p < 0.001$) and ENIGMA-FA ($r = 0.69$, $p < 0.001$). The reliability of FA derived from the two

pipelines is also positively correlated (r = 0.55, p = 0.01). As anticipated, the highest correlation of reliability was observed between outcomes derived from the same pipeline.

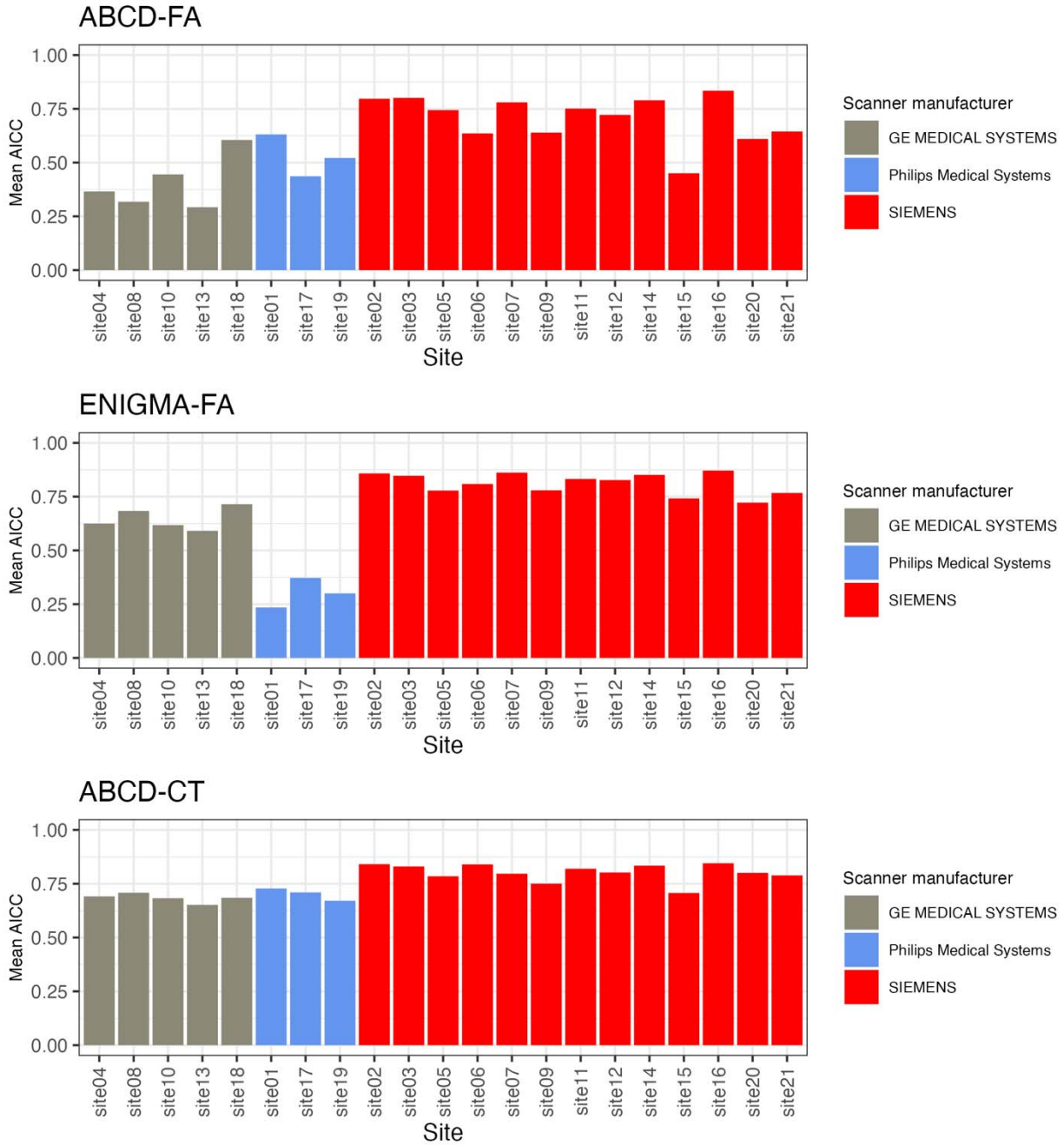

Figure 1. Mean AICC across brain regions for each site. The bar plots demonstrate the mean data reliability (i.e. AICC) of each site in the ABCD study. *Red* bars indicate sites using Siemens scanners, *blue* Philips, and *grey* GE. The top, middle, and bottom panel show the mean AICC across all brain regions of ABCD-FA, ENIGMA-FA, and ABCD-CT, respectively. Overall, reliability was better for sites using Siemens scanners (AICC=0.78±0.11 vs 0.62±0.21); reliability of FA values obtained from the ENIGMA-DTI pipeline was also better than those obtained from the ABCD recommended pipeline (AICC=0.70±0.19 vs 0.61±0.19). Morphometry

measures such as ABCD-CT has higher reliability (AICC=0.76±0.12) than FA measures and less variations across site-and-scanner.

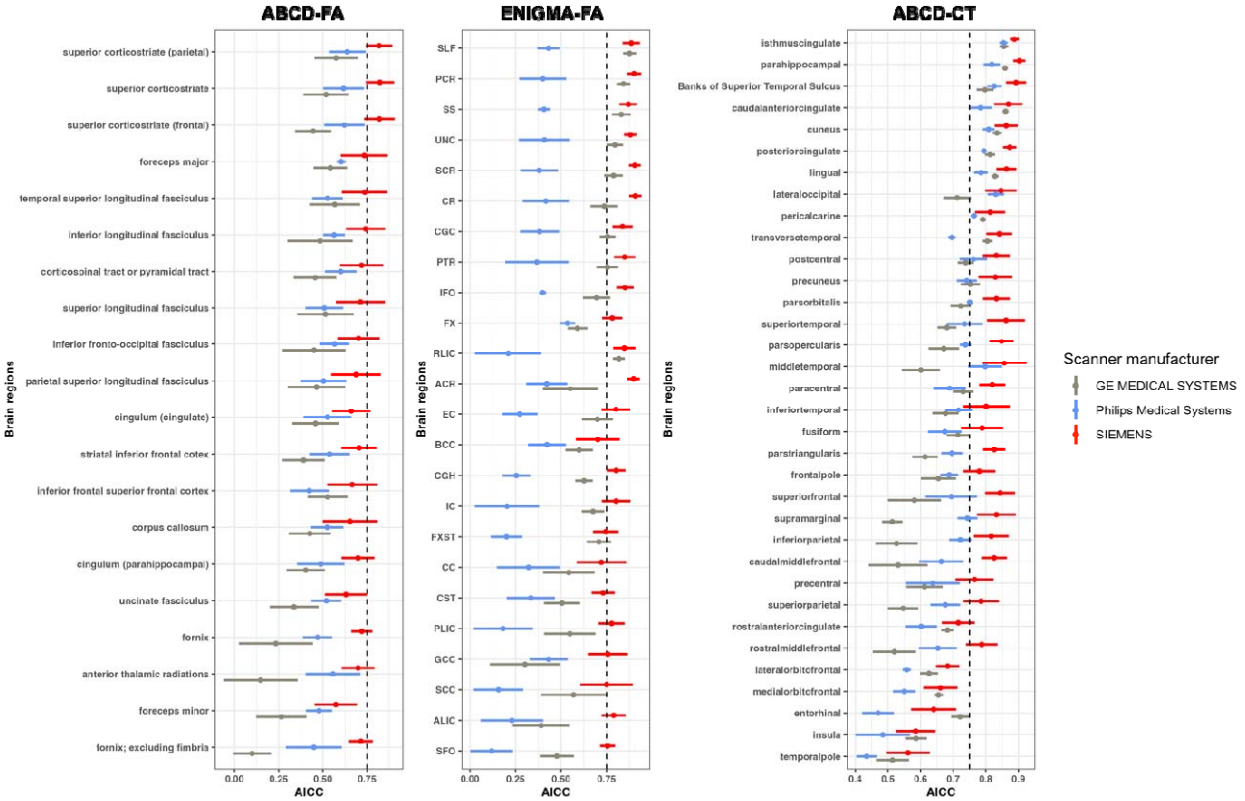

Figure 2. Regional differences of the mean AICC across sites, grouped by the type of scanners used. Dots represent the mean regional AICC across sites using scanners from the same manufacturer, and error bars indicate the standard deviation. *Red* color shows the measures obtained from Siemens scanners, *blue* Philips, and *grey* GE. The dotted line is at the conventionally good ICC score, 0.75. The full names of brain tracts from ENIGMA-FA were summarized in Table S2 from the supplementary material. Overall, FA in the fornix from both the ABCD and ENIGMA-DTI pipelines consistently show a relatively low reliability. The variations of FA reliability are smaller in Siemens sites compared to Phillips or GE sites. For ABCD-CT, differences in AICC related to scanners are notably smaller compared to both FA measures. Nevertheless, the temporal pole, insula, and entorhinal cortex consistently exhibited low mean AICC across sites, regardless of the scanner type employed.

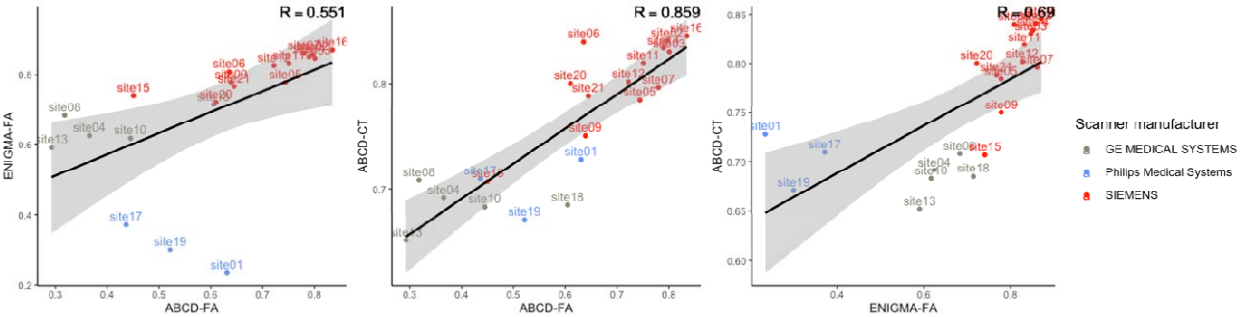

Figure 3. Correlation of mean AICC between each measure. Linear regression analysis

showed a significant and positive correlation between the site-wise reliability for all three imaging outcomes, ABCD-FA, ENIGMA-FA, and ABCD-CT. The correlation of reliability is the highest between outcomes derived from the same pipeline (ABCD-FA vs. ABCD-CT, r=0.86, p<0.001). The reliability of FA measurements derived from different pipelines displayed not as high but also significantly positive correlation (r=0.55, p=0.01).

Age trajectories

The age trajectories are displayed in Figure 4. White matter FAs derived from both the ABCD (partial ρ=0.36, p<0.001) and ENIGMA pipelines (partial ρ=0.30, p<0.001) at Siemens sites replicate the well-established age-related increase observed during adolescence (Barnea-Goraly, Menon et al. 2005, Lebel and Deoni 2018). However, data collected from GE and Philips scanners showed lower correlations (ABCD-FA: partial ρ=0.06, p<0.001; ENIGMA-FA: partial ρ=0.06, p<0.001). CT demonstrates higher reliability across all sites and show smaller differences between the trajectories obtained from Siemens (partial ρ=-0.35, p<0.001) and non-Siemens sites (partial ρ=-0.30, p<0.001). To iteratively improve the estimation of both age effects and site-wise AICC, it is essential to estimate age effects across all sites as opposed to within each site. Within this iterative process, sites characterized by inferior imaging data reliability exert diminished influence on the age effect estimate due to lower weights assigned based on the relatively low AICC obtained from the preceding iteration. Consequently, the iterative refinement process facilitates a progressively more accurate estimation of both the overall age effect and the site-wise development-corrected reliability measure.

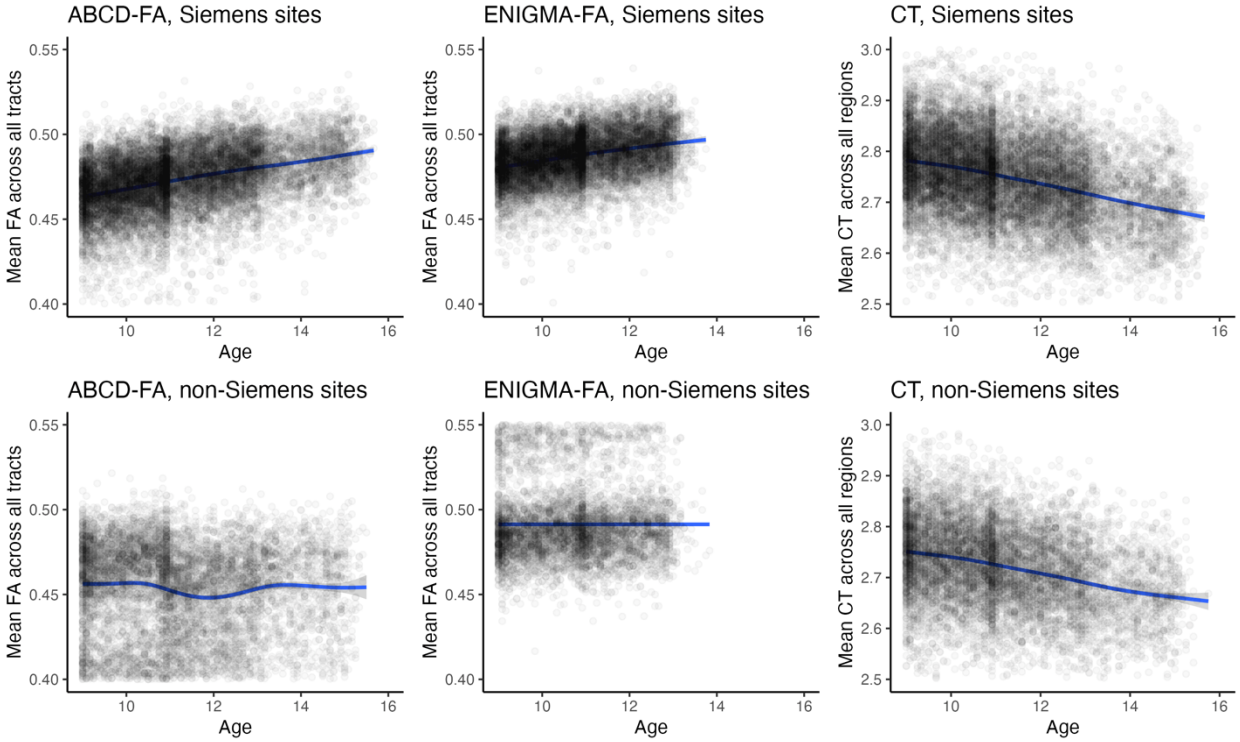

Figure 4. Correlation between age and imaging outcomes. Scatter plots of biological age and three imaging outcomes: 1) FA derived from the ABCD recommended pipeline (ABCD-FA), 2) FA derived from the ENIGMA-DTI pipeline (ENIGMA-FA), 3) CT derived the ABCD pipeline

(ABCD-CT), are shown with linear regression fits. The correlations between age and FA are positive in Siemens sites, indicating a myelination of WM in development as expected (ABCD-FA: partial $\rho=0.36$, $p<0.001$; ENIGMA-FA: partial $\rho=0.30$, $p<0.001$), while the effect sizes are attenuated in non-Siemens sites (ABCD-FA: partial $\rho=0.06$, $p<0.001$; ENIGMA-FA: partial $\rho=0.06$, $p<0.001$). The differences in the correlations between CT values obtained from different preprocessing pipelines are much smaller (Siemens sites: partial $\rho=-0.35$, $p<0.001$ vs. non-Siemens sites: partial $\rho=-0.30$, $p<0.001$).

Regional differences

Reliability measures also reveal regional differences (Figure 2). Superior corticostriate is the only region from the ABCD-FA pipeline that generated a mean AICC across sites that is higher than 0.7 (SD 0.16). The lowest AICC occurred in Foreceps minor and fornix (excluding fimbria) with a mean value of 0.49 (SD 0.17) and 0.53 (SD 0.28), respectively. In FA tracts from the ENIGMA pipeline, superior longitudinal fasciculus (SLF) showed the highest reliability across all sites, with a mean AICC of 0.82 (SD 0.17), and superior fronto-occipital fasciculus had the lowest overall reliability (AICC = 0.60, SD = 0.24). Notably, across sites using Siemens scanners, the mean regional AICC of ENIGMA-FA is higher than 0.7 for all tracts. The lowest AICC occurred in the body of corpus callosum (BCC), corticospinal tract (CST), and fornix/stria terminalis (FXST) in Siemens sites, ranging from 0.7 to 0.75. For the morphometry measure ABCD-CT, scanner-related AICC differences are much smaller than for both FA measures. However, temporal pole, insula, and the entorhinal cortex displayed consistently low mean AICC across sites using all three types of scanners.

4.Discussion

ABCD is the largest longitudinal cohort tracking adolescent brain development using standardized protocols for multi-site data collection. We examined reproducibility of fractional anisotropy values extracted using the standard ABCD and ENIGMA diffusion tensor imaging (DTI) workflows. ABCD participants were scanned two to three times within a 5-year period and therefore we expected that maturational change in white matter microstructure were reflected in FA values. Our study was focused on identifying and correcting the sources of non-biological variance that may enter the data from differences in scanners, sites, and data analyses workflows. Towards this end, we developed a novel iterative ICC approach that takes into account normal developmental trends. We show that despite the great efforts in ensuring consistent data collection across Siemens, Philips and GE scanners, the scanner related variations were large, with data collected on Philips and GE platforms showing significantly poorer reproducibility. We also showed that this manufacturer-specific poorer reproducibility is more detrimental to longitudinal diffusion data and less so on longitudinal cortical thickness data. We further show that the reproducibility metrics can be significantly improved by the use of ENIGMA-DTI workflow but it still could not fully correct for scanner differences. The sources of these scanner-related non-biological variances remain unknown but possible sources include previously reported deviation from sequencing protocol, where data collection parameters are altered from prescribed values on GE and Phillips scanners more so than on Siemens platform (Sinha and Raamana 2023).

In this study, we used a novel site-wise reliability measure and evaluated longitudinal fidelity for regional FA values derived using ABCD and ENIGMA-DTI pipelines. The use of ENIGMA-DTI pipeline improved the average reliability for white matter FA to AICC=0.70±0.19 from 0.61±0.19 for the standard ABCD pipeline. The ABCD and ENIGMA use different approaches for regional analysis of diffusion phenotypes. The ABCD pipeline is a multimodal approach that uses spatial registration information derived from the structural (T1w) images to perform inter-subject alignment of DTI data. ENIGMA-DTI is a single modality approach and performs non-linear registration using FA images. The ABCD pipeline uses a discrete cosine transform for non-linear subject alignment. ENIGMA-DTI uses a more powerful deformation field approach that warps individual images to the minimal deformation target brain(Kochunov, Lancaster et al. 2001, Kochunov, Lancaster et al. 2002, Jahanshad, Kochunov et al. 2013). ABCD pipeline uses probabilistic atlas-based method for automated segmentation of white matter fiber tracts(Hagler, Ahmadi et al. 2009). ENIGMA pipeline uses tract-based spatial statistics approach that encodes the medial trajectory of the individual white matter fiber-tracts for projection using group-wise skeleton of white matter structures. Overall, ENIGMA-DTI pipeline was specifically developed for post-hoc meta-analysis of DTI phenotypes collected using different protocol and scanners and previously demonstrated excellent test-retest performance(Acheson, Wijtenburg et al. 2017). However, even its performance was affected by difference in scanner platforms.

Despite the rigorous implementation procedures used by the ABCD investigators, we found that the scanner manufacturer differences had a large impact on the measured effects of biological age on longitudinal FA values in ABCD. The FA is hypothesized to have a positive trend with age to reflect adolescent maturation of cerebral white matter. The correlations between whole-brain average FA extracted by ABCD pipeline and age were positive for Siemens scanners as

expected. However, data collected on GE or Phillips scanners showed attenuated effect sizes, likely caused by measurement errors (Carroll, Ruppert et al. 1995). Using the ENIGMA-DTI pipelines, the variance of the estimated parameters became smaller and thus suggests an increased power for Siemens scanners in FA analyses. When examining cortical thickness, the age-slope difference between Siemens sites and non-Siemens sites is smaller. This may be explained by lower across-site variation in reliability (i.e., AICC) and reduced disparity in measurement errors between Siemens and non-Siemens sites. Therefore, the influence of measurement errors on effect size (i.e., correlation) estimation was attenuated. These results underscore the substantial influence of imaging data reliability on association analyses and the credibility of neurobiological findings. Quality assessment and subsequent control is critical for obtaining reproducible and reliable association analysis results.

Data collected using Siemens scanners also showed significantly higher reproducibility than these using GE/Phillips scanners for ABCD and ENIGMA-FA values, as well as cortical gray matter thickness (Wilcoxon test $p<0.001$). The findings for data collected using Siemens scanners showed good agreement with expected maturation trends and good-to-excellent reproducibility for ABCD imaging outcomes. The use of ENIGMA-DTI pipeline improved reproducibility and agreed with high ICCs reported in previous studies that were performed in adults where little change was expected (Acheson, Wijtenburg et al. 2017). The ENIGMA-DTI pipeline was specifically developed for meta-and-mega analysis of unharmonized data collected worldwide(Jahanshad, Kochunov et al. 2013). The effect sizes yielded from these analyses across the variety of the neuropsychiatric illnesses showed an excellent reproducibility in independent cohorts. The big variances in individual FA values due to scanner platform are concerning despite the effort invested into harmonization of the sequences. The scanner-related data differences in ABCD structural and functional data have been reported before (Nielson, Pereira et al. 2018, Sinha and Raamana 2023). We speculate on two likely causes for this.

The likely sources of non-biological variance that lowered the reproducibility metrics for GE and Philips scanners include protocol deviation and differences in machine hardware and software technology that cannot be homogenized across the manufacturers. The protocol deviation or non-compliance refers to imaging parameter differences between what are described in the protocol and what were actually input and executed during imaging. The deviations are more pervasive for data collected on Philips and GE versus Siemens scanners (Sinha and Raamana 2023). The deviation can be caused by operator errors. For example, the GE MR750 software may not show all slices for DTI sequence and to overcome this, operator must change the prescribed parameters and then change it back to the protocol mandated value. This can introduce a human error, compared to Siemens platform that does not require this step. The deviation of the protocol can also be caused by the differences in the automated scanner optimization algorithms built in by manufacturers that lead to changes in critical sequence parameters such as echo and repetition time, unbeknown to the operator (Sinha and Raamana 2023). The scanner related difference in long term reproducibility may also arise from complex methodological differences such as the overall stability of the magnet and gradient systems, differences in the k-space sampling trajectories and others. For example, Philips and GE scanners used by ABCD have slower gradient slew rates and therefore longer echo spacing. This led the ABCD protocol to use more aggressive partial Fourier imaging. Siemens DTI protocol collected 75% of the k-space versus 69% for GE and only 55% for Philips. The default direction in which k-space is transversed is also

different between Siemens and GE and Philips scanners. These differences are hard to harmonize and likely cause complex interaction with the multi-slice excitation, leading to differences in signal, noise, distortions, and Nyquist ghost locations that reduce the overall reproducibility.

We further observed significant differences in reproducibility in regional FA values that were consistent with the results from another reproducibility study of ENIGMA-FA (Acheson, Wijtenburg et al. 2017). For Siemens sites, FA values showed the lowest reliability in cortico-spinal (CST) and fornix-stria teminalis tracts (FXST). Acheson et al. also reported large year-to-year variances observed in CST and FX using the ENIGMA-FA pipeline in two independent cohorts (Acheson, Wijtenburg et al. 2017). Both CST and FXST are long and narrow bundles that are hard-to-correct for individual anatomical variability (Acheson, Wijtenburg et al. 2017). This implies that the low reproducibility observed in these tracts is more likely attributable to white matter template alignment-related methodological confounds rather than a failure to account for the developmental trajectory. ABCD pipeline likewise showed lowest reliability for CST and FX in Siemens sites, suggesting the misalignments in registration or partial voluming effects (M, FB et al. 2014). Additionally, ABCD-CT measurements also demonstrated consistently low AICC regions within the temporal lobe (i.e. temporal pole, insula, entorhinal cortex) and in the medial/lateral orbital frontal lobe. Seiger et. al's work showed that these regions have the largest difference in region-wise CT estimations using two different methods, suggesting these regions are susceptible to methodological confounds when estimating CT values (Seiger, Ganger et al. 2018).

The presence of measurement errors in neuroimaging data introduces biases in association analyses. Measurement error in imaging predictors can attenuate or weaken observed brain-behavior effect sizes, while measurement error in imaging outcomes can inflate estimates of standard errors (Kenny 1979). In the context of longitudinal data analysis, non-negligible measurement errors can also cause the ordinary maximum likelihood estimators to be inconsistent (Fuller 2009). To address measurement errors, potential strategies involve careful consideration of data quality based on both data-driven methods such as AICC and manual quality control (MQC) measures. MQC on the ABCD T1-weighted images suggested that 55.1% of the scans were rated as lower-quality images that led to bias in statistical analysis results(Elyounssi, Kunitoki et al. 2023). Therefore, the evaluation of data reliability becomes crucial for the robustness and accuracy of longitudinal multimodal imaging data analysis. While MQC is a valuable method for flexible quality assessment, its inherent labor-intensive nature poses practical challenges, particularly in the context of large-scale studies. In such scenarios, quantitative data consistency metrics like AICC can offer a more efficient and scalable alternative, allowing for data quality assessments with enhanced speed and consistency.

Commonly used test-retest reliability scores like ICC typically assume repeated measurements under similar conditions, and thus they are not applicable to the ABCD study due to the adolescent brain development that inflates intra-subject variability. Our proposed AICC integrates data from all sites to estimate age trajectories and addresses the need for age correction in unbiased site-wise testing-retesting measures. Using a weighted mixed model, AICC can be easily extended to account for demographic variables and random effects, effectively reducing bias and being robust to dropouts. Additionally, AICC can also be applied to other large multi-

site study that longitudinally collect participants with time-varying brain conditions (e.g., elder participants in aging).

Ensuring high data quality is critical for large-scale studies seeking meaningful insights from imaging data analysis. Both correction methods and rigorous quality control filtering are essential for reducing measurement errors. Excluding subjects or even sites with lower data quality decreases the sample size, while maintaining an optimal sample size could result in latent bias. Therefore, the challenge lies in striking a balance between sample size and noise. Towards these goals, we have devised a novel longitudinal reliability statistical method. This approach takes into account the expected normal age-related intra-subject variance, a crucial consideration to maximize the value of neurodevelopmental data, such as those provided by the ABCD study. We identified critical scanner-type-related confounds for specific imaging data, particularly longitudinal diffusion imaging data collected during rapid neurodevelopment, and have recommended at least partial solutions by using more robust pipeline and attention to the use of data from different sites for different types of data use.

Reference

Acheson, A., S. Wijtenburg, L. Rowland, A. Winkler, C. W. Mathias, L. Hong, N. Jahanshad, B. Patel, P. Thompson, S. McGuire, P. Sherman, P. Kochunov and D. D. Dougherty (2017). "Reproducibility of tract-based white matter microstructural measures using the ENIGMA-DTI protocol." Genes Brain Behav **7**(2): 10.1002/brb1003.1615.

Alexander, A. L., J. E. Lee, M. Lazar and A. S. Field (2007). "Diffusion tensor imaging of the brain." Neurotherapeutics **4**(3): 316-329.

Ashburner, J. and K. J. Friston (2000). "Voxel-based morphometry--the methods." NeuroImage **11**(6 Pt 1): 805-821.

Barnea-Goraly, N., V. Menon, M. Eckert, L. Tamm, R. Bammer, A. Karchemskiy, C. C. Dant and A. L. Reiss (2005). "White Matter Development During Childhood and Adolescence: A Cross-sectional Diffusion Tensor Imaging Study." Cerebral Cortex **15**(12): 1848-1854.

Barnhart, H. X., M. J. Haber and L. I. Lin (2007). "An Overview on Assessing Agreement with Continuous Measurements." Journal of Biopharmaceutical Statistics **17**(4): 529-569.

Bartzokis, G., P. H. Lu, K. Tingus, M. F. Mendez, A. Richard, D. G. Peters, B. Oluwadara, K. A. Barrall, J. P. Finn, P. Villablanca, P. M. Thompson and J. Mintz (2010). "Lifespan trajectory of myelin integrity and maximum motor speed." Neurobiology of aging **31**(9): 1554-1562.

Basser, P. J. (1994). "Focal magnetic stimulation of an axon." IEEE Transactions on Biomedical Engineering **41**(6): 601-606.

Beaulieu, C. (2002). "The basis of anisotropic water diffusion in the nervous system - a technical review." NMR Biomed **15**(7-8): 435-455.

Behrens, T. E., M. W. Woolrich, M. Jenkinson, H. Johansen-Berg, R. G. Nunes, S. Clare, P. M. Matthews, J. M. Brady and S. M. Smith (2003). "Characterization and propagation of uncertainty in diffusion-weighted MR imaging." Magn Reson Med **50**(5): 1077-1088.

Budde, M. D., J. H. Kim, H. F. Liang, R. E. Schmidt, J. H. Russell, A. H. Cross and S. K. Song (2007). "Toward accurate diagnosis of white matter pathology using diffusion tensor imaging." Magn Reson Med **57**(4): 688-695.

Carroll, R. J., D. Ruppert and L. A. Stefanski (1995). Measurement Error in Nonlinear Models, CRC Press.
Casey, B. J., T. Cannonier, M. I. Conley, A. O. Cohen, D. M. Barch, M. M. Heitzeg, M. E. Soules, T. Teslovich, D. V. Dellarco, H. Garavan, C. A. Orr, T. D. Wager, M. T. Banich, N. K. Speer, M. T. Sutherland, M. C. Riedel, A. S. Dick, J. M. Bjork, K. M. Thomas, ... and A. M. Dale (2018). "The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites." Developmental Cognitive Neuroscience **32**: 43-54.
Casey, B. J., R. M. Jones and T. A. Hare (2008). "The Adolescent Brain." Annals of the New York Academy of Sciences **1124**: 111-126.
Casey, B. J., J. T. Nigg and S. Durston (2007). "New potential leads in the biology and treatment of attention deficit-hyperactivity disorder." Curr Opin Neurol **20**(2): 119-124.
Chen, C.-H., E. D. Gutierrez, W. Thompson, M. S. Panizzon, T. L. Jernigan, L. T. Eyler, C. Fennema-Notestine, A. J. Jak, M. C. Neale, C. E. Franz, M. J. Lyons, M. D. Grant, B. Fischl, L. J. Seidman, M. T. Tsuang, W. S. Kremen and A. M. Dale (2012). "Hierarchical Genetic Organization of Human Cortical Surface Area." Science (New York, N.Y.) **335**(6076): 1634-1636.
Conturo, T. E., R. C. McKinstry, E. Akbudak and B. H. Robinson (1996). "Encoding of anisotropic diffusion with tetrahedral gradients: a general mathematical diffusion formalism and experimental results." Magn Reson Med **35**(3): 399-412.
Desikan, R. S., F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert and R. J. Killiany (2006). "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest." NeuroImage **31**(3): 968-980.
Elyounssi, S., K. Kunitoki, J. A. Clauss, E. Laurent, K. Kane, D. E. Hughes, C. E. Hopkinson, O. Bazer, R. F. Sussman, A. E. Doyle, H. Lee, B. Tervo-Clemmens, H. Eryilmaz, R. L. Gollub, D. M. Barch, T. D. Satterthwaite, K. F. Dowling and J. L. Roffman (2023). "Uncovering and mitigating bias in large, automated MRI analyses of brain development." bioRxiv: 2023.2002.2028.530498.
Flechsig, P. (1901). "Developmental (myelogenetic) localisation of the cerebral cortex in the human." Lancet **158**: 1027-1030.
Fortin, J.-P., N. Cullen, Y. I. Sheline, W. D. Taylor, I. Aselcioglu, P. A. Cook, P. Adams, C. Cooper, M. Fava, P. J. Mcgrath, M. Mcinnis, M. L. Phillips, M. H. Trivedi, M. M. Weissman and R. T. Shinohara (2018). "Harmonization of cortical thickness measurements across scanners and sites." NeuroImage **167**: 104-120.
Friston, K. J. and C. D. Frith (1995). "Schizophrenia: a disconnection syndrome?" Clin Neurosci **3**(2): 89-97.
Fuller, W. A. (2009). Measurement Error Models, John Wiley & Sons.
Garavan, H., H. Bartsch, K. Conway, A. Decastro, R. Z. Goldstein, S. Heeringa, T. Jernigan, A. Potter, W. Thompson and D. Zahs (2018). "Recruiting the ABCD sample: Design considerations and procedures." Developmental Cognitive Neuroscience **32**: 16-22.
Gogtay, N., J. N. Giedd, L. Lusk, K. M. Hayashi, D. Greenstein, A. C. Vaituzis, T. F. Nugent, 3rd, D. H. Herman, L. S. Clasen, A. W. Toga, J. L. Rapoport and P. M. Thompson (2004). "Dynamic mapping of human cortical development during childhood through early adulthood." Proc Natl Acad Sci U S A **101**(21): 8174-8179.
Gogtay, N. and P. M. Thompson (2009). "Mapping gray matter development: implications for typical development and vulnerability to psychopathology." Brain Cogn **72**(1): 6-15.

Hagler, D. J., M. E. Ahmadi, J. Kuperman, D. Holland, C. R. McDonald, E. Halgren and A. M. Dale (2008). "Automated white‐matter tractography using a probabilistic diffusion tensor atlas: Application to temporal lobe epilepsy." Human Brain Mapping **30**(5): 1535-1547.
Hagler, D. J., S. N. Hatton, M. D. Cornejo, C. Makowski, D. A. Fair, A. S. Dick, M. T. Sutherland, B. J. Casey, D. M. Barch, M. P. Harms, R. Watts, J. M. Bjork, H. P. Garavan, L. Hilmer, C. J. Pung, C. S. Sicat, J. Kuperman, H. Bartsch, F. Xue, ... and A. M. Dale (2019). "Image processing and analysis methods for the Adolescent Brain Cognitive Development Study." NeuroImage **202**: 116091.
Hagler, D. J., Jr., M. E. Ahmadi, J. Kuperman, D. Holland, C. R. McDonald, E. Halgren and A. M. Dale (2009). "Automated white-matter tractography using a probabilistic diffusion tensor atlas: Application to temporal lobe epilepsy." Hum Brain Mapp **30**(5): 1535-1547.
Jahanshad, N., P. Kochunov, E. Sprooten, R. C. Mandl, T. E. Nichols, L. Almassy, J. Blangero, R. M. Brouwer, J. E. Curran, G. I. de Zubicaray, R. Duggirala, P. T. Fox, L. E. Hong, B. A. Landman, N. G. Martin, K. L. McMahon, S. E. Medland, B. D. Mitchell, R. L. Olvera, ... and D. C. Glahn (2013). "Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: A pilot project of the ENIGMA-DTI working group." Neuroimage **doi:pii: S1053-8119(13)00408-4. 10.1016/j.neuroimage.2013.04.061.**
José C. Pinheiro , D. M. B. (2024). Mixed-Effects Models in S and S-PLUS.
Jovicich, J., S. Czanner, D. Greve, E. Haley, A. van der Kouwe, R. Gollub, D. Kennedy, F. Schmitt, G. Brown, J. Macfall, B. Fischl and A. Dale (2006). "Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data." NeuroImage **30**(2): 436-443.
Kalia, M. (2008). "Brain development: anatomy, connectivity, adaptive plasticity, and toxicity." Metabolism **57 Suppl 2**: S2-5.
Karcher, N. R. and D. M. Barch (2021). "The ABCD study: understanding the development of risk for mental and physical health outcomes." Neuropsychopharmacology **46**(1): 131-142.
Kenny, D. A. (1979). Correlation and causality.
Kochunov, P., T. Coyle, L. Rowland, N. Jahanshad, P. Thompson, S. Kelly, X. Du, H. Sampath, H. Bruce, J. Chiappelli, M. Ryan, F. Fisseha, A. Savransky, B. Adhikari, C. Chen, S. Paciga, C. Whelan, Z. Xie, C. Hyde, ... and H. LE. (2017). "White Matter and Core Cognitive Deficits in Schizophrenia." JAMA Psychiatry **74**(9): 958-966.
Kochunov, P., D. Glahn, T. Nichols, A. Winkler, E. Hong, H. Holcomb, J. Stein, P. Thompson, J. Curran, M. Carless, R. Olvera, M. Johnson, S. Cole, V. Kochunov, J. Kent and J. Blangero (2011). "Genetic Analysis of Cortical Thickness and Fractional Anisotropy of Water Diffusion in the Brain." Frontiers in Neuroscience **5**(120): 1-15.
Kochunov, P., D. C. Glahn, J. Lancaster, P. M. Thompson, V. Kochunov, B. Rogers, P. Fox, J. Blangero and D. E. Williamson (2011). "Fractional anisotropy of cerebral white matter and thickness of cortical gray matter across the lifespan." Neuroimage **58**(1): 41-49.
Kochunov, P., L. E. Hong, E. L. Dennis, R. A. Morey, D. F. Tate, E. A. Wilde, M. Logue, S. Kelly, G. Donohoe, P. Favre, J. Houenou, C. R. K. Ching, L. Holleran, O. A. Andreassen, L. S. van Velzen, L. Schmaal, J. E. Villalon-Reina, C. E. Bearden, F. Piras, ... and N. Jahanshad (2020). "ENIGMA-DTI: Translating reproducible white matter deficits into personalized vulnerability metrics in cross-diagnostic psychiatric research." Hum Brain Mapp.

Kochunov, P., J. Lancaster, P. Thompson, A. W. Toga, P. Brewer, J. Hardies and P. Fox (2002). "An optimized individual target brain in the Talairach coordinate system." Neuroimage **17**(2): 922-927.
Kochunov, P., J. L. Lancaster, P. Thompson, R. Woods, J. Mazziotta, J. Hardies and P. Fox (2001). "Regional spatial normalization: toward an optimal target." J Comput Assist Tomogr **25**(5): 805-816.
Kochunov, P., P. M. Thompson and L. E. Hong (2019). "Toward High Reproducibility and Accountable Heterogeneity in Schizophrenia Research." JAMA Psychiatry **76**(7): 680-681.
Kochunov, P., P. M. Thompson, A. Winkler, M. Morrissey, M. Fu, T. R. Coyle, X. Du, F. Muellerklein, A. Savransky, C. Gaudiot, H. Sampath, G. Eskandar, N. Jahanshad, B. Patel, L. Rowland, T. E. Nichols, J. R. O'Connell, A. R. Shuldiner, B. D. Mitchell and L. E. Hong (2015). "The common genetic influence over processing speed and white matter microstructure: Evidence from the Old Order Amish and Human Connectome Projects." Neuroimage **125**: 189-197.
Kochunov, P., D. E. Williamson, J. Lancaster, P. Fox, J. Cornell, J. Blangero and D. C. Glahn (2012). "Fractional anisotropy of water diffusion in cerebral white matter across the lifespan." Neurobiol Aging **33**(1): 9-20.
Kochunov, P. V., J. L. Lancaster and P. T. Fox (1999). "Accurate high-speed spatial normalization using an octree method." Neuroimage **10**(6): 724-737.
Konrad, K., C. Firk and P. J. Uhlhaas (2013). "Brain Development During Adolescence." Deutsches Ärzteblatt International **110**(25): 425-431.
Lebel, C. and S. Deoni (2018). "The Development of Brain White Matter Microstructure." NeuroImage **182**: 207-218.
Lillrank, S. M., B. K. Lipska and D. R. Weinberger (1995). "Neurodevelopmental animal models of schizophrenia." Clin Neurosci **3**(2): 98-104.
M, B., L. FB, L. A, T. CM, B. GJ, S. B and M.-H. KH (2014). "Methodological considerations on tract-based spatial statistics (TBSS)." NeuroImage **100**.
Madler, B., S. A. Drabycz, S. H. Kolind, K. P. Whittall and A. L. MacKay (2008). "Is diffusion anisotropy an accurate monitor of myelination? Correlation of multicomponent T2 relaxation and diffusion tensor anisotropy in human brain." Magn Reson Imaging **26**(7): 874-888.
Nielson, D. M., F. Pereira, C. Y. Zheng, N. Migineishvili, J. A. Lee, A. G. Thomas and P. A. Bandettini (2018). "Detecting and harmonizing scanner differences in the ABCD study - annual release 1.0." bioRxiv: 309260.
Pierpaoli, C. and P. J. Basser (1996). "Toward a quantitative assessment of diffusion anisotropy." Magn Reson Med **36**(6): 893-906.
Rapoport, J. L., A. Addington and S. Frangou (2005). "The neurodevelopmental model of schizophrenia: what can very early onset cases tell us?" Curr Psychiatry Rep **7**(2): 81-82.
Repovs, G., J. G. Csernansky and D. M. Barch (2011). "Brain network connectivity in individuals with schizophrenia and their siblings." Biol Psychiatry **69**(10): 967-973.
Ryan, M. C., P. Sherman, L. M. Rowland, S. A. Wijtenburg, A. Acheson, E. Fieremans, J. Veraart, D. S. Novikov, L. E. Hong, J. Sladky, P. D. Peralta, P. Kochunov and S. A. McGuire (2017). "Miniature pig model of human adolescent brain white matter development." Journal of Neuroscience Methods **296**: 99-108.

Ryan, M. C., P. Sherman, L. M. Rowland, S. A. Wijtenburg, A. Acheson, E. Fieremans, J. Veraart, D. S. Novikov, L. E. Hong, J. Sladky, P. D. Peralta, P. Kochunov and S. A. McGuire (2018). "Miniature pig magnetic resonance spectroscopy model of normal adolescent brain development " Journal of Neuroscience Methods **296**: 99-108.
S, M., O. K, J. H, J. L, L. X, A. K, H. K, F. AV, M. A, W. R, T. AW, P. GB, N. PR, E. A, Z. J, H. H, M. MI, v. Z. P and M. J (2008). "Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template." NeuroImage **40**(2).
Seiger, R., S. Ganger, G. S. Kranz, A. Hahn and R. Lanzenberger (2018). "Cortical Thickness Estimations of FreeSurfer and the CAT12 Toolbox in Patients with Alzheimer's Disease and Healthy Controls." Journal of Neuroimaging **28**(5): 515-523.
Shrout, P. E. and J. L. Fleiss (1979). "Intraclass correlations: uses in assessing rater reliability." Psychological Bulletin **86**(2): 420-428.
Sinha, H. and P. R. Raamana (2023). "Solving the Pervasive Problem of Protocol Non-Compliance in MRI using an Open-Source tool mrQA." bioRxiv: 2023.2007.2017.548591.
Sled, J. G., A. P. Zijdenbos and A. C. Evans (1998). "A nonparametric method for automatic correction of intensity nonuniformity in MRI data." IEEE transactions on medical imaging **17**(1): 87-97.
Smith, S., M. Jenkinson, M. Woolrich, C. Beckmann, T. Behrens, H. Johansen-Berg, P. Bannister, M. De Luca, I. Drobnjak, D. Flitney, R. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J. Brady and P. Matthews (2004). "Advances in functional and structural MR image analysis and implementation as FSL." Neuroimage **23**: 208-219.
Smith, S. M., M. Jenkinson, H. Johansen-Berg, D. Rueckert, T. E. Nichols, C. E. Mackay, K. E. Watkins, O. Ciccarelli, M. Z. Cader, P. M. Matthews and T. E. Behrens (2006). "Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data." Neuroimage **31**(4): 1487-1505.
Song, S. K., S. W. Sun, W. K. Ju, S. J. Lin, A. H. Cross and A. H. Neufeld (2003). "Diffusion tensor imaging detects and differentiates axon and myelin degeneration in mouse optic nerve after retinal ischemia." Neuroimage **20**(3): 1714-1722.
Song, S. K., J. Yoshino, T. Q. Le, S. J. Lin, S. W. Sun, A. H. Cross and R. C. Armstrong (2005). "Demyelination increases radial diffusivity in corpus callosum of mouse brain." Neuroimage **26**(1): 132-140.
Ulug, A. M., P. B. Barker and P. C. van Zijl (1995). "Correction of motional artifacts in diffusion-weighted images using a reference phase map." Magn Reson Med **34**(3): 476-480.
Wald, L., F. Schmitt and A. Dale (2001). "Systematic spatial distortion in MRI due to gradient non-linearities." NeuroImage **6 Supplement**(13): 50.
Weinberger, D. R. (1996). "On the plausibility of "the neurodevelopmental hypothesis" of schizophrenia." Neuropsychopharmacology **14**(3 Suppl): 1S-11S.
Wells, W. M., P. Viola, H. Atsumi, S. Nakajima and R. Kikinis (1996). "Multi-modal volume registration by maximization of mutual information." Medical Image Analysis **1**(1): 35-51.
Wijtenburg, S. A., F. E. Gaston, E. A. Spieker, S. A. Korenic, P. Kochunov, L. E. Hong and L. M. Rowland (2013). "Reproducibility of phase rotation STEAM at 3T: focus on glutathione." Magn Reson Med **72**(3): 603-609.

Xue, C., J. Yuan, G. G. Lo, A. T. Y. Chang, D. M. C. Poon, O. L. Wong, Y. Zhou and W. C. W. Chu (2021). "Radiomics feature reliability assessed by intraclass correlation coefficient: a systematic review." Quantitative Imaging in Medicine and Surgery **11**(10): 4431-4460.
Zhuang, J., J. Hrabe, A. Kangarlu, D. Xu, R. Bansal, C. A. Branch and B. S. Peterson (2006). "Correction of Eddy-Current Distortions in Diffusion Tensor Images Using the Known Directions and Strengths of Diffusion Gradients." Journal of magnetic resonance imaging : JMRI **24**(5): 1188.
Zuo, X.-N. and X.-X. Xing (2014). "Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: A systems neuroscience perspective." Neuroscience & Biobehavioral Reviews **45**: 100-118.