University of Pennsylvania
UPenn Biostatistics Working Papers

Year 2015 *Paper* 44

Statistical estimation of white matter microstructure from conventional MRI

Leah Suttner* Amanda Mejia† Blake Dewey‡
Pascal Sati** Daniel S. Reich†† Russell T. Shinohara‡‡

*Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania

†Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, amejia@jhsph.edu

‡Translational Neuroradiology Unit, Division of Neuroimmunology and Neurovirology, National Institute of Neurological Disorders and Stroke, National Institutes of Health

**Translational Neuroradiology Unit, Division of Neuroimmunology and Neurovirology, National Institute of Neurological Disorders and Stroke, National Institutes of Health

††Translational Neuroradiology Unit, Division of Neuroimmunology and Neurovirology, National Institute of Neurological Disorders and Stroke, National Institutes of Health; Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, reichds@ninds.nih.gov

‡‡Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, rshi@upenn.edu

Statistical estimation of white matter microstructure from conventional MRI

Leah Suttner, Amanda Mejia, Blake Dewey, Pascal Sati, Daniel S. Reich, and Russell T. Shinohara

Abstract

Diffusion tensor imaging (DTI) has become the predominant modality for studying white matter integrity in multiple sclerosis (MS) and other neurological disorders. Unfortunately, the use of DTI-based biomarkers in large multi-center studies is hindered by systematic biases that confound the study of disease-related changes. Furthermore, the site-to-site variability in multi-center studies is significantly higher for DTI than that for conventional MRI-based markers. In our study, we apply the Quantitative MR Estimation Employing Normalization (QuEEN) model to estimate the four DTI measures: MD, FA, RD, and AD. QuEEN uses a voxel-wise generalized additive regression model to relate the normalized intensities of one or more conventional MRI modalities to a quantitative modality, such as DTI. We assess the accuracy of the models by comparing the prediction error of estimated DTI images to the scan-rescan error in subjects with two sets of scans. Across the four DTI measures, the performance of the models is not consistent: Both MD and RD estimations appear to be quite accurate, while AD estimation is less accurate than MD and RD; the accuracy of FA estimation is poor. Thus, it some cases when assessing white matter integrity, it may sufficient to acquire conventional MRI sequences alone.

Statistical Estimation of White Matter Microstructure from Conventional MRI

Leah Suttner[1], Amanda Mejia[2], Blake Dewey[3], Pascal Sati[3], Daniel S Reich[2,3], and Russell T Shinohara[1]

[1]Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA 19104
[2]Department of Biostatistics, The Johns Hopkins University, Baltimore MD 21205
[3]Translational Neuroradiology Unit, Division of Neuroimmunology and Neurovirology, National Institute of Neurological Disease and Stroke, National Institute of Health, Bethesda MD 20892

Abstract
Diffusion tensor imaging (DTI) has become the predominant modality for studying white matter integrity in multiple sclerosis (MS) and other neurological disorders. Unfortunately, the use of DTI-based biomarkers in large multi-center studies is hindered by systematic biases that confound the study of disease-related changes. Furthermore, the site-to-site variability in multi-center studies is significantly higher for DTI than that for conventional MRI-based markers. In our study, we apply the Quantitative MR Estimation Employing Normalization (QuEEN) model to estimate the four DTI measures: MD, FA, RD, and AD. QuEEN uses a voxel-wise generalized additive regression model to relate the normalized intensities of one or more conventional MRI modalities to a quantitative modality, such as DTI. We assess the accuracy of the models by comparing the prediction error of estimated DTI images to the scan-rescan error in subjects with two sets of scans. Across the four DTI measures, the performance of the models is not consistent: Both MD and RD estimations appear to be quite accurate, while AD estimation is less accurate than MD and RD; the accuracy of FA estimation is poor. Thus, it some cases when assessing white matter integrity, it may sufficient to acquire conventional MRI sequences alone.

Introduction

Diffusion tensor imaging (DTI) has become the predominant modality for studying white matter integrity in multiple sclerosis (MS) and other neurological disorders, including Alzheimer's disease, traumatic brain injury, epilepsy, depression, and stroke. In MS, DTI is generally understood to provide information about diffuse changes in the microstructure and integrity in the white matter of patients. Numerous studies have shown that certain DTI measures differ between MS subjects and controls. In particular, researchers have found that patients with MS have increased mean diffusivity (MD) and decreased fractional anisotropy (FA) in normal appearing white matter (NAWM) compared to healthy controls [1]–[4]. As changes in NAWM are frequently used to study and monitor the progression of MS, new biomarkers based on these differences could offer better disease management for patients. Moreover, many of these studies promote the use of DTI by suggesting that DTI can measure microstructural changes, whereas conventional magnetic resonance imaging (MRI) is only able identify changes at the macrostructural level [3], [5], [6].

Unfortunately, the use of DTI-based biomarkers in large multi-center studies is hindered by systematic biases [7], [8], which confound the study of disease-related changes. Furthermore, the site-to-site variability in multi-center studies is significantly higher for DTI [9], [10] than that for conventional MRI-based markers [11], [12]. These issues necessitate larger sample size in multi-center studies using DTI and highlight the potential for less generalizable findings from single-scanner studies. Additionally, DTI sequences require additional scanning time, which increases the cost of MRI in both clinical and research settings. In this paper, we investigate whether conventional MRI can be used to detect microstructural changes and estimate DTI measures.

Synthesizing or estimating a specific imaging modality or sequence from other imaging sequences can be advantageous for several reasons. Doing so allows for larger multicenter imaging studies in which patients, or even entire centers, are missing particular types of images. It also allows researchers to capitalize on imaging datasets that have already been acquired. For example, the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [13] contains thousands of conventional MRIs, and the Comprehensive Longitudinal Investigation of Multiple sclerosis at the Brigham and Women's Hospital (CLIMB) contains conventional MRI from thousands of MS patients [14]; there are also numerous well controlled phase 3 studies that have acquired conventional MRI data. Using image estimation, these and other databases could be used to conduct large-scale studies with quantitative imaging not acquired at every study site. Image synthesis also has significant applications in multi-modal image registration [15],[16],[17] and normalization [16]. Thus, there has been a growing interest in this field in recent years. A recent study on image synthesis showed that it is possible to synthesize T_2-weighted contrasts from T_1-weighted images and to synthesize 3-tesla T_1-weighted magnetization prepared rapid gradient echo (MPRAGE) images from 1.5-tesla MPRAGEs using an atlas image-based nonlinear regression [18]. [17] uses a generalized patch-based label propagation method to synthesis arbitrary target image modalities. Specifically, they use T_1-wighted MRI to synthesize T_2 MRI and, with much less accuracy, DTI-fractional anisotropy.

In our study, we adopt the methods introduced by [19], in which quantitative T_1 maps (qT_1) are estimated from the conventional MRI sequences: T_1-weighted (T_1w), T_2-weighted (T_2w), proton density-weighted (PDw), and T_2-weighted fluid attenuated inversion recovery (FLAIR) images. We apply their Quantitative MR Estimation Employing Normalization (QuEEN) model to estimate the four DTI measures: MD, FA, radial diffusivity (RD) and axial diffusivity (AD).

Methods

Study Sample

The dataset we use in this study includes 73 subjects with clinically diagnosed MS and at least one set of MRI and DTI scans. After performing quality control to exclude images that exhibit subject movement, major segmentation errors, and registration problems, we have 50 subjects, including 14 with relapsing-remitting MS (RRMS), 12 with secondary-progressive MS (SPMS), and 24 with primary-progressive MS (PPMS). The resulting

study sample is 64 % female with a mean age of 53 (S.D. 11). Additionally, a second set of MRI and DTI scans was acquired for 19 subjects, which we use to evaluate the predictive accuracy of the model. The median time between scans is 181 days (range 147 – 301). Additional summary statistics are provided in Table 1.

	All MS	**PPMS**	**RRMS**	**SPMS**
n	50	24	14	12
% Female	64.0	50.0	85.7	66.7
Mean Age (SD)	52.9 (11.4)	56.6 (7.4)	44.4 (15.0)	55.4 (8.3)
Disease Duration (SD)	14.9 (10.8)	13.5(9.5)	15.0 (7.9)	25.7 (9.2)
Median EDSS Score (range)	3.5 (1.0 – 7.5)	5.5 (1.5 – 7.5)	1.5 (1.0 – 6.0)	6.25 (1.5 – 7.0)

Table 1: Study Sample Summary Statistics

Imaging Protocol

Each MRI study includes the following volumes, all collected on a Siemens Skyra 3T scanner equipped with a 32-channel receive coil: T1-MPRAGE (Magnetization-Prepared RApid Gradient Echo) [TR=3000ms, TE=3.03ms, TI=900ms, FA=9]; PD*w* and T2w images from a dual-echo turbo-spin-echo (TSE) sequence [TR=3000ms, TE=11ms/101ms, FA=150, ETL=14]; and a 3D T2-weighted FLAIR image acquired using a T2-selective inversion pulse optimized for T2 of 120ms [TR=4800ms, TE=354ms, TI=1800ms, variable FA]. All scans are acquired at 1.0mm isotropic resolution except the PD*w*/T2*w* TSE sequence, which is acquired at 0.93 x 0.93 x 3.0mm resolution.

Each DTI study was obtained using a two-acquisition method with opposite phase encode directions to minimize influence of B_0-related image distortions. Each acquisition acquired 1 reference image without diffusion weighting and 30 images with radially spaced diffusion directions and a diffusion weighting of b=1000s/mm^2. Diffusion weighted images were acquired at 2 x 2 x 2mm isotropic resolution.

Image Processing

We perform image preprocessing as described in [19]. Briefly, we rigidly align all images to the MNI152 1.0 mm nonlinear template. To remove extracerebral voxels, we use the SPECTRE skull-stripping algorithm [20]. To segment tissue classes, we use Lesion Topology Preserving Anatomy Driven Segmentation (Lesion-TOADS) [21]. We correct for ventricular segmentation errors using non-topologically constrained maximum membership classes. We generate brain masks by excluding cerebrospinal fluid (CSF), hypointense voxels in FLAIR, and any voxels outside the field of view on any image. To create conservative tissue class masks, we use a 3x3x3 diamond-shaped kernel for erosion of the Lesion-TOADS segmentation. We use the conservative masks for training and validation, but we use the full masks for whole-image predictions.

Image Normalization

As conventional structural MRI is acquired in arbitrary units, intensity normalization is

required to compare values across images acquired at different sites; using different hardware, software or protocols; or even on different dates using the same scanner and protocol. Historically, image normalization has been performed with respect to NAWM [22] or CSF [23]–[28] as a reference region. However, both of these tissue classes are less than optimal for this study. CSF is highly variable in intensity, which can lead to inconsistent normalization. Changes in NAWM are of primary interest, and thus normalizing with respect to NAWM could obscure disease-related changes.

Instead, we utilize the adaptation of the z-score normalization method [22], [29] proposed by [19]. The z-score method subtracts a measure of location and divides by a measure of scale; we refer to these two steps as *scaling* and *shifting*, respectively. The adapted method proposed by [13] utilizes a combination of NAWM and cerebellar gray matter (CBGM). Although using NAWM alone as a reference can confound the normalization, the standard deviation of NAWM is well estimated and can be used for scaling. Abnormalities of CBGM in MS are poorly detected by conventional MRI methods, and thus CBGM signal intensity should be similar between subjects and disease groups. Thus, CBGM can be used for shifting. Previously, CBMG has been used for normalization in positron emission tomography (PET) in MS [30] and Alzheimer's disease [31].

The normalized units are given by

$$M_i^N(v) = \frac{M_i(v) - \mu_{i,m}^{(CBGM)}}{\sigma_{i,m}^{(NAWM)}}$$

$M_i(v)$ denotes the intensity of voxel v for subject i in modality M ∈ {FLAIR, PDw, T_1w, T_2w}. For subject i in modality M, $\mu_{i,M}{}^{(CBGM)}$is the mean intensity in CBGM, and $\sigma_{i,M}{}^{(NAWM)}$is the standard deviation of intensities within NAWM.

Statistical Prediction Model

To model each DTI modality, we apply the QuEEN statistical model to generate "statistical DTI" measures of MD, FA, RD and AD. QuEEN uses a voxel-wise generalized additive regression model to relate the normalized intensities of one or more conventional MRI modalities to a quantitative modality, such as qT1 or DTI. We specify a set of models to use the standard MRI modalities, T_1w, T_2w, PDw, and FLAIR, to predict the DTI measures MD, FA, RD, and AD. As in the QuEEN model, we fit a separate model for each tissue class, c. Thus, for each DTI measure, we relate the value at each voxel v in class c of subject i to the corresponding predictor modalities by:

$$DTI_{ji}(v) \sim f_{j1}^c\left(T_1w_i^N(v)\right) + f_{j2}^c\left(T_2w_i^N(v)\right) + f_{j1}^c\left(PDw_i^N(v)\right) + f_{j1}^c\left(FLAIR_i^N(v)\right) + \epsilon_i(v)$$

where $j = 1,2,3,4$ indexes the four DTI measures we wish to predict. The smooth functions $f_{jk}^c(\cdot)$, $k = 1,2,3,4$, relate predictor modality k to DTI measure j within tissue class

c. We implement the models in R using the `gam` function from the `mgcv` package (version 1.7-28, [32], [33]) as described in [19]. The `gam` function represents the smooth curves as penalized regression splines. The degree of smoothness is estimated using generalized cross-validation, and the smoothing parameter estimation criterion is optimized using the Newton method [34]. The statistical DTI measures (stat-MD, stat-FA, stat-RD, and stat-AD) are estimated by applying the estimated regression curves to the respective predictor modalities.

Model Validation

To evaluate each model, we assess the accuracy and utility of the statistical DTI measures. First, we assess the accuracy of the models by comparing the statistical prediction mean square error (MSE) to the scan-rescan MSE in subjects with two sets of scans. The statistical prediction MSE is defined as the voxel-wise squared difference in intensity between the acquired and statistical DTI measures, averaged across voxels; the scan-rescan MSE is defined as the voxel-wise squared difference in intensity between two acquired images, averaged across voxels. We summarize both measures by tissue class.

Second, we assess the utility of the model by comparing the mean DTI values in NAWM across MS subtypes for the statistical and acquired images. For each DTI measure, we conduct two-sided Wilcoxon rank sum tests between each subtype. [35] found evidence that RRMS patients had decreased FA and increased MD, RD, and AD in white matter regions compared to healthy controls.

Associations with Clinical Measures

To investigate the use of theses images as potential biomarkers we conduct a preliminary assessment of the association of the observed and predicted DTI measures and two commonly used scores for disability and cognition. First, we test the correlations between mean MD and mean RD in NAWM and the Expanded Disability Status Scale (EDSS) score [36], a common measure designed to assess neurological impairment in MS. We also test the correlation of mean MD and mean RD in NAWM with the Symbol Digit Modality Test (SDMT) [37] score, a tool for measuring cognitive function. We use the nonparametric Kendall's Tau coefficient for assessing correlations.

Results

Sample images from two randomly selected subjects are shown in Figure 1. The subject in Figure 1a has PPMS and the subject in Figure 1b has SPMS. The top rows in Figures 1a/b show the acquired DTI images, the middle rows show the statistical DTI images estimated using the leave-one-subject-out cross-validation (LOOCV), and the bottom rows show the absolute value of the difference between the acquired and statistical DTI images. Both stat-MD and stat-RD appear to be quite accurate estimates of MD and RD, respectively, while stat-AD is less accurate than stat-MD and stat-RD, and the accuracy of stat-FA is poor (although stat-FA still captures the strong gray/white differentiation typical of FA maps).

To assess the accuracy of our statistical DTI models, we compare the prediction MSE of the statistical DTI images to the scan-rescan MSE of the acquired images. Figure 2 shows boxplots of the scan-rescan log-MSE for the acquired DTI images (shown in green) and the prediction log-MSE for the statistical estimate of each DTI image (shown in orange) for three tissue classes of interest: NAWM, GM, and lesions. Each boxplot was constructed based on the MSE from 19 subjects with two sets of DTI scans. In MD and RD, the MSEs are comparable in NAWM and lesions. In MD, the prediction log-MSE in the statistical images and the scan-rescan log-MSE in the acquired images have means in NAWM of -7.4 mm^2s^{-1} (25^{th} percentile = -7.5, 75^{th} percentile = -7.3) and -7.4 mm^2s^{-1} (-7.5, -7.3), respectively, and means in lesions of -7.4 mm^2 s^{-1} (-7.5, -7.1) and -7.5 mm^2s^{-1} (-7.6, -7.3), respectively. In RD, the prediction log-MSE and the scan-rescan log-MSE have means in NAWM of -7.3 mm^2s^{-1} (-7.5, -7.3) and -7.3 mm^2s^{-1} (-7.5, -7.2), respectively, and means in lesions of -7.2 mm^2s^{-1} (-7.3, -7.1) and -7.4 mm^2s^{-1} (-7.6, -7.1), respectively. Across all four DTI measures, the statistical prediction log-MSE of the estimated images is smaller in GM compared to the scan-rescan log-MSE of the acquired images, indicating that our models can accurately estimate GM. However, in FA and AD, the prediction log-MSE is much larger in NAWM and lesions compared to the scan-rescan log-MSE of the acquired images, suggesting that the FA and AD models may be incorrect.

Figures 3 and 4 show the MS subtype group differences for mean MD and mean RD, respectively, in NAWM. In all comparisons, the differences between subgroups were not significant. However, the spread of the data is tighter in the statistical images compared to the acquired images. As shown in Figure 4, the level of significance between groups was similar for the statistical and acquired RD images. Figure 3 shows more pronounced differences in statistical MD between PPMS and each of the other subtypes compared to the acquired images as well as a smaller spread.

Table 2a provides the results of the exploratory analyses of the correlations between mean MD in NAWM and EDSS and SDMT scores, respectively. Table 2b provides the correlations between mean RD in NAWM and the EDSS and SDMT scores. None of the results withstand the multiple comparisons correction, however these results suggest that higher RD values may be associated with lower SDMT scores, and this association is stronger using the statistical RD measure.

MD	Score	Correlation	p-value
Acquired	EDSS	0.133	0.20
Statistical	EDSS	0.0378	0.72
Acquired	SDMT	-0.19	0.07
Statistical	SDMT	-0.141	0.17

Table 2a: Mean MD correlations

RD	Score	Correlation	p-value
Acquired	EDSS	0.133	0.20
Statistical	EDSS	0.171	0.10
Acquired	SDMT	-0.184	0.07
Statistical	SDMT	-0.248	0.02

Table 2b: Mean RD correlations

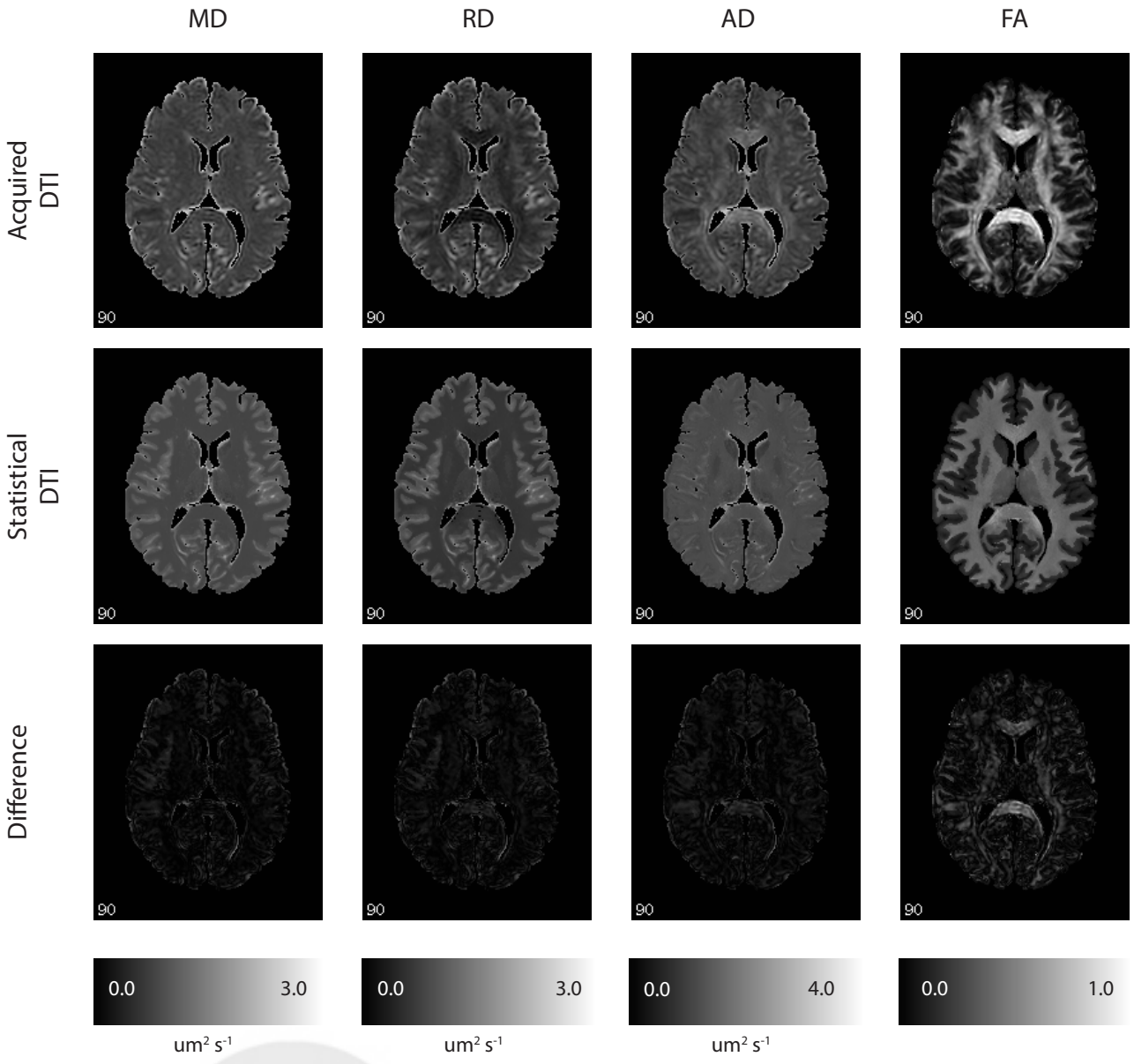

Figure 1a: The four DTI measures are shown for a PPMS subject. The top row shows the acquired DTI, the middle row shows the statistical DTI, and the bottom row shows the absolute value of the difference between the acquired and statistical DTI. The models for MD and RD perform relatively well, while the models for AD and FA appear to fail.

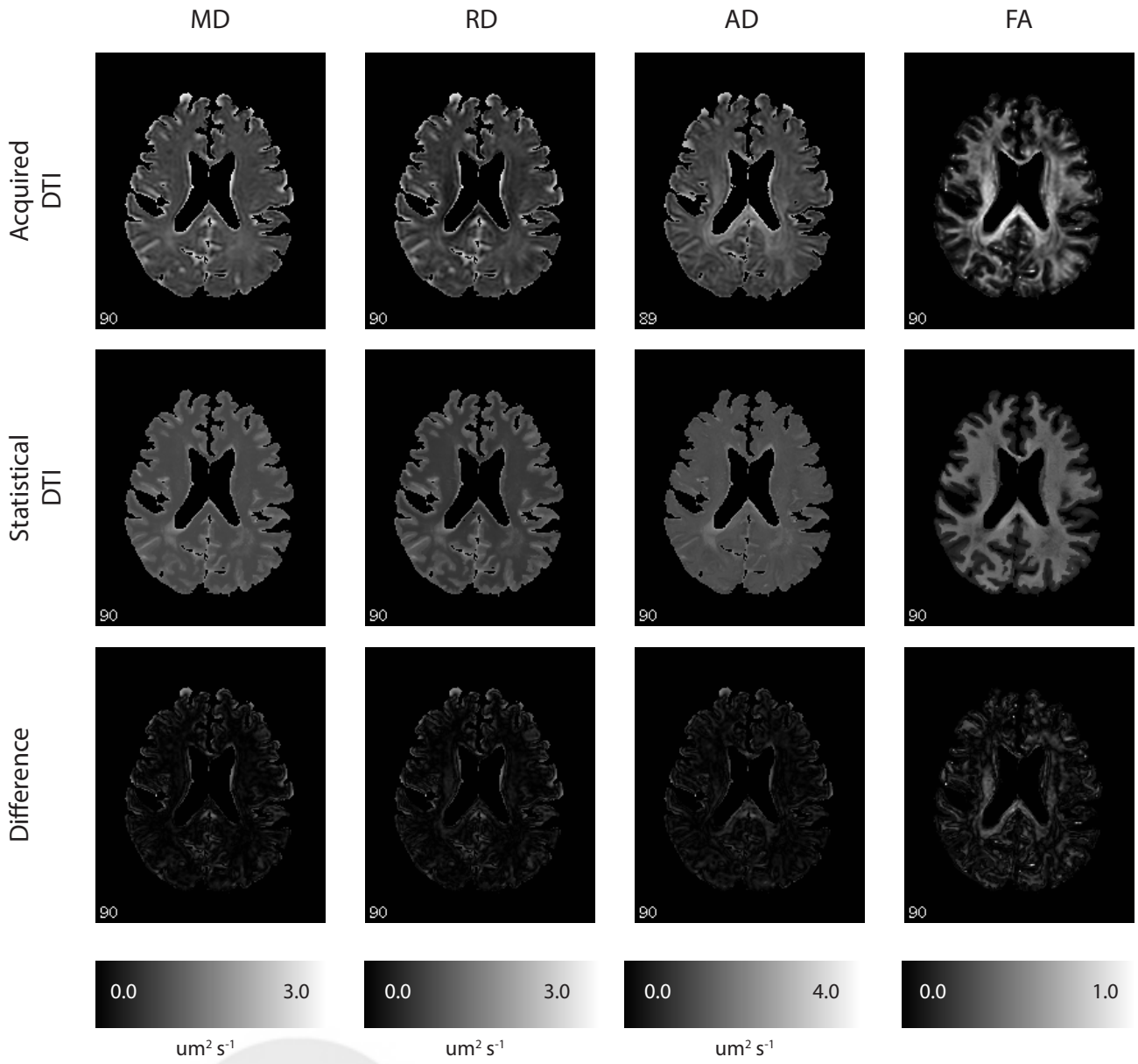

Figure 1b: The four DTI measures are shown for a SPMS subject. The top row shows the acquired DTI, the middle row shows the statistical DTI, and the bottom row shows the absolute value of the difference between the acquired and statistical DTI. The models for MD and RD perform relatively well, while the models for AD and FA appear to fail.

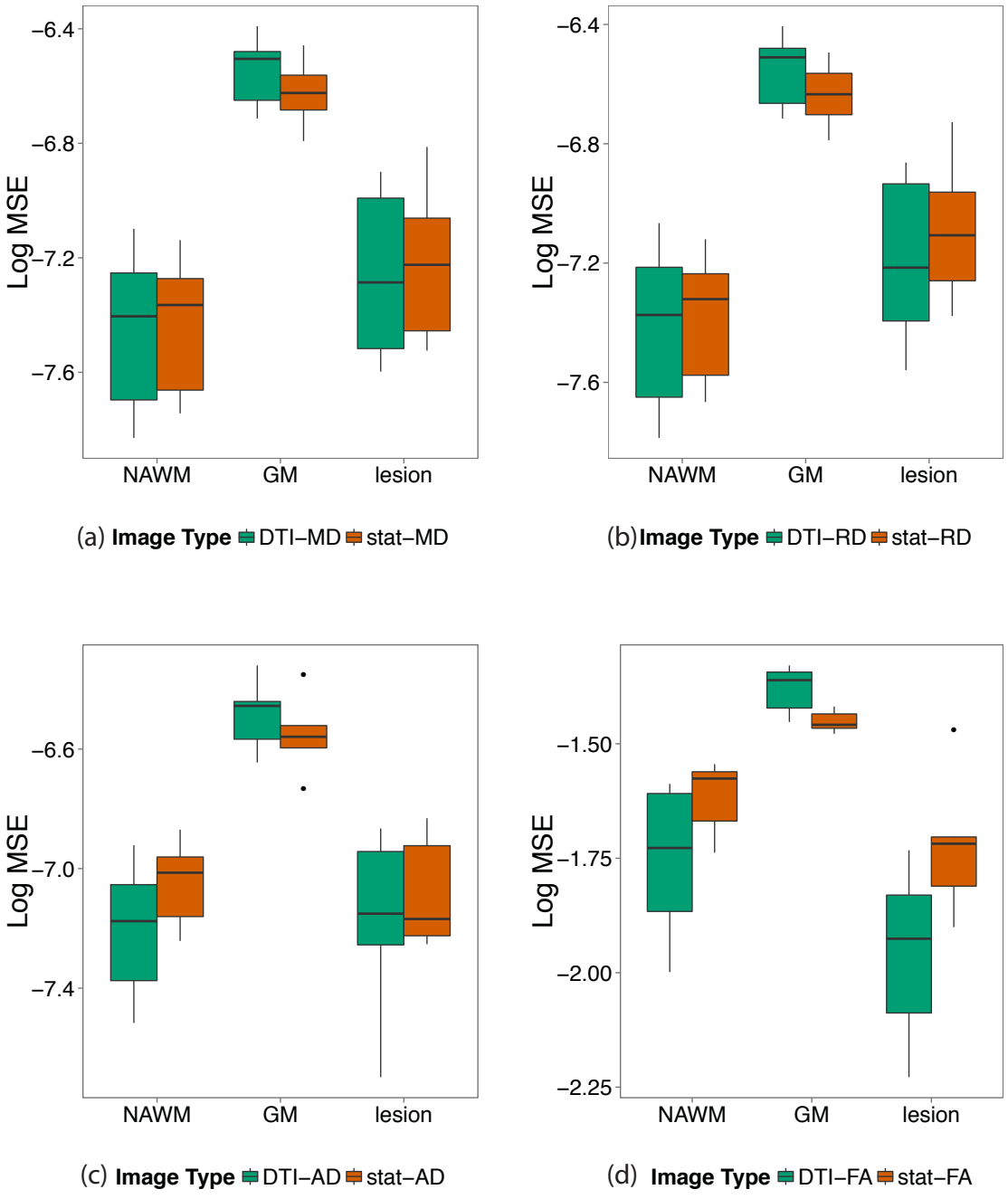

Figure 2: For each DTI measure, acquired scan-rescan MSE (green) and statistical prediction MSE (orange) are shown in lesion, GM, and NAWM. In NAWM the prediction MSEs in MD (a) and RD (b) are similar to scan-rescan MSEs, but these are slightly larger in AD (c) and FA (d).

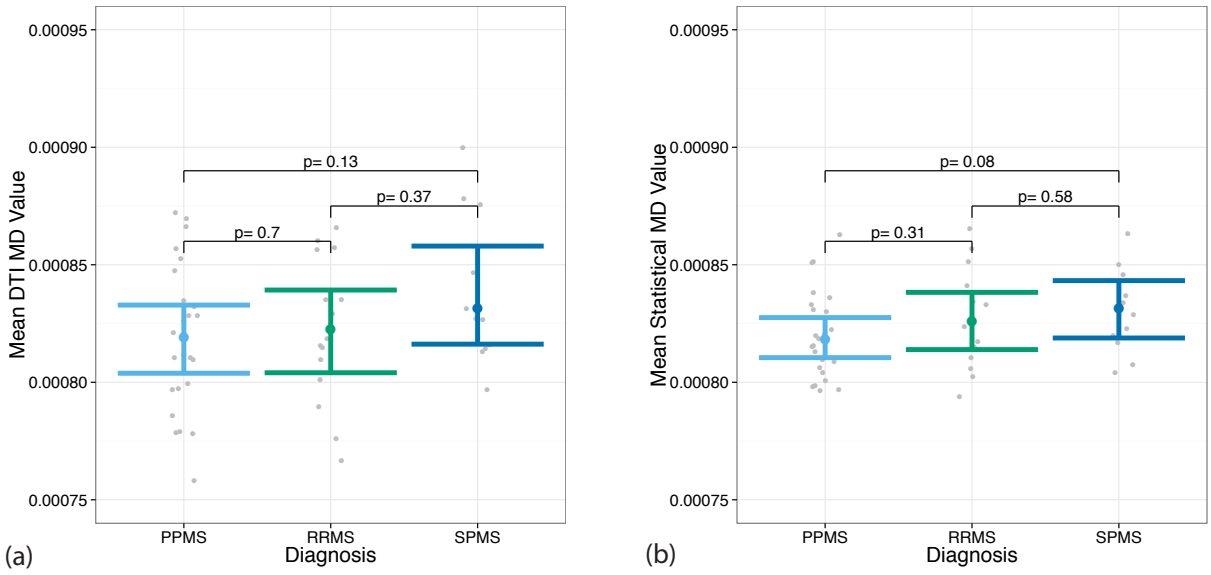

Figure 3: Differences in MD in NAWM between MS subtypes were tested using (a) the acquired DTI images and (b) the statistical MD images. The plots show the mean MD in NAWM for each subject (gray) and the Wilcoxon 95 % confidence intervals for each subtype. The p-values are reported for the two-sided Wilcoxon tests. There are no significant differences between the subgroups in either DTI MD or statistical MD. However, the variance of the data is smaller in the statistical MD images compared to the acquired MD images.

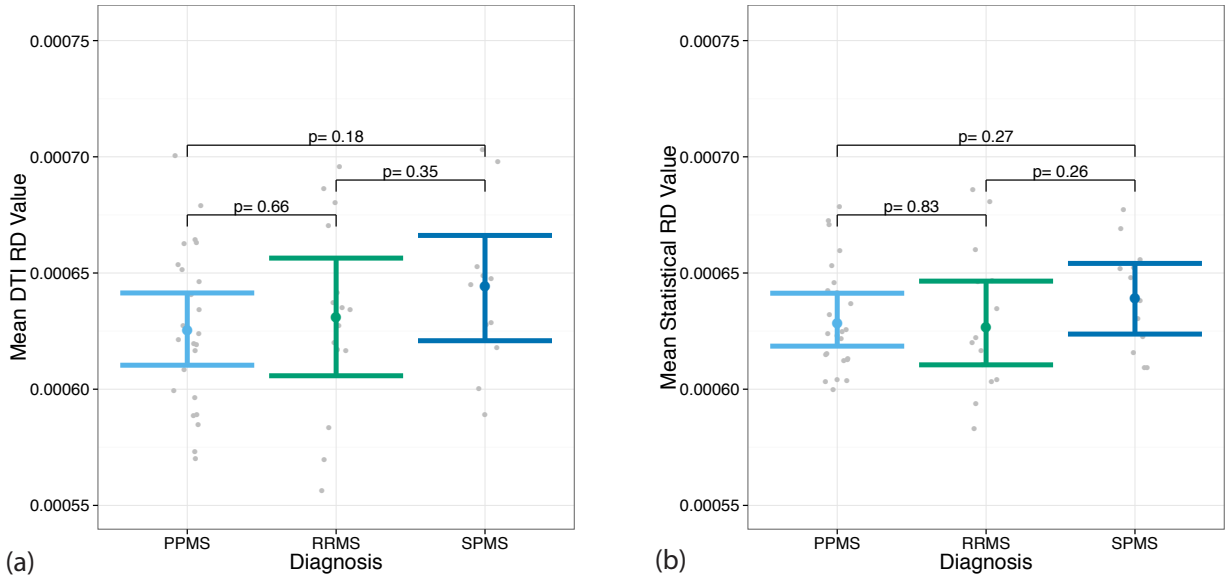

Figure 4: Differences in RD in NAWM between MS subtypes were tested using (a) the acquired DTI images and (b) the statistical RD images. The plots show the mean RD in NAWM for each subject (gray) and the Wilcoxon 95 % confidence intervals for each subtype. The p-values are reported for the two-sided Wilcoxon tests. There are no significant differences between the subgroups in either DTI RD or statistical RD.

Discussion

Here, we have used conventional MRI to estimate several common DTI measures, namely MD, RD, AD, and FA. Across the four DTI measures, the performance of the models is not consistent: while MD and RD are accurately estimated within NAWM using statistical methods, AD and FA are less so. The difference in the performance of the models may be explained by what each DTI measure describes. MD is a global measure sensitive to tissue microstructure, and in practice it is highly correlated to RD, which, in NAWM, represents diffusivity perpendicular to white matter bundles. On the other hand, AD quantifies diffusivity along the major orientation of the axon bundles and is substantially less sensitive to pathology, such as diffuse inflammation, demyelination, and edema – all of which are detectable by the conventional MRI approaches. Finally, FA represents tissue anisotropy as a measure of contrast between AD and RD; thus, any prediction errors for either AD or RD are likely to be magnified in the FA prediction [38], [39]. Furthermore, the four MRI contrasts that we use in the estimation are not sensitive to fiber orientation, but another contrast, T2*, can be quite sensitive to fiber orientation. Thus, including T2* in our statistical model may improve the accuracy of our AD and FA estimations.

As described above, we fit a separate model within each tissue class for each DTI prediction model. Model performance differs by tissue class both within and across DTI measures, as seen by the prediction MSE plots. Across all DTI measures, we predict GM well with respect to the acquired DTI scans. This is likely due to the lack of variation in structured diffusion across the brain in GM. In NAWM, we can accurately predict MD and RD but not AD and FA, again indicating that the model is insufficient for the latter DTI measures or that additional information exists in the AD and FA maps. In lesions, the statistical DTI predictions are generally worse than the acquired DTI predictions for each measure, which may be due to the greater amount of variability in DTI measures within lesions.

In addition, we show that in NAWM the group differences between the MS subtypes are more pronounced in the statistical MD images than in the acquired MD images, however none of the differences are significant. This finding needs further examination in a larger sample. In the present sample, the PPMS subjects are older, have shorter disease duration, and have fewer females compared to the subjects in the RRMS and SPMS groups.

We have discussed the differences in the performance of our statistical estimation across the four DTI measures and various tissue classes. Although we do not present accurate estimation models for AD or FA maps, we can accurately estimate MD and RD. Thus, we have shown that the information in the latter image types is not unique, and that the information may even be better captured by using statistical methods on conventional imaging. Therefore, in studies that are not interested in anisotropy, or only interested in MD, it may be sufficient to acquire only conventional MRI sequences.

Acknowledgments

We thank Dr. Govind Nair for his insightful comments and discussions about this work. We acknowledge the contribution of the NINDS Neuroimmunology Clinic, which recruited the patients and performed clinical evaluations, and the NIH Functional MRI Facility, where scanning took place. Suttner, Mejia, and Shinohara are partially funded by the NIH grant RO1 NS085211 from the National Institute of Neurological Disorders and Stroke (NINDS). The study was supported in part by the Intramural Research Program of NINDS. This work represents the opinions of the researchers and not necessarily that of the granting organizations.

Reference:

[1] H. Vrenken, P. J. W. Pouwels, J. J. G. Geurts, D. L. Knol, C. H. Polman, F. Barkhof, and J. A. Castelijns, "Altered diffusion tensor in multiple sclerosis normal-appearing brain tissue: cortical diffusion changes seem related to clinical deterioration," *J. Magn. Reson. Imaging JMRI*, vol. 23, no. 5, pp. 628–636, May 2006.

[2] O. Ciccarelli, D. J. Werring, G. J. Barker, C. M. Griffin, C. A. M. Wheeler-Kingshott, D. H. Miller, and A. J. Thompson, "A study of the mechanisms of normal-appearing white matter damage in multiple sclerosis using diffusion tensor imaging--evidence of Wallerian degeneration," *J. Neurol.*, vol. 250, no. 3, pp. 287–292, Mar. 2003.

[3] J. Senda, H. Watanabe, T. Tsuboi, K. Hara, H. Watanabe, R. Nakamura, M. Ito, N. Atsuta, F. Tanaka, S. Naganawa, and G. Sobue, "MRI mean diffusivity detects widespread brain degeneration in multiple sclerosis," *J. Neurol. Sci.*, vol. 319, no. 1–2, pp. 105–110, Aug. 2012.

[4] D. J. Werring, C. A. Clark, G. J. Barker, A. J. Thompson, and D. H. Miller, "Diffusion tensor imaging of lesions and normal-appearing white matter in multiple sclerosis," *Neurology*, vol. 52, no. 8, pp. 1626–1632, May 1999.

[5] M. A. Rocca, A. Ceccarelli, A. Falini, P. Tortorella, B. Colombo, E. Pagani, G. Comi, G. Scotti, and M. Filippi, "Diffusion tensor magnetic resonance imaging at 3.0 tesla shows subtle cerebral grey matter abnormalities in patients with migraine," *J. Neurol. Neurosurg. Psychiatry*, vol. 77, no. 5, pp. 686–689, May 2006.

[6] M. Wilson, C. R. Tench, P. S. Morgan, and L. D. Blumhardt, "Pyramidal tract mapping by diffusion tensor magnetic resonance imaging in multiple sclerosis: improving correlations with disability," *J. Neurol. Neurosurg. Psychiatry*, vol. 74, no. 2, pp. 203–207, Feb. 2003.

[7] A. Pfefferbaum, E. Adalsteinsson, and E. V. Sullivan, "Replicability of diffusion tensor imaging measurements of fractional anisotropy and trace in brain," *J. Magn. Reson. Imaging*, vol. 18, no. 4, pp. 427–433, Oct. 2003.

[8] C. Vollmar, J. O'Muircheartaigh, G. J. Barker, M. R. Symms, P. Thompson, V. Kumari, J. S. Duncan, M. P. Richardson, and M. J. Koepp, "Identical, but not the same: Intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0T scanners," *NeuroImage*, vol. 51, no. 4, pp. 1384–1394, Jul. 2010.

[9] T. Zhu, R. Hu, X. Qiu, M. Taylor, Y. Tso, C. Yiannoutsos, B. Navia, S. Mori, S. Ekholm, G. Schifitto, and J. Zhong, "Quantification of accuracy and precision of multi-center DTI measurements: a diffusion phantom and human brain study," *NeuroImage*, vol. 56, no. 3, pp. 1398–1411, Jun. 2011.

[10] S. J. Teipel, M. Wegrzyn, T. Meindl, G. Frisoni, A. L. W. Bokde, A. Fellgiebel, M. Filippi, H. Hampel, S. Klöppel, K. Hauenstein, M. Ewers, and EDSD study group, "Anatomical MRI and DTI in the diagnosis of Alzheimer's disease: a European multicenter study," *J. Alzheimers Dis. JAD*, vol. 31 Suppl 3, pp. S33–47, 2012.
[11] S. J. Teipel, M. Ewers, S. Wolf, F. Jessen, H. Kölsch, S. Arlt, C. Luckhaus, P. Schönknecht, K. Schmidtke, I. Heuser, L. Frölich, G. Ende, J. Pantel, J. Wiltfang, F. Rakebrandt, O. Peters, C. Born, J. Kornhuber, and H. Hampel, "Multicentre variability of MRI-based medial temporal lobe volumetry in Alzheimer's disease," *Psychiatry Res.*, vol. 182, no. 3, pp. 244–250, Jun. 2010.
[12] M. Ewers, S. J. Teipel, O. Dietrich, S. O. Schönberg, F. Jessen, R. Heun, P. Scheltens, L. van de Pol, N. R. Freymann, H.-J. Moeller, and H. Hampel, "Multicenter assessment of reliability of cranial MRI," *Neurobiol. Aging*, vol. 27, no. 8, pp. 1051–1059, Aug. 2006.
[13] S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, and L. Beckett, "The Alzheimer's disease neuroimaging initiative," *Neuroimaging Clin. N. Am.*, vol. 15, no. 4, pp. 869–877, xi–xii, Nov. 2005.
[14] S. A. Gauthier, B. I. Glanz, M. Mandel, and H. L. Weiner, "A model for the comprehensive investigation of a chronic autoimmune disease: the multiple sclerosis CLIMB study," *Autoimmun. Rev.*, vol. 5, no. 8, pp. 532–536, Oct. 2006.
[15] F. Michel and N. Paragios, "Image transport regression using mixture of experts and discrete Markov Random Fields," 2010, pp. 1229–1232.
[16] A. Jog, S. Roy, A. Carass, and J. L. Prince, "Pulse Sequence based Multi-acquisition MR Intensity Normalization," *Proc. SPIE-- Int. Soc. Opt. Eng.*, vol. 8669, Mar. 2013.
[17] D. H. Ye, D. Zikic, B. Glocker, A. Criminisi, and E. Konukoglu, "Modality propagation: coherent synthesis of subject-specific scans with data-driven regularization," *Med. Image Comput. Comput.-Assist. Interv. MICCAI Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, vol. 16, no. Pt 1, pp. 606–613, 2013.
[18] A. Jog, S. Roy, A. Carass, and J. L. Prince, "MAGNETIC RESONANCE IMAGE SYNTHESIS THROUGH PATCH REGRESSION," *Proc. IEEE Int. Symp. Biomed. Imaging Nano Macro IEEE Int. Symp. Biomed. Imaging*, vol. 2013, pp. 350–353, Dec. 2013.
[19] A. Mejia, E. Sweeney, B. Dewey, G. Nair, P. Sati, C. Shea, D. S. Reich, and R. T. Shinohara, "Statistical Estimation of T1 Relaxation Time Using Conventional Magnetic Resonance Imaging," *NeuroImage*, 2016.
[20] A. Carass, J. Cuzzocreo, M. B. Wheeler, P.-L. Bazin, S. M. Resnick, and J. L. Prince, "Simple paradigm for extra-cerebral tissue removal: algorithm and analysis," *NeuroImage*, vol. 56, no. 4, pp. 1982–1992, Jun. 2011.
[21] N. Shiee, P.-L. Bazin, A. Ozturk, D. S. Reich, P. A. Calabresi, and D. L. Pham, "A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions," *NeuroImage*, vol. 49, no. 2, pp. 1524–1535, Jan. 2010.
[22] R. T. Shinohara, E. M. Sweeney, J. Goldsmith, N. Shiee, F. J. Mateen, P. A. Calabresi, S. Jarso, D. L. Pham, D. S. Reich, C. M. Crainiceanu, Australian Imaging Biomarkers Lifestyle Flagship Study of Ageing, and Alzheimer's Disease Neuroimag-

ing Initiative, “Statistical normalization techniques for magnetic resonance imaging,” *NeuroImage Clin.*, vol. 6, pp. 9–19, 2014.
[23] J. Pujol, C. Junqué, P. Vendrell, J. M. Grau, J. L. Martí-Vilalta, C. Olivé, and J. Gili, “Biological significance of iron-related magnetic resonance imaging changes in the brain,” *Arch. Neurol.*, vol. 49, no. 7, pp. 711–717, Jul. 1992.
[24] J. H. van Waesberghe, M. A. van Walderveen, J. A. Castelijns, P. Scheltens, G. J. Lycklama à Nijeholt, C. H. Polman, and F. Barkhof, “Patterns of lesion development in multiple sclerosis: longitudinal observations with T1-weighted spin-echo and magnetization transfer MR,” *AJNR Am. J. Neuroradiol.*, vol. 19, no. 4, pp. 675–683, Apr. 1998.
[25] R. Bakshi, R. H. B. Benedict, R. A. Bermel, S. D. Caruthers, S. R. Puli, C. W. Tjoa, A. J. Fabiano, and L. Jacobs, “T2 hypointensity in the deep gray matter of patients with multiple sclerosis: a quantitative magnetic resonance imaging study,” *Arch. Neurol.*, vol. 59, no. 1, pp. 62–68, Jan. 2002.
[26] C. W. Tjoa, R. H. B. Benedict, B. Weinstock-Guttman, A. J. Fabiano, and R. Bakshi, “MRI T2 hypointensity of the dentate nucleus is related to ambulatory impairment in multiple sclerosis,” *J. Neurol. Sci.*, vol. 234, no. 1–2, pp. 17–24, Jul. 2005.
[27] S. D. Brass, R. H. B. Benedict, B. Weinstock-Guttman, F. Munschauer, and R. Bakshi, “Cognitive impairment is associated with subcortical magnetic resonance imaging grey matter T2 hypointensity in multiple sclerosis,” *Mult. Scler. Houndmills Basingstoke Engl.*, vol. 12, no. 4, pp. 437–444, Aug. 2006.
[28] M. Neema, A. Arora, B. C. Healy, Z. D. Guss, S. D. Brass, Y. Duan, G. J. Buckle, B. I. Glanz, L. Stazzone, S. J. Khoury, H. L. Weiner, C. R. G. Guttmann, and R. Bakshi, “Deep gray matter involvement on brain MRI scans is associated with clinical progression in multiple sclerosis,” *J. Neuroimaging Off. J. Am. Soc. Neuroimaging*, vol. 19, no. 1, pp. 3–8, Jan. 2009.
[29] R. T. Shinohara, C. M. Crainiceanu, B. S. Caffo, M. I. Gaitán, and D. S. Reich, “Population-wide principal component-based quantification of blood-brain-barrier dynamics in multiple sclerosis,” *NeuroImage*, vol. 57, no. 4, pp. 1430–1446, Aug. 2011.
[30] J. N. Ratchford, C. J. Endres, D. A. Hammoud, M. G. Pomper, N. Shiee, J. McGready, D. L. Pham, and P. A. Calabresi, “Decreased microglial activation in MS patients treated with glatiramer acetate,” *J. Neurol.*, vol. 259, no. 6, pp. 1199–1205, Jun. 2012.
[31] M. A. Kropholler, R. Boellaard, B. N. M. van Berckel, A. Schuitemaker, R. W. Kloet, M. J. Lubberink, C. Jonker, P. Scheltens, and A. A. Lammertsma, “Evaluation of reference regions for (R)-[(11)C]PK11195 studies in Alzheimer’s disease and mild cognitive impairment,” *J. Cereb. Blood Flow Metab. Off. J. Int. Soc. Cereb. Blood Flow Metab.*, vol. 27, no. 12, pp. 1965–1974, Dec. 2007.
[32] S. N. Wood, *Generalized additive models: an introduction with R*. Boca Raton, FL: Chapman & Hall/CRC, 2006.
[33] S. N. Wood, “Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models: Estimation of Semiparametric Generalized Linear Models,” *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 73, no. 1, pp. 3–36, Jan. 2011.

[34] Chong Gu and Grace Wahba, “Minimizing gcv/gml scores with multiple smoothing parameters via the newton method,” *SIAM J. Sci. Stat. Comput.*, vol. 12, no. 2, pp. 383–398, 1991.
[35] Y. Liu, Y. Duan, Y. He, C. Yu, J. Wang, J. Huang, J. Ye, P. M. Parizel, K. Li, and N. Shu, “Whole brain white matter changes revealed by multiple diffusion metrics in multiple sclerosis: a TBSS study,” *Eur. J. Radiol.*, vol. 81, no. 10, pp. 2826–2832, Oct. 2012.
[36] J. F. Kurtzke, “Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS),” *Neurology*, vol. 33, no. 11, pp. 1444–1452, Nov. 1983.
[37] A. Smith, *Symbol Digit Modalities Test: Manual*. Los Angeles: Western Psychological Services., 1982.
[38] S.-K. Song, S.-W. Sun, M. J. Ramsbottom, C. Chang, J. Russell, and A. H. Cross, “Dysmyelination revealed through MRI as increased radial (but unchanged axial) diffusion of water,” *NeuroImage*, vol. 17, no. 3, pp. 1429–1436, Nov. 2002.
[39] A. L. Alexander, J. E. Lee, M. Lazar, and A. S. Field, “Diffusion tensor imaging of the brain,” *Neurotherapeutics*, vol. 4, no. 3, pp. 316–329, Jul. 2007.