Discover Generics
Cost-Effective CT & MRI Contrast Agents
FRESENIUS KABI
WATCH VIDEO

AJNR

A Validation Study of Multicenter Diffusion Tensor Imaging: Reliability of Fractional Anisotropy and Diffusivity Values

R.J. Fox, K. Sakaie, J.-C. Lee, J.P. Debbins, Y. Liu, D.L. Arnold, E.R. Melhem, C.H. Smith, M.D. Philips, M. Lowe and E. Fisher

This information is current as of June 29, 2025.

AJNR Am J Neuroradiol 2012, 33 (4) 695-700
doi: https://doi.org/10.3174/ajnr.A2844
http://www.ajnr.org/content/33/4/695

ORIGINAL RESEARCH

A Validation Study of Multicenter Diffusion Tensor Imaging: Reliability of Fractional Anisotropy and Diffusivity Values

R.J. Fox
K. Sakaie
J.-C. Lee
J.P. Debbins
Y. Liu
D.L. Arnold
E.R. Melhem
C.H. Smith
M.D. Philips
M. Lowe
E. Fisher

BACKGROUND AND PURPOSE: DTI is increasingly being used as a measure to study tissue damage in several neurologic diseases. Our aim was to investigate the comparability of DTI measures between different MR imaging magnets and platforms.

MATERIALS AND METHODS: Two healthy volunteers underwent DTI on five 3T MR imaging scanners (3 Trios and 2 Signas) by using a matched 33 noncollinear diffusion-direction pulse sequence. Within each subject, a total of 16 white matter (corpus callosum, periventricular, and deep white matter) and gray matter (cortical and deep gray) ROIs were drawn on a single image set and then were coregistered to the other images. Mean FA, ADC, and longitudinal and transverse diffusivities were calculated within each ROI. Concordance correlations were derived by comparing ROI DTI values among each of the 5 magnets.

RESULTS: Mean concordance for FA was 0.96; for both longitudinal and transverse diffusivities, it was 0.93; and for ADC, it was 0.88. Mean scan-rescan concordance was 0.96–0.97 for all DTI measures. Concordance correlations within platforms were, in general, better than those between platforms for all DTI measures (mean concordance of 0.96).

CONCLUSIONS: We found that a 3T magnet and high-angular-resolution pulse sequence yielded comparable DTI measurements across different MR imaging magnets and platforms. Our results indicate that FA is the most comparable measure across magnets, followed by individual diffusivities. The comparability of DTI measures between different magnets supports the feasibility of multicentered clinical trials by using DTI as an outcome measure.

ABBREVIATIONS: CV = coefficient of variation; FA = fractional anisotropy; $\lambda_\perp$ = transverse diffusivity; $\lambda_\parallel$ = longitudinal diffusivity; MPRAGE = magnetization-prepared rapid acquisition of gradient echo; TIM = Total Imaging Matrix

DTI provides a quantitative measurement of the magnitude and direction of water diffusion within tissues.[1,2] Changes in water diffusion reflect alterations in tissue microstructure[3] and thus make DTI a noninvasive technique to study diseases caused by tissue destruction, such as MS.[2,4] The extent of tissue injury is commonly evaluated by using DTI-derived measures such as FA, $\lambda_\parallel$, $\lambda_\perp$, and ADC.[2,5] Studies in patients with MS revealed changes in DTI measures not only within MS lesions but also in normal-appearing white matter and gray matter, which are generally not discernible by using conventional T1- and T2-weighted imaging techniques.[6-9]

The sensitivity of DTI metrics to ongoing changes in the brain makes it an attractive outcome measure for MS therapeutic studies.[10] However, little is known about the validity of DTI measures between scanners. Recently, 2 independent studies[11,12] have attempted to address the reproducibility of DTI measures. They assessed FA[11,12] and $\lambda_\perp$[11] in the corpus callosum across different centers, all by using 3T scanners. They both found DTI measures to generally be homogeneous across centers when measured with scanners from the same manufacturer.

However, 2 major issues that remain to be resolved are the following: the comparability of DTI measures across different brain regions and the comparability of DTI measures across different scanners. We thus sought to address these issues by comparing DTI measures obtained from 5 centers from 2 subjects. We addressed the following issues: 1) scanners from different manufacturers, 2) multiple platforms from the same manufacturer, 3) DTI measures from different brain regions with a wide range of anisotropy, and 4) software upgrades.

Materials and Methods

MR Imaging Scanners

Two volunteers (both men and age 35, healthy controls) were imaged on five 3T MR imaging magnets located in different locations: 3 Trios (Siemens, Erlangen, Germany) and 2 Signas (GE Healthcare, Milwaukee, Wisconsin). In addition, the same volunteers were imaged annually during the course of 2 years (year 2 and year 3 on one of the Siemens magnets, Trio-1). During the course of the study, a scanner upgrade (TIM upgrade) on the Trio-1 alone between years 1 and 2

Received April 21, 2011; accepted after revision June 24.

From the Mellen Center for Multiple Sclerosis (R.J.F.), Diagnostic Radiology (K.S., M.D.P., M.L.), Quantitative Health Sciences (J.-C.L.), and Department of Biomedical Engineering (E.F.), Cleveland Clinic, Cleveland, Ohio; Barrow Neurological Institute (J.P.D.), Phoenix, Arizona; St. Luke's Medical Center (Y.L.), Milwaukee, Wisconsin; Montreal Neurological Institute (D.L.A.), McGill University, Montreal, Quebec, Canada; and Genentech Inc. (C.H.S.), Seattle, Washington.

This work was supported by NIH K23 NS 47211 (R.J.F.), NMSS RG 4091 A3/1 (R.J.F.), and a research grant from Genentech.

Please address correspondence to Robert J. Fox, MD, Mellen Center for Multiple Sclerosis Treatment and Research, Cleveland Clinic Foundation, 9500 Euclid Ave, U-10, Cleveland, OH; e-mail: foxr@ccf.org

Indicates open access to non-subscribers at www.ajnr.org

Indicates article with supplemental on-line figures.

http://dx.doi.org/10.3174/ajnr.A2844

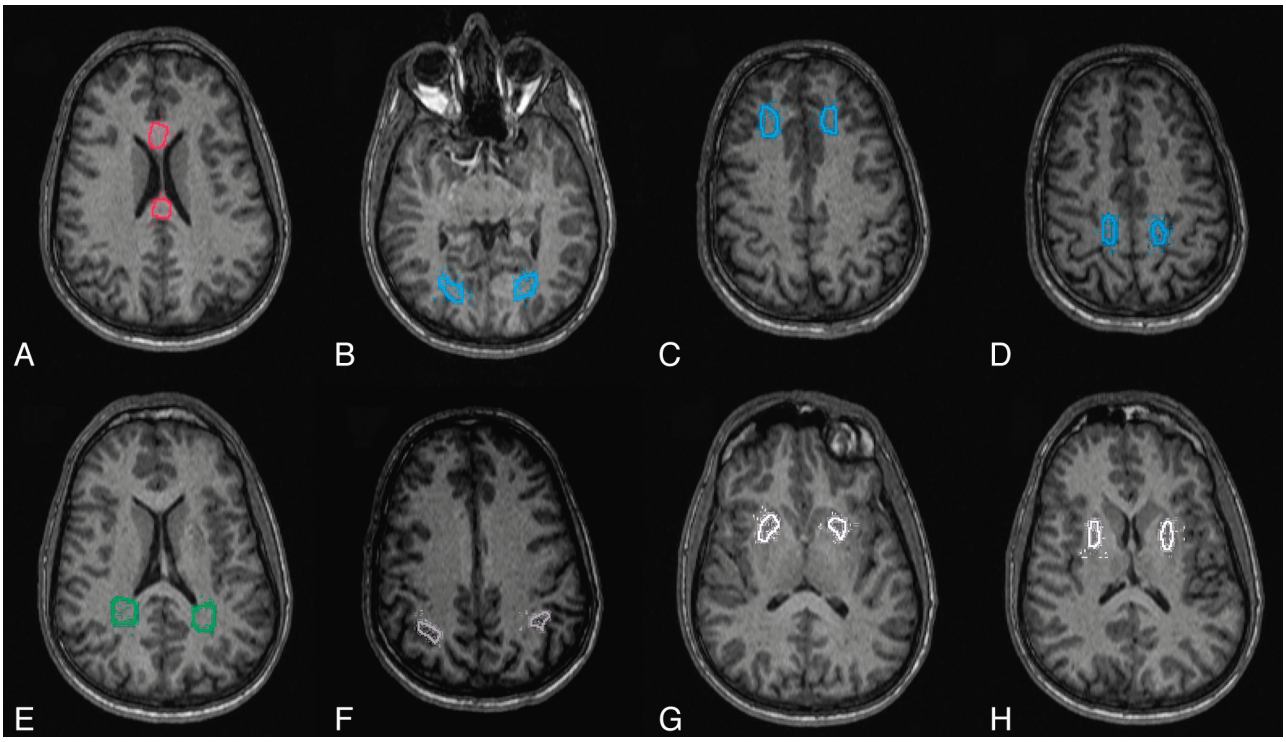

Fig 1. White and gray matter ROIs used for calculating DTI metrics in a single subject. Similar ROIs were used in the second subject. *A,* Anterior and posterior corpus callosum (red). *B–D,* Deep white matter (left and right occipital, frontal, and parietal, respectively; blue). *E,* Left and right periventricular white matter (green). *F,* Left and right cortical gray matter (gray). *G* and *H,* Left and right deep gray matter (putamen, white).

allowed comparisons across the upgrade (year 1 versus year 2) as well as independent of the upgrade (year 2 versus year 3). The other 2 Trio scanners (year 1) were used before the TIM upgrade.

Imaging Protocol

On the Siemens scanners, imaging included anatomic T1 MPRAGE (256 × 256 mm FOV, 256 × 128 matrix, 120 1.2-mm-thick sections, TE = 1.71 ms, TR = 1900 ms, TI = 900 ms, bandwidth = 490 Hz/Px) and twice-refocused spin-echo DTI with an EPI readout (256 × 256 mm FOV, 104 × 104 matrix, 3 slabs of sixteen 2.5-mm thick sections, partial Fourier factor = 6/8, TE = 100 ms, TR = 2500 ms, bandwidth = 1456 Hz/Px, 33 noncollinear diffusion-weighting directions with $b = 1000$ s/mm^2 and 9 $b = 0$ acquisitions, 4 repeats for each slab).

We adapted these pulse sequences to the GE scanners, endeavoring to match the parameters. The imaging included anatomic T1 MPRAGE (spoiled gradient-recalled 256 × 256 FOV, 128 × 128 matrix, 124 sections, 1.2 mm thick, TE = 1.3 ms, TR = 2800 ms, TI = 900 ms, bandwidth = 977 Hz/Px) and twice-refocused spin-echo DTI with EPI readout (same as for Siemens, except TE = 94 ms, TR = 4200 ms, bandwidth = 1954 Hz/Px). The diffusion-gradient directions were explicitly matched to be the same on each scanner. The voxel size of DTI for both Siemens and GE scanners was 2.5 × 2.5 × 2.5 mm.

Data Analysis

Image Registration. Images were registered by using a surface-based method, a modified version of the Iterated Closest Point algorithm,[13] which minimized the distance between automatically segmented brain surfaces to determine the optimal rigid transformation (3 rotations + 3 translations). Transformation parameters were determined between the DTI and MPRAGE images acquired at the same site and between the MPRAGE images acquired at each of the 5 sites. These parameters were further combined to obtain a set of final registration transformation parameters, which were used to map ROIs as described below.

Mapping ROIs. For each subject, 16 ROIs were drawn on the MPRAGE image from 1 of the magnets (Trio-1, Fig 1), which was used as a reference mask. The ROIs encompassed 5 major areas with differing anisotropic properties: 1) the corpus callosum (genu, splenium), 2) periventricular white matter (parietal, occipital), 3) deep white matter (frontal, parietal, occipital), 4) cortical gray matter (posterior parietal, occipital), and 5) deep gray matter (putamen). ROI volumes ranged from approximately 182 to 742 mm^3, with the smallest ROIs located in the cortex. The final registration transformations (described above) were applied to coregister the reference ROI mask to each set of DTI images obtained from the other 4 centers. The ROIs were verified visually and adjusted manually to account for nonlinear distortions inherent to EPI data.

Calculation of FA, $\lambda_{\|}$, $\lambda_{\perp}$, and ADC. Mean values for FA, $\lambda_{\|}$ and $\lambda_{\perp}$, and ADC were calculated within each ROI. The diffusion tensor and its properties (eigenvalues, FA, and ADC) were calculated by using software developed in-house. First, the diffusion profile for the 33 diffusion directions was calculated on a voxel-by-voxel basis by a standard linear-log fit.[14] The diffusion profile was fit to the 9-element diffusion tensor by using least-squares. The tensor was then diagonalized to obtain the eigenvalues and eigenvectors by using standard routines to derive $\lambda_{\|}$ and $\lambda_{\perp}$.[15] ADC was calculated by taking the mean of the eigenvalues, and FA was derived accordingly.[1]

Statistical Analysis

All analyses were performed by using SAS (SAS Institute, Cary, North Carolina). The Lin concordance correlation[16] was used to compute

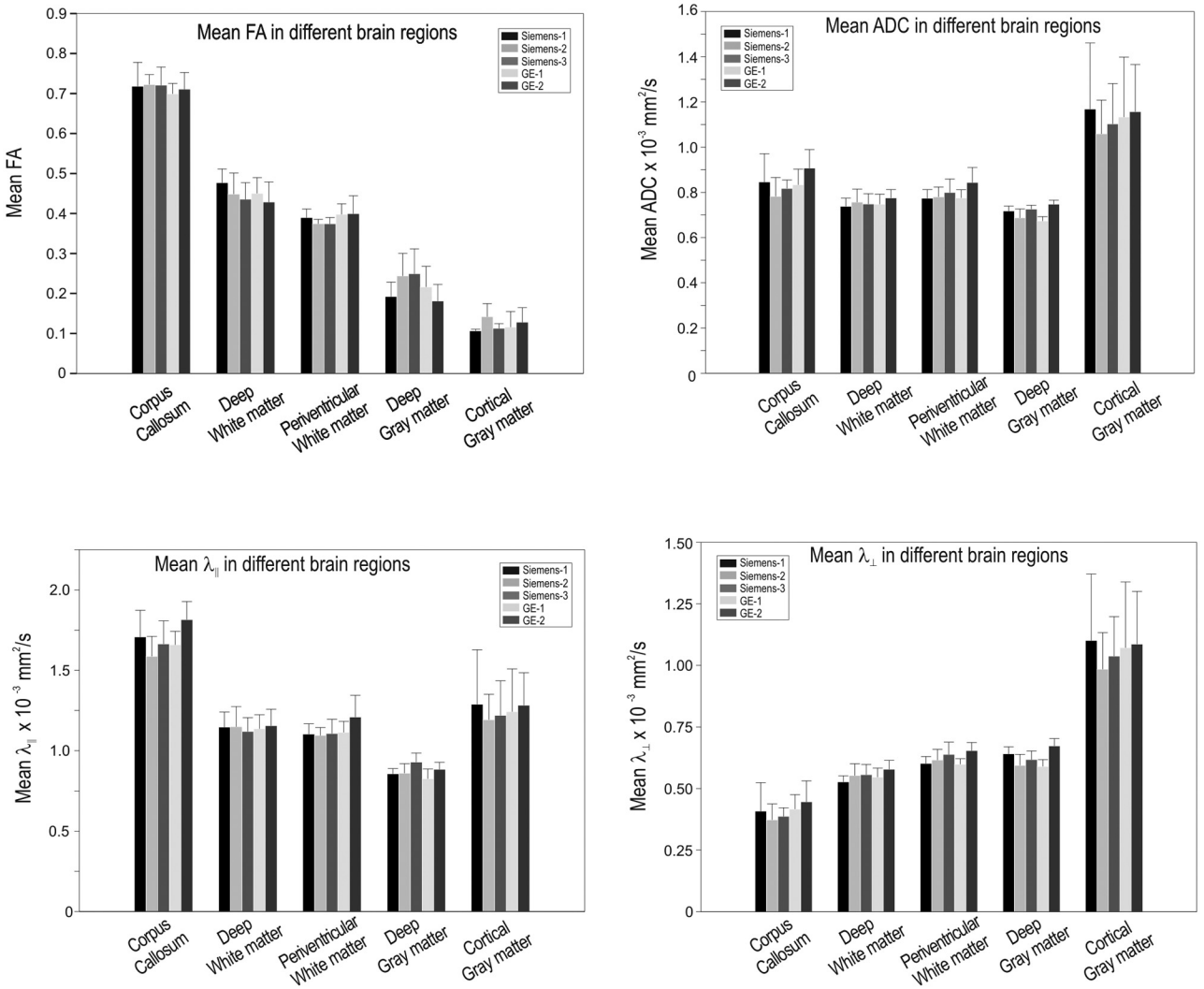

Fig 2. Mean DTI measures for different brain regions from each of the 5 magnets. Error bars are SDs.

concordance among DTI measures from all 16 ROIs. The following comparisons were made to evaluate the relationship between scanners: 1) overall mean concordance (ie, the mean of each paired concordance analysis) among each of the 5 magnets, 2) overall mean concordance within the 3 Siemens and within the 2 GE magnets (same-platform comparison), 3) overall mean concordance between the Siemens and GE magnets (cross-platform comparison), 4) longitudinal comparison (year 1–year 3) in Trio-1, and 5) before and after TIM upgrade in Trio-1. The CV was derived within each ROI among the 5 different scanners and longitudinally within the single scanner, and it was then averaged together across the ROIs and 2 study subjects.

Table 1: Mean concordance coefficients of DTI measures[a]

Type of Concordance	FA	$\lambda_{\|}$	$\lambda_{\perp}$	ADC
Overall mean concordance	0.96	0.93	0.93	0.88
Cross-platform mean concordance	0.96	0.92	0.91	0.86
Siemens mean concordance	0.96	0.93	0.93	0.89
GE correlation (1 comparison only)	0.97	0.94	0.94	0.90
Longitudinal mean concordance	0.96	0.96	0.97	0.96
Across-system upgrade	0.95	0.96	0.97	0.96
After system upgrade	0.97	0.95	0.97	0.96

[a] Average Lin's concordance coefficients comparing different imaging conditions across all regions from both subjects. Each value represents the mean of the different between-magnet concordance coefficients.

Results

Overall Comparison among the 5 Magnets

The regional mean DTI values were comparable among all 5 scanners (Fig 2). The FA, $\lambda_{\|}$, $\lambda_{\perp}$, and ADC values obtained from regions of low (gray matter), medium (periventricular and deep white matter), and high (corpus callosum) anisotropy were very similar among the 5 scanners. As indicated by error bars in Fig 2, the $\lambda_{\|}$, $\lambda_{\perp}$, and ADC values were more variable in the cortical gray matter compared with other regions, possibly because these ROIs were smaller and, therefore, more sensitive to noise. Because FA is calculated as a ratio with diffusivities in the numerator and denominator, noise effects may partially cancel out. The overall mean concordance (Table 1) was highest for FA (0.96) and lowest for ADC (0.88), with the values for both $\lambda_{\perp}$ and $\lambda_{\|}$ being 0.93.

Table 2: Mean CVs across ROIs

Scanner Comparison	CV across ROIs	FA	$\lambda_{\parallel}$	$\lambda_{\perp}$	ADC
Across 5 scanners	Overall CV (%), white matter only CV (%)	9.1 6.8	5.0 4.3	6.6 4.3	4.8 4.1
Same scanner with time	Overall CV (%), white matter only CV (%)	9.6 8.7	3.7 3.5	4.2 4.6	2.2 2.2

Average CV averaged across ROIs for the 5 different scanners and within a single scanner with time.

Cross-Platform Comparison

To evaluate manufacturer differences, we compared the DTI measures obtained from the 3 Siemens platforms with those from the 2 GE platforms (On-line Fig 1 and Table 1). The concordance was strongest for FA (0.96), followed by $\lambda_{\parallel}$ (0.92) and $\lambda_{\perp}$ (0.91), and weakest for ADC (0.86).

Within-Platform Comparison

We found better concordance in the within-platform than the cross-platform comparison (Table 1, On-line Fig 1). We found concordance to be the strongest for FA (0.96 for Siemens; 0.97 for GE). In addition, we also determined concordance for the other 3 DTI measures and found strong concordance for $\lambda_{\perp}$ and $\lambda_{\parallel}$ (0.93 for Siemens, 0.94 for GE) as well. As in earlier cases, we observed a weak concordance for ADC (0.89 for Siemens, 0.90 for GE).

Longitudinal Scan Comparison

There was a modestly higher concordance with $\lambda_{\parallel}$, $\lambda_{\perp}$, and ADC when the 2 subjects were rescanned on the same Siemens magnet (Table 1 and On-line Fig 2), compared with findings between magnets.

Software Upgrade Comparison

The first longitudinal scan (year 1) was obtained before the Siemens TIM upgrade, and the subsequent 2 scans (year 2 and year 3) were obtained after the upgrade. There was excellent correlation both across and after the software upgrade (Table 1 and On-line Fig 2). We observed a small decrease in mean FA in periventricular and deep white matter regions, which was driven predominantly by a corresponding increase in $\lambda_{\perp}$ (On-line Fig 2). There was also a small increase in ADC values following the TIM upgrade (On-line Fig 2); however, that did not appear to have any effect on concordance values.

CV

The CV across scanners varied from 4.8% to 9.1%, depending upon the DTI measure (Table 2). CV was less within-scanner with time, compared with across different scanners. CV was generally lower when evaluating white matter only.

Discussion

We evaluated the applicability of DTI in multicenter clinical trials by systematically comparing DTI measures obtained by scanning 2 subjects in 5 different 3T MR imaging scanners by using a standardized pulse sequence. Comparability of DTI measures is important when combining values from different magnets and different platforms together into a single dataset. We observed very strong concordances (Fig 2 and On-line Fig 1 and Table 1) for FA, $\lambda_{\perp}$, and $\lambda_{\parallel}$ values across different magnets within different brain regions encompassing both white and gray matter.

Reliability of FA and $\lambda_{\perp}$

We found the highest concordance for FA (0.96) followed by the 2 component diffusivities and the weakest for ADC. One explanation for the lower ADC concordance is that because ADC is the mean of the diffusion tensor eigenvalues, it can reflect variability due to changes in the overall magnitude of diffusivity, anisotropy of diffusivity, or a combination of the 2 effects. Therefore, fitting the diffusion profile to a single isotropic coefficient may lead to considerable variability. Moreover, diffusion in the brain is not isotropic, thereby making the derivation of a precise diffusion coefficient value more complex. A previous simulation study had also predicted that cross-scanner comparisons would find greater variability in ADC than FA, due to the underlying properties of ADC calculations as mentioned above.[17] The results of our study suggest that ADC may provide less statistical power than the other metrics examined. However, our CV estimates found the greatest variability in FA and the least in ADC (Table 2). Because FA is a unitless ratio of eigenvalues, variability in individual eigenvalues may compound when eigenvalues are combined to derive FA, leading to greater variability. Concordance values are a function not only of the absolute difference in values but also the range of values. As evident from Fig 2, there are more significant variations in FA values throughout the brain compared with ADC, thus having a positive impact on concordance, despite the higher CV. In addition, FA becomes non-Gaussian at high values, whereby the same amount of noise in individual eigenvalues will lead to less variability with high FAs compared with low FAs. This non-Gaussian behavior of FA may have contributed to its stronger concordance.

Two potential sources of error in our study are inaccuracies in image coregistration and variability in different diffusion pulse sequences across platforms. Inaccuracies in coregistration would decrease the observed correlations between ROIs, which is expected to be greater between platforms. FA, $\lambda_{\parallel}$, and $\lambda_{\perp}$ have greater regional variability than ADC, which would suggest that FA and component diffusivities would be more susceptible to coregistration error and thus have weaker concordance correlations. We have, in fact, observed the opposite—stronger correlations with regional FA and component diffusivities than ADC. We used identical pulse sequences within platforms and matched the pulse sequences between platforms. The effect of a less well-matched pulse sequence is not known.

These observations regarding comparability suggest that in multicentered studies, it may be best to rank regional FA followed by component diffusivities as a higher priority outcome measure than regional ADC. Furthermore, our observations suggest that ROI analyses are sufficiently reproducible to apply in multicentered studies. This is of significance especially in MS, because subtle changes within small regions of the brain may go undetected in a whole-brain histogram analysis. Furthermore, changes in DTI values have been shown pathologically to reflect alterations in tissue integrity.[18] We have evaluated 16 ROIs in 2 healthy controls and found similar concordance in both individuals (Fig 2). Although it is unlikely that our observations would change with additional

healthy controls, the cross-scanner comparability of DTI measures in individuals with MS is not known. Most important, all the measures from 16 ROIs were used to estimate Lin concordance coefficient. Due to small sample sizes, we did not perform statistical comparisons among the concordance coefficients.

Partial volume averaging between fiber populations of different orientations can lead to systematic dependence of diffusion tensor properties on parameters such as voxel volume, shape, or location.[19] Although this study suggests the feasibility of a multicenter DTI trial, the concordances may vary with changes in scan parameters.

Longitudinal Scan and Software Upgrade Comparison

We included a longitudinal scan-rescan analysis, in which we found FA concordance similar to that observed between magnets. Component diffusivities and ADC values were modestly more reproducible on scan-rescan than between scanners. This difference likely reflects cross-scanner differences in the absolute measures of diffusivity, a difference that is greatly reduced when using a scalar measure such as FA. The similarity of FA concordance values between and within magnets suggests that the remaining variability is likely secondary to either biologic variability or errors in co-registration. Scan-rescan reproducibility of DTI measures with 3T scanners has been previously published and has been satisfactory for FA and ADC values ($\leq$5%).[20,21] In our study, we have additionally determined regional values for diffusivity measures and found them to be comparable.

In the longitudinal scan-rescan component of our study, scanner upgrade led to a slight shift in FA values, which appeared to be driven predominantly by a small change in $\lambda_{\perp}$. Despite slight changes in diffusion values, DTI scans retained good concordance following the system upgrade. This is important because it is difficult to maintain specific platform software and hardware configurations during the course of a prolonged longitudinal study. This observation suggests that longitudinal control subjects may be useful in longitudinal studies of patients with disease, particularly across system upgrades.

Our studies used a high-field magnet (3T) and a high-angular-resolution pulse sequence (33 noncollinear directions), which have been found, in previous studies, to provide robust diffusion tensor estimates.[17] Because low-angular-resolution (eg, 6-direction DTI) decreases the reliability of tensor estimates, it is likely that concordance would be lower. A 1.5T system will, in general, demonstrate a lower signal intensity–to-noise ratio but less susceptibility-related artifacts than a 3T system. The choice of field may, therefore, hinge on whether the ROI is located in areas of brain that have susceptibility artifacts, such as the medial temporal lobe.[9] A previous scan-rescan study of 8 healthy adults by using a single 1.5T magnet and a 60-direction DTI pulse sequence found 3%–6% within-patient CV in FA and MD in different organized fiber tracts.[22] When the authors re-analyzed their data by using only 12 directions, there was little change in CV, though the tracts were significantly smaller and had a higher FA and lower MD.

Another scan-rescan study on 10 healthy adults by using two 1.5T scanners, a 6-direction DTI pulse sequence, and an ROI approach found a within-scanner CV of 1.9% for FA in the corpus callosum.[23] The non-Gaussian distribution of high FA values in the corpus callosum may have contributed to the low CV. In the same study, CV increased to 4.5% when comparing between scanners.[23] With the exception of FA, our CVs across scanners within white matter are similar to those reported in these other studies.[22,23] Our higher CV for FA may relate to inclusion of a broad spectrum of FA values, whereas previous studies focused only on highly organized white matter regions, which have high FA values.

Our results are in agreement with a recent small study reporting greater homogeneity in corpus callosum FA and $\lambda_{\perp}$ values compared with $\lambda_{\parallel}$ and mean diffusivity across different centers.[11] The authors' observations were based on the comparison of 5 subjects on 2 Siemens scanners. Our experimental design expanded these observations by including the platforms of 2 different manufacturers in 5 different centers. Furthermore, in the same analysis, we have incorporated a longitudinal study and software upgrade, a situation that closely mimics real-life scenarios. We have also not limited our analysis to the corpus callosum (high anisotropic region) alone but included other areas of the brain with a diverse range of anisotropic properties. A second study,[12] which compared both scan-rescan and intercenter reliability, also validated the homogeneity of FA values. As with the previous one, this study did not compare platforms from different manufacturers, and the authors' analysis was restricted to white matter regions of the brain.

A recent study that evaluated brain atrophy measurements in patients with MS, as estimated by 5 different centers, also reported an intercenter concordance value of 0.95 for the percentage of brain volume change and 0.94 for normalized brain volume.[24] The concordance values for brain atrophy are very similar to those observed in our analysis for DTI measures. Adequate characterization of variability among scanner platforms will prove important for multicenter DTI-based trials. Given a larger number of subjects, voxel-based methods such as those used by Takao et al[25] may prove to be more powerful. However, this study shows that even with a limited dataset, it is possible to assess the reproducibility of diffusion tensor measures across a number of platforms.

Conclusions

We have observed strong correlations in DTI measures from different magnets and different manufacturers, with the strongest correlation observed with FA, followed by $\lambda_{\perp}$, and $\lambda_{\parallel}$. Despite significant pulse sequence differences between platforms, DTI values appeared to vary little between platforms. In scan-rescan comparisons on the same magnet, we have found excellent correlations in all DTI values. Our study provides strong evidence for the feasibility of multicentered DTI studies by using a 3T MR imaging and a high-angular-resolution pulse sequence.

Disclosures: Robert Fox—*UNRELATED: Consultancy:* Biogen Idec, Genentech, Novartis, *Grants/Grants Pending:* Biogen Idec, Genentech, *Payment for Lectures (including service on Speakers Bureaus):* Biogen Idec, Genentech, Novartis, Teva Neuroscience, Comments: Speaking fees for lectures; I am not on any speakers bureaus. Douglas Arnold—*UNRELATED: Consultancy:* Bayer Healthcare, Eli Lilly, Genentech, GlaxoSmithKline, Opexa Therapeutics; *Grants/Grants Pending:* Bayer Healthcare,* Biogen Idec,* *Stock/Stock Options:* NeuroRx Research, Comments: Dr Arnold holds stock in NeuroRx Research, the imaging Contract Research Organization for the study. Elizabeth Fisher—*UNRELATED: Consultancy:*

Biogen Idec, Pfizer, Genzyme, Wyeth; *Grants/Grants Pending*: Biogen Idec,* Genzyme,* Wyeth;* *Payment for Lectures (including service on Speaker Bureaus)*: Biogen Idec, Teva.
*Money paid to the institution.

References

1. Basser PJ, Pierpaoli C. **Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI.** *J Magn Reson* 1996;111: 209–19
2. Fox RJ. **Picturing multiple sclerosis: conventional and diffusion tensor imaging.** *Semin Neurol* 2008;28:453–66
3. Cassol E, Ranjeva JP, Ibarrola D, et al. **Diffusion tensor imaging in multiple sclerosis: a tool for monitoring changes in normal-appearing white matter.** *Mult Scler* 2004;10:188–96
4. Ge Y, Law M, Grossman RI. **Applications of diffusion tensor MR imaging in multiple sclerosis.** *Ann N Y Acad Sci* 2005;1064:202–19
5. Pierpaoli C, Basser PJ. **Toward a quantitative assessment of diffusion anisotropy.** *Magn Reson Med* 1996;36:893–906
6. Bammer R, Augustin M, Strasser-Fuchs S, et al. **Magnetic resonance diffusion tensor imaging for characterizing diffuse and focal white matter abnormalities in multiple sclerosis.** *Magn Reson Med* 2000;44:583–91
7. Werring DJ, Clark CA, Barker GJ, et al. **Diffusion tensor imaging of lesions and normal-appearing white matter in multiple sclerosis.** *Neurology* 1999;52: 1626–32
8. Filippi M, Iannucci G, Cercignani M, et al. **A quantitative study of water diffusion in multiple sclerosis lesions and normal-appearing white matter using echo-planar imaging.** *Arch Neurol* 2000;57:1017–21
9. Ciccarelli O, Werring DJ, Wheeler-Kingshott CA, et al. **Investigation of MS normal-appearing brain using diffusion tensor MRI with clinical correlations.** *Neurology* 2001;56:926–33
10. Maldjian JA, Grossman RI. **Future applications of DWI in MS.** *J Neurol. Sci* 2001;186(suppl 1):S55–57
11. Pagani E, Hirsch JG, Pouwels PJ, et al. **Intercenter differences in diffusion tensor MRI acquisition.** *J Magn Reson Imaging* 2010;31:1458–68
12. Vollmar C, O'Muircheartaigh J, Barker GJ, et al. **Identical, but not the same: intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0T scanners.** *Neuroimage* 2010;51:1384–94
13. Besl PJ, McKay ND. **A method for registration of 3-D shapes.** *IEEE Trans Pattern Analysis* 1992;14:239–56
14. Basser PJ, Mattiello J, LeBihan D. **Estimation of the effective self-diffusion tensor from the NMR spin echo.** *J Magn Reson B* 1994;103:247–54
15. Press WH. *Numerical Recipes in C: The Art of Scientific Computing.* 2nd ed., New York: Cambridge University Press; 1992
16. Lin LI-K. **A note on the concordance correlation coefficient.** *Biometrics* 2000; 56:324–25
17. Jones DK. **The effect of gradient sampling schemes on measures derived from diffusion tensor MRI: a Monte Carlo study.** *Magn Reson Med* 2004;51:807–15
18. Fox RJ, Cronin T, Lin J, et al. **Measuring myelin repair and axonal loss with diffusion tensor imaging.** *AJNR Am J Neuroradiol* 2010;32:85–91. Epub 2010 Oct 14
19. Oouchi H, Yamada K, Sakai K, et al. **Diffusion anisotropy measurement of brain white matter is affected by voxel size: underestimation occurs in areas with crossing fibers.** *AJNR Am J Neuroradiol* 2007;28:1102–06
20. Bisdas S, Bohning DE, Besenski N, et al. **Reproducibility, interrater agreement, and age-related changes of fractional anisotropy measures at 3T in healthy subjects: effect of the applied b-value.** *AJNR Am J Neuroradiol* 2008;29:1128–33
21. Jansen JF, Kooi ME, Kessels AG, et al. **Reproducibility of quantitative cerebral T2 relaxometry, diffusion tensor imaging, and 1H magnetic resonance spectroscopy at 3.0 Tesla.** *Invest Radiol* 2007;42:327–37
22. Heiervang E, Behrens TE, Mackay CE, et al. **Between session reproducibility and between subject variability of diffusion MR and tractography measures.** *Neuroimage* 2006;33:867–77
23. Pfefferbaum A, Adalsteinsson E, Sullivan EV. **Replicability of diffusion tensor imaging measurements of fractional anisotropy and trace in brain.** *J Magn Reson Imaging* 2003;18:427–33
24. Jasperse B, Valsasina P, Neacsu V, et al. **Intercenter agreement of brain atrophy measurement in multiple sclerosis patients using manually-edited SIENA and SIENAX.** *J Magn Reson Imaging* 2007;26:881–85
25. Takao H, Hayashi N, Kabasawa H, et al. **Effect of scanner in longitudinal diffusion tensor imaging studies.** *Hum Brain Mapp* 2012;33:466–77