Characterizing patterns of diffusion tensor imaging variance in aging brains

Chenyu Gao,[a,*] Qi Yang,[b] Michael E. Kim,[b] Nazirah Mohd Khairi,[a] Leon Y. Cai,[c] Nancy R. Newlin,[b] Praitayini Kanakaraj,[b] Lucas W. Remedios,[b] Aravind R. Krishnan,[a] Xin Yu,[b] Tianyuan Yao,[b] Panpan Zhang,[d] Kurt G. Schilling,[e,f] Daniel Moyer,[b] Derek B. Archer,[g,h] Susan M. Resnick,[i] Bennett A. Landman[a,b,c,e,f] for the Alzheimer's Disease Neuroimaging Initiative[1] and the BIOCARD Study team[2]

[a]Vanderbilt University, Department of Electrical and Computer Engineering, Nashville, United States
[b]Vanderbilt University, Department of Computer Science, Nashville, United States
[c]Vanderbilt University, Department of Biomedical Engineering, Nashville, United States
[d]Vanderbilt University Medical Center, Department of Biostatistics, Nashville, United States
[e]Vanderbilt University Medical Center, Department of Radiology and Radiological Sciences, Nashville, USA
[f]Vanderbilt University, Vanderbilt University Institute of Imaging Science, Nashville, USA
[g]Vanderbilt Memory and Alzheimer's Center, Vanderbilt University School of Medicine, Nashville, USA
[h]Vanderbilt University Medical Center, Vanderbilt Genetics Institute, Nashville, USA
[i]National Institute on Aging, Laboratory of Behavioral Neuroscience, Baltimore, United States

[1]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

[2]Data used in preparation of this article were derived from BIOCARD study data, supported by grant U19 – AG033655 from the National Institute on Aging. The BIOCARD study team did not participate in the analysis or writing of this report, however, they contributed to the design and implementation of the study. A listing of BIOCARD investigators may be accessed at: https://www.biocard-se.org/public/Core%20Groups.html

Abstract

Purpose: As large analyses merge data across sites, a deeper understanding of variance in statistical assessment across the sources of data becomes critical for valid analyses. Diffusion tensor imaging (DTI) exhibits spatially varying and correlated noise, so care must be taken with distributional assumptions. Here we characterize the role of physiology, subject compliance, and the interaction of subject with the scanner in the understanding of DTI variability, as modeled in spatial variance of derived metrics in homogeneous regions.

Approach: We analyze DTI data from 1035 subjects in the Baltimore Longitudinal Study of Aging (BLSA), with ages ranging from 22.4 to 103 years old. For each subject, up to 12 longitudinal sessions were conducted. We assess variance of DTI scalars within regions of interest (ROIs) defined by four segmentation methods and investigate the relationships between the variance and covariates, including baseline age, time from the baseline (referred to as "interval"), motion, sex, and whether it is the first scan or the second scan in the session.

Results: Covariate effects are heterogeneous and bilaterally symmetric across ROIs. Inter-session interval is positively related ($p \ll 0.001$) to FA variance in the cuneus and occipital gyrus, but negatively ($p \ll 0.001$) in the caudate nucleus. Males show significantly ($p \ll 0.001$) higher FA variance in the right putamen, thalamus, body of the corpus callosum, and cingulate gyrus. In 62 out of 176 ROIs defined by the Eve type-1 atlas, an increase in motion is associated ($p < 0.05$) with a decrease in FA variance. Head motion increases during the rescan of DTI ($\Delta\mu = 0.045$ millimeters per volume).

Conclusions: The effects of each covariate on DTI variance, and their relationships across ROIs are complex. Ultimately, we encourage researchers to include estimates of variance when sharing data and consider models of heteroscedasticity in analysis. This work provides a foundation for study planning to account for regional variations in metric variance.

Keywords: brain, aging, DTI, variance, motion.

*Address all correspondence to Chenyu Gao, e-mail address: chenyu.gao@vanderbilt.edu, full postal address: Vanderbilt University EECS, 2301 Vanderbilt Pl., PO Box 351679 Station B, Nashville, TN 37235-1679

1 Introduction

Large datasets enable exploration of questions that would be impractical with smaller- or moderate-sized datasets while giving rise to the development and application of deep learning models which can assimilate complex data. One prevalent challenge is that large datasets often comprise samples aggregated from distinct sources at different time points using diverse technologies, causing data heterogeneity, experimental variations, and statistical biases if the analysis is not executed appropriately.[1] In such scenarios, understanding the variance and variability of data is of great importance. The general linear model[2], a structured and widely used framework for relationship modeling, allows us to illustrate the importance. The general linear model is assessed through regression. A linear regression can be expressed by $\mathrm{Y} = \mathrm{X}\beta + \varepsilon$, where the response variable Y, the covariate matrix X, and the regression coefficients β are conventionally represented in matrix forms given by:

$$\mathrm{Y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ y^{(3)} \\ \vdots \\ y^{(M)} \end{bmatrix} \qquad \mathrm{X} = \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ x^{(3)} \\ \vdots \\ x^{(M)} \end{bmatrix} = \begin{bmatrix} 1 & x_1^{(1)} & \cdots & x_N^{(1)} \\ 1 & x_1^{(2)} & \cdots & x_N^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(M)} & \cdots & x_N^{(M)} \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix} \tag{1}$$

where we use M to denote the number of samples, and N to denote the number of independent variables. The error term ε is given by:

$$\varepsilon \sim \mathcal{N}(0, \Sigma) \tag{2}$$

where Σ represents the covariance matrix. If we assume the errors are uncorrelated, Σ is simplified to a diagonal matrix. We can simplify estimation of Eq. (1) with a whitening matrix[3], W:

$$\mathrm{WY} = \mathrm{WX}\beta + \mathrm{W}\varepsilon, \text{ where } \mathrm{W} = \mathrm{W}^{\mathrm{T}} = \Sigma^{-1/2} \tag{3}$$

Note $\mathrm{W}\varepsilon \sim \mathcal{N}(0, \mathrm{I}_M)$, where I_M denotes the identity matrix of dimension $M \times M$. We illustrate the practical importance of understanding the variance structure for reducing statistical errors (Fig. 1).

Diffusion tensor imaging (DTI)[4–6] is a modeling approach used in diffusion-weighted imaging [7–9], a variant of conventional magnetic resonance imaging (MRI) based on the tissue water diffusion rate.[10] DTI allows for visualization and measurement of the degree of anisotropy and structural orientation of fibers in the brain and has been widely used in studies.[11–13] DTI is inherently subject to low signal-to-noise ratio, and the noise structure exhibits spatial variability and correlation, primarily attributed to fast imaging and noise suppression techniques.[14,15] Understanding the statistical nature of DTI variance or noise has been proven to be beneficial for diffusion tensor estimation[15,16], outlier detection[17], reproducibility assessment[18–20].

Methods have been proposed for DTI variance (or noise) estimation, among which we recognize four types. The first type requires multiple repeated acquisitions (therefore, we refer to it as the "multiple acquisition method"). After the repeated acquisitions, we could take the standard deviation of the measurements, or perform data resampling, such as bootstrap or jackknife,[21] to quantify uncertainties of DTI parameters. The second type involves two repeated acquisitions, which we call the "scan-rescan method". We compute the difference between the images from each acquisition and then calculate the standard deviation of the difference across the space.[22,23] Note that the standard deviation must be divided by the square root of 2 to account for the combination of two random variables (noise in each image). The third and fourth types are used when we have only one acquisition. For the third type, we select a homogeneous ROI and compute

the standard deviation of the measurements within this ROI ("ROI-based method").[24] The fourth type, often referred to as "model-based resampling", involves fitting a model (e.g., diffusion tensor) locally to the observed data. The residuals from the fitted model, along with the original observed data, are then used by data resampling techniques such as wild bootstrap to generate random subsets.[21,25] From each subset, we obtain an estimate of a specific parameter. Across all subsets, we get the distribution of the estimates and thus quantify the uncertainty of the DTI parameter. The first type (multiple acquisition method) makes no assumptions about the noise properties at the cost of requiring multiple acquisitions (for each diffusion gradient in the DTI scenario).[21] The second type (scan-rescan method) assumes that the noise is constant across space, or across the region from which the standard deviation is computed.[23] The third method (ROI-based method) assumes that both the noise and the signal are constant across the ROI.[24] The fourth method (model-based resampling) assumes that the non-constant variance of measured signals can be captured by the chosen model.[21] In this study, we choose the ROI-based method for estimating noise across brain regions in DTI. This is because it does not require repeated acquisitions, and one advantage is that we can compute noise from each individual scan within an imaging session, enabling inter-scan comparisons. Also, this method can be applied to datasets that do not have scan-rescan data available, thus enabling validation of our findings using other datasets. Secondly, by using an individual scan for noise estimation, we can mitigate errors caused by motion and inter-scan misalignment of the brain, which could be problematic when using the multiple acquisition method or the scan-rescan method. Thirdly, we want to avoid using the assumption of the fourth type (model-based resampling).[21]

Up to this point, we have been using the terms "variance" and "noise" interchangeably. In the following text, we use "variance" when referring to the measure of the dispersion of data, and "noise" when referring to the imaging noise such as MRI noise—primarily caused by thermal fluctuations and electrical noise—to avoid misinterpretation.

To gain a better understanding of DTI variance or noise, it is important to characterize the role of physiology, subject compliance, and the interaction between the subject and the scanner. Our approach is driven by two fundamental questions (Fig. 2): Which factors are associated with DTI variance? Where and how does this association manifest? We assess variance of DTI scalars, including fractional anisotropy (FA), axial diffusivity (AD), mean diffusivity (MD), and radial diffusivity (RD), within ROIs, and investigate the associations between the variance and covariates, including baseline age, time from the baseline (referred to as "interval"), motion, sex, and whether it is the first or the second scan within the session, using linear mixed effects models[26].

2 Methods

We use the PreQual[27] pipeline for preprocessing and quality assurance of the DTI data. PreQual is an end-to-end pipeline that applies denoising, inter-scan normalization, susceptibility-induced distortion correction, eddy current-induced distortion correction, inter-volume motion correction, slice-wise signal dropout imputation, and more. PreQual also provides a summary of the data and preprocessing in a PDF report for more efficient quality assurance.

We use data acquired from the neuroimaging substudy of the Baltimore Longitudinal Study of Aging (BLSA).[28,29] The BLSA is an extensive, ongoing research project that began in 1958, enrolling healthy volunteers aged 20 years and older to study normal aging through a longitudinal approach by following participants for their entire lives. We consider all subjects with at least one session comprising both T1-weighted (T1w) magnetization-prepared rapid gradient-recalled echo

(MPRAGE) MRI data and DTI data. We exclude 49 DTI images exhibiting one or more of the following characteristics according to their potential impact on subsequent analyses:

(1) The presence of extreme susceptibility-induced distortion, motion artifacts, or eddy currents that resists correction.
(2) The failure of the preprocessed data to be fitted by the tensor model.
(3) An exceptionally low signal-to-noise ratio in the FA and MD images.

The exclusion of these cases results in the dataset depicted in Fig. 3. We identify 1035 subjects (562 F/ 473 M, 22.4 to 94.4 y/o at baseline) with 2751 sessions (1497 F/ 1254 M). Detailed demographic information can be found in the supplementary material. The 2751 sessions of MRI data were acquired by four different scanners, including a 1.5 Tesla Philips Intera scanner (scanner A, 83 sessions) at the Kennedy Krieger Institute (KKI), two 3 Tesla Philips Achieva scanners using the same platform and protocol (scanner B and C, 16 sessions and 59 sessions, respectively) at the KKI, and a 3 Tesla Philips Achieva scanner (scanner D, 2593 sessions) at the National Institute on Aging. Detailed scanner information and protocol is provided in Table 1, which was previously reported[30]. Among the 1035 subjects, 59 switched to a different scanner in follow-up scans. During subsequent sessions, 4 female subjects and 10 male subjects were diagnosed with Alzheimer's disease, while the remaining subjects remained cognitively normal throughout all sessions.

2.1 ROI-Based DTI Variance Estimation

We use a registration-based approach for brain segmentation in the b0 (minimally weighted) volume (Fig. 4). We initiate the process with brain segmentations for the T1w images obtained through manual parcellations provided by the JHU-MNI-ss atlas ("Eve atlas")[31,32] and automated whole-brain segmentation by SLANT[33]. For the Eve atlas, there are three types of parcellations available, each with different regional focus.[31] For SLANT segmentation, labels for 132 regions covering the whole brain are provided.[33] We use the method by Hansen et al.[34] to transfer these labels from T1w to b0 space. After label transferring, we manually review the segmentation to see if the labels align with the anatomical regions.

2.2 Linear Mixed-Effects Model

We use linear mixed-effects models[26] to analyze the association between DTI scalar standard deviation and covariates. (R program, version 4.2.2 [35]; Ubuntu 20.04.5 LTS; R package lme4, version 1.1.31 [36]; R package lmerTest, version 3.1.3 [37].)

We study linear mixed-effects models of the form:

$$\sigma_{ijkl} \sim \mathrm{Age}_{\mathrm{baseline},i} + \mathrm{Age}_{\mathrm{interval},ij} + \mathrm{Motion}_{ijl} + \mathrm{Sex}_i + \mathrm{Rescan}_{ijl} + r_{1,i} + r_{2,k} + \varepsilon_{ijkl} \quad (4)$$

where σ_{ijkl} represents the standard deviation of a DTI scalar (FA, AD, MD, or RD) in a specific brain region of subject i at session j via scanner k in acquisition l, $\mathrm{Age}_{\mathrm{baseline},i}$ (hereafter referred to as "baseline") is the age of subject i at baseline session (unit: decade), $\mathrm{Age}_{\mathrm{interval},ij}$ (hereafter referred to as "interval") is the time between the current session, j, and the baseline session (unit: decade). Motion_{ijl} is a scalar value reflecting the degree of head movement of subject i at session j during acquisition l (calculated based on eddy movement, unit: millimeters)[38], Sex_i is the gender of subject i (0 for female and 1 for male), and Rescan_{ijl} is a binary variable indicating if the acquisition l is the first scan (coded 0) or the rescan (coded 1) of session j. We consider subject and scanner as two random intercepts, respectively denoted by $r_{1,i}$ and $r_{2,k}$. Prior to fitting the models, we standardize the dependent variable σ. The results for the models are based on the standardized σ.

We have a total of 2224 models, derived from the four DTI scalars (FA, AD, MD, or RD), across varying ROIs defined by Eve Type 1 (176 ROIs), Eve Type 2 (130 ROIs), Eve Type 3 (118 ROIs),[31,32] and SLANT (132 ROIs)[33]. Each model starts with a full model, with all fixed effects and random effects, followed by an implementation of backward model selection.[37] The p-values for the fixed-effect terms are calculated based on the associated F tests.[37] To account for multiple comparisons, we adjust the p-values across the pairs of DTI scalar and ROI for a false discovery rate (FDR) of 0.05 using the Benjamini-Hochberg method.[39]

3 Results

The magnitude and direction of the effects of each covariate on DTI variance exhibit heterogeneous patterns across ROIs (Fig. 5). Specifically, interval is positively related to FA variance in ROIs such as the cuneus, middle occipital gyrus, superior occipital gyrus, medulla, precuneus white matter, but negatively related in ROIs such as the caudate nucleus, posterior thalamic radiation, and superior fronto-occipital fasciculus (Table 2). Males have higher FA variance in the right putamen, thalamus, body of corpus callosum, and cingulum (cingulate gyrus), but lower FA variance in the middle frontal gyrus (Table 3). In the right inferior temporal gyrus, an increase of 1 millimeter in motion is associated with an increase of 2.211 standard deviations in the z-scored standard deviation (σ) of FA values ($\beta = 2.211, p \ll 0.001$). Interestingly and counterintuitively, in several ROIs, including the medulla, middle occipital white matter, cingulum (cingulate gyrus), an increase in motion is linked with a decrease in FA variance (Table 2).

In the lateral fronto-orbital gyrus, left insular, gyrus rectus, and inferior occipital gyrus, both motion and interval exhibit a positive association with FA variance (Table 2). In the left caudate nucleus, and right posterior thalamic radiation, they both show a negative association with FA variance (Table 2). In many other ROIs, such as the cuneus, lingual white matter, and middle occipital white matter, motion is negatively related to FA variance while interval is positively related (Table 2). Results from the left ROI closely align with those from the corresponding right ROI (Table 2). There are some ROIs where interval is significantly ($p \ll 0.001$) associated with FA variance, while motion either gets removed during the model selection or shows weak associations ($p \geq 0.05$) (Fig. 5).

On data extracted from ROIs defined by SLANT segmentation, which has different regional focus and delineation than Eve type-1 segmentation, the aforementioned patterns of effects can also be observed (Fig. 6, Fig. 7). For instance, in the left cerebellum exterior, both motion and interval are positively associated with FA variance, with motion's coefficient ($\beta = 0.960, p \ll 0.001$) higher than that of interval ($\beta = 0.485, p \ll 0.001$). This parallels the relationship observed between the motion and interval coefficients in the left cerebellum defined by the Eve type-1 segmentation ($\beta = 0.993, p \ll 0.001$ for motion; $\beta = 0.471, p \ll 0.001$ for interval). Similarly, in ROIs such as the right cuneus, left precuneus, right superior occipital gyrus, and right middle occipital gyrus, motion shows a negative association with FA variance, while interval shows a positive association. The exact coefficients and p-values are in the supplementary materials, where we also provide the results from other pairs of DTI scalar and segmentation method.

In the model selection process, we observe that the rescan and the motion terms appear mutually exclusive, with only one preserved post-selection in most models. This pattern is echoed in the heatmaps of coefficients (Fig. 5, Fig. 6), where the cell in either the motion or the rescan column is colored grey. This hints at a correlation between rescan and motion. Supporting this observation, we detect an increase in head motion in the rescan of DTI acquired right after the first

scan of DTI in the same session (mean shift $\Delta\mu = 0.045$ millimeters per volume, relative mean shift $\Delta\mu/\mu = 17.0\%$, coefficient of determination $R^2 = 0.065$).

To examine the differences between scanners, we measure the signal-to-noise ratio (SNR) of the FA images based on ROIs. There are differences between scanners regarding SNR across ROIs (Fig. 8). To validate our findings, we include two additional datasets, ADNI[40] and BIOCARD[41]. We exclude data points from subjects with cognitive impairment and use the remaining 1808 subjects from the three datasets for the experiment. Detailed information about the data is included in the supplementary materials. We reproduce the experiments of the linear mixed-effects models, except that the rescan term in equation (4) is omitted because the definitions of rescan in ADNI and BIOCARD differ from the definition of rescan in BLSA. The coefficients and p-values, visualized in Fig. 9, show a similar pattern to those derived from the BLSA, although the effect sizes differ, and the hierarchical clustering differs partly due to the omission of one covariate.

4 Discussion

While many studies have estimated and shown the spatial variability of DTI variance (or noise),[14,15,42] we characterize how DTI variance is associated with physiological and behavioral factors across brain regions. We answer the questions: Which factors are associated with DTI variance? Where and how does this association manifest? We found region-specific and bidirectional effects of covariates—including interval (which captures the within-individual longitudinal change over time), motion, and sex—on FA variance across brain regions. For instance, FA variance is positively associated with interval in cuneus, but negatively associated in caudate nucleus. Long-standing research has demonstrated that there is a decline in white matter microstructure with aging,[43–48] with the consensus being that frontal and parietal areas are particularly vulnerable and the occipital and motor areas are mostly preserved. The frontal lobe exhibits the most pronounced decline, with FA declining by approximately 3% per decade starting at ~35 years of age.[49] Although our study focuses on the standard deviation of FA, our results converge with these prior research studies as we have shown high sensitivity to aging in the frontal, parietal, and temporal areas. While it is unclear what mechanisms are driving the changes in these areas, potential culprits include the change of uniformity of fiber orientations and fiber density.[50–52]

Previous studies[53–55] have shown differences in FA between genders across brain regions. Oh et al. found that males have significantly higher FA values in global corpus callosum structure areas, while they exhibit lower FA values than females in the partial areas of the rostrum, genu, and splenium.[53] Menzler et al. found that males show higher FA values in the thalamus, corpus callosum and cingulum.[55] Most of these regions previously identified in the literature also show significant ($p \ll 0.001$) associations between FA variance and sex in our study. While previous studies have reported changes in mean FA values, we offer a different perspective by depicting the variance of FA values.

The negative association observed between motion and FA variance in multiple regions, while counterintuitive, is not unreasonable. One might naturally expect that as motion increases, the uncertainty (reflected as variance) in the image should increase, given that motion leads to lower image quality, signal-to-noise ratio, and artifacts that can mislead image interpretation.[56,57] However, the images we use for analysis have undergone motion correction during preprocessing. Although in practice, motion artifacts cannot be fully eliminated from the image, the recorded motion value doesn't reflect the motion's impact in the image after preprocessing. Instead, it reflects the subject's motion during image acquisition. A higher motion value does not necessarily

correspond to a noisier image post-preprocessing. Furthermore, Zeng et al. found that head motion during brain imaging is not merely a technical artifact but a reflection of a neurobiological trait. Specifically, individuals with stronger distant connectivity in the default network could consistently refrain from moving and such "head motion tendency" remains consistent within individuals.[58] These points, taken together, provide explanations from image processing and biological perspectives, respectively, for why FA variance can decrease as motion increases.

This study underscores the significance of heteroscedasticity in diffusion-weighted MRI mega-analyses and provides a relatively straightforward approach to addressing this issue. Despite the longstanding recognition of heteroscedasticity in statistical analyses,[59–62] its application to diffusion-weighted MRI is still in its nascent stages. Recent advancements have started to bridge this gap, with emerging studies illustrating the importance of accounting for heteroscedasticity in MRI data.[63–65] These pioneering efforts are pivotal, yet they remain underutilized in the broader research community. To enhance the precision and reliability of findings in mega-analyses, it is imperative to disseminate these methodologies more widely and integrate them into commonly used analytical tools. Future research should focus on developing and employing increasingly sophisticated techniques to model and understand heteroscedasticity, thereby improving the robustness of statistical assessments in large-scale neuroimaging studies.

4.1 Limitations of Current Study

First, this study relies on a registration-based method for brain segmentation in the b0 space. Despite rigorous quality assurance, the labels for each brain region may not correspond flawlessly with the true anatomical regions. Consequently, the standard deviation of DTI scalars extracted from each region combines both voxel-wise modeling factors and image analysis factors from neighboring regions. Second, we used backward model selection for the fixed-effect terms of the linear mixed-effects models. Such method can be unstable according to Breiman et. al.[66] Third, we use the variance of DTI scalar values in the ROI as a proxy for measuring noise. This is not an ideal proxy, because signal intensities may not be homogeneously distributed within each ROI, and the ROI-based variance captures not only the image noise but also the spatial variability in voxel intensity due to microstructural variations. This makes it suboptimal to reflect DTI noise. Fourth, the motion value used in this study is based on movement calculated by FSL's eddy[38], which approximates true head motion. Additional sensors or motion tracking sequences might be necessary to quantify head motion during scanning more accurately.

5 Conclusion

The notion of harnessing variance to enhance the reliability of analysis is universally applicable. Having a better understanding of variance is pivotal in mega-analyses, where heteroscedasticity is often an inherent challenge. Our study illuminates the complex and heterogeneous effects of covariates including baseline age, interval, motion, sex, and rescan on DTI variance across ROIs. More comprehensive efforts are required to fully characterize the variance. In the meantime, we encourage researchers to consider models of heteroscedasticity in their analyses and to include their estimates of variance when sharing data. As highlighted in the introduction, the application of the whitening matrix, constructed using the variance of the data, significantly reduces statistical errors. We anticipate that more sophisticated methods can further unlock the potential benefits derived from a nuanced understanding of variance, thereby bolstering the accuracy and reliability of future research.

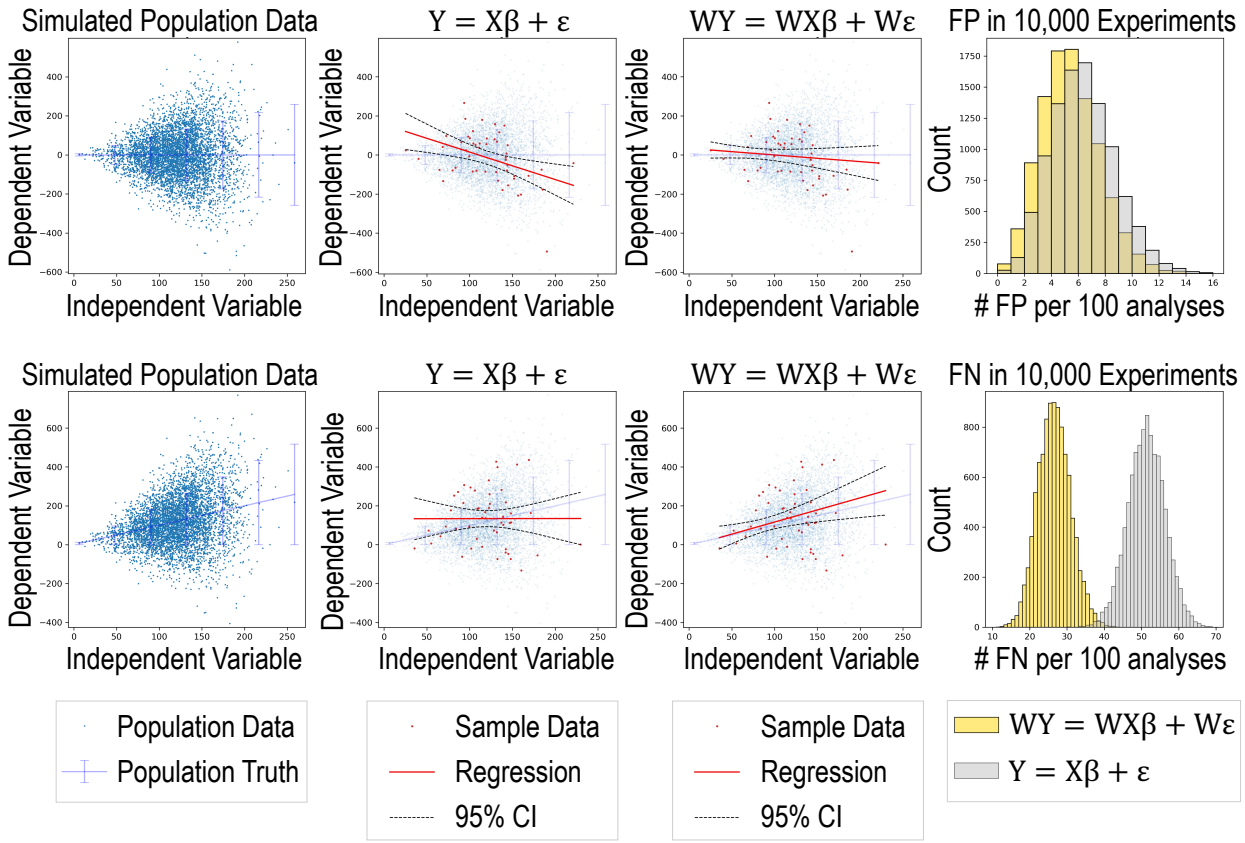


Fig. 1 Simulation shows that applying the whitening matrix to the standard linear regression equation reduces the number of false positives (FP) and false negatives (FN) under heteroskedasticity. In the top row, the population truth has zero slope. In data sampled from the synthetic population data, ordinary least square (OLS) regression using the standard equation generates FP, while the solution with whitening, W, does not falsely reject the null hypothesis (the horizontal line). After 10,000 experiments, the FP frequency is lower with whitening, centering at 5 per 100. In the second row, the population truth has a positive slope. In data sampled from the synthetic population data, OLS regression using the standard equation generates FN, while the solution with whitening, W, does not. After 10,000 experiments, the FN frequency with whitening is half that of the one without whitening.

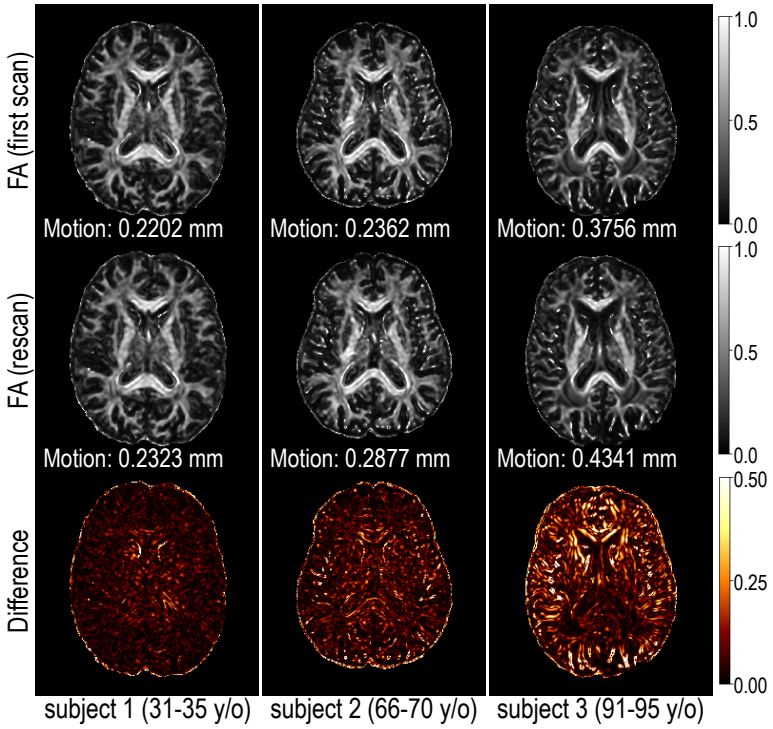


Fig. 2 We observe that the noise (approximated by the difference between the scan and rescan acquired within the same imaging session) in DTI scalar images, such as fractional anisotropy (FA) images, generally increase with age. (Subjects' ages are grouped into five-year bins to respect privacy.) But motion is also considered to increase with age.[67,68] We would like to know: Which factor is associated with DTI variance? Where and how does this association manifest?

	Female	Male
# Subjects	562	473
Age at baseline		
Range	24.2 - 94.4	22.4 - 92.4
Mean (SD)	69.8 (12.5)	71.5 (13.2)
# Sessions	1497	1254
w/ rescan DTI	1471	1207
w/o rescan DTI	26	47
# Subjects by the total number of sessions		
1	205	172
2	112	94
3	86	77
4	69	60
5	51	33
6	17	20
7	11	10
8	6	1
9	2	3
10	1	0
11	0	3
12	2	0

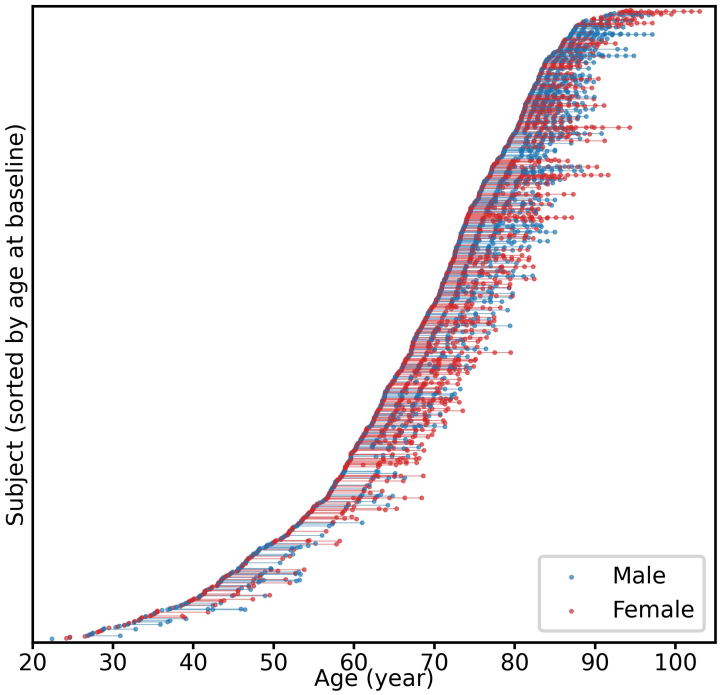
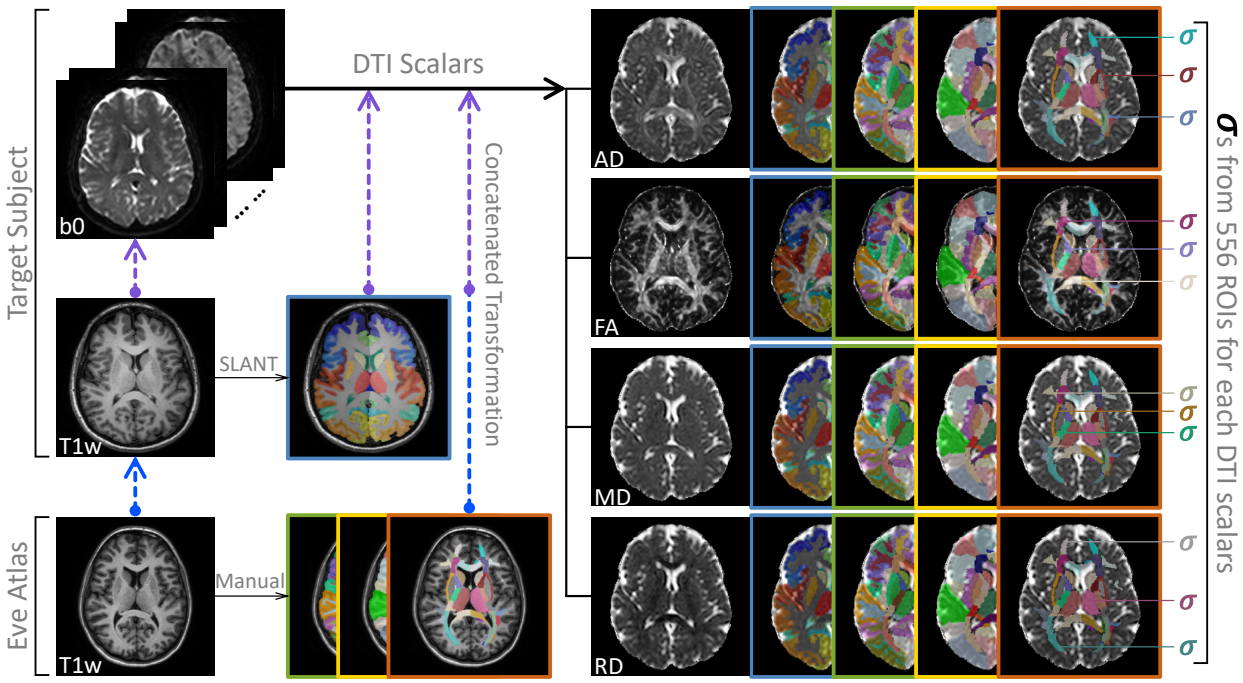


Fig. 3 The BLSA dataset we use has a slight imbalance between the number of females and males, but it is well-matched and appropriate for our research objectives in other aspects: i) the age ranges of females and males align closely; ii) rescan DTI data were acquired in most sessions, enabling inter-scan comparisons; iii) the distributions of sessions of females and males align closely.

Parameter	Scanner A	Scanners B/C	Scanner D
MPRAGE			
Head coil	Philips 8-ch	Philips 8-ch	Philips 8-ch
Scan time (mins:secs)	3:58	10:52	10:52
Slice thickness (mm)	1.5	1.2	1.2
Number of slices	124	170	170
Flip angle (deg)	8	8	8
TR/TE (ms)	6.6/3.3	6.8/3.1	6.5/3.1
Field of view (mm)	240x240	256x240	256x240
Acquisition matrix	208x208	256x240	256x240
Reconstruction matrix	256x256	256x256	256x256
Reconstructed voxel size (mm)	0.94x0.94	1.00x1.00	1.00x1.00
DTI			
Head coil	Philips 8-ch	Philips 8-ch	Philips 8-ch
Scan time (mins:secs)	3:56	3:58	4:20
Number of gradients	30	32	32
Number of b0 images	1	1	1
Max b-factor (s/mm²)	700	700	700
Number of signal averages (NSA)	1	1	1
Diffusion gradient timing DELTA/delta (ms)	39.2/15.1	36.3/16	36.3/13.5
Slice thickness (mm)	2.5	2.2	2.2
Number of slices	50	65	70
Flip angle (deg)	90	90	90
TR/TE (ms)	6210/80	6801/75	7454/75
Field of view (mm)	240x240	212x212	260x260
Acquisition matrix	96x96	96x95	116x115
Reconstruction matrix	256x256	256x256	320x320
Reconstructed voxel size (mm)	0.94x0.94	0.83x0.83	0.81x0.81

Table 1 Protocol for the T1w MPRAGE scan and DTI scan.


Target Subject
b0
T1w
SLANT
Eve Atlas
T1w
Manual
DTI Scalars
Concatenated Transformation
AD
FA
MD
RD
σ
σ
σ

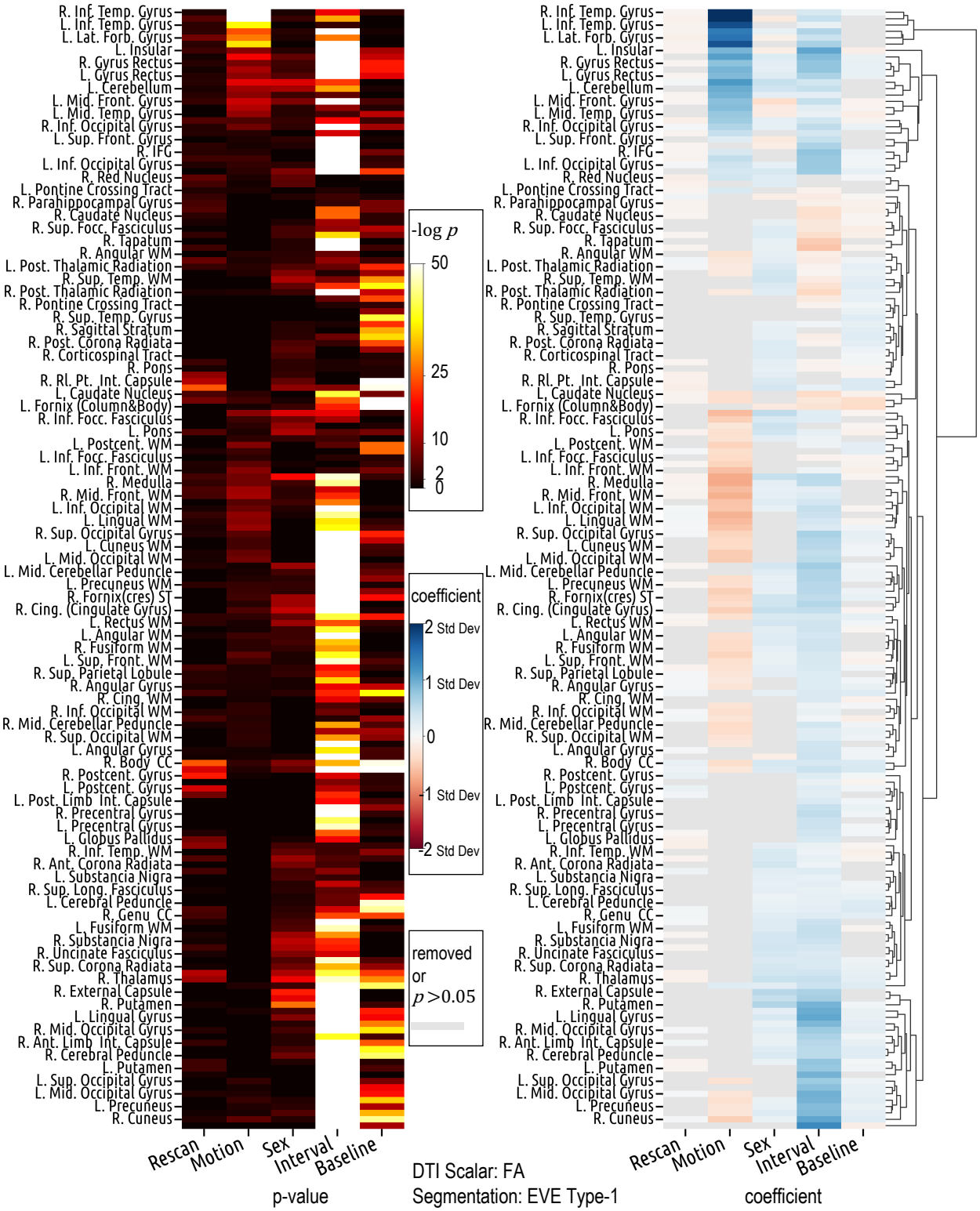


Fig. 5 Covariate effects on FA standard deviation (standardized) are region-specific. Motion and interval exhibit opposite effect directions in many ROIs. Gender differences exist in multiple ROIs. Counterintuitively, motion is negatively related to FA standard deviation in many ROIs. The lookup table for the abbreviation of ROI name is in the Supplementary Materials.

Covariate	$Age_{interval}$				Motion			
ROI	Left		Right		Left		Right	
	β	p-value	β	p-value	β	p-value	β	p-value
Caudate Nucleus	-0.383	4.0e-41	-0.317	2.4e-26	-0.337	2.1e-03	NA	NA
Posterior Thalamic Radiation	-0.091	1.1e-07	-0.414	7.7e-94	-0.186	5.0e-03	-0.188	1.8e-02
Superior Fronto-occipital Fasciculus	-0.358	2.5e-35	-0.220	6.4e-11	NA	NA	NA	NA
Middle Occipital Gyrus	0.957	7.2e-218	0.726	1.1e-173	-0.266	2.3e-02	NA	NA
Lateral Fronto-orbital Gyrus	0.427	5.5e-28	0.457	1.4e-31	1.481	3.6e-28	2.011	2.3e-51
Insular	1.055	1.7e-132	1.276	1.5e-208	0.935	5.8e-10	NA	NA
Gyrus Rectus	0.636	4.6e-59	0.787	1.2e-88	0.887	2.1e-10	0.818	5.2e-09
Inferior Occipital Gyrus	0.746	1.6e-109	0.505	2.6e-62	0.422	9.3e-4	0.597	3.4e-08
Body of Corpus Callosum	0.439	7.6e-60	0.336	2.6e-32	-0.242	2.4e-02	-0.357	1.1e-03
Inferior Cerebellar Peduncle	0.560	2.9e-55	0.605	1.7e-61	-0.469	2.6e-04	-0.346	1.1e-02
Cingulum (Cingulate Gyrus)	0.247	1.9e-17	0.498	3.0e-57	-0.630	2.0e-10	-0.476	1.2e-05
Cuneus	0.951	1.1e-288	1.061	≤ 2.3e-308	-0.318	3.2e-04	-0.471	5.4e-07
Cuneus WM	0.555	2.0e-96	0.585	1.4e-95	-0.383	4.0e-05	-0.393	7.9e-05
Fornix(cres) Stria Terminalis	0.340	5.4e-41	0.509	9.0e-60	-0.313	1.4e-03	-0.433	9.9e-05
Middle Frontal WM	0.203	1.2e-15	0.275	5.4e-22	-0.623	8.0e-12	-0.716	2.5e-12
Superior Frontal WM	0.341	2.4e-48	0.338	3.5e-39	-0.316	1.6e-04	-0.440	7.3e-07
Lateral Fronto-orbital WM	0.372	7.3e-28	0.410	1.0e-33	-0.566	1.7e-06	-0.376	2.5e-03
Lingual WM	0.366	2.3e-36	0.426	6.1e-45	-0.640	6.4e-10	-0.695	1.5e-10
Medulla	0.537	1.9e-48	0.519	3.9e-45	-0.744	1.6e-08	-0.750	1.6e-08
Superior Occipital Gyrus	0.734	4.7e-163	0.644	1.3e-134	-0.311	1.1e-03	-0.445	1.7e-06
Middle Occipital WM	0.544	3.5e-102	0.308	2.3e-38	-0.501	2.2e-08	-0.564	3.5e-11
Precuneus WM	0.549	3.0e-96	0.499	1.7e-76	-0.341	6.2e-04	-0.508	1.4e-07

Table 2 Covariate effects on FA standard deviation (standardized) in selected ROIs from Eve Type 1 atlas.

Covariate	Sex			
ROI	Left		Right	
	β	p-value	β	p-value
Putamen	NA	NA	0.602	2.6e-27
Thalamus	0.388	5.9e-13	0.450	4.3e-18
Body of Corpus Callosum	0.287	4.1e-07	0.353	1.3e-09
Cingulum (Cingulate Gyrus)	0.509	1.7e-16	0.471	1.4e-14
Middle Frontal Gyrus	-0.370	1.8e-09	-0.148	1.9e-02

Table 3 Effects of sex on FA standard deviation (standardized) in selected ROIs from Eve Type 1 atlas.

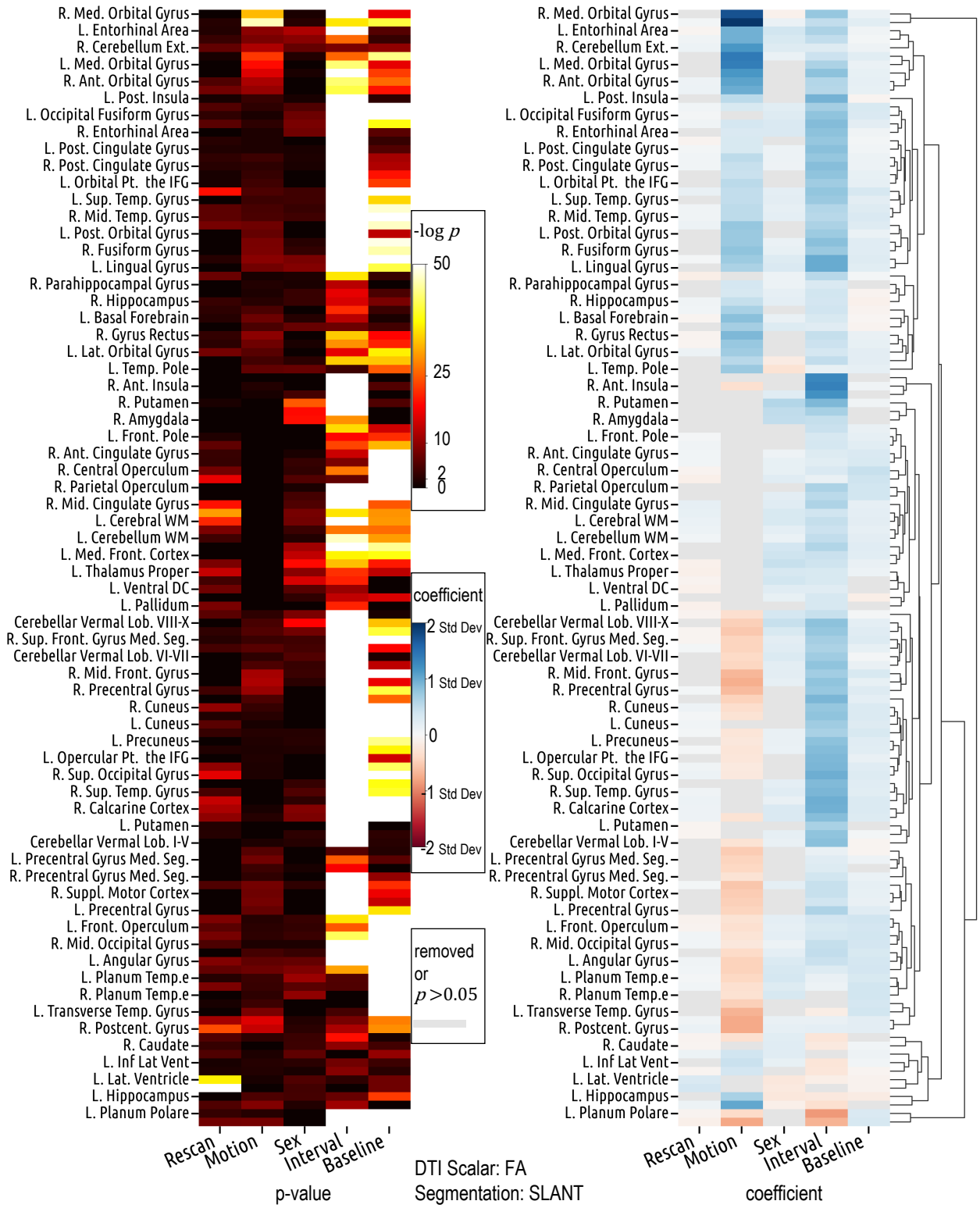


Fig. 6 The region-specific and bidirectional patterns of covariate effects are similarly observed in the results derived from SLANT segmentation, despite its differing definitions and delineations of ROIs compared to Eve type-1 segmentation (Fig. 5).

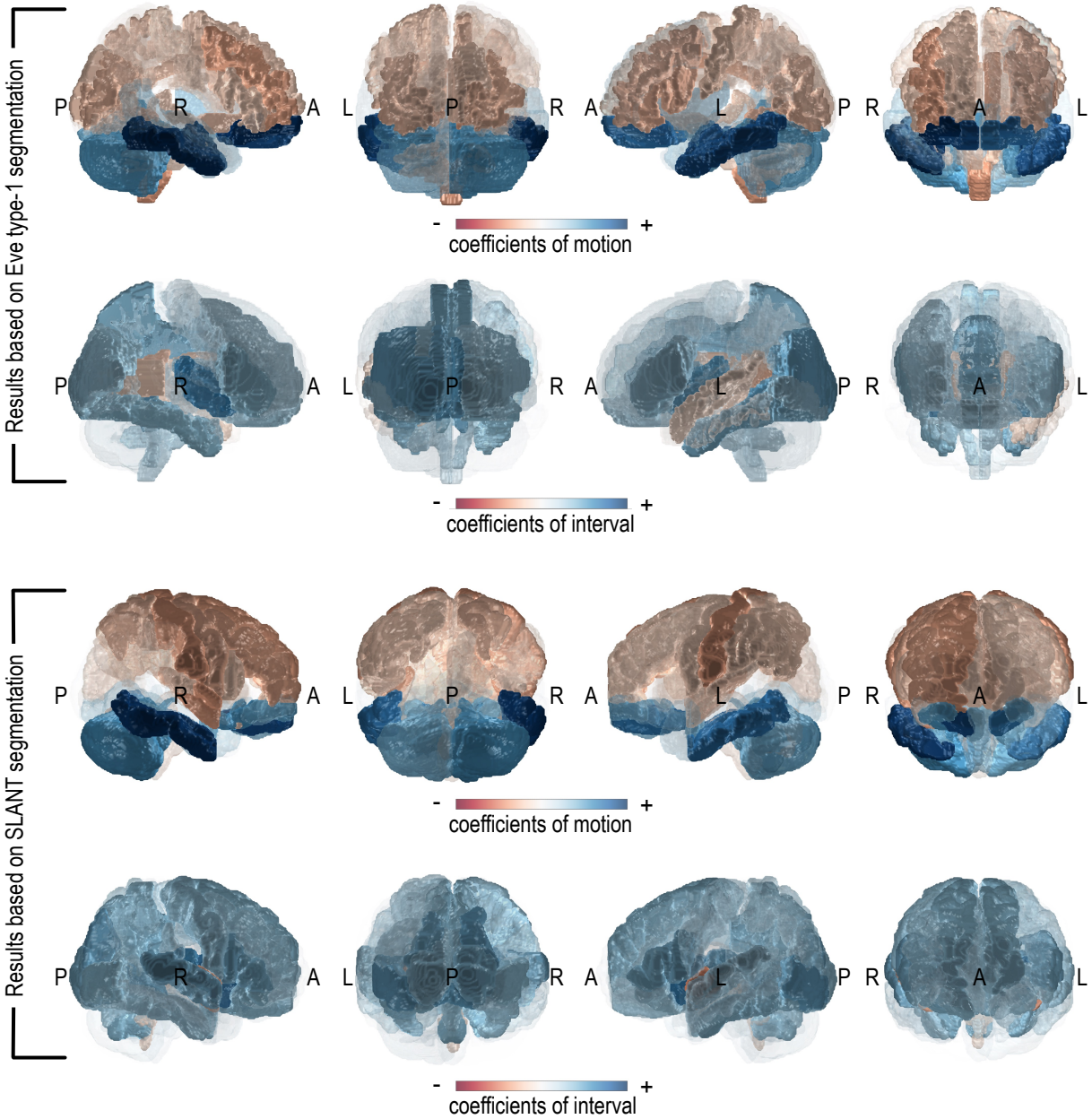


Fig. 7 Despite the different definitions and delineations of ROIs between Eve type-1 and SLANT segmentations, results based on the two segmentation methods are largely similar (comparable regions are colored similarly) and both show that the effects of motion and interval on FA variance vary across ROIs.

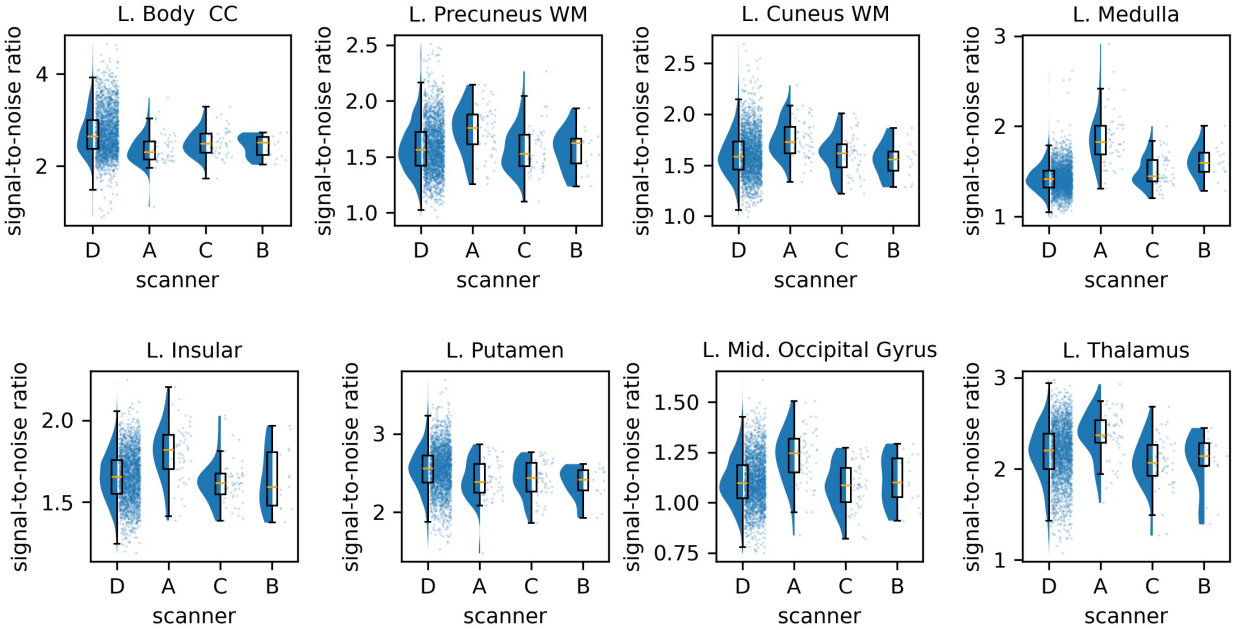


Fig. 8 Signal-to-noise ratio (SNR) of the FA images across 8 Eve type-1 atlas-defined ROIs, including white matter (WM) regions (body of the corpus callosum, precuneus WM, cuneus WM), gray matter regions (insular, putamen, middle occipital gyrus, thalamus), and mixed regions (medulla), in 4 different scanners of BLSA (where scanner A is the 1.5 Tesla Philips Intera scanner, and scanners B/C/D are the 3 Tesla Philips Achieva scanners). "L." stands for the left hemisphere of the brain.

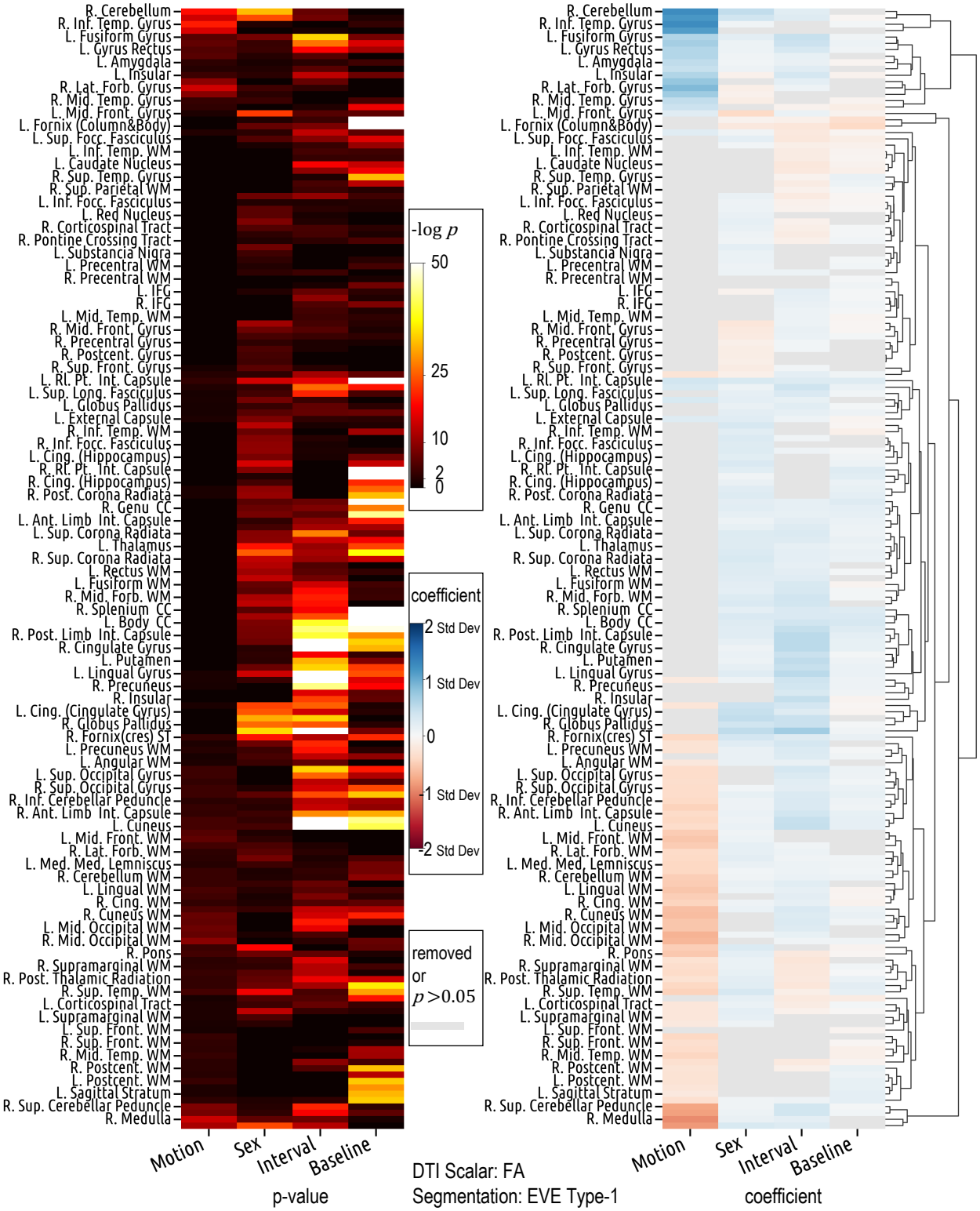


Fig. 9 To assess the generalizability of our findings, we include two additional datasets, ADNI and BIOCARD and fit the linear mixed-effects models. The coefficients and p-values show similar patterns of those from BLSA alone, despite that the effect sizes and the hierarchical clustering are different, partly due to the omission of the rescan covariate.

Disclosures

No conflict of interest.

Code, Data, and Materials Availability

The dataset used in this study can be accessed upon approval from the Baltimore Longitudinal Study of Aging team at https://www.blsa.nih.gov/. The code for the experiments can be found on GitHub at https://github.com/MASILab/Variance-Aging-Diffusion.

Acknowledgements

This work was supported in part by the National Institute of Health through grants 1R01EB017230-01A1 (Bennett A. Landman), 5-K01-EB032898-02 (Kurt Schilling), 1U24AG074855-01 (Timothy J. Hohman), and ViSE/VICTR VR3029 and by the Intramural Research Program of the National Institute on Aging, NIH. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University, Nashville, TN. We appreciate the National Institute of Health S10 Shared Instrumentation grant 1S10OD020154-01 (Smith), and grant 1S10OD023680-01 (Vanderbilt's High-Performance Computer Cluster for Biomedical Research). The Vanderbilt Institute for Clinical and Translational Research (VICTR) is funded by the National Center for Advancing Translational Sciences (NCATS) Clinical Translational Science Award (CTSA) Program, Award Number 5UL1TR002243-03. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.
The BLSA is supported by the Intramural Research Program, National Institute on Aging, NIH. The BIOCARD study is supported by a grant from the National Institute on Aging (NIA): U19-AG03365. The BIOCARD Study consists of 7 Cores and 2 projects with the following members: (1) The Administrative Core (Marilyn Albert, Corinne Pettigrew, Barbara Rodzon); (2) the Clinical Core (Marilyn Albert, Anja Soldan, Rebecca Gottesman, Corinne Pettigrew, Leonie Farrington, Maura Grega, Gay Rudow, Rostislav Brichko, Scott Rudow, Jules Giles, Ned Sacktor); (3) the Imaging Core (Michael Miller, Susumu Mori, Anthony Kolasny, Hanzhang Lu, Kenichi Oishi, Tilak Ratnanather, Peter vanZijl, Laurent Younes); (4) the Biospecimen Core (Abhay Moghekar, Jacqueline Darrow, Alexandria Lewis, Richard O'Brien); (5) the Informatics Core (Roberta Scherer, Ann Ervin, David Shade, Jennifer Jones, Hamadou Coulibaly, Kathy Moser, Courtney Potter); the (6) Biostatistics Core (Mei-Cheng Wang, Yuxin Zhu, Jiangxia Wang); (7) the Neuropathology Core (Juan Troncoso, David Nauen, Olga Pletnikova, Karen Fisher); (8) Project 1 (Paul Worley, Jeremy Walston, Mei-Fang Xiao), and (9) Project 2 (Mei-Cheng Wang, Yifei Sun, Yanxun Xu. Data collection and sharing for ADNI were supported by National Institutes of Health Grant U01-AG024904 and Department of Defense (award number W81XWH-12-2-0012). ADNI is also funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.;

Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Ethics approval

IRB of Vanderbilt University waived ethical approval for de-identified access of the human subject data. (IRB #172072, PI: Bennett A. Landman)

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author Chenyu Gao used ChatGPT to check grammar and improve readability of the original draft. After using this tool, the author reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

1. Fan J, Han F, Liu H. Challenges of Big Data analysis. *Natl Sci Rev*. 2014;1(2):293-314. doi:10.1093/nsr/nwt032
2. Monti M. Statistical Analysis of fMRI Time-Series: A Critical Review of the GLM Approach. *Front Hum Neurosci*. 2011;5. https://www.frontiersin.org/articles/10.3389/fnhum.2011.00028
3. Kessy A, Lewin A, Strimmer K. Optimal Whitening and Decorrelation. *Am Stat*. 2018;72(4):309-314. doi:10.1080/00031305.2016.1277159
4. Pierpaoli C, Jezzard P, Basser PJ, Barnett A, Di Chiro G. Diffusion tensor MR imaging of the human brain. *Radiology*. 1996;201(3):637-648. doi:10.1148/radiology.201.3.8939209
5. Beaulieu C. The basis of anisotropic water diffusion in the nervous system – a technical review. *NMR Biomed*. 2002;15(7-8):435-455. doi:https://doi.org/10.1002/nbm.782
6. Basser PJ, Mattiello J, LeBihan D. MR diffusion tensor spectroscopy and imaging. *Biophys J*. 1994;66(1):259-267. doi:https://doi.org/10.1016/S0006-3495(94)80775-1
7. Merboldt KD, Hanicke W, Frahm J. Self-diffusion NMR imaging using stimulated echoes. *Journal of Magnetic Resonance (1969)*. 1985;64(3):479-486. doi:https://doi.org/10.1016/0022-2364(85)90111-8
8. D G Taylor, M C Bushell. The spatial mapping of translational diffusion coefficients by the NMR imaging technique. *Phys Med Biol*. 1985;30(4):345. doi:10.1088/0031-9155/30/4/009
9. Le Bihan D, Breton E, Lallemand D, Grenier P, Cabanis E, Laval-Jeantet M. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology*. 1986;161(2):401-407. doi:10.1148/radiology.161.2.3763909
10. Soares JM, Marques P, Alves V, Sousa N. A hitchhiker's guide to diffusion tensor imaging. *Front Neurosci*. 2013;(7 MAR). doi:10.3389/fnins.2013.00031
11. CASCIO CJ, GERIG G, PIVEN J. Diffusion Tensor Imaging: Application to the Study of the Developing Brain. *J Am Acad Child Adolesc Psychiatry*. 2007;46(2):213-223. doi:https://doi.org/10.1097/01.chi.0000246064.93200.e8
12. Metzler-Baddeley C, Hunt S, Jones DK, Leemans A, Aggleton JP, O'Sullivan MJ. Temporal association tracts and the breakdown of episodic memory in mild cognitive impairment. *Neurology*. 2012;79(23):2233. doi:10.1212/WNL.0b013e31827689e8
13. Schilling KG, Archer D, Yeh FC, et al. Aging and white matter microstructure and macrostructure: a longitudinal multi-site diffusion MRI study of 1218 participants. *Brain Struct Funct*. 2022;227(6):2111-2125. doi:10.1007/s00429-022-02503-z
14. Landman BA, Bazin PL, Prince JL. Estimation and application of spatially variable noise fields in diffusion tensor imaging. *Magn Reson Imaging*. 2009;27(6):741-751. doi:https://doi.org/10.1016/j.mri.2009.01.001
15. Tabelow K, Voss HU, Polzehl J. Local estimation of the noise level in MRI using structural adaptation. *Med Image Anal*. 2015;20(1):76-86. doi:https://doi.org/10.1016/j.media.2014.10.008
16. Landman B, Bazin PL, Prince J. Diffusion Tensor Estimation by Maximizing Rician Likelihood. In: *2007 IEEE 11th International Conference on Computer Vision*. ; 2007:1-8. doi:10.1109/ICCV.2007.4409140
17. Chang LC, Jones DK, Pierpaoli C. RESTORE: Robust estimation of tensors by outlier rejection. *Magn Reson Med*. 2005;53(5):1088-1095. doi:https://doi.org/10.1002/mrm.20426

18. Jones DK, Basser PJ. “Squashing peanuts and smashing pumpkins”: How noise distorts diffusion-weighted MR data. *Magn Reson Med*. 2004;52(5):979-993. doi:https://doi.org/10.1002/mrm.20283
19. Bastin ME, Armitage PA, Marshall I. A theoretical study of the effect of experimental noise on the measurement of anisotropy in diffusion imaging. *Magn Reson Imaging*. 1998;16(7):773-785. doi:https://doi.org/10.1016/S0730-725X(98)00098-8
20. Farrell JAD, Landman BA, Jones CK, et al. Effects of signal-to-noise ratio on the accuracy and reproducibility of diffusion tensor imaging–derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5T. *Journal of Magnetic Resonance Imaging*. 2007;26(3):756-767. doi:https://doi.org/10.1002/jmri.21053
21. Chung S, Lu Y, Henry RG. Comparison of bootstrap approaches for estimation of uncertainties of DTI parameters. *Neuroimage*. 2006;33(2):531-541. doi:https://doi.org/10.1016/j.neuroimage.2006.07.001
22. Covell MM, Hearshen DO, Carson PL, et al. Automated analysis of multiple performance characteristics in magnetic resonance imaging systems. *Med Phys*. 1986;13(6):815-823. doi:https://doi.org/10.1118/1.595804
23. Landman BA, Bazin PL, Smith SA, Prince JL. Robust estimation of spatially variable noise fields. *Magn Reson Med*. 2009;62(2):500-509. doi:https://doi.org/10.1002/mrm.22013
24. Kaufman L, Kramer DM, Crooks LE, Ortendahl DA. Measuring signal-to-noise ratios in MR imaging. *Radiology*. 1989;173(1):265-267. doi:10.1148/radiology.173.1.2781018
25. Whitcher B, Tuch DS, Wisco JJ, Sorensen AG, Wang L. Using the wild bootstrap to quantify uncertainty in diffusion tensor imaging. *Hum Brain Mapp*. 2008;29(3):346-362. doi:https://doi.org/10.1002/hbm.20395
26. Laird NM, Ware JH. Random-Effects Models for Longitudinal Data. *Biometrics*. 1982;38(4):963-974. doi:10.2307/2529876
27. Cai LY, Yang Q, Hansen CB, et al. PreQual: An automated pipeline for integrated preprocessing and quality assurance of diffusion weighted MRI images. *Magn Reson Med*. 2021;86(1):456-470. doi:https://doi.org/10.1002/mrm.28678
28. Shock NW. *Normal Human Aging: The Baltimore Longitudinal Study of Aging*. US Department of Health and Human Services, Public Health Service, National …; 1984.
29. Ferrucci L. The Baltimore Longitudinal Study of Aging (BLSA): a 50-year-long journey and plans for the future. *J Gerontol A Biol Sci Med Sci*. 2008;63(12):1416-1419.
30. Venkatraman VK, Gonzalez CE, Landman B, et al. Region of interest correction factors improve reliability of diffusion imaging measures within and across scanners and field strengths. *Neuroimage*. 2015;119:406-416. doi:https://doi.org/10.1016/j.neuroimage.2015.06.078
31. Mori S, Oishi K, Jiang H, et al. Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. *Neuroimage*. 2008;40(2):570-582. doi:10.1016/J.NEUROIMAGE.2007.12.035
32. Oishi K, Faria A, Jiang H, et al. Atlas-based whole brain white matter analysis using large deformation diffeomorphic metric mapping: Application to normal elderly and Alzheimer’s disease participants. *Neuroimage*. 2009;46(2):486-499. doi:10.1016/J.NEUROIMAGE.2009.01.002

33. Huo Y, Xu Z, Xiong Y, et al. 3D whole brain segmentation using spatially localized atlas network tiles. *Neuroimage*. 2019;194:105-119. doi:https://doi.org/10.1016/j.neuroimage.2019.03.041
34. Hansen CB, Yang Q, Lyu I, et al. Pandora: 4-D White Matter Bundle Population-Based Atlases Derived from Diffusion MRI Fiber Tractography. *Neuroinformatics*. 2021;19(3):447-460. doi:10.1007/s12021-020-09497-1
35. R Core Team. R: A Language and Environment for Statistical Computing. Published online 2022. https://www.R-project.org
36. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw*. 2015;67(1):1-48. doi:10.18637/jss.v067.i01
37. Kuznetsova A, Brockhoff PB, B. Christensen RH. lmerTest Package: Tests in Linear Mixed Effects Models. *J Stat Softw*. 2017;82(13):1-26. doi:10.18637/jss.v082.i13
38. Andersson JLR, Sotiropoulos SN. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage*. 2016;125:1063-1078. doi:https://doi.org/10.1016/j.neuroimage.2015.10.019
39. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1995;57(1):289-300. doi:https://doi.org/10.1111/j.2517-6161.1995.tb02031.x
40. Jack Jr. CR, Bernstein MA, Fox NC, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*. 2008;27(4):685-691. doi:https://doi.org/10.1002/jmri.21049
41. Sacktor N, Soldan A, Grega M, et al. The BIOCARD Index: A Summary Measure to Predict Onset of Mild Cognitive Impairment (P1.095). *Neurology*. 2017;88(16 Supplement):P1.095. http://n.neurology.org/content/88/16_Supplement/P1.095.abstract
42. Zhu T, Liu X, Connelly PR, Zhong J. An optimized wild bootstrap method for evaluation of measurement uncertainties of DTI-derived parameters in human brain. *Neuroimage*. 2008;40(3):1144-1156. doi:https://doi.org/10.1016/j.neuroimage.2008.01.016
43. Salat DH, Tuch DS, Greve DN, et al. Age-related alterations in white matter microstructure measured by diffusion tensor imaging. *Neurobiol Aging*. 2005;26(8):1215-1227. doi:https://doi.org/10.1016/j.neurobiolaging.2004.09.017
44. Davis SW, Dennis NA, Buchler NG, White LE, Madden DJ, Cabeza R. Assessing the effects of age on long white matter tracts using diffusion tensor tractography. *Neuroimage*. 2009;46(2):530-541. doi:https://doi.org/10.1016/j.neuroimage.2009.01.068
45. Bennett IJ, Madden DJ, Vaidya CJ, Howard D V, Howard Jr. JH. Age-related differences in multiple measures of white matter integrity: A diffusion tensor imaging study of healthy aging. *Hum Brain Mapp*. 2010;31(3):378-390. doi:https://doi.org/10.1002/hbm.20872
46. Sexton CE, Kalu UG, Filippini N, Mackay CE, Ebmeier KP. A meta-analysis of diffusion tensor imaging in mild cognitive impairment and Alzheimer's disease. *Neurobiol Aging*. 2011;32(12):2322.e5-2322.e18. doi:https://doi.org/10.1016/j.neurobiolaging.2010.05.019
47. Lebel C, Gee M, Camicioli R, Wieler M, Martin W, Beaulieu C. Diffusion tensor imaging of white matter tract evolution over the lifespan. *Neuroimage*. 2012;60(1):340-352. doi:https://doi.org/10.1016/j.neuroimage.2011.11.094

48. de Groot M, Cremers LGM, Ikram MA, et al. White Matter Degeneration with Aging: Longitudinal Diffusion MR Imaging Analysis. *Radiology*. 2015;279(2):532-541. doi:10.1148/radiol.2015150103
49. Grieve SM, Williams LM, Paul RH, Clark CR, Gordon E. Cognitive Aging, Executive Function, and Fractional Anisotropy: A Diffusion Tensor MR Imaging Study. *American Journal of Neuroradiology*. 2007;28(2):226. http://www.ajnr.org/content/28/2/226.abstract
50. MEIER-RUGE W, ULRICH J, BRÜHLMANN M, MEIER E. Age-Related White Matter Atrophy in the Human Brain. *Ann N Y Acad Sci*. 1992;673(1):260-269. doi:https://doi.org/10.1111/j.1749-6632.1992.tb27462.x
51. Bronge L, Bogdanovic N, Wahlund LO. Postmortem MRI and Histopathology of White Matter Changes in Alzheimer Brains: A Quantitative, Comparative Study. *Dement Geriatr Cogn Disord*. 2002;13(4):205-212. doi:10.1159/000057698
52. Stricker NH, Schweinsburg BC, Delano-Wood L, et al. Decreased white matter integrity in late-myelinating fiber pathways in Alzheimer's disease supports retrogenesis. *Neuroimage*. 2009;45(1):10-16. doi:https://doi.org/10.1016/j.neuroimage.2008.11.027
53. Oh JS, Song IC, Lee JS, et al. Tractography-guided statistics (TGIS) in diffusion tensor imaging for the detection of gender difference of fiber integrity in the midsagittal and parasagittal corpora callosa. *Neuroimage*. 2007;36(3):606-616. doi:https://doi.org/10.1016/j.neuroimage.2007.03.020
54. Liu F, Vidarsson L, Winter JD, Tran H, Kassner A. Sex differences in the human corpus callosum microstructure: A combined T2 myelin-water and diffusion tensor magnetic resonance imaging study. *Brain Res*. 2010;1343:37-45. doi:https://doi.org/10.1016/j.brainres.2010.04.064
55. Menzler K, Belke M, Wehrmann E, et al. Men and women are different: Diffusion tensor imaging reveals sexual dimorphism in the microstructure of the thalamus, corpus callosum and cingulum. *Neuroimage*. 2011;54(4):2557-2562. doi:https://doi.org/10.1016/j.neuroimage.2010.11.029
56. Yendiki A, Koldewyn K, Kakunoori S, Kanwisher N, Fischl B. Spurious group differences due to head motion in a diffusion MRI study. *Neuroimage*. 2014;88:79-90. doi:10.1016/J.NEUROIMAGE.2013.11.027
57. Baum GL, Roalf DR, Cook PA, et al. The impact of in-scanner head motion on structural connectivity derived from diffusion MRI. *Neuroimage*. 2018;173:275-286. doi:10.1016/j.neuroimage.2018.02.041
58. Zeng LL, Wang D, Fox MD, et al. Neurobiological basis of head motion in brain imaging. *Proceedings of the National Academy of Sciences*. 2014;111(16):6058-6062. doi:10.1073/pnas.1317424111
59. COOK RD, WEISBERG S. Diagnostics for heteroscedasticity in regression. *Biometrika*. 1983;70(1):1-10. doi:10.1093/biomet/70.1.1
60. Müller HG, Stadtmüller U. Estimation of Heteroscedasticity in Regression Analysis. *The Annals of Statistics*. 1987;15(2):610-625. http://www.jstor.org/stable/2241329
61. Dette H, Munk A. Testing heteroscedasticity in nonparametric regression. *J R Stat Soc Series B Stat Methodol*. 1998;60(4):693-708. doi:https://doi.org/10.1111/1467-9868.00149
62. Long JS, Ervin LH. Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. *Am Stat*. 2000;54(3):217-224. doi:10.1080/00031305.2000.10474549

63. Seiler C, Holmes S. Multivariate Heteroscedasticity Models for Functional Brain Connectivity. *Front Neurosci*. 2017;11. https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2017.00696
64. Zugman A, Harrewijn A, Cardinale EM, et al. Mega-analysis methods in ENIGMA: The experience of the generalized anxiety disorder working group. *Hum Brain Mapp*. 2022;43(1):255-277. doi:https://doi.org/10.1002/hbm.25096
65. Lutti A, Corbin N, Ashburner J, et al. Restoring statistical validity in group analyses of motion-corrupted MRI data. *Hum Brain Mapp*. 2022;43(6):1973-1983. doi:https://doi.org/10.1002/hbm.25767
66. Breiman L. Heuristics of instability and stabilization in model selection. *The Annals of Statistics*. 1996;24(6):2350-2383. doi:10.1214/aos/1032181158
67. Pardoe HR, Kucharsky Hiess R, Kuzniecky R. Motion and morphometry in clinical and nonclinical populations. *Neuroimage*. 2016;135:177-185. doi:https://doi.org/10.1016/j.neuroimage.2016.05.005
68. Hausman HK, Hardcastle C, Kraft JN, et al. The association between head motion during functional magnetic resonance imaging and executive functioning in older adults. *Neuroimage: Reports*. 2022;2(2):100085. doi:https://doi.org/10.1016/j.ynirp.2022.100085

Chenyu Gao is a Ph.D. student in electrical and computer engineering at Vanderbilt University, working with Prof. Bennett Landman on the harmonization of diffusion MRI. His research interest is focused on image processing and computer vision with application to medical image analysis. He received his BS degree in biomedical engineering from Sun Yat-sen University, and his MS degree in biomedical engineering from Johns Hopkins University, working with Prof. Jerry Prince.

Caption List

Fig. 1 Simulation shows that applying the whitening matrix to the standard linear regression equation reduces the number of false positives (FP) and false negatives (FN) under heteroskedasticity. In the top row, the population truth has zero slope. In data sampled from the synthetic population data, ordinary least square (OLS) regression using the standard equation generates FP, while the solution with whitening, W, does not falsely reject the null hypothesis (the horizontal line). After 10,000 experiments, the FP frequency is lower with whitening, centering at 5 per 100. In the second row, the population truth has a positive slope. In data sampled from the synthetic population data, OLS regression using the standard equation generates FN, while the solution with whitening, W, does not. After 10,000 experiments, the FN frequency with whitening is half that of the one without whitening.

Fig. 2 We observe that the noise (approximated by the difference between the scan and rescan acquired within the same imaging session) in DTI scalar images, such as fractional anisotropy (FA) images, generally increase with age. (Subjects' ages are grouped into five-year bins to respect privacy.) But motion is also considered to increase with age.67,68 We would like to know: Which factor is associated with DTI variance? Where and how does this association manifest?

Fig. 3 The BLSA dataset we use has a slight imbalance between the number of females and males, but it is well-matched and appropriate for our research objectives in other aspects: i) the age ranges of females and males align closely; ii) rescan DTI data were acquired in most

sessions, enabling inter-scan comparisons; iii) the distributions of sessions of females and males align closely.

Table 1 Protocol for the T1w MPRAGE scan and DTI scan.

Fig. 4 Brain segmentation labels are obtained using the SLANT segmentation of the target subject's T1w image, and using three types of manual parcellations provided by the Eve atlas. To generate transformation matrices for transferring these labels to DTI scalar images, intra- and inter- modality registrations are performed. Standard deviations of DTI scalars within each ROI are computed.

Fig. 5 Covariate effects on FA standard deviation (standardized) are region-specific. Motion and interval exhibit opposite effect directions in many ROIs. Gender differences exist in multiple ROIs. Counterintuitively, motion is negatively related to FA standard deviation in many ROIs. The lookup table for the abbreviation of ROI name is in the Supplementary Materials.

Table 2 Covariate effects on FA standard deviation (standardized) in selected ROIs from Eve Type 1 atlas.

Table 3 Effects of sex on FA standard deviation (standardized) in selected ROIs from Eve Type 1 atlas.

Fig. 6 The region-specific and bidirectional patterns of covariate effects are similarly observed in the results derived from SLANT segmentation, despite its differing definitions and delineations of ROIs compared to Eve type-1 segmentation (Fig. 5).

Fig. 7 Despite the different definitions and delineations of ROIs between Eve type-1 and SLANT segmentations, results based on the two segmentation methods are largely similar (comparable regions are colored similarly) and both show that the effects of motion and interval on FA variance vary across ROIs.

Fig. 8 Signal-to-noise ratio (SNR) of the FA images across 8 Eve type-1 atlas-defined ROIs, including white matter (WM) regions (body of the corpus callosum, precuneus WM, cuneus WM), gray matter regions (insular, putamen, middle occipital gyrus, thalamus), and mixed regions (medulla), in 4 different scanners of BLSA (where scanner A is the 1.5 Tesla Philips Intera scanner, and scanners B/C/D are the 3 Tesla Philips Achieva scanners). "L." stands for the left hemisphere of the brain.

Fig. 9 To assess the generalizability of our findings, we include two additional datasets, ADNI and BIOCARD and fit the linear mixed-effects models. The coefficients and p-values show similar patterns of those from BLSA alone, despite that the effect sizes and the hierarchical clustering are different, partly due to the omission of the rescan covariate.