Multicenter Reproducibility Study of Diffusion MRI and Fiber Tractography of the Lumbosacral Nerves

Wieke Haakma, PhD,[1,2,3]* Jeroen Hendrikse, MD, PhD,[1] Lars Uhrenholt, PhD,[2] Alexander Leemans, PhD,[4] Lene Warner Thorup Boel, MD, PhD,[2] Michael Pedersen, PhD,[3] and Martijn Froeling, PhD [1]

Background: Diffusion tensor imaging (DTI) has been applied in the lumbar and sacral nerves in vivo, but information about the reproducibility of this method is needed before DTI can be used reliably in clinical practice across centers.
Purpose: In this multicenter study the reproducibility of DTI of the lumbosacral nerves in healthy volunteers was investigated.
Study Type: Prospective control series.
Subjects: Twenty healthy subjects.
Field Strength/Sequence: 3T MRI. 3D turbo spin echo, and 3.0 mm isotropic DTI scan.
Assessment: The DTI scan was performed three times (twice in the same session, intrascan reproducibility, and once after an hour, interscan reproducibility). At site 2, 1 week later, the protocol was repeated (interweek reproducibility). Fiber tractography (FT) of the lumbar and sacral nerves (L3–S2) was performed to obtain values for fractional anisotropy, mean, axial, and radial diffusivity.
Statistical Tests: Reproducibility was determined using the intraclass correlation coefficient (ICC), and power calculations were performed.
Results: FT was successful and reproducible in all datasets. ICCs for all diffusion parameters were high for intrascan (ranging from 0.70–0.85), intermediate for interscan (ranging from 0.61–0.73), and interweek reliability (ranging from 0.58–0.62). There were small but significant differences between the interweek diffusivity values ($P < 0.0005$). Depending on the effect size, nerve location, and parameter of interest, power calculations showed that sample sizes between 10 and 232 subjects are needed for cross-sectional studies.
Data Conclusion: We found that DTI and FT of the lumbosacral nerves have intermediate to high reproducibility within and between scans. Based on these results, 10–58 subjects are needed to find a 10% change in parameters in cross-sectional studies of the lumbar and sacral nerves. The small significant differences of the interweek comparison suggest that results from longitudinal studies need to be interpreted carefully, since small differences may also be caused by factors other than disease progression or therapeutic effects.
Level of Evidence: 1
Technical Efficacy: Stage 2

J. MAGN. RESON. IMAGING 2018;48:951–963.

Three-dimensional magnetic resonance imaging (MRI) and diffusion tensor imaging (DTI) are important imaging modalities to visualize nerve tissue[1,2] and potential nerve damage.[3,4] In nervous tissue, diffusion of water molecules is greater along the nerve than perpendicular to it. With DTI, this diffusion profile can be quantified by measuring the diffusion signal along multiple diffusion gradient orientations and estimating a diffusion tensor.[5] From the diffusion tensor

View this article online at wileyonlinelibrary.com. DOI: 10.1002/jmri.25964

Received Dec 7, 2017, Accepted for publication Jan 20, 2018.

*Address reprint requests to: W.H., Department of Radiology, University Medical Center Utrecht, Room E.01.126, P.O. Box 85500, 3508 GA, Utrecht, the Netherlands. E-mail: w.haakma@umcutrecht.nl

From the [1]Department of Radiology, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands; [2]Department of Forensic Medicine, Aarhus University, Aarhus, Denmark; [3]Comparative Medicine Lab, Department of Clinical Medicine, Aarhus University, Aarhus, Denmark; and [4]Image Sciences Institute, University Medical Center Utrecht, Utrecht, the Netherlands

Additional supporting information may be found in the online version of this article.

a 3D reconstruction of nervous tissue architecture can be obtained with fiber tractography (FT). DTI has been used in various peripheral nerve studies[6,7] including the lumbar and sacral nerves.[8–10]

MRI and DTI data of the spine is relevant to investigate potential damage due to, for example, disc herniation[11–14] or other degenerative lumbar and sacral disorders.[9,15] These studies have shown that diffusion parameters can provide additional information regarding abnormalities in nervous tissue compared to traditional anatomical scans. However, the use of DTI to investigate peripheral nervous tissue, such as the lumbar and sacral nerves, has not been widely adopted and there are no existing standards of how such data should be analyzed. Furthermore, there are different factors that can affect the diffusion measures, such as the moment of performing the scan, the MR scanner, the scanning resolution, and the manual FT segmentation.[16–18] Although several studies have demonstrated high scan–rescan reproducibility of manual FT segmentation in the brain[19–21] and in peripheral nervous tissue,[22–25] essential information regarding the reproducibility of peripheral nerve DTI measurements is still missing. Such information is needed before DTI can be implemented and used reliably in clinical practice across centers.

The aim of this multicenter study was to investigate and describe different aspects of reproducibility of DTI measurements of the lumbar and sacral nerves using DTI and FT in healthy volunteers.

Materials and Methods

The local Institutional Review Boards approved this study and written informed consent was obtained prior to inclusion. The scans were acquired with two 3T Philips Achieva MRI systems (Philips Healthcare, Best, the Netherlands) at two centers: site 1 and site 2. At each site 10 healthy volunteers were included (in total, 20 healthy volunteers, six females; mean age of 36 years, range 25–60 years). Healthy volunteers were asymptomatic and did not have any previous history related to spinal diseases.

Scan Parameters

At the start of each scan session an anatomical 3D turbo spin echo (3D-TSE) scan was obtained. The scan parameters were: repetition time (TR) = 3000 msec, echo time (TE) = 272 msec, TSE factor 180, startup echoes 4, field of view (FOV) 250 × 250 × 100 mm^3 selected at the level of L3–S2, slice orientation coronal, matrix size 250 × 250 × 100, reconstruction matrix 512 × 512 × 200, resulting in a voxel size of 0.49 × 0.49 × 1 mm^3, number of excitations 2, SENSE factor 2, SPAIR fat suppression with inversion delay = 240 msec, and TR = 3000 msec; total acquisition time 5:03 min. Next, at both sites a series of DTI acquisitions were performed. Due to differences in hardware specifications of the MRI scanners between both sites, the acquisition parameters were also slightly different, as shown in Table 1. First, a 3.0 mm isotropic DTI scan was performed (scan 1) which was repeated once in the same session (scan 2) and once after 1 hour (scan 3). Additionally, and directly after scan 2, a 2.5 mm isotropic DTI scan was performed (scan 4). At site 2, scan 1, scan 2, and scan 3 were repeated after 1 week (week 2) in the same volunteers.

Data Processing and Analysis

First, all data were inspected visually to identify artifacts and evaluate the data quality by two researchers with 4 years of experience (W.H.), and 10 years of experience (M.F.). The DTI datasets were processed using the ExploreDTI diffusion MRI toolbox (www.ExploreDTI.com).[26] Data processing comprised the following steps: 1) A reduced FOV was obtained by selecting a fixed (25) number of slices on both sides of the spine in the coronal plane, which included the spine and its nerve roots from the level of L3 to the sacral nerves; 2) motion and eddy current induced geometrical distortions were corrected, where the diffusion gradients were adjusted using the b-matrix rotation[27]; 3) diffusion tensors and subsequently diffusion parameters (fractional anisotropy [FA], mean diffusivity [MD], axial diffusivity [AD], and radial diffusivity [RD]) were calculated with the REKINDLE approach[28]; and 4) a standardized fiber tractography approach was used[1] to reconstruct the 3D nerve architecture. Fiber tract pathways were generated by whole volume seeding (seed distance 2.0 × 2.0 × 2.0 mm^3). Stopping criteria of the tractography were FA >0.05, maximum fiber angle change per 1 mm step of 30 degrees, with a minimum fiber length of 10 mm. Two "AND" ROIs were used to select a segment of 3 cm at each level at each side (L3–S2) starting at the level where the nerves were branching from the spine (Fig. 1, "segments with ROIs"). The first ROI was placed at the position where the nerve branched from the spine and the second 10 slices lower. The specified FA range together with two "AND" ROIs allowed to track the nerve roots in a reproducible and reliable manner.[8] For each of the 3 cm segments (10 nerves per scan) average values of the diffusion parameters (FA, MD, AD, and RD) were calculated.

Experimental Procedures

FT results of the scans were visually compared with the maximal intensity projection (MIP) of the 3D TSE to evaluate their description of the 3D lumbar and sacral nerve anatomy (Fig. 1) by two researchers with 4 years of experience (W.H.), and 10 years of experience (M.F.). In total, four different experiments were conducted to test for 1) interresolution (scan 2 vs. scan 4), 2) intrascan (scan 1 vs. scan 2), 3) interscan (scan 2 vs. scan 3), and 4) interweek (scans of week 1 vs. scans of week 2) reproducibility. As mentioned before, only at site 2 data for experiment 4 was acquired, whereas for the other experiments data from both sites were available and used for analysis. For experiments 1, 2, and 3, where data from two sites were available, we present the results for both sites separately and both sites combined. To emphasize, all experiments in this study concern longitudinal assessment and no intersystem comparisons were made.

Statistical Analysis

Statistical analyses were performed using the SPSS Statistics v. 21.0 (Chicago, IL) and comprised two steps.

First, any variation in diffusion parameters were investigated using one-way repeated measures analysis of covariance (ANCOVA) with site and subject ID as covariates. Three different repeated measures ANCOVAs were performed: 1) interresolution

TABLE 1. DTI Sequences at Site 1 and Site 2 Used to Visualize the Lower Lumbar and Sacral Nerve Roots

Sequence parameters	DTI 3.0 mm Site 1	DTI 3.0 mm Site 2	DTI 2.5 mm Site 1	DTI 2.5 mm Site 2
Sequence	**SS-EPI**	**SS-EPI**	**SS-EPI**	**SS-EPI**
Acquisition plane	Coronal	Coronal	Coronal	Coronal
FOV (mm^2)	336 × 216	336 × 336	320 × 200	320 × 320
TR/TE (ms)	3573 / 45	3451/47	4537 / 45	4400/47
b-value (s/mm^2)	800	800	800	800
Gradient direction	15	15	15	15
EPI factor (ETL)	43	35	47	39
Acquisition matrix	112 × 71	112 × 112	128 × 78	128 × 128
Acquisition voxel size (mm^3) (mm)(mm)	3.0 × 3.0 × 3.0	3.0 × 3.0 × 3.0	2.5 × 2.56 × 2.5	2.5 × 2.5 × 2.5
Half Fourier scan factor	0.62	0.62	0.62	0.62
Slice thickness/ gap (mm)	3.0 / 0	3.0 / 0	2.5 / 0	2.5 / 0
Number of slices	25	25	30	30
Number of excitations	2	2	2	2
Type of fat suppression	SPIR	SPIR	SPIR	SPIR
Total acquisition time (min)	4:20	4:11	5:31	5:21

analysis, 2) combined intra and interscan analysis, and 3) interweek analysis, where $P < 0.05$ was considered significant. Second, the reproducibility of the four different experiments was tested. This was done by means of the interclass correlation coefficient (ICC) together with the 95% confidence interval (CI), to check for reliability, and by means of the Bland–Altman analysis, to check for agreement. ICCs were defined as the single measures two-way random model to test for reliability between DTI scans, ie, how well the measures resemble each other, despite measurement errors. An ICC of <0.40 was considered a low reproducibility, 0.40-0.75 an intermediate to good reproducibility, and >0.75 high reproducibility. Bland–Altman analysis was used to test for agreement between DTI measures, ie, assesses how similar the parameters of repeated measures were. The application conditions of Bland–Altman were verified by a visual check for the absence of correlation between the absolute differences and the mean, and whether the differences were normally distributed for which histogram plots were used.[29] The 95% limits of agreement (LoA) per diffusion parameter were defined as the mean of paired differences ± 1.96 × its standard deviation (SD). Coefficient of variation (CoV) was calculated ($CoV = 100\% \frac{SD}{Mean}$) to determine both intersubject and interscan variability.

Experiment 1: Interresolution Analysis

In this experiment the FT results obtained from the 2.5 mm and the 3.0 mm isotropic scan were compared to each other (scan 2 vs. scan 4). Next, diffusion parameters (ie, FA, MD, AD, and RD) of the scans with different resolution of both sites were compared to each other (interresolution). For all of the diffusion parameters, ICC's were calculated for each site separately and for both sites combined. Finally, Bland–Altman plots were obtained displaying the results of both sites in one graph per diffusion parameter.

Experiment 2: Intrascan Analysis

In this experiment, FT results of scan 1 were compared with scan 2. Diffusion parameters (ie, FA, MD, AD, and RD) of scan 1 and scan 2 with 3 mm isotropic resolution were compared with each other to investigate the intrascan reproducibility, ie, to what extent scans within the same session vary. Then ICCs were calculated for each site and for both sites combined for all of the diffusion

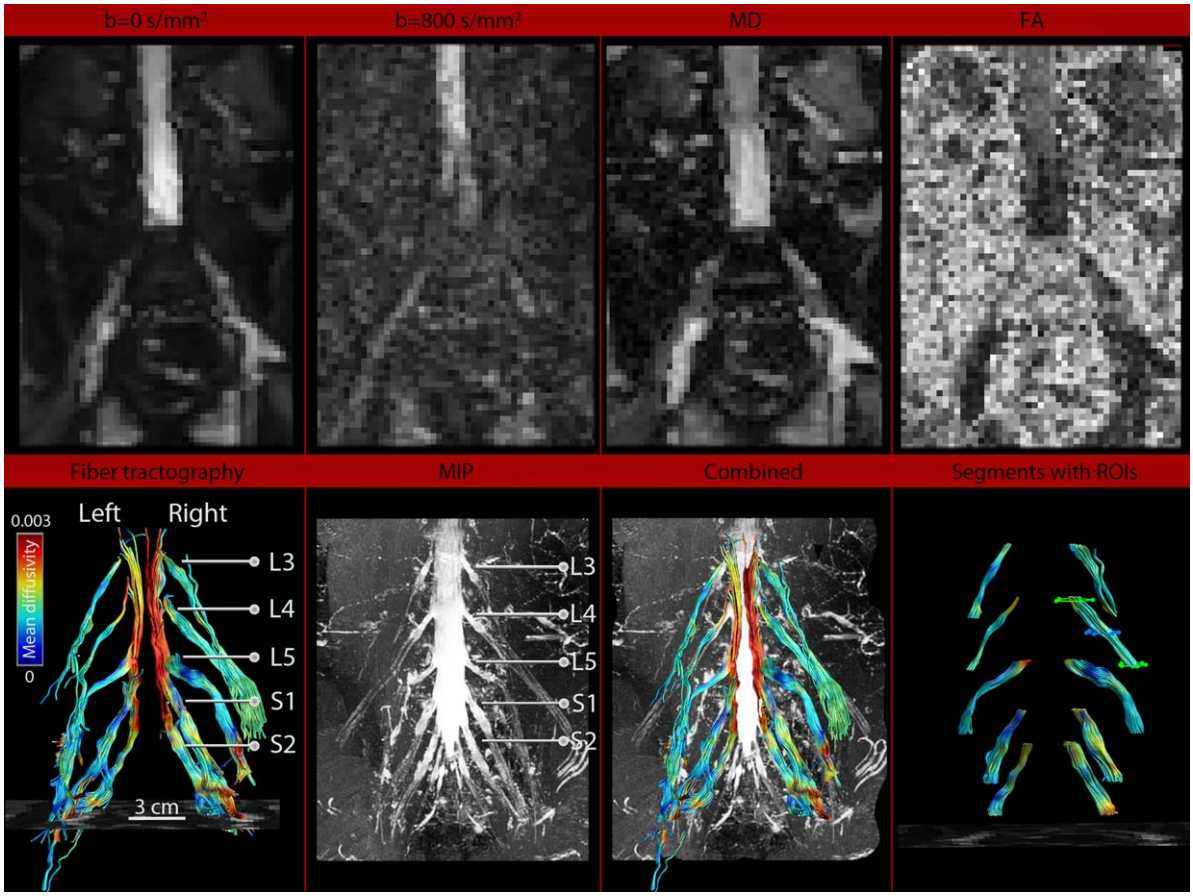

FIGURE 1: Upper row: b = 0 s/mm^2, b = 800 s/mm^2, mean diffusivity (MD), and fractional anisotropy (FA) map. Bottom row: Fiber tractography (FT) results of the lumbosacral nerves, a maximal intensity projection (MIP), FT and MIP combined, and display of the segments at each level and "SEED" (blue) and "AND" (green) region of interest (ROI) placement.

parameters. Finally, Bland–Altman plots were obtained displaying the results of both sites in one graph per diffusion parameter.

Experiment 3: Interscan Analysis

In this experiment the FT results of scan 2 were compared with scan 3. Diffusion parameters (ie, FA, MD, AD, and RD) of scan 2 and scan 3 were compared with each other, which were made with a timeframe of 1 hour between each other to check the interscan reproducibility, ie, to what extent scans on two timepoints vary. ICCs were calculated for each site and for both sites combined for all of the diffusion parameters. Finally, Bland–Altman plots were obtained displaying the results of both sites in one graph per diffusion parameter.

Experiment 4: Interweek Analysis

The FT results were compared between scans of week 1 and scans of week 2. Diffusion parameters (ie, FA, MD, AD, and RD) of the scans obtained in week 1 and week 2 (obtained 1 week later) were compared at site 2. Scan 1 of week 1 was compared with scan 1 of week 2, scan 2 of week 1 with scan 2 of week 2, and scan 3 of week 1 with scan 3 of week 2 to investigate the interweek reproducibility. ICCs were calculated for each site and for both sites combined for all of the diffusion parameters. Finally, Bland–Altman plots were obtained displaying the results of both sites in one graph per diffusion parameter.

Intersubject vs. Interscan Variability and Sample Size Calculation

Box-and-whisker plots were made to determine the variability between subjects and between scans for each diffusion parameter (ie, FA, MD, AD, and RD) and each level (L3–S2). Plots were obtained both for site 1 where scan 1, scan 2, and scan 3 were compared, and for site 2 where the results of the scans of week 1 and week 2 were compared. CoV was calculated for each level (ie, L3–S2) for both subjects (intersubject variability) and scans (ie, scan 1, scan 2, and scan 3 for site 1, and the scans of week 1 and week 2 for site 2). These values were then compared. Finally, the sample size was calculated using $sample\ size = \frac{(r+1)}{r}\frac{CoV^2(z_1-\beta+z_{\alpha/2})^2}{E^2}$, with a significance level of $\alpha = 0.05$ ($z_{\alpha/2} = 1.96$), a desired power of $\beta = 0.8$ ($z_{1-\beta} = 0.84$), the assumption of an equal number of cases and controls ($r = 1$), and an effect size E of 5% and 10%.[30] The CoVs used are the intersubject CoV of site 1 reported in Table 3.

Results

Data Processing and Analyzing

All scans were completed successfully. Based on visual inspection, all data had sufficient data quality and did not contain any artifacts that could interfere with the analysis. Raw data of the diffusion scans are shown in Fig. 1, upper

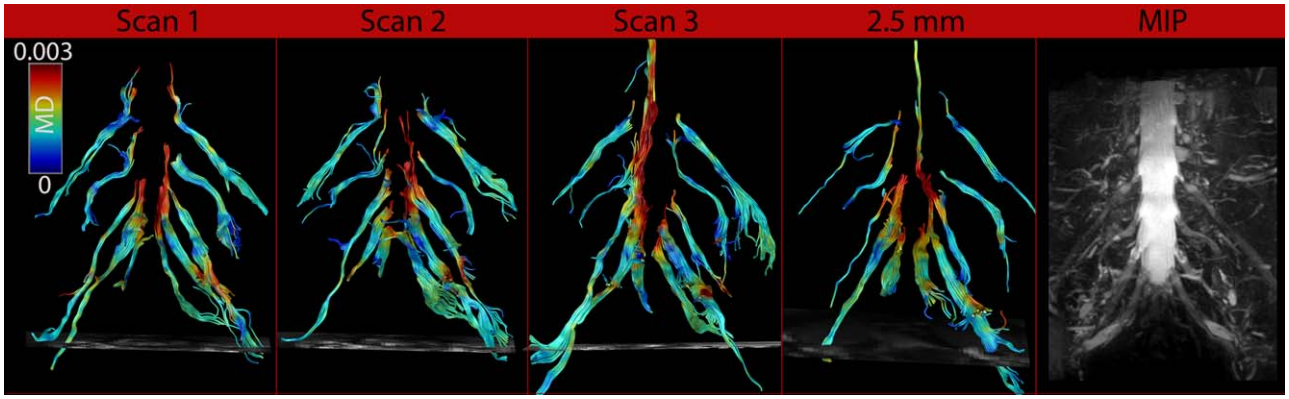

FIGURE 2: Fiber tractography results of DTI data of scan 1, scan 2, scan 3, scan of 2.5 mm isotropic protocol, and a corresponding intensity projection (MIP) of site 1 displayed as mean diffusivity (MD) color-encoded maps.

row. Lumbar (L3–L5) and sacral nerves (S1–S2) were identified in each subject in the anatomical MIP scans and the locations of the nerves matched the locations of the nerves on the DTI scans. An example is shown in Fig. 1, lower row. In two subjects at site 1, it was not possible to reconstruct S2, since it was not included in the FOV. Therefore, in total 96 nerve segments at site 1 and 100 nerve segments at site 2 were reconstructed for each of the four scans. The application conditions for Bland–Altman were met, as there was no correlation between the mean and the differences, and histogram plots showed normal distributions. However, in eight cases the histogram plots showed long tails (ie, present in interresolution, interscan, and interweek).

Experiment 1: Interresolution Analysis

Figure 2 shows an example of FT of the 2.5 mm isotropic and the 3.0 mm isotropic DTI data. Overall, FT results of the 2.5 mm isotropic protocol were comparable to the 3.0 mm isotropic protocol (Supplemental Materials 1 and 2). However, in some cases the nerves of scans obtained from the 2.5 mm isotropic protocol were more difficult to track compared to those from the 3.0 mm isotropic scan (site 1: P2, P9, and P10, site 2: P6, P7, and P10). Interresolution mean diffusion parameters and ICCs are displayed in Table 2. There were small differences in MD, AD, and RD between the two resolutions ($P < 0.0005$) with higher values (up to 0.10×10^{-3} mm^2/s) for the 3.0 mm isotropic scan. ICCs were intermediate to good and on average 0.61. The lowest ICC was for AD at site 1 (0.50, with a 95% CI of 0.34–0.64), and the highest for RD at site 2 (0.75, with a 95% CI of 0.65–0.83). The top row of Fig. 3 shows the Bland–Altman plots displaying the interresolution agreement. LoA were −0.078–0.055, $(-0.2245–0.308) \times 10^{-3}$ mm^2/s, $(-0.286–0.382) \times 10^{-3}$ mm^2/s, and $(-0.203–0.291) \times 10^{-3}$ mm^2/s, for FA, MD, AD, and RD, respectively.

Experiment 2: Intrascan Analysis

Figures 2 and 4 show similar FT results for DTI scan 1 and scan 2 of two subjects at site 1 (Fig. 2) and site 2 (Fig. 4, week 1: A, and B, week 2: E and F). The architectural configuration was similar to the MIP (Fig. 4D,H). This was also confirmed in all the other subjects (Supplemental Materials 1 and 2). Intrascan mean diffusion parameters and ICCs are displayed in Table 2. There were no significant differences in any of the diffusion values between scan 1 and scan 2. The ICCs for all diffusion parameters for both sites were high, on average 0.81. The lowest ICC was for FA at site 2 (0.70, with a 95% CI of 0.58–0.78), and the highest was for MD at site 2 (0.85, with a 95% CI of 0.78–0.90). The second row of Fig. 3 shows the Bland–Altman plots displaying the intrascan agreement. LoA were $-0.0591–0.0623$, $(-0.152–0.157) \times 10^{-3}$ mm^2/s, $(-0.185–0.193) \times 10^{-3}$ mm^2/s, and $(-0.155–0.158) \times 10^{-3}$ mm^2/s for FA, MD, AD, and RD, respectively.

Experiment 3: Interscan Analysis

Figures 2 and 4 show similar FT results for DTI scan 2 and scan 3 of two subjects at site 1 (Fig. 2) and site 2 (Fig. 4, week 1: B, and C, week 2: F and G). The architectural configuration was similar to the MIP (Fig. 4D,H). This was also confirmed in all the other subjects (Supplemental Materials 1 and 2). Interscan mean diffusion parameters and ICCs are displayed in Table 2. There were no significant differences in any of the diffusion values between scan 2 and scan 3. However, the analysis showed significant differences in MD and AD between sites ($P = 0.035$, and $P = 0.016$, respectively). The ICCs for all diffusion parameters for both sites were intermediate to good, on average 0.68. The lowest ICC was for AD at site 2 (0.62 with a 95% CI of 0.48–0.73), and the highest was for MD at site 1 (0.73, with a 95% CI of 0.61–0.81). The third row of Fig. 3 shows the Bland–Altman plots displaying the interscan agreement on the third line. LoA were $-0.0747–0.0777$, $(-0.225–0.308) \times 10^{-3}$ mm^2/s, $(-0.286–0.382)$

TABLE 2. Experiments 1, 2, 3, and 4: Interresolution, Intrascan, Interscan, and Interweek Reproducibility

				Diffusivity ($\times 10^{-3}$ mm^2/s)		
		Site	FA (mean ± SD)	MD (mean ± SD)	AD (mean ± SD)	RD (mean ± SD)
	Week 1					
	Scan 1	1	0.30 ± 0.04	1.30 ± 0.14	1.72 ± 0.17	1.08 ± 0.13
		2	0.31 ± 0.04	1.31 ± 0.14	1.76 ± 0.16	1.09 ± 0.13
		c	0.30 ± 0.04	1.30 ± 0.14	1.74 ± 0.16	1.08 ± 0.13
	Scan 2	1	0.30 ± 0.05	1.30 ± 0.14	1.73 ± 0.16	1.09 ± 0.13
		2	0.31 ± 0.04	1.31 ± 0.14	1.76 ± 0.16	1.08 ± 0.13
		c	0.31 ± 0.04	1.31 ± 0.14	1.75 ± 0.16	1.09 ± 0.13
	Scan 3	1	0.30 ± 0.04	1.27 ± 0.14	1.70 ± 0.17	1.06 ± 0.14
		2	0.31 ± 0.05	1.28 ± 0.12	1.72 ± 0.15	1.06 ± 0.12
		c	0.31 ± 0.05	1.28 ± 0.13	1.71 ± 0.16	1.06 ± 0.13
	Scan 4	1	0.31 ± 0.04	1.30 ± 0.19	1.74 ± 0.21	1.08 ± 0.18
		2	0.32 ± 0.04	1.23 ± 0.13	1.66 ± 0.15	1.01 ± 0.13
		c	0.32 ± 0.04	1.26 ± 0.17	1.70 ± 0.19	1.04 ± 0.16
	Week 2					
	Scan 1	2	0.32 ± 0.04	1.28 ± 0.15	1.72 ± 0.18	1.05 ± 0.14
	Scan 2	2	0.32 ± 0.04	1.27 ± 0.15	1.71 ± 0.18	1.04 ± 0.15
	Scan 3	2	0.32 ± 0.04	1.26 ± 0.14	1.71 ± 0.18	1.04 ± 0.14
Exp. 1	ICC	1	0.69 (0.56–0.78)	0.55 (0.39–0.67)	0.50 (0.34–0.64)	0.58 (0.43–0.70)
		2	0.69 (0.57–0.78)	0.74 (0.63–0.81)	0.67 (0.55–0.77)	0.75 (0.65–0.83)
		c	0.69 (0.61–0.76)	0.60 (0.50–0.68)	0.53 (0.42–0.62)	0.63 (0.54–0.71)
Exp. 2	ICC	1	0.79 (0.69–0.85)	0.81 (0.73–0.87)	0.82 (0.74–0.88)	0.80 (0.72–0.86)
		2	0.70 (0.58–0.78)	0.85 (0.78–0.90)	0.83 (0.75–0.88)	0.84 (0.77–0.89)
		c	0.75 (0.68–0.80)	0.83 (0.78–0.87)	0.82 (0.77–0.87)	0.82 (0.77–0.86)
Exp. 3	ICC	1	0.64 (0.51–0.74)	0.73 (0.61–0.81)	0.72 (0.61–0.81)	0.72 (0.61–0.80)
		2	0.61 (0.47–0.72)	0.69 (0.59–0.78)	0.62 (0.48–0.73)	0.70 (0.59–0.79)
		c	0.63 (0.53–0.70)	0.71 (0.63–0.77)	0.68 (0.59–0.75)	0.71 (0.63–0.77)
Exp. 4	ICC	2	0.58 (0.50–0.65)	0.60 (0.52–0.67)	0.57 (0.49–0.64)	0.62 (0.54–0.68)

Mean diffusion parameters (FA, MD, AD, and RD) of DTI scans (scan 1 = 2.5 mm isotropic and, scan 2, 3, and 4 = 3.0 mm isotropic scanned within the same scan session and after 1 hour) of two sites based on measurements obtained from the level of L3-S2 (per scan; $n = 96$ at site 1 ('1'), $n = 100$ at site 2 ('2'), and $n = 196$ at site 1 and site 2 combined ('c')) and intraclass correlation coefficients (ICC) with 95% confidence interval.

$\times 10^{-3}$ mm^2/s, and (−0.203–0.291) $\times 10^{-3}$ mm^2/s for FA, MD, AD, and RD respectively.

Experiment 4: Interweek Analysis

Figure 4 shows similar FT results for a DTI scan of a control subject at site 2 between weeks (week 1: A, B, and C, week 2: E, F, and G). The architectural configuration was similar to the MIP (D and H, respectively). This was also confirmed in the other subjects at site 2 (Supplemental Material 2). Interweek mean diffusion parameters and ICCs are displayed in Table 2. There were small but significant differences between scans obtained in week 1 and week 2 for MD, AD, and RD ($P < 0.0005$), where scans from week two had slightly lower diffusion values (−0.03 to −0.05 $\times 10^{-3}$ mm^2/s). ICCs were intermediate and on average 0.59. AD had the lowest ICC (0.57, with a 95% CI of 0.49–0.64), and the highest was that of RD (0.62, with a 95% CI of 0.54–0.68). The bottom row of Fig. 3 shows the Bland–Altman plots displaying the interweek agreement.

TABLE 3. Coefficient of Variation (CoV) for Each Nerve Level (L3-S2) and for All Nerve Levels Together (Overall) for Both Intersubject and Interscan Variability

	CoV (%)											
	L3		L4		L5		S1		S2		Overall	
	Intersubject	Interscan	Intersubject	Interscan	Intersubject	Interscan	Intersubject	Interscan	Intersubject	Interscan	Intersubject	Interscan
Site 1												
FA	11.1	6.9	10.1	7.2	15.3	7.2	10.4	5.6	19.2	8.5	14.9	7.0
MD	8.9	4.3	8.7	5.3	10.5	4.7	8.5	3.8	11.6	5.8	10.8	4.7
AD	8.1	3.7	7.8	4.6	9.2	4.3	8.4	4.0	12.3	5.3	9.8	4.3
RD	10.3	5.4	10.1	6.3	12.9	5.7	9.2	4.1	12.4	6.6	12.5	5.6
Site 2 interweek												
FA	11.5	6.7	11.7	5.3	12.0	8.0	14.7	7.4	13.1	6.9	13.7	7.3
MD	11.7	5.3	8.5	5.1	9.2	5.8	10.1	5.4	9.9	6.0	10.9	5.5
AD	11.4	4.9	8.6	5.0	7.8	4.7	7.9	4.4	9.0	6.0	9.7	5.0
RD	12.7	5.8	9.8	6.0	11.2	7.2	12.5	6.6	11.3	6.4	12.8	6.4

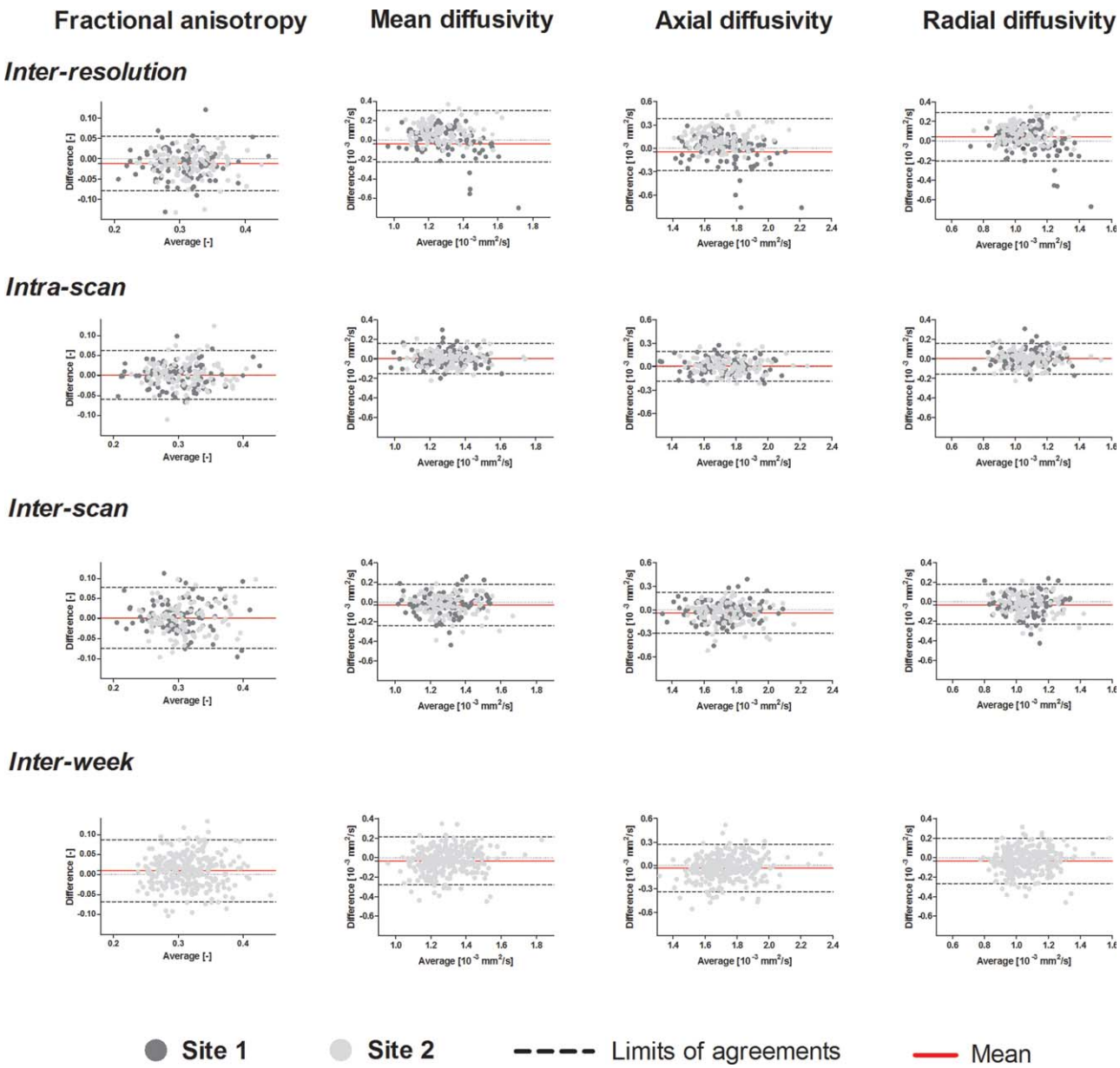

FIGURE 3: Bland–Altman plots of fiber tractography results of DTI data of the lumbosacral nerves of the interresolution, the intrascan, the interscan, and the interweek agreement of the fractional anisotropy (FA), the mean diffusivity (MD), the axial diffusivity (AD), and the radial diffusivity (RD).

LoA were -0.068–0.087, $(-0.279$–$0.213) \times 10^{-3}$ mm^2/s, $(-0.338$–$0.273) \times 10^{-3}$ mm^2/s, and $(-0.267$–$0.200) \times 10^{-3}$ mm^2/s, for FA, MD, AD, and RD, respectively.

Intersubject vs. Interscan Variability and Sample Size Calculation

Figure 5 shows the box-and-whisker plots of site 1 for all diffusion parameters per scan for each of the nerve levels. The range of the boxplots (variability between subjects) is large, whereas the mean of the measurements remains the same over measurements (ie, scan 1, scan 2, and scan 3). Figure 6 shows similar results for the interweek variability between subjects and scans, variability between subjects is large and similar to those of site 1, whereas the mean remains similar between weeks. Table 3 shows that intersubject CoV measurements are ~2 times higher than the interscan CoV measurements. Intersubject CoV measurements for L3–S2 were on average respectively 9.6%, 9.2%, 12.0%, 9.1%, and 13.9% for site 1, and 11.8%, 9.7%, 10.1%, 11.3%, and 10.8% for the interweek comparison at site 2. Interscan CoV measurements for L3–S2 were on average respectively 5.1%, 5.9%, 5.5%, 4.4%, and 6.6% for site 1, and 5.7%, 5.4%, 6.4%, 6.0%, and 6.3% for the interweek comparison. Table 4 shows the calculated sample sizes for an effect size of 5% and 10%. Overall, for FA most samples are needed (ie, 35 samples for an effect size of 10%) whereas AD needs the least (16 samples for an effect size of 10%). The calculated sample sizes vary per nerve, where for

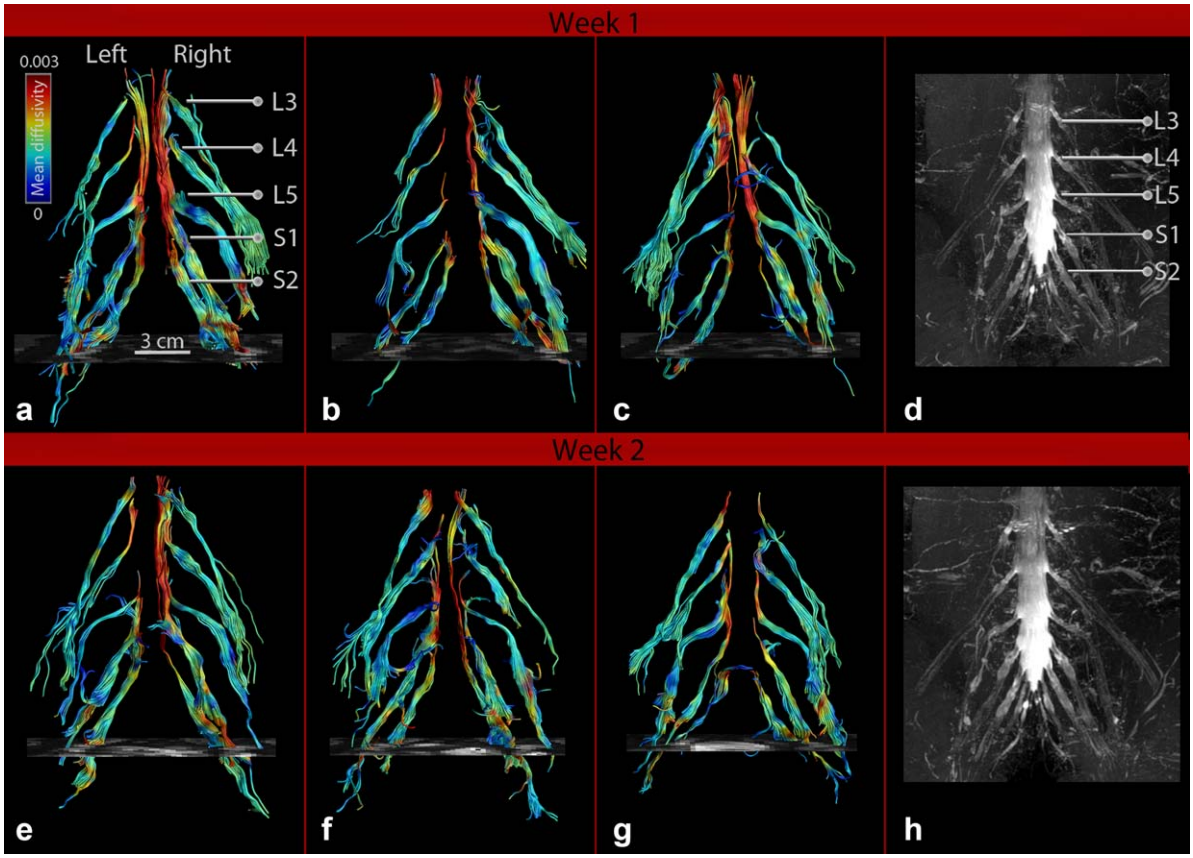

FIGURE 4: Fiber tractography results of DTI data of scan 1, scan 2, and scan 3 (A–C) in week 1, and (D) the corresponding maximal intensity projection (MIP), and for week 2 (E–H, respectively) displayed as mean diffusivity color-encoded maps.

L3 the number of needed samples was lowest and for S2 it was highest, ie, 11–20 and 22–58 for an effect size of 10%, respectively.

Discussion

This study describes the reproducibility of DTI of the lumbosacral nerves in healthy volunteers, which is essential to appreciate its potential for research and applications in a clinical arena. Intrascan reliability was high, whereas interresolution, interscan, and interweek reliability were intermediate to good. However, there were small but significant differences between the diffusivity values of the interresolution scans, and interweek scans. Bland–Altman plots showed high agreement between the intrascan session results, and LoA were larger for interresolution, interscan, and interweek agreement, which may be explained by the long tails of the histogram plots of the differences.

Interresolution reliability, investigated in experiment 1, was intermediate to high and small, but significant differences in diffusion parameters were found (mean value of sites 1 and 2 combined for MD was $1.26 \pm 0.17 \times 10^{-3}$ mm^2/s and $1.31 \pm 0.14 \times 10^{-3}$ mm^2/s, for AD was $1.70 \pm 0.19 \times 10^{-3}$ mm^2/s and $1.75 \pm 0.16 \times 10^{-3}$ mm^2/s, and for RD was $1.04 \pm 0.16 \times 10^{-3}$ mm^2/s and $1.09 \pm 0.13 \times 10^{-3}$ mm^2/s for 2.5 and 3.0 mm isotropic, respectively, $P < 0.0005$). DTI is very sensitive to the effects of noise, which affect the eigenvalues estimation[31] and reliability of tractography.[32] As such, with a voxel size of 2.5 mm isotropic in some cases it proved challenging to reliably track the nerves. We believe that differences in partial volume effects are the cause of the differences we found between the two protocols,[33,34] since the higher resolution showed consistently higher diffusion values. Although partial volume of nerve tissue with the surrounding tissue is different for both resolutions (larger voxel size results in more partial volume effects), we experienced that a voxel size of 3.0 mm isotropic is still sufficient to reconstruct and characterize the lumbar and sacral nerves. As the 3.0 mm isotropic resolution protocol has approximately a 2-fold higher signal-to-noise ratio (SNR), facilitating more reliable estimation of the tensor eigenvalues and FT, we used this protocol for further analyses.

The intrascan reproducibility, investigated in experiment 2, of all diffusion measures was high, implying that the DTI sequence is stable, which is in line with previous findings.[22] Compared to our results, Simon et al showed higher reliability (ICCs were on average FA = 0.93, AD = 0.77, RD = 0.83) in the peroneal and tibial nerve based on three scans obtained at three different timepoints on 1 day.[22] This could be due to that the peroneal and

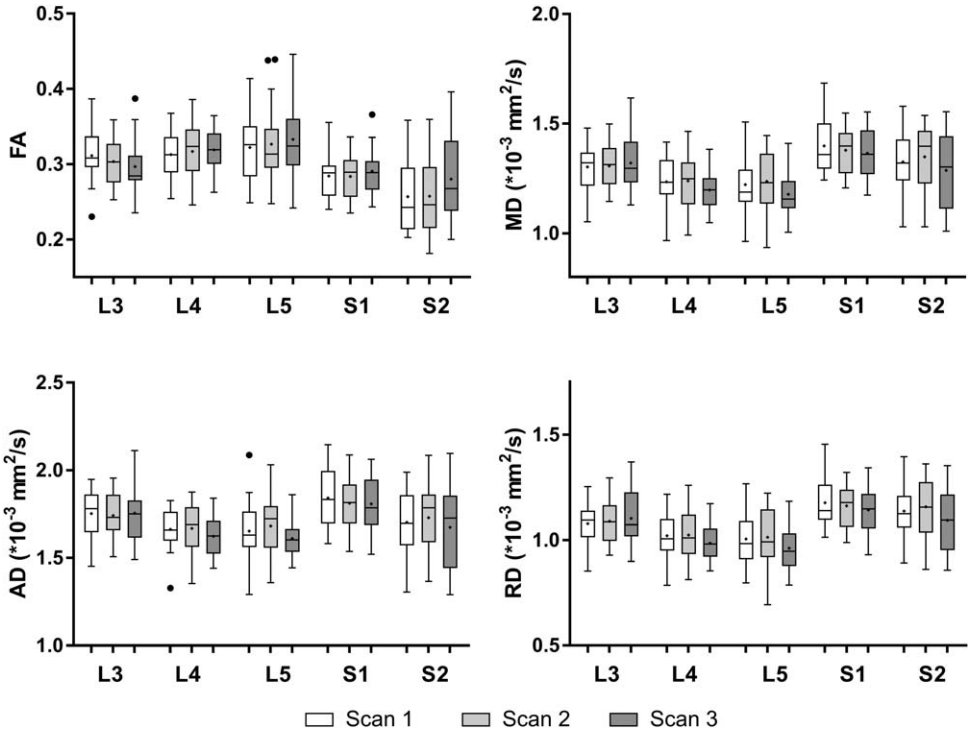

FIGURE 5: Boxplots with whiskers, outliers, medians, and means (indicated with "+") of the intersubject variability over measurements (ie, scan 1, scan 2, and scan 3) of site 1 displayed for all nerve levels (L3–S2) for all diffusion parameters (fractional anisotropy [FA], the mean diffusivity [MD], the axial diffusivity [AD], and the radial diffusivity [RD]).

tibial nerve are approximately two times larger in cross-sectional area than the lumbar and sacral nerves.[22,33] Therefore, it is less sensitive to changes in partial volume effects. Furthermore, the knee region is not affected by bowel movement and respiration, in contrast to the lumbosacral region, which makes the knee region less sensitive to distortions and motion artifacts, which may influence the reproducibility. Our study showed some diversity in ICC values between different positions along the lumbar and sacral nerves (with highest range in AD 0.57–0.95 along different points). In general, the ICC values of the FA in our study were lowest of all parameters, which is to be expected since the FA contains higher-order terms in the eigenvalues, causing higher degrees of uncertainty.[35] Bland–Altman plots of the intrascan and interscan agreement showed similar LoA, as earlier described in a postmortem study investigating the lumbar and sacral nerves, taking into account that diffusion values are ~3–4 times lower than in vivo results.[24]

Interscan reproducibility (experiment 3) was intermediate to high, suggesting that the technique is performing sufficiently for application in research and clinical studies, which was also confirmed by previous findings.[23] Although interweek reproducibility (experiment 4) showed intermediate to high ICCs, we did find small but significant differences between the scan measurements obtained in week 1 and week 2 (MD was $1.30 \pm 0.13 \times 10^{-3}$ mm^2/s and $1.27 \pm 0.15 \times 10^{-3}$ mm^2/s, AD was $1.75 \pm 0.16 \times 10^{-3}$ mm^2/s and $1.71 \pm 0.18 \times 10^{-3}$ mm^2/s, and RD was $1.08 \pm 0.13 \times 10^{-3}$ mm^2/s and $1.04 \pm 0.14 \times 10^{-3}$ mm^2/s for week 1 and week 2, respectively). A study investigating the tibial and sciatic nerve in diabetic patients showed no significant differences between scans performed in week 1 and week 2.[25] However, that study investigated a different anatomical region with different cross-sectional areas of the nerves.

The results of our study suggest that for longitudinal studies investigating a small effect size, it is very important to consider how potential findings should be interpreted, since small differences may also be caused by factors other than disease progression or therapeutic effects. Diffusion parameters are very sensitive to change for many different reasons. Repositioning the subject or the coils can induce small changes to the magnetic field or overall data quality. Partial volume effects are likely to occur, as the nerve size cross-section is generally smaller than the voxel size. Therefore, differences in subject positioning, and changes in surrounding muscle tissue properties,[36] may affect the overall diffusion measures in the nerves.[34] The interweek reproducibility is likely to be more affected compared to the interscan reproducibility, which is also represented by the lower ICC values and wider LoA. Further physiological contributions such as the presence of pulsations of arterial vessels, breathing, and bowel movement could affect the diffusion parameters, since DTI is sensitive to motion.[22,37]

There have been several studies reporting intra- or interscan reproducibility of the application of DTI in

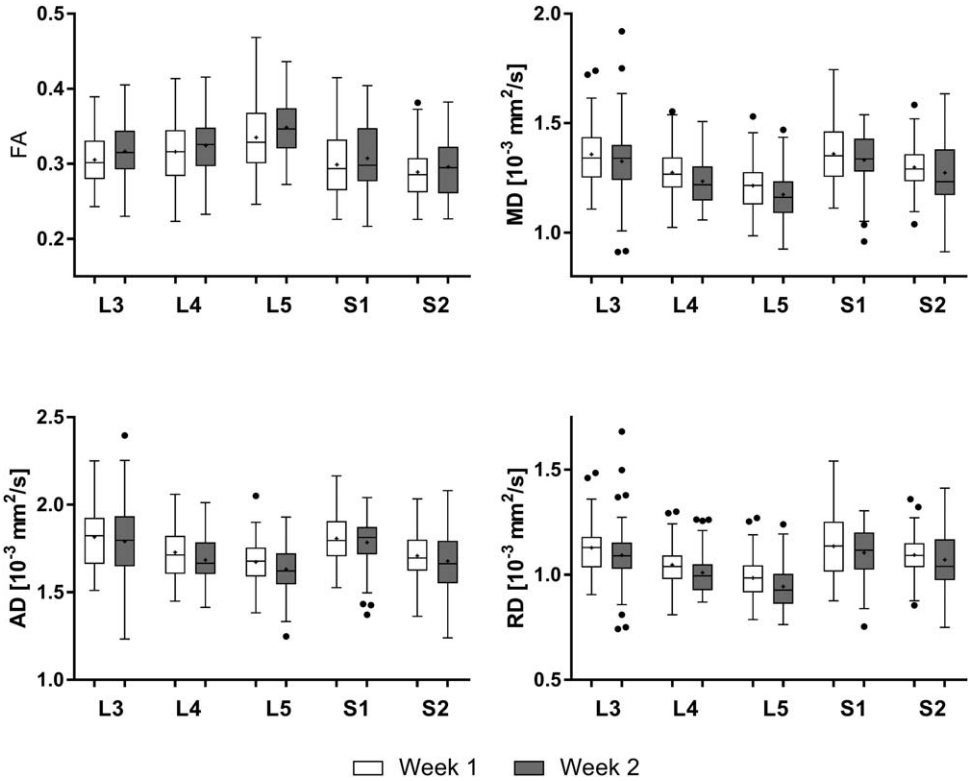

FIGURE 6: Boxplots with whiskers, outliers, medians, and means (indicated with "+") of the intersubject variability over measurements of scans of week 1 vs. week 2 at site 2 displayed for all nerve levels (L3–S2) for all diffusion parameters (fractional anisotropy [FA], the mean diffusivity [MD], the axial diffusivity [AD], and the radial diffusivity [RD]).

peripheral nervous tissue.[22–24] For future studies investigating the lumbosacral nerves with DTI, it is essential to have information regarding the variability in diffusion parameters between subjects vs. the variability between scans within one subject. If the variability between subjects is smaller than the variability between scans within one subject, the technique is not sufficient to use in clinical practice. This study shows that the CoV values between subjects are approximately two times higher than those between scans, indicating that DTI of the lumbosacral nerves is reproducible and sensitive to detect potential changes in cross-sectional studies. The calculated CoV and sample sizes are higher for AD than for RD, and highest for FA, which is in accordance with the expected variation due to noise.[35,38,39] We have shown that large sample sizes are needed to detect an effect size of 5%. Future studies investigating the lumbosacral nerves should consider this in their study design. However, with improvements in data acquisition and data processing, these sample sizes are likely to decrease.

In this study the nerves were manually segmented, which may induce a rater bias. Nevertheless, since the two AND ROIs do not have to be precise and only tracts spanning the 3 cm distance are selected it is unlikely that ROI placement would affect the measures. In future applications,

TABLE 4. Sample Size Calculations for Each Nerve Level (L3-S2) and for All Nerve Levels Together (Overall) for an Effect Size of 5% and 10%

	Sample size											
	L3		L4		L5		S1		S2		Overall	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
FA	78	20	65	17	147	37	68	17	232	58	140	35
MD	50	13	48	12	70	18	46	12	85	22	73	19
AD	42	11	39	10	54	14	45	12	95	24	61	16
RD	67	17	65	17	105	27	54	14	97	25	99	25

automatic segmentation methods may even further improve the reproducibility of these DTI measures. However, at this point there is no robust automated method to accurately segment the nerves running from the level of L3–S2. Qualitative assessment further shows that if the FT results of the different scans are comparable, it supports that the experimental design used in this study is sufficient to segment the lumbar and sacral nerves and sample their diffusion parameters in a reliable way.

Several studies have investigated the lumbosacral nerves in a clinical setting using DTI with applications related to lumbar disc herniation[11–14] and neurogenic bladder disorders.[9] Our study confirms that the application of DTI and FT in the lumbosacral nerves is reliable for such cross-sectional studies.

We included two sites with 3T MRI systems provided by the same vendors. However, we could not assess the intersystem reproducibility, since our subjects differed between sites. However, Table 2 indicates that the diffusion measures are very similar between the two sites. Future studies could determine whether DTI and FT are reproducible even when MRI systems of similar or different vendors are used for the same participants, which is currently unknown for neuro(muscular) applications using diffusion MRI. In addition, it may also be important to determine whether our applied methods are still reproducible when nerves are affected (eg, due to polyneuropathy or nerve trauma also in postmortem applications).[24]

In conclusion, this multicenter study showed that the reproducibility of diffusion measures of the lumbar and sacral nerves was intermediate to high, and that FT results were comparable to each other and to the anatomical scans. This confirms that for cross-sectional studies of lumbar and sacral nerves, DTI can be used reliably in a clinical setting. Sample sizes for cross-sectional studies depend on nerve location, parameter of interest, and effect size and can range between 10 to 232 subjects. The small but significant differences of the interweek comparison highlight that one needs to be careful when interpreting differences in longitudinal studies, since small differences may also be caused by factors other than disease progression or therapeutic effects.

Acknowledgment

Contract grant sponsor: VIDI, Netherlands Organisation for Scientific Research (NWO); contract grant number: 639.072.411 (to A.L.); Contract grant sponsor: NWO; contract grant number: 91712322 (to J.H.); Contract grant sponsor: European Research Council; contract grant number: 637024

We thank all the volunteers who participated in this study.

References

1. Basser PJ, Pajevic S, Pierpaoli C, Duda J, Aldroubi A. In vivo fiber tractography using DT-MRI data. Magn Reson Med 2000;44:625–632.
2. Jones DK, Leemans A. Diffusion tensor imaging. In: Magnetic Resonance Neuroimaging, vol. 711. Modo M, Bulte JWM (eds). New York: Humana Press; 2011:127–144.
3. Takagi T, Nakamura M, Yamada M, et al. Visualization of peripheral nerve degeneration and regeneration: monitoring with diffusion tensor tractography. Neuroimage 2009;44:884–892.
4. Stein D, Neufeld A, Pasternak O, et al. Diffusion tensor imaging of the median nerve in healthy and carpal tunnel syndrome subjects. J Magn Reson Imaging 2009;29:657–662.
5. Basser PJ, Mattiello J, LeBihan D. MR diffusion tensor spectroscopy and imaging. Biophys J 1994;66:259–267.
6. Hiltunen J, Suortti T, Arvela S, Seppä M, Joensuu R, Hari R. Diffusion tensor imaging and tractography of distal peripheral nerves at 3 T. Clin Neurophysiol 2005;116:2315–2323.
7. Heckel a., Weiler M, Xia A, et al. Peripheral nerve diffusion tensor imaging: assessment of axon and myelin sheath integrity. PLoS One 2015;10:1–13.
8. van der Jagt PKN, Dik P, Froeling M, et al. Architectural configuration and microstructural properties of the sacral plexus: a diffusion tensor MRI and fiber tractography study. Neuroimage 2012;62:1792–1799.
9. Haakma W, Dik P, Ten Haken B, et al. Diffusion tensor MRI and fiber tractography of the sacral plexus in children with spina bifida. J Urol 2014;192:927–933.
10. Karampinos DC, Melkus G, Shepherd TM, et al. Diffusion tensor imaging and T2 relaxometry of bilateral lumbar nerve roots: feasibility of in-plane imaging. NMR Biomed 2013;26:630–637.
11. Balbi V, Budzik J-F, Duhamel A, Bera-Louville A, Thuc V, Cotten A. Tractography of lumbar nerve roots: initial results. Eur Radiol 2011;21: 1153–1159.
12. Shi Y, Zong M, Xu X, et al. Diffusion tensor imaging with quantitative evaluation and fiber tractography of lumbar nerve roots in sciatica. Eur J Radiol 2015;84:690–695.
13. Eguchi Y, Oikawa Y, Suzuki M, et al. Diffusion tensor imaging of radiculopathy in patients with lumbar disc herniation: preliminary results. Bone Joint J 2016;98–B:387-394.
14. Chuanting L, Qingzheng W, Wenfeng X, Yiyi H, Bin Z. 3.0T MRI tractography of lumbar nerve roots in disc herniation. Acta Radiol 2014; 55:969–975.
15. Oikawa Y, Eguchi Y, Inoue G, et al. Diffusion tensor imaging of lumbar spinal nerve in subjects with degenerative lumbar disorders. Magn Reson Imaging 2015;33:956–961.
16. Hakulinen U, Brander A, Ryymin P, et al. Intra- and inter-observer variation of region-of-interest methods in quantitative clinical diffusion tensor imaging. In: 5th Eur Conf Int Fed Med Biol Eng; 2011:551–554.
17. Clerici AM, Bono G, Delodovici ML, Azan G, Cafasso G, Micieli G. A rare association of early-onset inclusion body myositis, rheumatoid arthritis and autoimmune thyroiditis: a case report and literature review. Funct Neurol 2013;28:127–132.
18. Li K, Dortch RD, Welch EB, et al. Multi-parametric MRI characterization of healthy human thigh muscles at 3.0 T-relaxation, magnetization transfer, fat/water, and diffusion tensor imaging. NMR Biomed 2014;27:1070–1084.
19. Heiervang E, Behrens TEJ, Mackay CE, Robson MD, Johansen-Berg H. Between session reproducibility and between subject variability of diffusion MR and tractography measures. Neuroimage 2006;33:867–877.
20. Kristo G, Leemans A, De Gelder B, Raemaekers M, Rutten GJ, Ramsey N. Reliability of the corticospinal tract and arcuate fasciculus reconstructed with DTI-based tractography: Implications for clinical practice. Eur Radiol 2013;23:28–36.

21. Bonilha L, Gleichgerrcht E, Fridriksson J, et al. Reproducibility of the structural brain connectome derived from diffusion tensor imaging. PLoS One 2015;10.

22. Simon NG, Lagopoulos J, Gallagher T, Kliot M, Kiernan MC. Peripheral nerve diffusion tensor imaging is reliable and reproducible. J Magn Reson Imaging 2015;43:962–969.

23. Ho MJ, Manoliu A, Kuhn FP, et al. Evaluation of reproducibility of diffusion tensor imaging in the brachial plexus at 3.0 T. Invest Radiol 2017;52:482–487.

24. Haakma W, Pedersen M, Froeling M, Uhrenholt L, Leemans A, Boel LWT. Diffusion tensor imaging of peripheral nerves in non-fixed postmortem subjects. Forensic Sci Int 2016;263:139–146.

25. Vaeggemose M, Pham M, Ringgaard S, et al. Diffusion tensor imaging MR neurography for the detection of polyneuropathy in type 1 diabetes. J Magn Reson Imaging 2017;45:1125–1134.

26. Leemans A, Jeurissen B, Sijbers J, Jones DK. ExploreDTI: a graphical toolbox for processing, analyzing, and visualizing diffusion MR data. Proc Intl Soc Mag Reson Med 2009;17:3536.

27. Leemans A, Jones DK. The B-matrix must be rotated when correcting for subject motion in DTI data. Magn Reson Med 2009;61:1336–1349.

28. Tax CMW, Otte WM, Viergever M a, Dijkhuizen RM, Leemans A. REKINDLE: Robust extraction of kurtosis INDices with linear estimation. Magn Reson Med 2014;73:794–808.

29. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;135–160.

30. Rosner, B. Fundamentals of biostatistics, 7th Ed. Boston: Brooks/Cole; 2011.

31. 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;16:773–785.

32. Froeling M, Nederveen AJ, Nicolay K, Strijkers GJ. DTI of human skeletal muscle: the effects of diffusion encoding parameters, signal-to-noise ratio and T2 on tensor indices and fiber tracts. NMR Biomed 2013;26:1339–1352.

33. Hogan Q. Size of human lower thoracic and lumbosacral nerve roots. Anesthesiology 1996;85:37–42.

34. Vos SB, Jones DK, Viergever MA, Leemans A. Partial volume effect as a hidden covariate in DTI analyses. Neuroimage 2011;55:1566–1576.

35. 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. J Magn Reson Imaging 2007;26:756–767.

36. Froeling M, Oudeman J, Strijkers GJ, et al. Muscle changes detected with diffusion-tensor imaging after long-distance running. Radiology 2015;274:548–562.

37. Ries M, Jones RA, Basseau F, Moonen CT, Grenier N. Diffusion tensor MRI of the human kidney. J Magn Reson Imaging 2001;14:42–49.

38. Froeling M, Oudeman J, van den Berg S, et al. Reproducibility of diffusion tensor imaging in human forearm muscles at 3.0 T in a clinical setting. Magn Reson Med 2010;64:1182–1190.

39. Jones DK, Basser PJ. "Squashing peanuts and smashing pumpkins": How noise distorts diffusion-weighted MR data. Magn Reson Med 2004;52:979–993.