Magnetic Resonance Imaging 109 (2024) 56–66

Contents lists available at ScienceDirect

Magnetic Resonance Imaging

journal homepage: www.elsevier.com/locate/mri

Diffusion tensor imaging within the healthy cervical spinal cord: Within-participants reliability and measurement error

Hussein Al-shaari [a,b,*], Jon Fulford [b], C.J Heales [b]

[a] Diagnostic Radiology Department, College of Applied Medical Sciences, Najran University, Najran 61441, Saudi Arabia
[b] Department of Medical Imaging, Faculty of Health and Life Sciences, The University of Exeter, South Cloisters, University of Exeter, St Luke's Campus, Heavitree Road, Exeter EX1 2LU, UK

ARTICLE INFO

Keywords:
Diffusion tensor imaging
Mean diffusivity
Axial diffusivity
Radial diffusivity
Fractional anisotropy, reliability, spinal cord, within-participants

ABSTRACT

Background: Diffusion tensor imaging (DTI) is a promising technique for the visualization of the cervical spinal cord (CSC) in vivo. It provides information about the tissue structure of axonal white matter, and it is thought to be more sensitive than other MR imaging techniques for the evaluation of damage to tracts in the spinal cord.
Aim: The purpose of this study was to determine the within-participants reliability and error magnitude of measurements of DTI metrics in healthy human CSC.
Methods: A total of twenty healthy controls (10 male, mean age: 33.9 ± 3.5 years, 10 females, mean age: 47.5 ± 14.4 years), with no family history of any neurological disorders or a contraindication to MRI scanning were recruited over a period of two months. Each participant was scanned twice with an MRI 3 T scanner using standard DTI sequences. Spinal Cord Toolbox (SCT) software was used for image post-processing. Data were first corrected for motion artefact, then segmented, registered to a template, and then the DTI metrics were computed. The within-participants coefficients of variation (CV%), the single and average within-participants intraclass correlation coefficients (ICC) and Bland-Altman plots for WM, VC, DC and LC fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD), and radial diffusivity (RD) were determined for the cervical spinal cord (between the 2nd and 5th cervical vertebrae).
Results: DTI metrics showed poor to excellent within-participants reliability for both single and average ICC and moderate to high reproducibility for CV%, all variation dependent on the location of the ROI. The BA plots showed good within-participants agreement between the scan-rescan values.
Conclusion: Results from this reliability study demonstrate that clinical trials using the DTI technique are feasible and that DTI, in particular regions of the cord is suitable for use for the monitoring of degenerative WM changes.

1. Introduction

Quantitative magnetic resonance (MRI) measurements provide information about pathological changes underpinning progression and disability in neurological disease. Diffusion tensor imaging (DTI) is a non-invasive method that can be used for visualizing the spinal cord in vivo. DTI provides information about the tissue structure of axonal white matter, and is thought to be more sensitive in evaluating damage to tracts in the spinal cord than conventional MRI approaches [1,2]. DTI measures the molecular motion of water molecules that diffuse across each voxel of an image in directions both parallel and transverse to the orientation of neural axons and provides a number of metrics that provide different information about the spinal cord tracts. Fractional anisotropy (FA) assesses the degree of anisotropy in the diffusivity of water protons in tissues and is affected by many factors such as changes in water content and the presence of crossing fibers. Mean diffusivity (MD) provides an overall measure of water translational diffusivity. Axial diffusivity (AD) specifies the magnitude and the direction of maximum water diffusion being influenced by longitudinal axonal integrity and radial diffusivity (RD) evaluates the diffusivity properties

Abbreviations: AD, Axial diffusivity; BA, Bland-Altman analysis; CSC, Cervical spinal cord; CV, Coefficient of variation; DTI, Diffusion Tensor Imaging; FA, Fractional anisotropy; ICC, Intraclass correlation coefficient; MD, Mean diffusivity; RD, Radial diffusivity; LOA, Limits of Agreement; CI, Confidence Interval; SCT, Spinal cord toolbox.

* Corresponding author at: College of Health and Life Sciences, Department of Medical Imaging, South Cloisters, St Luke's Campus, Heavitree Road, Exeter, Devon EX1 2LU, UK.
E-mail addresses: ha457@exeter.ac.uk (H. Al-shaari), J.Fulford@exeter.ac.uk (J. Fulford), c.j.heales@exeter.ac.uk (C.J Heales).

https://doi.org/10.1016/j.mri.2024.03.005
Received 6 May 2023; Received in revised form 5 March 2024; Accepted 5 March 2024
Available online 7 March 2024

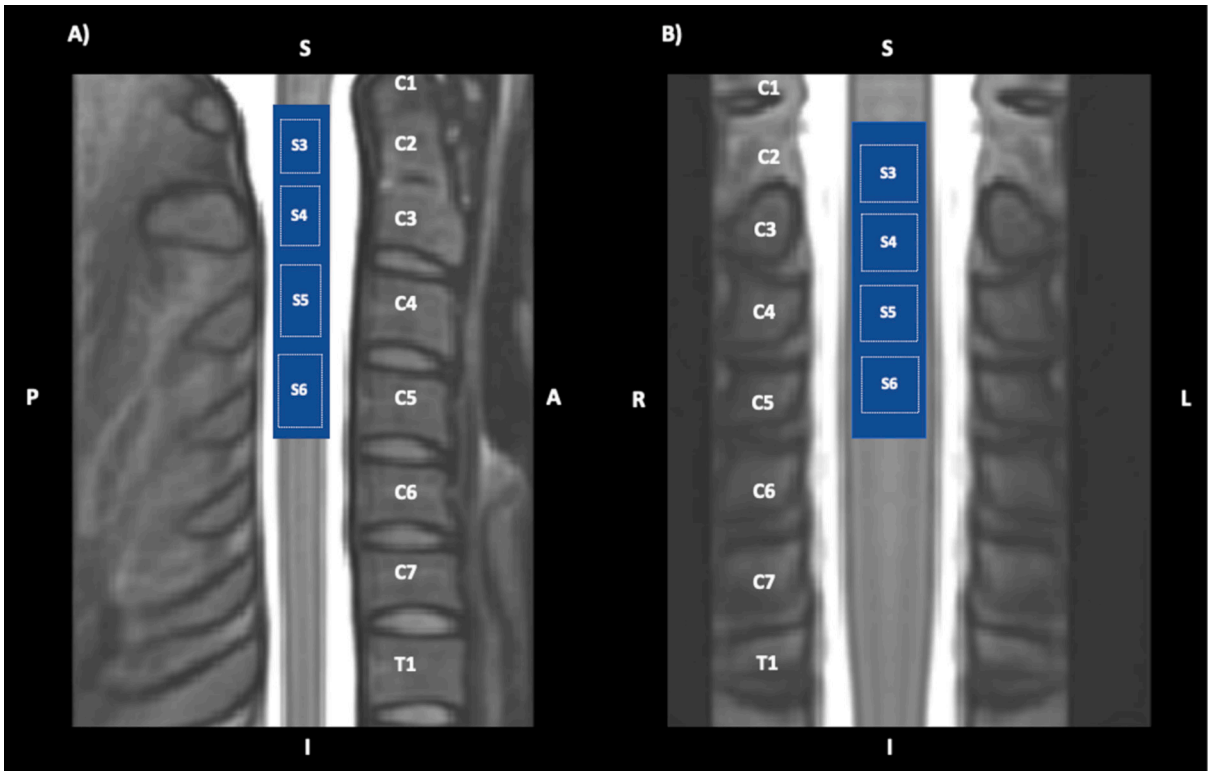

Fig. 1. Illustration indicating the approximate location of C2, C3, C4, and C5 spinal cord vertebral levels with their corresponding cord segments S3, S4, S5, and S6 on the PAM50 T2-weighted spinal cord template (a, midsagittal and b, midcoronal) slices. Using spinal cord toolbox (SCT) probability maps, spinal cord segments were discovered. A = anterior, P = posterior, L = left, R = right, S = superior, I = inferior.

of tissue in the perpendicular axonal structure [3].

Due to the unique anisotropic properties of the spinal cord, DTI may be able to differentiate white matter (WM) from grey matter (GM) and evaluate structural damage to the cord in certain cases [1]. Prior to use clinically it is therefore important to investigate the test–retest reliability of DTI measures in order to ensure the technique is accurately demonstrating pathological changes. Reliability (the variability between repeated measurements on the same subject under the same conditions [4]) of DTI in the cervical spinal cord (CSC) has previously been investigated in a number of studies observing control and clinical groups examining the whole cervical cord and/or across cervical vertebrae levels [5–7]. However, data evaluating the reliability of DTI parameters of the averaged cord segments from the top border of the 2nd cervical vertebra (C2) to the lower border of the fifth cervical vertebra (C5) are still lacking and those regions are commonly impacted in conditions such as multiple sclerosis. An advantage of averaging over a number of vertebrae rather than looking at them individually is the decreased susceptibility to small regions, localised signal variations and any errors associated with the segmentation process. Increasing the volume further to include C6-C7 however, risked signal drop-off as a result of limited coverage of the head-neck coils typically employed. The aim of this study was therefore to assess the within-participants reliability and error magnitude of measurements of all DTI metrics in healthy human CSC over the total WM as well as within subregions, (specifically the ventral column (VC), dorsal column (DC) and lateral column (LC)), for the cord segments spanning from the upper border of C2 to the lower border of C5 to better understand DTI reliability in this key cervical spinal cord locations.

2. Material and methods

The study was approved by the local ethics committee at the University of Exeter. Written informed consent was obtained from each participant prior to data acquisition and collection. Following MRI safety screening each participant underwent MRI scanning twice, with the visits two weeks apart. Every participant was re-scanned at approximately the same time of day as the previous scan.

2.1. Study subjects

A total of twenty healthy controls (10 males, mean age ± sd: 33.9 ± 3.5 years, 10 females, mean age ± sd: 47.5 ± 14.4 years), were recruited over a period of two months. Recruiting was achieved through advertising (posters, University news and social media). Considering the purpose of the study was to examine the reliability of the DTI technique rather than its utility for disease monitoring, participants were recruited from the general population. Inclusion criteria were participants aged between 18 and 70 years. Exclusion criteria were participants who have had any family history of any neurological disorders or contraindications to MRI scanning.

2.2. Image acquisition

All MR data acquisition were performed on a Siemens 3.0 T MR scanner (MAGNETOM Prisma, Siemens Healthineers, Erlangen, Germany) based at the MGNC. This scanner is equipped with 80 mT/m gradients with maximum 200 mT/m/ms slew rate. A 64-channel head/neck was used. A peripheral pulse unit (PPU) oximetry trigger was used to diminish physiological motion artefacts. Patients were scanned in the supine position. The field-of-view (FOV) was centred at the level of the C3/C4 intervertebral disc, spanning four vertebral segments from the upper boarder of C2 to the lower border of C5 in all participants as illustrated in Fig. 1. Thirty-five non-collinear directions of diffusion weighted images (DWI) were acquired at b = 800 s·mm $^{-2}$, one non-DWI (b = 0) measurement and four non-DWI (b = 0) measurements with the opposite phase encoding direction (PED). Axial scans and an anterior-

Table 1
Illustration of the image acquisition parameters of Diffusion tensor image (DTI) and T2*-weighted sequences.

Imaging parameters	Diffusion tensor image	T2*-weighted image
TR (ms)	810	663
TE (ms)	59	14
Image resolution (mm^2)	0.9 × 0.9	0.5 × 0.5
Image matrix (mm^2)	80 × 80	320 × 320
FOV (mm^2)	86	224
Slice thickness (mm)	5	5
Bandwidth (Hz/voxel)	1279	260
Number of slices	20	15
Scan time (min:sec)	2:39	4:45
Flip angle (°)	90	30
Signal averages	5 for b = 0, 1 for b = 800	2

posterior (AP) PED were used. To decrease the sensitivity of the Echo Planar Imaging (EPI) sequence used for diffusion MRI to susceptibility artefacts due to bone: tissue interfaces, a reduced field-of-view (ZOOMit) method was applied using the outer volume suppression (OVS) technique [8], with second order shimming. Additionally, a high-resolution T2*-weighted image was included to provide an anatomical reference using a gradient echo multi-echo (GEME) sequence to allow template-specific analysis of DTI data, in two processing steps with co-registration to the template performed for the anatomical images. Further details of the imaging acquisition parameters are listed in Table 1. An example of the resultant images is shown in Fig. 2.

2.3. Image post processing

Raw data (MAGNETOM Prisma, Siemens Healthineers, Erlangen, Germany) in Digital Imaging and Communication in Medicine (DICOM) format, for T2WI* and DTI were transformed into Neuroimaging Informatics Technology Initiative (NIFTI) as a first step before the analysis. The whole processing pipeline is illustrated in Fig. 3 and detailed below:

The following steps were applied: The first step was correcting the motion of DTI datasets using a motion correction function implemented in the SCT. The *se*cond step for DWI data processing was CSC segmentation using the segmentation function implemented in SCT in which the

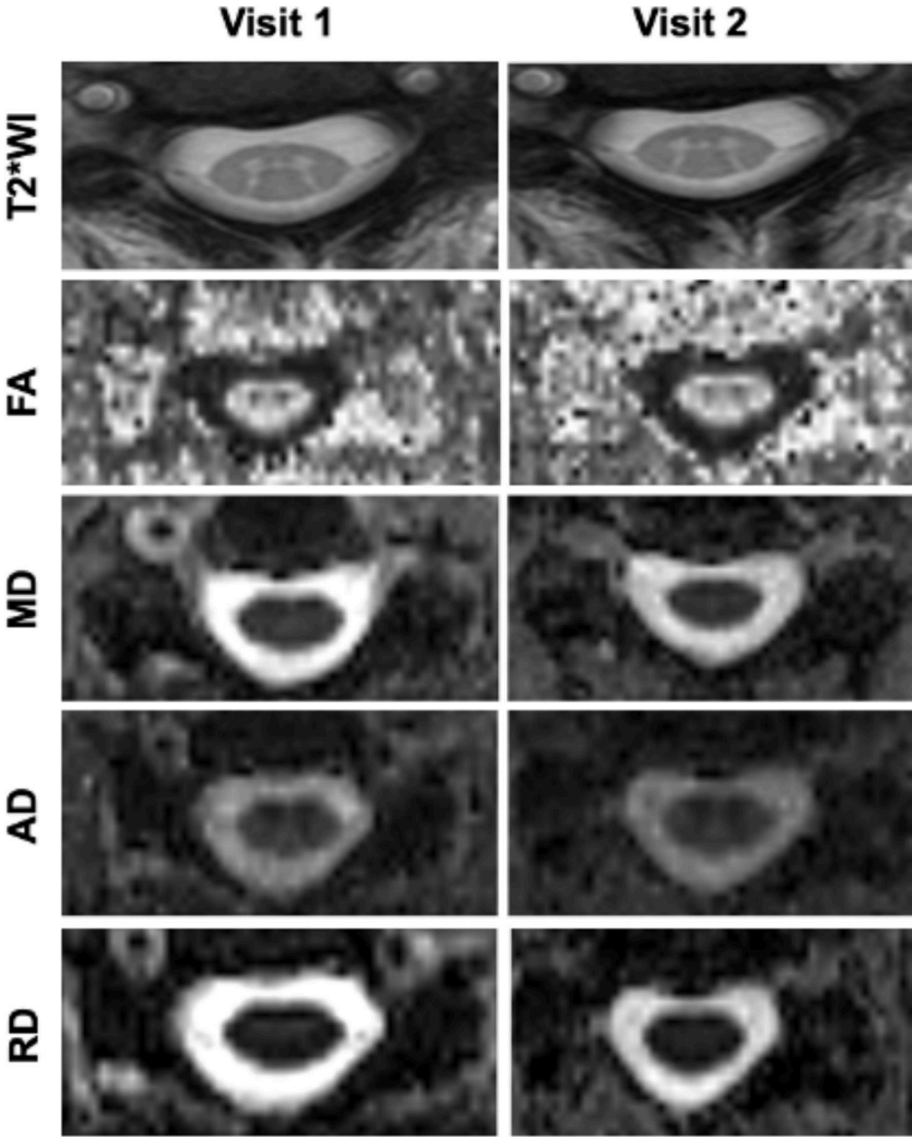

Fig. 2. Representative examples of T2* weighted images (T2*WI), and diffusion tensor image (DTI) metrics maps of one healthy participant at the C3 vertebral-body level (slice number 9). WM is darker than GM on T2*WI, mean diffusivity (MD), and perpendicular diffusivity (RD) maps and brighter than GM on fractional anisotropy (FA) and parallel diffusivity (AD) maps. Cerebrospinal fluid (CSF) is bright on all maps except for FA. Note that contrast between WM and GM can best be appreciated in the T2*WI, FA, and AD maps.

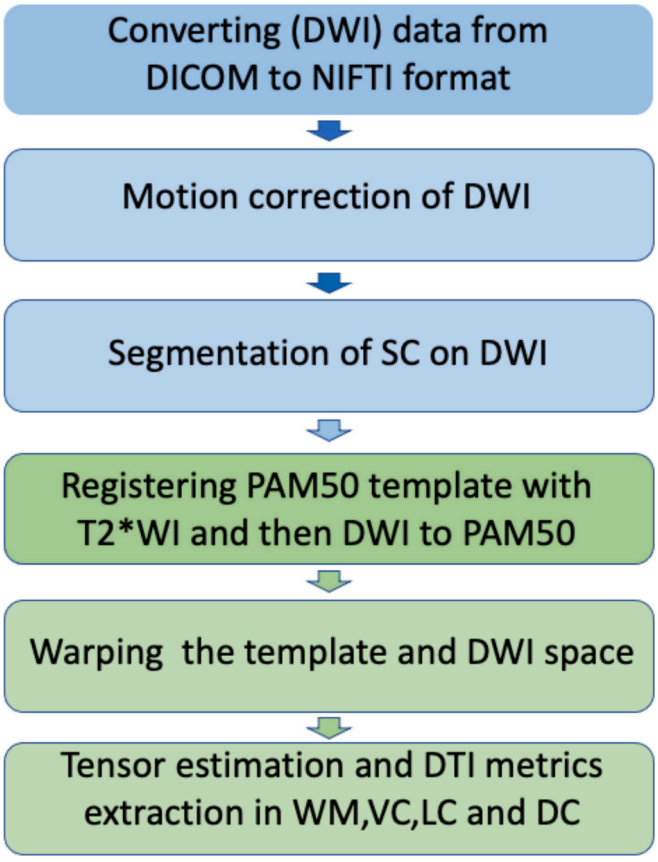

Fig. 3. Illustration of the processing pipeline for diffusion weighted images data (DWI). Metrics were extracted between C2-C5 for white mater (WM), ventral column (VC), right and left lateral columns (LC) and dorsal column (DC) of cervical spinal cord.

entire CSC, spanning four vertebral segments levels from the upper boarder of C2 to the lower boarder of C5 (as in Fig. 1 A and B), was segmented based on the mean DWI image generated from all the different b-values (b = 800 s.mm^{-2}) images for each participant [9]. The third step was mapping the cervical vertebral levels onto the DWI data. This was done by cross-modal registration of the T2*-weighted anatomical data, due to its high resolution, to the mean DWI images using a multi-modal registration method implemented in the SCT. The process was done by applying different affine and non-rigid transformations between the mean of the b = 0 data, T2*-weighted anatomical data, and the PAM50 template [10]. This allowed matching of the vertebral levels of the subject cord to the template shape [9]. The alignment of data to the template offered a robust definition of the intervertebral levels of the spine which then allowed for the evaluation of the average metrics in the spinal cord utilising the atlas-specific (or template-based) method [11], which helped to reduce bias related to PVE. In addition, this method is less sensible to susceptibility distortions [12]. The fourth step was warping the PAM50 spinal cord template to the DWI data using a function implemented in SCT, which finds the subject-to-template composite transformation (forward warping field), then applying the inverse transformation (backward warping field) to the template, thereby mapping the atlas ROIs into individual subject space. The final step was diffusion tensor estimation and metrics calculation using "sct_dmri_compute_dti" and sct_extract_metrics functions, respectively, which implemented in SCT [9]. The calculated DTI metrics included: mean diffusivity which represents the diffusion independent from the orientation of tissue (MD = λ1 + λ2 + λ3/3); axial diffusivity which reveals diffusion in the longitudinal direction of the WM fibers (AD = λ1); radial diffusivity which represents diffusion in the

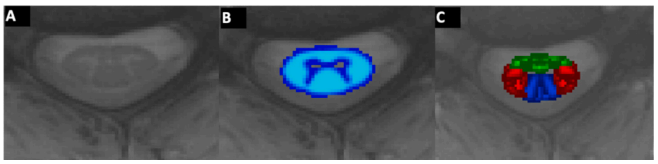

Fig. 4. Representative images revealing (A) T2*WI (A) with (B) probabilistic maps of the white matter (light blue) and (C) WM regions which include ventral columns (green), lateral columns (red), and dorsal columns (blue) following registration to the SCT atlas. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Interpretation of the classification of single and average intra-class correlation coefficient (ICC) values as suggested by Shrout and Fleiss [240], Cicchetti [28].

ICC values	ICC classification
< 0.4	Poor
0.4 < ICC < 0.59	Fair
0.60 < ICC < 0.74	Good
0.75 < ICC < 1	Excellent

perpendicular directions of the WM fibers (RD = λ1 + λ2/2), as well as fractional anisotropy (FA) which measures the degree of the orientation of water protons diffusivity in tissues, as given in the following equation:

$$FA = \sqrt{\frac{(\lambda_1 - \lambda_2)^2 + (\lambda_1 - \lambda_3)^2 + (\lambda_2 - \lambda_3)^2}{2\left(\lambda_1^2 + \lambda_2^2 + \lambda_3^2\right)^2}} \qquad (\text{FA} =)$$

These metrics were averaged for each inter-vertebral level, spanning four vertebral segments between C2-C5 for the WM, VC, LC, and DC. The selected regions of interest (WM, VC, LC, and DC) are shown in Fig. 4 B and C. The DTI metrics were assessed utilising a maximum a posteriori method, which allowed biases related to PVE to be diminished [11].

2.4. Statistical analysis

Data analyses was undertaken using SPSS 13.0 (SPSS Inc., Chicago, USA). Mean and standard deviation (SD) values were calculated by each visit and pooled over the two visits, together with the mean Bland-Altman (BA) difference, limit of agreement (LOA), and differences as percentage, all with 95% Confidence Interval (CI). DTI metrics for WM, VC, DC and LC were measured for the entire volume segment of spinal cord contained between the vertebral levels of the upper limit of C2 and lower limit of C5. A Paired samples *t*-test was used to assess whether there were any differences in participant parameter values between visits, with a significance level of $P \leq 0.05$.

The single and average within-participants intra-class correlation coefficient (ICC) of DTI metrics were analysed to determine the reliability of the repeated scans. The single ICC value illustrates the situation where measurements are performed only once, as is typical in clinical settings. The average ICC value refers to the situation where the same region is evaluated twice, and the final value is the mean of the two measures. The formula for calculating the ICC is [true variance/(true variance + error variance)] [13]. It quantifies the actual proportion of variation associated with the "true" error-free values of individuals (the participant's variance) in comparison to the overall variability. The reliability of test and retest of DTI metrics was assessed based on a single rater, consistent agreement and a two-way mixed-effect model of intra-class correlation (ICC). Higher ICC indicates greater reliability as proposed by Shrout and Fleiss [13], Cicchetti [14] and as presented in Table 2 for both single and average ICC values.

For measurement errors, the coefficient of variation (CV) was measured for each DTI variable to clarify the relative variability of each

Table 3
DTI measurements in whole WM and among different regions of the WM.

	ROI	Visit		Pooled (95%)	Bland-Altman analysis		Within-participants effects t (P)
		1 Mean (95%)	2 Mean (95%)		Difference (D)	LOA	
MD [μm^2/ms]	WM	0.91 (0.87–0.95)	0.94 (0.90–0.97)	0.92 (0.89–0.96)	−0.03 (−0.06 to 0.01)	-0.17 to 0.12	−1.51 (0.15)
	DC	0.95 (0.91–0.99)	0.98 (0.95–1)	0.97 (0.94–1)	−0.03 (−0.06 to 0.01)	−0.18 to 0.13	−1.53 (0.14)
	VC	0.94 (0.89–0.99)	0.95 (0.87–1)	0.94 (0.89–0.99)	0.01 (−0.07 to 0.09)	−0.31to 0.31	0.23 (0.81)
	LC	0.87 (0.82–0.91)	0.90 (0.86–0.94)	0.88 (0.85–0.92)	−0.04 (−0.09 to −0.00)	−0.18 to 0.11	−2.20 (0.04)*
AD [μm^2/ms]	WM	1.88 (1.84–1.92)	1.91 (1.87–1.94)	1.89 (1.86–1.93)	−0.03 (−0.06 to 0.00)	-0.17 to 0.10	−2.06 (0.05)*
	DC	1.99 (1.95–2.04)	2.03 (1.99–2.06)	2.01 (1.98–2.05)	−0.03 (−0.07 to 0.01)	−0.20 to 0.14	−1.64 (0.12)
	VC	1.83 (1.75–1.90)	1.82 (1.76–1.89)	1.82 (1.77–1.88)	0.00 (−0.09 to 0.09)	−0.38 to 0.39	0.06(0.96)
	LC	1.82 (1.77–1.86)	1.86 (1.81–1.91)	1.84 (1.79–1.88)	−0.04 (−0.07 to −0.01)	−0.17 to 0.09	−2.83 (0.01)*
RD [μm^2/ms]	WM	0.43 (0.38–0.48)	0.45 (0.41–0.49)	0.44 (0.40–0.48)	−0.02 (−0.06 to 0.02)	-0.19 to 0.16	−0.95 (0.35)
	DC	0.43 (0.38–0.48)	0.46 (0.42–0.49)	0.44 (0.41–0.48)	−0.03 (−0.07 to 0.07)	-0.21 to 0.16	−1.28 (0.22)
	VC	0.51 (0.43–0.59)	0.50 (0.45–0.55)	0.51 (0.45–0.56)	−0.03 (−0.07 to 0.01)	−0.30 to 0.32	0.34 (0.74)
	LC	0.39 (0.35–0.44)	0.42 (0.38–0.47)	0.41 (0.37–0.45)	−0.03 (−0.07 to 0.01)	−0.20 to 0.14	−1.63 (0.12)
FA	WM	0.71 (0.69–0.74)	0.70 (0.69–0.72)	0.71 (0.69–0.73)	0.01 (−0.02 to 0.03)	−0.09 to 0.10	0.79 (0.44)
	DC	0.73 (0.71–0.76)	0.73 (0.71–0.75)	0.73 (0.71–0.75)	0.01 (−0.02 to 0.03)	−0.09 to 0.10	0.61 (0.55)
	VC	0.65 (0.62–0.69)	0.64 (0.63–0.67)	0.65 (0.72–0.67)	0.00 (−0.03 to 0.03)	−0.12 to 0.13	0.21 (0.83)
	LC	0.72 (0.70–0.75)	0.71 (0.69–0.73)	0.72 (0.70–0.74)	0.01 (−0.01 to 0.03)	−0.09 to 0.11	1.01 (0.33)

measurement as follows: CV = [the within-participants standard deviation]/ within-participants mean ×100%. A CV of <10% is considered to be acceptable indicating that the dependent variable has a relatively small amount of variation [15]. A CV between 11% and 20% is considered to be adequate and indicates a moderate variation. A CV >20% is considered to represent a high amount of variability [16]. To examine the agreement level and evaluate the variation between the two measurements, Bland–Altman (BA) plots with 95% CI were generated by plotting the mean measurement (visit1 + visit2)/2 against the difference in measurements (visit1− visit2). The smaller the differences between repeated measurements the better the agreement.

The signal-to-noise- ratio was estimated using a "difference method" (SNR_{diff}) described by Dietrich et al. (2007) [17] using the "sct_compute_snr" function implemented in SCT [9]. This relies on the calculation of the difference between two images from two identical (repeated) acquisitions [18–22]. In this study, the SNR of both DWI and non-DWI data was measured in the segmented CSC between the 2nd and 5th vertebral levels based on two non-DWI images from the DTI sequence (b = 0 $s.mm^{-2}$) and two DTI images at (b = 800 $s.mm^{-2}$) according to the equation: SNR = [S_{CSC} /SD_{CSC}], where S_{CSC} is the average of the signal mean from the two images (S_1 + S_2/2) and SD_{CSC} is a measure of the differences in standard deviation of signal of the two images given by SD_1-$SD_2/\sqrt{2}$ where SD_1 and SD_2 are the standard deviations of the individual images.

Table 4
Within-participants reliability metrics utilised to evaluate the reliability of DTI metrics in whole WM and among different regions of the WM ($N = 20$).

	ROI	Correlation between -within-participants mean and standard deviation R(P)	Within-participants standard deviation, SD (range)	Within-participants coefficient of variation, CV% (range)	Intraclass correlation coefficient, ICC	
					Single (range)	Average (range)
MD [μm^2/ms]	WM	−0.01 (0.96)	4 (2–6)	4 (2–6)	0.57 (0.18–0.80)	0.72 (0.30–0.89)
	DC	0.01(0.95)	5 (3–6)	4 (3–7)	0.45 (0.02–0.74)	0.62 (0.03–0.85)
	VC	0.32 (0.17)	8 (5–12)	9 (5–12)	0.33 (−0.12–0.67)	0.49 (−0.28–0.80)
	LC	0.16 (0.50)	4 (3–6)	5 (3–7)	0.66 (0.32–0.85)	0.79 (0.48–0.92)
AD [μm^2/ms]	WM	0.13 (0.58)	5 (4–6)	2 (2–3)	0.65 (0.30–0.84)	0.79 (0.46–0.92)
	DC	0.44 (0.06)	5 (4–7)	3 (2–3)	0.46 (0.04–0.75)	0.63 (0.08–0.86)
	VC	0.31 (0.18)	11(7–15)	6 (4–8)	0.18 (−0.27–0.57)	0.31 (−0.74–0.73)
	LC	0.20 (0.39)	5 (3–6)	3 (2–3)	0.79 (0.54–0.91)	0.88 (0.70–0.95)
RD [μm^2/ms]	WM	−0.26 (0.26)	5(3–7)	12 (6–17)	0.52 (0.12–0.78)	0.69 (0.21–0.88)
	DC	−0.21 (0.38)	5(3–7)	12 (618)	0.49 (0.07–0.76)	0.66 (0.13–0.86)
	VC	0.18 (0.44)	8(5–12)	17 (11−23)	0.41 (−0.02–0.72)	0.59 (−0.05–0.84)
	LC	−0.15 (0.53)	5(4–7)	13 (8–18)	0.57 (0.19–0.81)	0.73 (0.32–0.89)
FA	WM	0.45 (0.05) *	3(2–4)	4 (2–5)	0.42 (−0.02–0.72)	0.59 (−0.05–0.84)
	DC	0.42 (0.07)	3(2–4)	4 (2–5	0.43 (−0.003–0.73)	0.60 (−0.05–0.84)
	VC	0.31 (0.19)	4(2–5)	5 (4–7)	0.49 (0.07–0.76)	0.66 (0.14–0.87)
	LC	0.37 (0.11)	3(2–4)	4 (2–5)	0.46 (0.03–0.74)	0.63 (0.06–0.85)

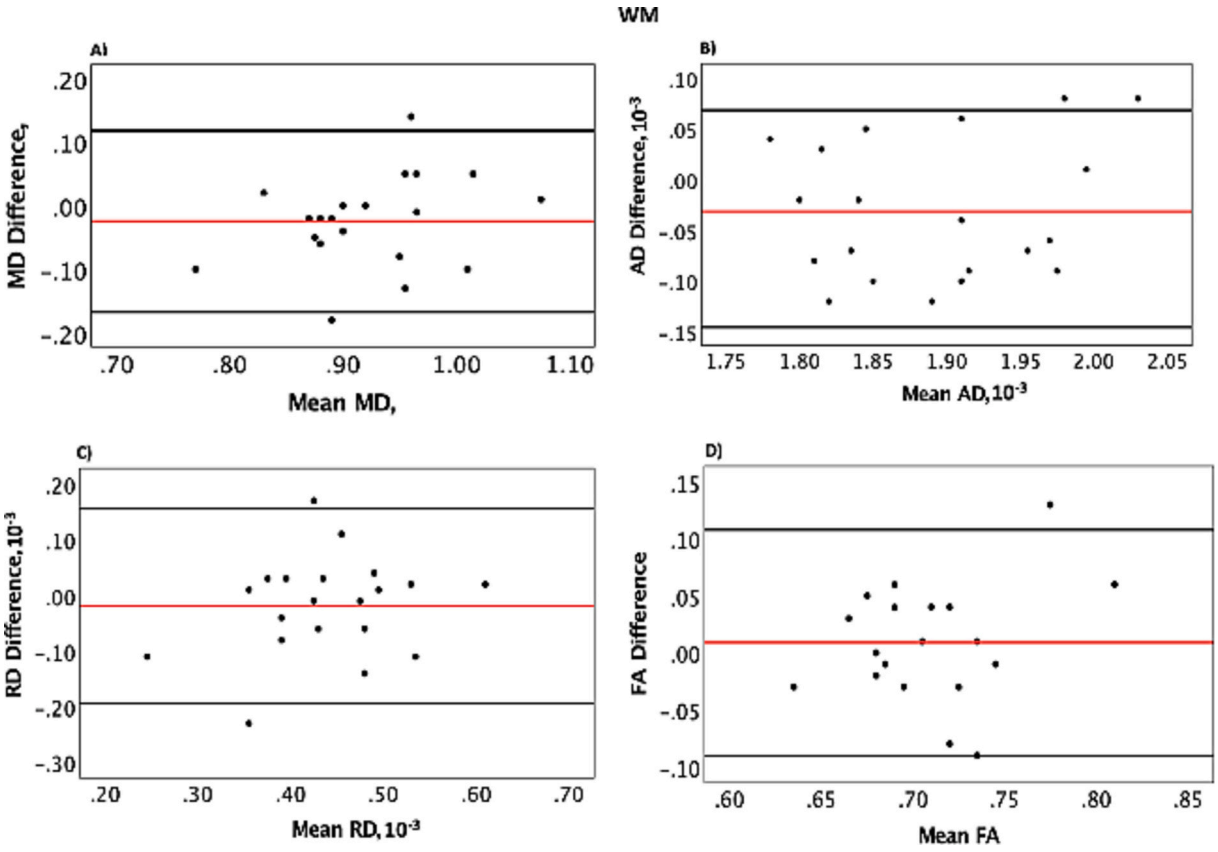

Fig. 5. Bland–Altman plots comparing within-participants scan-rescan values exhibiting the difference (solid-red lines) and limit of agreement 95% confidence interval (bold-black lines) of WM region in each subject ($N = 20$). The differences in diffusivity metrics, A) mean diffusivity (MD), B) axial diffusivity (AD), C) radial diffusivity (RD) and D) fractional anisotropy (FA) for pairs of scans were plotted against the mean for diffusivity metrics. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Results

3.1. Mean values

The mean value of DTI metrics for visits 1 and 2 with the Bland-Altman difference (D), Bland-Altman limits of agreement (LOA), 95% CI, and within-participants effects and paired samples *t*-tests (t-statistic and *P*-value) to assess for differences across visits are shown in Table 3. The mean SNR value of the CSC of non-DWI and DWI images was 9.6 ± 1.90 and 6.80 ± 1.5, respectively.

Mean values given by visits and pooled over the two visits, together Bland-Altman difference, limit of agreement (LOA), and differences as percentage, all with the 95% Confidence Interval and the within-participants effects (t-statistic and *P*-value). The diffusivity metrics are, mean diffusivity (MD), axial diffusivity (AD), radial diffusivity (RD), and fractional anisotropy (FA). ROI indicates the region of interest which includes WM; white matter, VC; ventral columns, LC; lateral columns and DC; dorsal columns.

Significant within-participants mean differences in the paired *t*-test are reported according to the resulting *P*-value ($^*P \leq 0.05$).

3.2. Within-participants intra-class correlation

Table 4 shows the single and average measures ICCs of DTI metrics, which showed variable reliability, depending on the ROI. For single ICC, the MD had fair to good reliability, while average ICC showed fair to excellent reliability, with the LC region yielding the best reliability (single/average ICC = 0.66/0.76). The AD revealed poor to excellent reliability for single and average ICC, with the VC having the lowest (single/average ICC = 0.18/0.31), and LC the best (single/average ICC = 0.69/0.88). The RD showed fair reliability for single ICC ranging between 0.41 and 0.57, and fair to good reliability for average ICC, with the LC having the highest values (single/average ICC = 0.57/0.73). For the FA, the results showed fair reliability for single ICC whatever the ROI, ranging from 0.42 to 0.49, while the average ICC reported fair to good reliability from 0.59 to 0.66, with the VC reporting the highest values (single/average ICC = 0.49/0.66).

Mean values given with the 95% Confidence Interval; R(P), Correlation between within-participants mean and standard deviation; SD, within participants standard deviation; CV, within participants coefficient of variation; ICC, within participants Intraclass correlation coefficient. The diffusivity metrics are, mean diffusivity (MD), axial diffusivity (AD), radial diffusivity (RD), and fractional anisotropy (FA). ROI indicates the region of interest which includes WM; white matter, VC; ventral columns, LC; lateral columns and DC; dorsal columns. Significant correlation between within-participants mean and standard deviation in the paired *t*-test are reported according to the resulting *P*-value ($^*P \leq 0.05$).

3.3. Within-participants variation

The within-participants variations (CV%) are shown in Table 4. In general, the CV of all DTI metrics were low for the WM, DC, VC and LC regions demonstrating the relative error on each parameter with acceptable to adequate reproducibility values, ranging from 4% to 17%, depending on the ROIs. For the MD, the smallest variation was observed in the WM, and DC regions, with 4% in both. The AD variation ranged from 2% to 6% with the smallest variability in the WM. For FA, the CV was 4% in WM, DC and LC and 5% in the VC area showing high reproducibility. For the RD values, adequate variations were reported

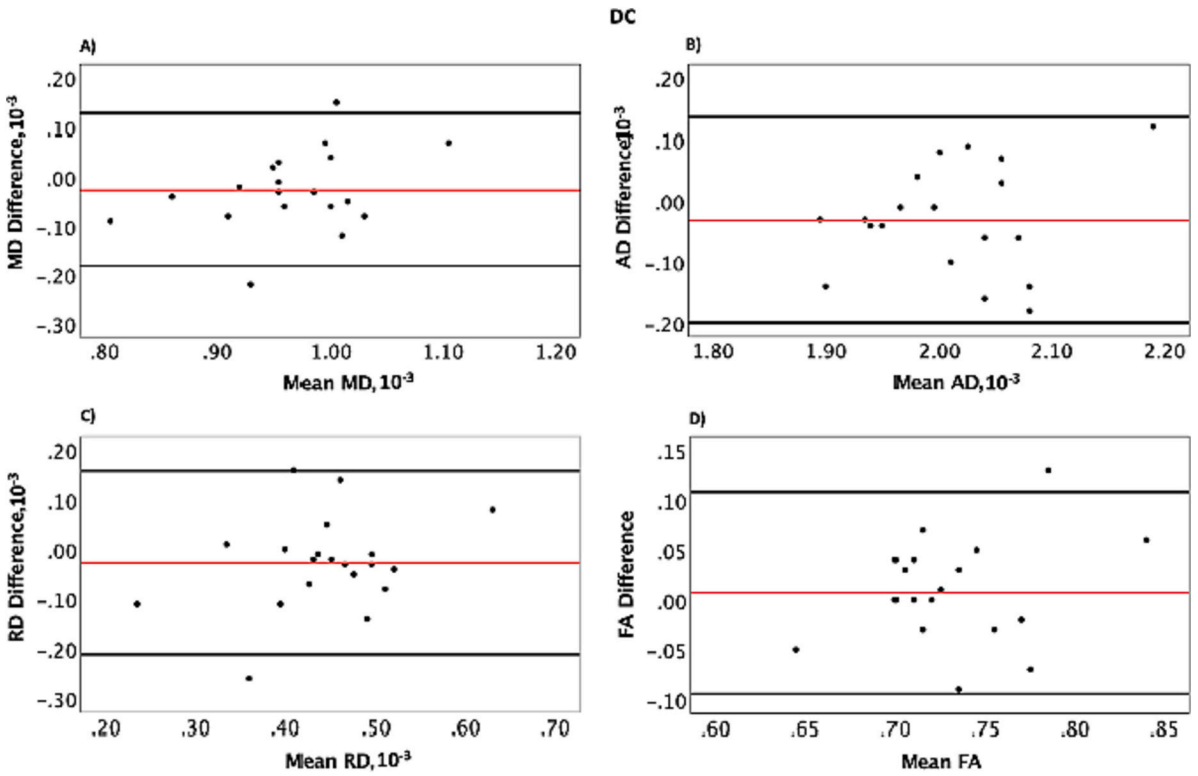

Fig. 6. Bland–Altman plots comparing within-participants scan-rescan values exhibiting the difference (solid-red lines) and limit of agreement 95% confidence interval (bold-black lines) of DC region in each subject ($N = 20$). The differences in diffusivity metrics, A) mean diffusivity (MD), B) axial diffusivity (AD), C) radial diffusivity (RD) and D) fractional anisotropy (FA) for pairs of scans were plotted against the mean for diffusivity metrics. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

for WM, DC and LC with values of 12–13% whereas a value of 17% was found for the VC region. Table 4 also shows the within-participants correlation between subject mean and standard deviation (SD). There was a significant correlation between the within-participants mean and SD for FA in WM region indicating that the error in this measure was proportional instead of constant.

3.4. Bland-Altman (BA) analysis

Bland-Altman (BA) plots for within-participants values for DTI metrics in the whole WM, and other regions (DC, VC, and LC), are shown in Figs. 5, 6, 7 and 8 respectively. The perfect Bland–Altman plot would be a horizontal line at y equal to 0, demonstrating that there is no difference between test-retest measurements. The mean differences (D) for the two scans were small between the scans indicating a good agreement between visits of each DTI parameter, for example, the 95% CI for mean differences of all metrics overlaps zero. Data points falling outside the 95% CI correspond to significant deviation between visit 1 and visit 2. Overall mean differences of DTI metrics for the whole WM, and DC, VC, and LC were − 0.025, −0.027, 0.009, and 0.035 respectively for MD, −0.032, −0.032, 0.003, and − 0.043, respectively for AD, −0.019, −0.027, −0.027, and − 0.031, respectively for RD and 0.01, 0.01, 0, and 0.011 respectively, for FA. In addition, there was a statistically significant difference in the mean differences of MD in LC (t-test = −2.20, P = 0.04), AD in WM and LC (t-test = −2.06, P = 0.05 and t-test = −2.83, P = 0.01, respectively) whereas for other regions P-values showed non-significant differences.

4. Discussion

A quantitative, reliable method is required to evaluate spinal cord damage. If DTI can be demonstrated to be a reproducible and valid approach for assessing healthy neural tissue in the spinal cord, it is likely to be a valuable neuro-diagnostic complement to the standard clinical evaluation. DTI in the spine is challenging however due to factors such as motion artefacts from cerebrospinal fluid (CSF) pulsations and breathing, volume averaging, susceptibility artefacts from lungs and bones, lower signal-to-noise ratio (SNR) and partial volume effect (PVE) contamination between WM and surrounding CSF [23]. These factors have limited the application of DTI to the cervical spinal cord and the reproducibility of DTI metrics. In spite of these challenges, improvements in image resolution and quality for DTI in the spine have resulted due to advancements in scanner hardware and software, such as reduced FOV techniques and the use of cardiac gating methods. Combining this with automated post-processing software can result in the acquisition of more reproducible DTI metrics.

The purpose of this investigation was to evaluate the within-participants reliability of DTI measures of the healthy cervical spinal cord. To our best knowledge, this report is the first one that examines the within-participants reliability over the total WM and WM sub-regions spanning from the upper border of C2 to the lower border of C5 by reporting both the single and the average ICC measures, the percent CV, as well as the BA analysis. The relative within-participants reliability varied depending on location of the ROIs. Both single and average ICC varied from poor to excellent agreement with a CV% revealing acceptable to moderate variability. The mean differences of DTI metrics were small among the WM and WM sub-regions with the exception for MD in the LC, AD in the WM and LC, where a significant statistical difference was reported.

4.1. Within-participants intra-class correlation

The reliability of DTI in evaluating the CSC has been previously examined. Peterson, et al. [6] studied repeatability of DTI metrics in a

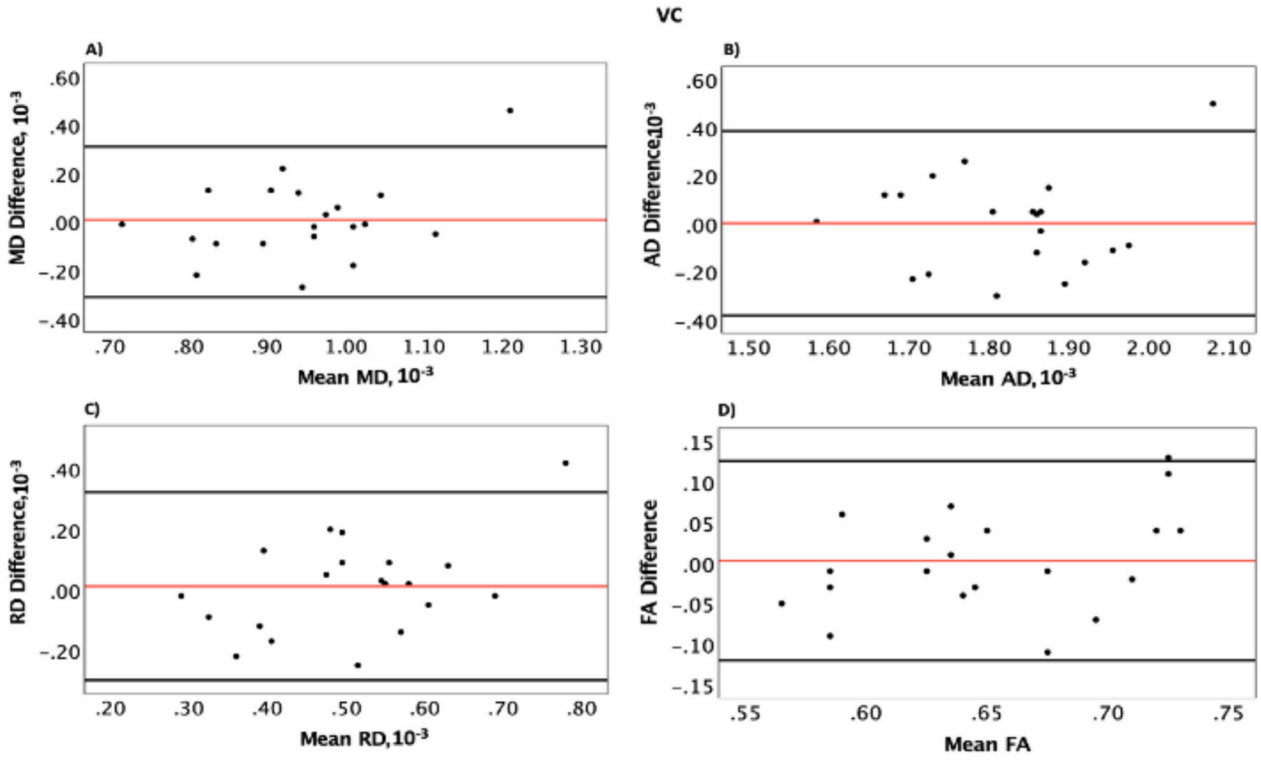

Fig. 7. Bland–Altman plots comparing within-participants scan-rescan values exhibiting the difference (solid-red lines) and limit of agreement 95% confidence interval (bold-black lines) of VC region in each subject ($N = 20$). The differences in diffusivity metrics, A) mean diffusivity (MD), B) axial diffusivity (AD), C) radial diffusivity (RD) and D) fractional anisotropy (FA) for pairs of scans were plotted against the mean for diffusivity metrics. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

group of 30 adult patients with acute spine trauma using a tract-specific segmentation method for WM, DC, VC, and LC using both CV and average ICC reliability measurements. The test-retest reproducibility of MD metrics had excellent reproducibility measurements in WM (ICC = 0.86) and VC (ICC = 0.75), respectively, and good agreement in both DC and LC (for both ICC =0.71). FA reported good test-retest reliability in VC (ICC = 0.72) and excellent reliability in WM, DC, and LC with ICC of 0.93, 0.80, and 0.84, respectively. The findings of the current study demonstrated a similar trend of reliability to Peterson, et al. study although generally lower ICC values. For MD VC had fair reliability (ICC = 0.49), WM and DC had good reliability (ICC = 0.72 and 0.62, respectively) and LC had an excellent reliability (ICC = 0.79). For FA WM had fair reliability (ICC = 0.59) and VC, DC, and LC illustrated good reliability (ICC = 0.66, 0.60, 0.63, respectively). Lower reliability in the VC region relative to other ROIs for MD and AD was found, as illustrated by average ICC values <0.50 for MD. This might be because of that the ventral column region was averaged from extremely small ROI tract volumes specified by the PAM50 atlas, this area might be more prone to PVE than the whole WM areas where such effects are more likely to be averaged out. Moreover, low reliability and high variability regions are typically associated with areas of crossing fibers which are densely packed fibers with specific orientations [24]. Future research is required to examine the contributions of fibers crossing to the reliability and variability of DTI indices. The results of the current study suggest that when analysing longitudinal changes in small ventral column regions, extra care should be given because these have higher intrinsic variability by means of atlas-specific registration approaches. Further, it is possible that the large divergence observed for a small number of participants, as indicated on the BA plots, had a large influence on the ICC values. Although the DTI data were cardiac-gated which helps reduce the CSF pulsation artefacts theoretically improving reliability, the resulting doubled scan time might lead to excessive motion artefacts in some cases, thereby impacting the reliability results. In addition, the ICC may be not an ideal measure because the comparatively low between-participants variation seen in a healthy group may not be representative of a patient group, in which varying degrees of pathologies will cause greater between-participants variability and thus potentially higher ICC values.

4.2. *Within-participants variation*

In terms of within-participants variation, the CV% is a useful statistical analysis that may also be used to determine the smallest quantifiable change. Because the CV% reveals the relative variability of the standard deviation relative to its mean, it is beneficial in comparing the errors across the DTI measures in this analysis. Peterson, et al. [6] reported a CV of 5.6%, 11.3%, 12% and 10.4%, in MD for WM, DC, VC and LC, respectively, and of 3.8%, 7.7%, 10.5%, and 6.2% in FA for WM, DC, VC and LC, respectively. The current study found low CV values of 4%, 4%, 9%, and 5% in MD and 4%, 4%, 5%, and 4% in FA for WM, DC, VC and LC, respectively, indicating higher reproducibility than in Peterson, et al. [6] study. The possible sources of lower variation in the current study compared to the Peterson, et al. [6] study may be that the latter was undertaken with patients with acute spine trauma. Patients with acute spine injuries may have DTI metrics which vary in time due to changes in tissue integrity, for example, inflammation, which can impact water molecule diffusion. Further, in some atrophied spinal cord diseases such as cervical spondylotic myelopathy (CSM) disease, drawing ROIs accurately is challenging due to the low spatial resolution of DTI images in comparison to healthy spinal cord thus potentially increasing the variability of DTI indices [25]. Another study on 3 healthy individuals by Taso, Manuel, et al. [26], reported the repeatability of DTI using atlas-based automatic ROI segmentation in C2 and C5 of the whole WM, as well as over both left and right WM pathways and anterior

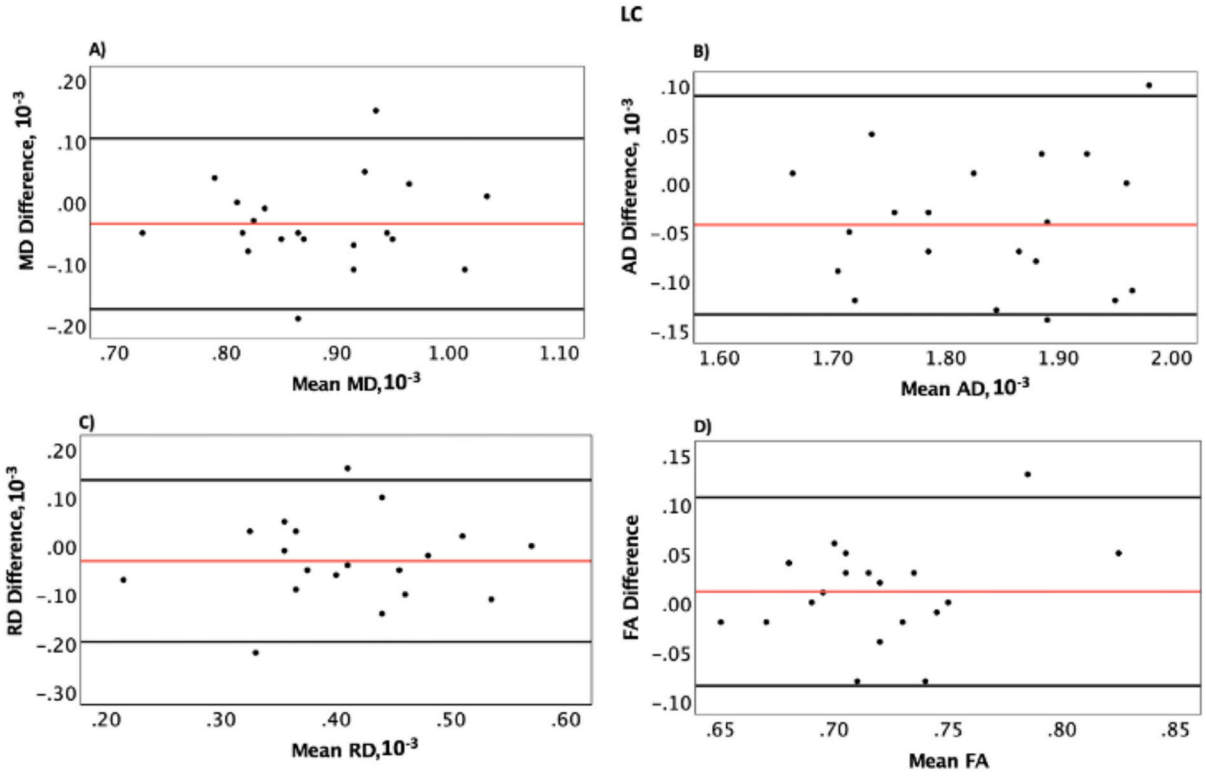

Fig. 8. Bland–Altman plots comparing within-participants scan-rescan values exhibiting the difference (solid-red lines) and limit of agreement 95% confidence interval (bold-black lines) of LC region in each subject (N = 20). The differences in diffusivity metrics, A) mean diffusivity (MD), B) axial diffusivity (AD), C) radial diffusivity (RD) and D) fractional anisotropy (FA) for pairs of scans were plotted against the mean for diffusivity metrics. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

grey matter (GM). They found the CV to be lower than 10% for all DTI metrics, 3.4%, 3.5%, 9.7% and 5.1% respectively, for MD, AD, RD and FA. The current study reported generally comparable or less variability, with CV values of 4% in MD, 2% in AD and 4% in FA for WM, although RD only had moderate reproducibility with a CV of 12%.

Overall, the current study revealed lowest variability for AD compared to the other DTI metrics. One possible reason may be that AD, unlike the other metrics, is based on only one component of the diffusion tensor, the primary eigenvalue, whereas for MD, RD, and FA, there is some averaging across tensor elements (i.e., eigenvalues), which may increase the variability. Further, as higher CV values were observed for RD, with the highest seen in the VC region, one might be cautious about using RD for longitudinal investigations, despite a reasonably high reliability (ICC = 0.73) in LC.

4.3. Bland-Altman analysis

With regard to the within-participants differences, Smith et al. [7] assessed the test-retest and inter-reader reproducibility of DTI in 9 healthy participants spanning from top limit of C2 to the end limit of C7 using measuring mean difference between repeated measures and found no strong difference between scans for each ROI drawn manually. The mean differences were − 0.01%, −0.08%, 0.05%, and − 0.03% for LC and 0.01%, 0.01%, 0.02%, and 0.01% for DC for MD, AD, RD, and FA respectively. The mean differences of DTI metrics reported in the current study were found to be comparable with the Smith et al. [7] study, showing no major differences in most of the ROIs (good within-participants agreement between the two visits). Overall, these findings showed no systematic bias or consistent differences of one visit over the other. The equal spread of data points around zero suggests that the bias does not vary between low to high values, and the fact that the data spread stays reasonably constant across the x axis reveals that the variability between visits is similar over all measurement ranges. The presence of outliers and systematic differences may be caused by different factors, including differences in B1 inhomogeneity and gradient performance. Although our DTI data did not show any evidence of EPI-based distortions (see Fig. 2), careful review of images is necessary to identify potential sources of variance.

4.4. SNR assessment

Evaluation of the SNR in the spinal cord of DTI images using a “differences method” has been previously explored by Griffanti, et al. (2012) [27]. The SNR of the non-DWI and DWI data in the current study was better than that found in the Griffanti, et al. (2012) study with a value of 9.6 ± 1.90 and 6.80 ± 1.5, respectively, a level that allowed the differentiation between WM, GM and CSF regions for all DTI-related metrics, as shown in Fig. 2. Obtaining a good SNR is crucial for reliability studies and a number of factors, such as acquisition parameters, might impact the SNR and thus reliability [28,29]. Decreased SNR is well-known to introduce bias in DTI measurements [30] and affect the reliability of DTI metrics [31]. In addition, the number of diffusion-weighted directions, for example, may also influence the SNR of DTI data. Further, increasing the number of diffusion-weighted directions can enhance SNR and increase information regarding the spatial dependency of diffusivity. However, this may also contribute to longer acquisition times and increased participant motion, thus detrimentally impacting the SNR. In a similar fashion, modification of the b-value, which affects diffusion-weighted signal strength, can influence the SNR, with an increase in b-value decreasing the SNR. The current study protocol used a b-value of 800 s·mm $^{-2}$ which is recommended as long as SNR is maintained at adequate levels [32]. In addition, in DWI

sequences, TE plays a key factor in controlling SNR, with an increase in TE resulting in a reduction in signal because of T2 relaxation [33].The application of OVS techniques such as ZOOMit FOV along with the second order shimming technique has allowed for further improvement of SNR of the current study by decreasing the sensitivity of the EPI sequence to susceptibility artefacts due to bone: tissue interfaces, which is particularly important near the vertebrae where magnetic field inhomogeneity is greater than in other body locations.

4.5. Limitations

There are a number of limitations in the current study. First, although the overall WM and WM sub-sections were investigated, smaller tract-specific regions of WM such as cortico-spinal tracts and rubrospinal tracts should be investigated further since these regions may be more influenced by the PVE. Second, a relatively small sample size was employed in the current study, such that a small number of outliers may significantly affect the statistical results. Thus, a further study with larger sample size is needed to explore the impact on reliability measurements. Third, because this is a single-center, single-vendor study, measured parameters may be influenced by sequence parameters, hardware, and postprocessing processes, all of which are intrinsic restrictions in most quantitative imaging studies. Additional studies involving multiple centres or vendors would be beneficial in determining the reliability and measurement error of the DTI technique.

5. Conclusion

The assessment of the reliability of DTI in a healthy population has allowed exploration of the validity of this technique as a method to assess the cervical spinal cord changes in conjunction with current imaging techniques and clinical examinations. The data demonstrates that the reliability of DTI measurements varies depending on where the ROIs are located, demonstrating the importance of carefully selecting ROI locations in clinical DTI research. Findings of this study also support the use of fully automated post-processing software when undertaking spinal cord studies.

Funding

This study has been supported by Najran University, Kingdom of Saudi Arabia as part of PhD scholarship for the corresponding author. Sponsorship role: This funding organisation did not participate in the designing and conducting of the study; data collection, management, and data interpretation; and preparation, review, or approval of the manuscript; and decision for submitting the manuscript for publication.

CRediT authorship contribution statement

Hussein Al-shaari: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Christine J. Heales:** Writing – review & editing, Visualization, Validation, Supervision. **Jon Fulford:** Writing – review & editing, Visualization, Validation, Supervision.

Acknowledgement

Authors would like to thank radiographers team from Mireille Gilling's Neuroimaging Centre (MGNC) located at Royal Devon and Exeter Hospital for their help and support during data collection.

References

[1] Beaulieu C. The basis of anisotropic water diffusion in the nervous system–a technical review. NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In Vivo 2002;15(7–8):435–55.
[2] Basser PJ, Mattiello J, LeBihan D. Estimation of the effective self-diffusion tensor from the NMR spin echo. J Magn Reson B 1994;103(3):247–54.
[3] Mukherjee P, et al. Diffusion tensor MR imaging and fiber tractography: theoretic underpinnings. Am J Neuroradiol 2008;29(4):632–41.
[4] Bartlett J, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. Ultrasound in Obstetrics and Gynecol: The Official J Intern Soc Ultrasound in Obstet and Gynecol 2008;31(4):466–75.
[5] Al-shaari H, et al. A systematic review of repeatability and reproducibility studies of diffusion tensor imaging of cervical spinal cord. Br J Radiol 2023;96(1151).
[6] Peterson D, et al. Test-retest and Interreader reproducibility of semiautomated atlas-based analysis of diffusion tensor imaging data in acute cervical spine trauma in adult patients. Am J Neuroradiol 2017;38(10):2015–20.
[7] Smith SA, et al. Reproducibility of tract-specific magnetization transfer and diffusion tensor imaging in the cervical spinal cord at 3 tesla. NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magn Resonan In vivo 2010;23(2):207–17.
[8] Wilm BJ, et al. Reduced field-of-view MRI using outer volume suppression for spinal cord diffusion imaging. Magn Resonan Med: An Official J Intern Soc Magn Resonan Med 2007;57(3):625–30.
[9] De Leener B, et al. SCT: spinal cord toolbox, an open-source software for processing spinal cord MRI data. Neuroimage 2017;145:24–43.
[10] De Leener B, et al. PAM50: unbiased multimodal template of the brainstem and spinal cord aligned with the ICBM152 space. Neuroimage 2018;165:170–9.
[11] Lévy S, et al. White matter atlas of the human spinal cord with estimation of partial volume effect. Neuroimage 2015;119:262–71.
[12] De Leener B, et al. SCT: spinal cord toolbox, an open-source software for processing spinal cord MRI data. Neuroimage 2017;145(Pt A):24–43.
[13] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86(2):420.
[14] Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994;6(4):284.
[15] Marenco S, et al. Regional distribution of measurement error in diffusion tensor imaging. Psychiatry Research: Neuroimaging 2006;147(1):69–78.
[16] Carlson HL, et al. Reliability and variability of diffusion tensor imaging (DTI) tractography in pediatric epilepsy. Epilepsy Behav 2014;37:116–22.
[17] Dietrich O, et al. Measurement of signal-to-noise ratios in MR images: influence of multichannel coils, parallel imaging, and reconstruction filters. J Magn Resonan Imag: An Official J Intern Soc Magnet Resonan Med 2007;26(2):375–85.
[18] Murphy B, et al. Signal-to-noise measures for magnetic resonance imagers. Magn Reson Imaging 1993;11(3):425–8.
[19] Firbank M, et al. A comparison of two methods for measuring the signal to noise ratio on MR images. Phys Med Biol 1999;44(12):N261.
[20] Association, N.E.M. Determination of signal-to-noise ratio (SNR) in diagnostic magnetic resonance imaging. NEMA Standards Publication MS 2001;1-2001.
[21] Reeder SB, et al. Practical approaches to the evaluation of signal-to-noise ratio performance with parallel imaging: application with cardiac imaging and a 32-channel cardiac coil. Magn Resonan Med: An Official J Intern Soc Magn Resonan Med 2005;54(3):748–54.
[22] Kellman P, McVeigh ER. Image reconstruction in SNR units: a general method for SNR measurement. Magn Reson Med 2005;54(6):1439–47.
[23] Stroman PW, et al. The current state-of-the-art of spinal cord imaging: methods. Neuroimage 2014;84:1070–81.
[24] Stieltjes B, et al. Diffusion tensor imaging and axonal tracking in the human brainstem. Neuroimage 2001;14(3):723–35.
[25] Lee E, et al. Reliability of pre-operative diffusion tensor imaging parameter measurements of the cervical spine in patients with cervical spondylotic myelopathy. Sci Rep 2020;10(1):1–9.
[26] Taso M, et al. Tract-specific and age-related variations of the spinal cord microstructure: a multi-parametric MRI study using diffusion tensor imaging (DTI) and inhomogeneous magnetization transfer (ihMT). NMR Biomed 2016;29(6):817–32.
[27] Griffanti L, et al. Signal-to-noise ratio of diffusion weighted magnetic resonance imaging: estimation methods and in vivo application to spinal cord. Biomed Signal Proces and Control 2012;7(3):285–94.
[28] Alexander AL, et al. *Comparison of diffusion tensor imaging measurements at 3.0 T versus 1.5 T with and without parallel imaging.* Neuroimaging. Clinics 2006;16(2):299–309.
[29] Ni H, et al. Effects of number of diffusion gradient directions on derived diffusion tensor imaging indices in human brain. Am J Neuroradiol 2006;27(8):1776–81.
[30] Jones DK, Basser PJ. "Squashing peanuts and smashing pumpkins": how noise distorts diffusion-weighted MR data. Magn Resonan Med: An Official J Intern Soc Magn Resonan Med 2004;52(5):979–93.
[31] Farrell JA, 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 1.5 T. J Magnetic Resonan Imaging: an Official J Intern Soc Magn Reson Med 2007;26(3):756–67.

[32] Martín Noguerol T, et al. Optimizing diffusion-tensor imaging acquisition for spinal cord assessment: physical basis and technical adjustments. Radiographics 2020;40 (2):403–27.

[33] Qin W, et al. *Effects of echo time on diffusion quantification of brain white matter at 1.5 T and 3.0 T.* Magnetic resonance in medicine: an official journal of the international society for. Magn Reson Med 2009;61(4):755–60.