BJR https://doi.org/10.1259/bjr.20221019

Received: 31 October 2022 | Accepted: 10 August 2023 | Published online: 28 September 2023

© 2023 The Authors. Published by the British Institute of Radiology under the terms of the Creative Commons Attribution-NonCommercial 4.0 Unported License http://creativecommons.org/licenses/by-nc/4.0/, which permits unrestricted non-commercial reuse, provided the original author and source are credited.

Cite this article as:
Al-shaari H, J F, R M, CJ H. A systematic review of repeatability and reproducibility studies of diffusion tensor imaging of cervical spinal cord. *Br J Radiol* (2023) 10.1259/bjr.20221019.

SYSTEMATIC REVIEW

A systematic review of repeatability and reproducibility studies of diffusion tensor imaging of cervical spinal cord

[1,2]HUSSEIN AL-SHAARI, [1]FULFORD J, [1]MEERTENS R and [1]HEALES CJ

[1]Medical Imaging Department, Faculty of Health and Life Sciences, University of Exeter, Exeter, United Kingdom
[2]Radiological Sciences Department, College of Applied Medical Sciences, Najran University, Najran, Saudi Arabia

Address correspondence to: Mr Hussein Al-shaari
E-mail: *ha457@exeter.ac.uk*

Fulford J, Meertens R and Heales CJ should be considered as co-first authors.

Objectives: Diffusion tensor imaging (DTI) techniques are being studied as a possible diagnostic and predictive tool for the evaluation of cervical spinal cord disease. This systematic review aims to evaluate the previous DTI studies that specifically investigated the repeatability and reproducibility of DTI in the cervical spinal cord.
Methods and materials: A search in the PubMed, Scopus, Web of Science and Ovid electronic databases was conducted for articles published between January 1990 and February 2022 that related to the repeatability and reproducibility of DTI in evaluating the cervical spinal cord using one of the following measurements: the intra-class correlation coefficient (ICC) and/or the coefficient of variation (CV), and/or Bland-Altman (BA) differences analysis methods. DTI studies that presented full statistical analysis of repeatability and/or reproducibility tests of the cervical spinal cord in peer-reviewed full-text publications published in journals were included. Articles that included at least one of the keywords within the titles or abstracts were identified. Additional full-text papers were found by searching the citations and reference lists of related articles. This review has followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidance. Risk of bias was evaluated with 13 criteria weighted toward methodological quality of reported studies using the QuADS assessment criteria. This assessment only included full-text articles written in English.
Results: A total of 11 studies were included and assessed for different characteristics, including sample size,(3–34) re-test time interval (<1h to >3 months), test-retest reproducibility scores and acquisition method. Six studies used ICC which ranged from poor (ICC<0.37) to excellent reproducibility (ICC 0.91–0.99). Four studies reported an overall CV lower than 40% for all DTI metrics. Three studies reported the Bland-Altman (BA) differences and reported a minimum percentage showing no strong differences between repeated measurements. Quantitative analysis was not undertaken due to heterogeneity of methods. Repeatability and reproducibility measures were generally found to be good.
Conclusion: This study revealed that the application of DTI and its related measures in a clinical setting in the assessment of cervical spinal cord changes is feasible and reproducible. However, cervical spinal cord DTI suffers from some existing limitations that prevent it from being routinely used in research and clinical settings.
Advances in knowledge: DTI with its parametric maps provide broad evaluation of the tissue structure of axonal white matter and are being studied as a possible diagnostic and predictive tool for the assessment of cervical spinal cord (CSC) disease.

INTRODUCTION

Diffusion tensor imaging (DTI) is a non-invasive method for visualizing the spinal cord *in vivo*. DTI provides broad measures of the tissue structure of axonal white matter and is thought to be more specific than conventional MRI sequences in evaluating damage to tracts in the spinal cord.[1,2] It assesses the molecular motion of water molecules that diffuse within and across each measurement voxel, both parallel and transverse to the direction of neural axons. DTI generates a range of parameters that provide measures of diffusion characteristics. Fractional anisotropy (FA) assesses the degree to which water proton diffusivity varies with direction within tissues and is affected by many factors such as changes in water content and the presence of

human white matter (WM) crossing fibres. It has values ranging between 0 and 1, where values close to 0 represent isotropic molecular motion of water in tissue (water diffusion is equal in all directions), whereas values close to one represents an anisotropic diffusion of water particles (water diffusion is restricted in specific directions).[3] Mean diffusivity (MD) provides an overall measure of water translational diffusivity, axial diffusivity (AD) specifies the magnitude and the direction of maximum water diffusion and is influenced by longitudinal axonal integrity and radial diffusivity (RD) evaluates the diffusivity properties of tissue in the perpendicular axonal structure.[4] The terms repeatability and reproducibility are used interchangeably in different studies. The term "repeatability" is defined as the variation in the repeated measurements done on the same subject under the same conditions. This implies that data were collected with the same method or instrument, by the same rater (or observer) if human input is needed, and over a short period of time, during which the underlying value can be assumed to be constant. The term "reproducibility", on the other hand, consists of various concepts, many of which overlap. There are several ways in which this might be described, but it is most often defined as the variation of the same measurement done on the same subject under different conditions. The changing conditions could be a result of different measuring instruments or methods being applied, measurements being made by different raters or observers, or measurements being taken over a period of time, during which the variable's 'error-free' level could change significantly.[5] This systematic review aims to assess studies that specially examined the repeatability and reproducibility of DTI metrics in the evaluation of the cervical spinal cord (CSC).

METHODS AND MATERIALS

Eligibility criteria

The systematic review was carried out throughout the months of February and March of 2022 and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (see appendix A).[6] The articles included in this review fulfilled all the criteria outlined in the following paragraphs.

Inclusion criteria

Articles that included at least one of the search terms supplied in the search keywords (in their titles or abstracts) were identified. This evaluation only included reports with full-text content written in English. DTI studies that presented full statistical analysis of repeatability and/or reproducibility tests of the cervical spinal cord in peer-reviewed full-text publications were exclusively included. All studies that satisfied the inclusion criteria are included in the results section (Table 2).

Exclusion criteria

Opinion pieces, ideas, case studies of single patients/healthy participant, and editorials were excluded.

Population

Both healthy and non-healthy participants who underwent cervical spinal cord DTI were included in this review. Studies with animal subjects were excluded.

Information source

In order to find any past systematic studies dealing with the repeatability and/or reproducibility of DTI in the cervical spinal cord, the PROSPERO Database of International prospective register of Systematic Reviews was screened. No relevant systematic review was matched. PubMed, Scopus, Web of science and Ovid were used to do an electronic search. The bibliographic references of all articles for which the full-text papers as well as studies that cited the included papers were collected for data extraction were also checked for possibly relevant research.

Research strategy

A PubMed, Scopus, Web of Science and MEDLINE (Ovid) citation search were conducted using the general Haynes[7] and Ingui[8] criteria, as well as the changes suggested by Geersing et al.[9] (each linked using "OR"). The keywords used in this review were as follows: diffusion tensor imaging, diffusion tensor MRI, DTI, Spinal cord, cervical spinal cord, repeatability, reliability, and reproducibility. More details of keywords used in different databases applied in this review are in appendix B.[1–4]

Study records

Data management and selection process

Any study which included one of the keywords in the title or abstract and met the inclusion criteria was screened by two reviewers who worked independently to avoid bias during the screening process. In order to determine whether or not a paper was eligible for inclusion, the title and abstract were reviewed. Consensus was reached in the case of disagreements. All the articles that were thought to be eligible were downloaded.

The full-text reports were examined by two reviewers separately to see whether they were eligible for inclusion and synthesis based on the keywords. Disagreements were handled by consensus once again. If differences could not be resolved, a third reviewer was available, although this option was not used. The reasons for excluding a particular full-text article were reported.

Data extraction

We collected information regarding the study populations, such as sample size, whether participants were healthy as regards their cervical spinal cord or non-healthy (with pathology) into two tables (1 and 2). The first table included the technical details of DTI studies for both healthy and unhealthy cervical spinal cord studies such as the imaging acquisition parameters, the software used to quantitatively extract DTI metrics, as well as any specific pre-processing processes applied to the DTI data prior to measurements extraction. The second table included author names, publication year, statistical analysis used, c-spine level examined, region of interest and results of repeatability and reproducibility studies. In addition, the normative values of DTI metrics for WM and grey matter (GM) were collected together with gender, mean age details and the ROI selected, as reported in Table 4. As a last point, we examined the statistical methods utilised for these repeatability and reproducibility studies.

Data synthesis

The heterogeneous data from the literature review were then used to create a narrative empirical synthesis.[10] In systematic reviews, narrative synthesis is especially effective for understanding the impact of interventions and the factors that influence intervention implementation.[10] The narrative strategy was employed to combine the qualitative and quantitative findings, allowing for in-depth investigation and collective understanding from numerous studies, resulting in a broader perception of the phenomena being studied. In the development of the main synthesis, two authors independently reviewed each study. The reviewers discussed common concepts and examined data trends in order to identify consistent findings in relation to the study outcomes. Interrogation of the findings examined relationships between characteristics of the study and their conclusions; the findings of other studies; and the effect of varying outcome measures, methodologies, and settings on the resulting data. A meta-analysis was not conducted due to a variety of reasons, including that all included studies did not have comparable outcome measurements; different regions of interest and vertebral levels were selected; and not all DTI metrics had been evaluated.

Outcomes and prioritisations

Primary outcome

The primary outcome of this review was to assess the statistics and metrics utilised to report the reproducibility such as CV, ICC, and BA differences, in evaluating the cervical spinal cord with respect to any DTI-MRI scanner vendor-, subject-, observer-, or site-related cause, as well as any external validation performed to assess for repeatability and reproducibility.

Different outcome measures were quantified for analysing the reproducibility of DTI studies, which include the coefficient of variation (CV%), and/or the intraclass corelation coefficient (ICC), and/or and Bland-Altman differences (BA differences) (Table 2), all studies reporting these measures and meeting the search criteria were reported. The CV is usually calculated for each variable to clarify the relative variability of each measurement as ([the within-subjects standard deviation]/mean × 100%). CV less than 10% is considered to be acceptable and illustrated that the dependent variable had a relatively small amount of variation,[11] CV between 11 and 20% were considered to be adequate and indicated a moderate variation. CV more than 20% were considered to be a high amount of variability.[12] The ICC is widely recommended for test-retest reproducibility, because of its excellent assessment and measured as ([true variance/(true variance+error variance)]).[13] It quantifies the actual percentage of variation related to the "true" error-free values of individuals (the within-subject variance) relative to the overall variance. Higher ICC indicates higher reproducibility as suggested by Shrout and Fleiss,[13] Cicchetti.[14] ICC values less than 0.5, values ranging from 0.5 to 0.75, values ranging from 0.75 to 0.9, and greater than 0.90 shows poor, moderate, good, and excellent reproducibility, respectively. Bland–Altman plots are used to examine the agreement between two independent measures by plotting the mean measurement (scan1+scan2)/2 against the difference in measurements (scan1 – scan2).

Secondary outcomes

The secondary outcomes were to assess the effect of different factors such as image acquisition parameters, DTI extraction software and the impact of pre-processing steps applied prior to metrics extraction on the repeatability or reproducibility of DTI.

Evaluation of study quality

Three independent reviewers assessed the quality of the included papers and arbitrated any discrepancies in scores using the updated version of the Quality Assessment for Diverse Studies tool (QuADS), which has revealed reliability and validity (in total 13 criteria were applied and are shown in appendix C).[15,16] The QuADS was used to assess the methodological and reporting quality of the included studies and the overall body of evidence due to the variety of the research types considered. The QuADS is specially designed to appraise qualitative, mixed, and multimethod studies in health services analysis. The tool involved 13 items (Table 3) and assigned a score of 0–3 to each item, with 0 being the lowest and three being the highest.[15] The maximum score was set to be 39. The final score was recorded as a percentage [final score = the total score of each study/ total score of criteria × 100%].

RESULTS

Literature search results

Figure 1depicts a flow diagram summary as recommended by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidance.[6] The PubMed, Scopus, Web of Science and Ovid search revealed 427 abstracts for screening according to our selection criteria, reduced to 179 after eliminating duplication. One hundred and one articles were excluded after title screening. This resulted in 78 abstracts being screened for eligibility criteria, and 61 of these reports were excluded because they did not meet the eligibility criteria. The 17 full-text articles which matched the eligibility criteria, were then fully screened after which 6 papers were excluded. The final qualitative synthesis was obtained from 11 studies, which included both healthy and non-healthy participants.

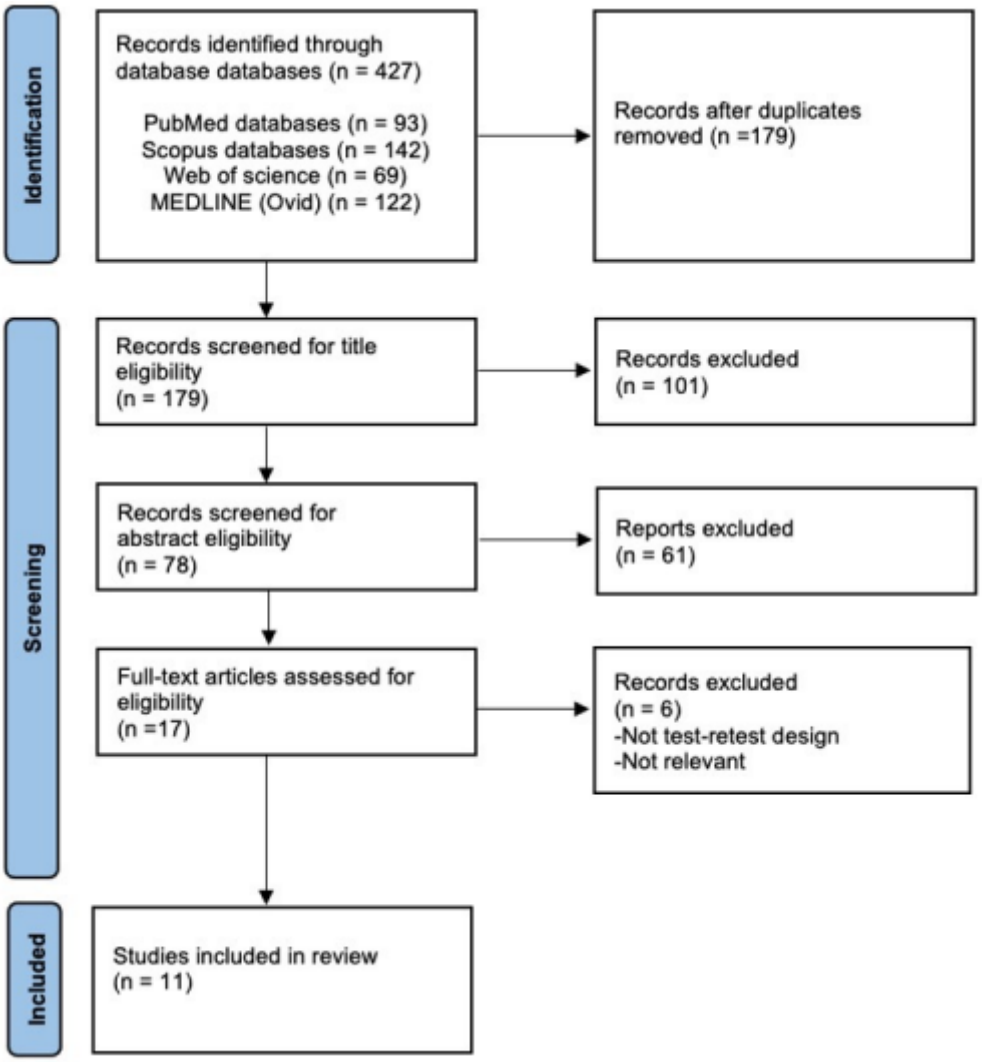
Excluded studies

Six studies were excluded at the full-text screening step because they did not meet the inclusion criteria for the following reasons: four studies reported an intervention that is not focused on test-retest design, one study focused on the repeatability and reproducibility of segmentation methods, which was not relevant to the inclusion criteria and one study focused on the reproducibility of DTI acquisition methods which was beyond the primary context of this review.

Characteristics of included repeatability and reproducibility studies of DTI

Table 1 shows the following DTI acquisition parameters for each study: scanner make, field strength (T), gradient directions applied, no. of signal averages, voxel size (mm^3), (FOV) field of view (mm), b-values (mm^2/s), TR (ms), TE (ms), acquisition type, ROI method, type of software used for DTI metrics extraction and no. of repeated scans [time interval]. Different time intervals between scan and rescan were applied in studies

Figure 1. Shows a flow diagram summary as recommended by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidance[17] .



(ranging from at least 1 h to 3 months) either with scans taking place during different visits[18–20,22–24,27,28] or during a single visit with repositioning after the first scan with an average time interval of 1–2 h between scan and rescan.[25,26] The direct scan–rescan evaluation potentially eliminates anatomical changes that could happen in the subject over time and narrows the causes of measurement variability to those related to the scanner and subject repositioning.[29,30] Moreover, the direct scan–rescan assessment with repositioning can be used to evaluate the sensitivity of the method to changes in orientation. Our findings found no obvious trend on the influence of time interval on the reproducibility of DTI metrics.

A summary of characteristics of the 11 repeatability and reproducibility studies of DTI are shown in (Table 2).[18–28,31] All these studies examined the repeatability and reproducibility of cervical spinal cord DTI measures using one of the following measurements: the intraclass correlation coefficient (ICC) (n = 6) and/or the coefficient of variation (CV) (n = 4) and/or Bland-Altman differences (n = 3).

The QuADS criterion (Table 3) gave the included research a wide range of ratings. The quality of studies ranged from a low to high percentage. All papers have a relatively high-quality percentage ranging from 66 to 87%,[18–26,28] except for,[27] which was given a low score (35%). All assessed studies performed badly in reporting the involvement of stakeholders in research design or conduct and were scored with zero.

The normative values of DTI metrics in GM and WM were reported in five studies as shown in Table 4. Three studies reported the population gender[20,23,24] whereas two studies did not.[19,22] Age of examined populations was reported in all studies[19,20,22–24] with three of them examining paediatrics groups and two adults. The effect of age on DTI metrics was investigated by Taso, et al[23] at C2 and C5, and found that the age group did not systematically impact the DTI indices. The DTI measures differed depending on the cervical level examined, sample group (healthy and non-healthy), gender, age, as well as acquisition parameters such as gradient directions.

Table 1. Technical details of Diffusion Tensor Imaging (DTI) studies on healthy and nonhealthy cervical spinal cord

Study	Scanner Make	Field strength (T)	Gradient directions	Voxel size (mm3)	FOV (mm)	b-values (mm2/s)	Signal averages	TR (ms)	TE (ms)	Acquisition	No. of repeated scans [time interval]	ROI method	Software for DTI metrics extraction
Smith et al[18]	Philips	3	16	NA	145 × 120	500	3	6300	63	Axial	Twice [1 week]	Manual	CATBAP
Mohamed, Feroze B., et al.[19]	General Electric	1.5	6	NA	128 × 128	700	2	6000	60	Axial	Twice [4–5 weeks]	Manual	DTI Studio
Barakat, N., et al.[20]	Siemens	3	20	1.2 × 1.2×3	NA	1000	3	(6100–8000)	115	Axial	Twice [9 h]	Manual	MedINRIA
Mulcahey et al.[21]	Siemens	3	20	1.2 × 1.2×3	NA	1000	3	(6100–8000)	115	Axial	Twice [9 h]	Manual	NA
Barakat et al.[22]	Siemens	3	20	1.2 × 1.2×3	NA	1000	3	(6100–8000)	115	Axial	Twice [3 months]	Manual	MedINRIA
Taso, Manuel, et al.[23]	Siemens	3	30	NA	128 × 128	800	4	3RR	79.2	Axial	NA	Automatic	Neuro 3D (Siemens)
By, Samantha, et al.[24]	Philips	3	6,15 and 32	NA	64 × 48	750	9,4 and 2	5000	50	Axial	Three [10 days]	Automatic	Camino
Peterson, D. J., et al.[25]	Siemens	3	10	NA	200 × 100	750	6	2600	90	Axial	Twice [afterwards]	Automatic	Spinal Cord Toolbox (SCT)
Martin, A. R., et al.[26]	General Electric	3	25	NA	80 × 80	800	NA	4050	91.2	Axial	Twice [afterwards]	Automatic	Spinal Cord Toolbox (SCT)
Guan, Li, et al.[27]	Philips	3	15	NA	180 × 180	600	1	2500	82	Axial	Twice [10 days]	Manual	DTI Studio
Lee, Eugene, et al.[28]	Philips	3	15	2 × 2×2	250	600	4	3400	60	Axial	Twice [two weeks]	Manual	DTI task card (Philips)

FOV, field of view; NA, not available; ROI, region of interest; RR, Corresponds to a minimum time of TE (3 seconds); T, Tesla; TE, echo time; TR, repetition time.

Table 2. Repeatability and Reproducibility of Diffusion Tensor Imaging of Healthy and Unhealthy studies of Cervical Spinal Cord

Authors (year)	Subject No.	Statistical analysis	C-spine levels examined	Region of interest	Repeatability	Reproducibility
Smith et al.[18]	9 Healthy participants	BA	Between C2-C7	Right, left, and posterior columns	Intrarater **Lateral columns (Rt and Lt) using BA differences** **LOA [95% CI], (D_{BA}%)** MD: −0.58,0.60 [-0.22,0.24], (0.9%); AD: −0.97, 0.81 [-0.04, 0.03], (3.8%); RD: −0.48, 0.59 [-0.16, 0.27], (8.4%); FA: −0.18, 0.12 [-0.09, 0.03], (4.6%) Intrarater **Dorsal columns using BA differences** **LOA [95% CI], (D_{BA}%)** MD: −0.30, 0.32 [0.11, 0.14], (1.3%); AD: −0.35, 0.36 [-0.13, 0.14], (0.1%); RD: −0.28, 0.33 [-0.09, 0.14], (3.6%); FA: −0.09, 0.11 [-0.03, 0.05], (1.6%).	Inter-rater **Lateral columns (Rt and Lt) using D_{BA} differences** **LOA [95% CI], (D_{BA}%)** MD: −0.25, 0.23 [-0.01, 0.08], (1.0%); AD: −0.17, 0.17 [-0.06, 0.07], (0.2%); RD: −0.31, 0.27 [-0.00, 0.09], (2.9%); FA: −0.10, 0.14 [-0.03, 0.07], (3.0%) Interrater **Dorsal columns using BA differences** **LOA [95% CI], (D_{BA}%)** MD: −0.20, 0.16 [-0.09, 0.05], (1.7%); AD: −0.20, 0.17 [-0.1, 0.05], (1.1%); RD: −0.20, 0.18 [-0.09, 0.06], (2.8%); FA: −0.07, 0.09 [-0.02, 0.04], (1.5%)
Mohamed, Feroze B., et al.[19]	5 Children with SCI	ICC	Between C1-C7	Whole WM and GM area	**Intrascans, ICC [95% CI]** MD: 0.77 [0.0–0.98]; AD: 0.80 [0.0–0.98]; RD: 0.77 [0.0–0.98]; FA: 0.97 [0.0–0.98].	-
Barakat, N., et al,[22]	25 Healthy paediatric	ICC	Between C1-T1 at each disk and mid-level of cervical vertebral body	Whole WM and GM area	Intrascans **C1-T1** **ICC [95% CI]** MD-ICC: 0.95 [0.92–0.98]; AD-ICC: 0.97 [0.92–0.99]; RD-ICC: 0.91 [0.83–0.98]; FA ICC: 0.87 [0.78–0.97] Intrascans **Per cord level: ICC [95%, CI not reported]** **C1**: MD = 0.68, AD = 0.73, RD = 0.70, FA = 0.59)**; Mid dens**: MD = 0.70, AD = 0.70, RD = 0.69, FA = 0.55)**; Mid C2 body:** MD = 0.71, AD = 0.73, RD = 0.73, FA = 0.55) C2-C3: MD = 0.76, AD = 0.76, RD = 0.76, FA = 0.56)**; Mid C3**: MD = 0.76, AD = 0.77, RD = 0.76, FA = 0.61) C3-C4: MD = 0.76, AD = 0.76, RD = 0.74, FA = 0.58)**; Mid C4**: MD = 0.73, AD = 0.75, RD = 0.71, FA = 0.59)**; C4-C5**: MD = 0.72, AD = 0.71, RD = 0.69, FA = 0.63)**; Mid C5**: MD = 0.68, AD = 0.65, RD = 0.64, FA = 0.57)**; C5-C6**: MD = 0.61, AD = 0.59, RD = 0.61, FA = 0,58)**; Mid C6**: MD = 0.39, AD = 0.51, RD = 0.39, FA = 0.62)**; C6-C7**: MD = 0.51, AD = 0.48, RD = 0.46, FA = 0.68) Mid C7: MD = 0.54, AD = 0.56, RD = 0.45; FA = 0.70)**; C7-T1**: MD = 0.57, AD = 0.61, RD = 0.48, FA = 0.43)	-

(Continued)

Table 2. (Continued)

Authors (year)	Subject No.	Statistical analysis	C-spine levels examined	Region of interest	Repeatability	Reproducibility
Mulcahey et al.[21]	10 Paediatric patients with SCI	ICC	Between C1-T1 at each disk and mid-level of cervical vertebral body	Whole cord at each level of the CSC	Intrascans **Per cord level: ICC [95% CI]: C1**: MD = 0.94[0.80–0.99], AD = 0.90[0.65–0.97], RD = 0.92[0.72–0.98], FA = 0.50[−.018 to 0.85]; **Mid dens**: MD = 0.94[0.77–0.98], AD = 0.86[0.51–0.96], RD = 0.91[0.69–0.98], FA = 0.62[0.07–0.89]; **Mid C2 body**: MD = 0.93 [0.74–0.98], AD = 0.87 [0.57–0.97], RD = 0.93 [0.74–0.98], FA = 0.60 [−0.05 to 0.88] **C2-C3**: MD = 0.89 [0.62–0.97], AD = 0.87 [0.51–0.97], RD = 0.93 (0.74–0.98), FA = 0.61 [−0.01 to 0.88]; **Mid C3**: MD = 0.87 [0.57–0.97], AD = 0.87 [0.52–0.97], RD = 0.88 [0.59–0.97], FA = 00.69 [0.14–0.91] **C3-C4**: MD = 0.87 [0.55–0.97], AD = 0.90 [0.64–0.97], RD = 0.82 [0.42–0.95], FA = 0.69 [0.12–0.91]; **Mid C4**: MD = 0.87 [0.54–0.97], AD = 0.92 [0.69–0.98], RD = 0.83 [0.45–0.96], FA = 0.75 [0.26–0.93] **; C4-C5**: MD = 0.92 [0.71–0.98], AD = 0.89 [0.61–0.97], RD = 0.87 [0.54–0.97], FA = 0.71 [0.19–0.92]**; Mid C5**: MD = 0.90 [0.68–0.98], AD = 0.89 [0.61–0.97], RD = 0.88 [0.60–0.97], FA = 0.58 [−0.04 to 0.88] **C5-C6**: MD = 0.95 [0.82–0.99], AD = 0.94 [0.73–0.99], RD = 0.94[(0.79–0.99], FA = 0.82 [0.41–0.95]; **Mid C6**: MD = 0.93 [0.75–0.98], AD = 0.91 [0.64–0.97], RD = 0.93 [0.75–0.98], FA = 0.89 [0.61–0.97]; C6-C7: MD = 0.88 [0.46–0.97], AD = 0.88 [0.48–0.97], RD = 0.90 [0.56–0.97], FA = 0.82 [0.43–0.95] **Mid C7**: MD = 0.87 [0.33–0.97], AD = 0.82 [0.33–0.96], RD = 0.89 [0.40–0.98], FA = 0.81 [0.35–0.95]**; C7-T1**: MD = 0.80 [0.15–0.95], AD = 0.82 [0.27–0.95], RD = 0.84 (0.44–0.96], FA = 0.82 [0.36–0.96]	-
Barakat et al.[20]	10 Normal paediatric	ICC	Between C1-T1 at each disk and mid-level of cervical vertebral body	Free-hand and fixed-size ROIs methods on the entire cord	-	-
Taso, Manuel, et al.[23]	3 Healthy volunteers	CV	At C2 and C5	Whole WM, Lt and Rt WM pathways, and anterior GM	**Intrascan CV % [averaged for the whole cervical cord regions]** MD: 3.4%; AD: 3.5%; RD: 9.7%; FA: 5.1%	-
By, Samantha, et al.[24]	5 Healthy volunteers	BA	Between C2-C5	Whole WM and GM regions	**WM region: Intrascan using BA differences** **LOA [95% CI], (D_{BA}%); WM at six directions** MD: −0.18,0,24 [-0.02,0.08], (3.15); AD: −0.27,0.37 [-0.02,0.13], (2.48); RD: −0.21,0.26 [-0.03,0.08], (4.63); FA: −0.15,0.13 [-0.04,0.02], (1.09)**; WM at 15 directions**: MD: −0.11,0.15 [-0.01,0.05], (2.28); AD: −0.19,0.26 [-0.02,0.09], (1.82); RD: −0.17,0.020 [-0.03,0.06], (3.1); FA: −0.12,0.12 [-0.03,0.03], (0.09)**; WM at 32 directions**: MD: −0.31,0.20 [-0.11,0.01], (5.3) ; AD: −0.26,0.23 [-0.07,0.04], (0.75); RD: −0.38,0.24 [-0.14,0.00], (13.36); FA: −0.12,0.09 [0.0,0.07], (5.4) **GM region: Intrascan, using BA differences** **LOA [95% CI], (D_{BA}%): GM at six directions**: MD: −0.16,0.17 [-0.03,0.04], (0.42) ;AD: −0.27,0.25 [-0.07,0.05], (0.48); RD: −0.15,0.18 [-0.03,0.05], (1.78); FA: −0.12,0.09 [-0.04,0.01], (2.84)**;** **GM at 15 directions:** MD: −0.16,0.16 [-0.04,0.04], (0.07); AD: −0.30,0.35 [-0.05,0.11], (1.83) ; RD: −0.16,0.12 [-0.05,02], (2.54); FA: −0.10,0.12 [-0.01,0.04], (2.43)**; GM at 32 directions**: MD: −0.26,0.22 [-0.08,0.03], (2.34); AD: −0.43,0.45 [-0.09,0.11], (0.8); RD: −0.24,0.17 [-0.09,0.01], (6.57); FA: −0.10,0.13 [-0.01,0.04], (2.53)	-

(Continued)

Table 2. (Continued)

Authors (year)	Subject No.	Statistical analysis	C-spine levels examined	Region of interest	Repeatability	Reproducibility
Peterson, D. J., et al.[25]	30 patients with acute cervical spine	ICC, CV	Extended from the foramen magnum to the C7-T1	Whole WM, GM and other 24 individual tracts	**Intrareader of MD: (CV%, ICC) and FA: (CV%, ICC)** **GM**: (8.1%,0.66), (6.1% ,0.83)**; WM**: (5.6%,0.86), (3.8%,0.93) **; WM regions:** Ventral: (12.0%,0.75), (10.5%,0.72); Lateral: (10.4%,0.71), (6.2% ,0.84) ; Dorsal: (11.3%,0.71), (7.7%, 0.80) **; Ventral WM tracts:** Ventral reticulospinal tract: (18.3%,0.75), (18.2%, 0.47); Ventral corticospinal tract: (13.5%,0.69), (15.2%, 0.56) ; Lateral vestibulospinal tract: (17.0%,0.62), (14.0%,0.58) ; Spino-olivary tract: (13.9%,0.74), (12.7% ,0.68) ; Tectospinal tract: (13.2%,0.58), (10.5%, 0.69) **; Lateral WM tracts:** Ventral spinocerebellar tract: (13.5%,0.72), (11.4%, 0.72);Ventrolateral reticulospinal tract: (14.6%,0.66), (12.1%, 0.69) ; Spinal lemniscus: (13.9%,0.67), (8.6%, 0.78) ; Rubrospinal tract: (18.0%,0.56), (10.9%, 0.66) ; Lateral corticospinal tract: (12.9%,0.69), (8.2%, 0.80) **;Dorsal WM tracts:** Fasciculus cuneatus: (11.6%,0.70), (8.0%, 0.79) Fasciculus gracilis: (12.5%,0.66), (8.9%, 0.75) All individual WM tracts: (14.5%,0.75), (10.8%, 0.81)	**Interreader of MD: (CV%, ICC) and FA: (CV%, ICC)** GM: (3.8%,0.93), (4.8% ,0.90) ; WM: (5.3%,0.88), (3.0%,0.95) **WM**: **regions:** Ventral: (7.6%,0.90), (6.8% ,0.88); Lateral: (9.2%, 0.78), (4.5%, 0.92) ; Dorsal: (9.4%,0.80), (5.6%,0.89) ; **Ventral WM tracts:** Ventral reticulospinal tract: (11.9%,0.82), (13.0%, 0.73) ; Ventral corticospinal tract: (10.6%, 0.81), (11.2%, 0.76); Lateral vestibulospinal tract: (10.0%,0.87), (9.4% ,0.81); Spino-olivary tract: (9.7%,0.87), (8.3% ,0.87) ; Tectospinal tract: (9.7%,0.85), (6.5%, 0.88) **Lateral WM tracts:** Ventral spinocerebellar tract: (11.3%,0.81), (8.8%, 0.83); Ventrolateral reticulospinal tract: (10.3%,0.83), (9.3% ,0.81) Spinal lemniscus: (9.8%,0.84), (5.4% ,0.91) Rubrospinal tract: (16.1%,0.65), (8.3%, 0.81) Lateral corticospinal tract: (11.6%,0.75), (6.4%,0.88); **Dorsal WM tracts:** Fasciculus cuneatus: (9.8%,0.79), (5.8%, 0.89); Fasciculus gracilis: (10.4%,0.77), (6.6%, 0.86) All individual WM tracts: (11.0%, 0.86), (7.8%,0.90)
Martin, A. R., et al.[26]	17 Healthy volunteers and 9 patients with degenerative cervical myelopathy (DCM)	CV	Between C1-C7: Rostral(C1-C3), middle (C4-C5), caudal (C6-C7) levels	Whole cord, WM, GM over both right, left lateral corticospinal tract, fasciculus gracilis, fasciculus cuneatus and spinal lemniscus.	**Intrasubjects of FA across restrocaudal level: (CV %): Healthy Rostral(C1-C3):** (2.5%): **Mid cervical or MCL** (C4-C5): (3.0%) **Caudal** (C6-C7): (2.2%)**; DCM**: **Rostral(C1-C3):** (2.8%); Mid cervical or MCL (C4-C5): (5.0%); **Caudal (C6-C7**): (4.6%)	-
Guan, Li, et al.[27]	10 Healthy volunteers	CV, BA	Between C1 and C7	Whole cord	-	**Interscan: CV [averaged for the whole regions] and BA not reported as numbers**: FA:7.1%, AD:7.6%; RD:16.6%,; MD:9.7%.

(Continued)

Table 2. (Continued)

Authors (year)	Subject No.	Statistical analysis	C-spine levels examined	Region of interest	Repeatability	Reproducibility
Lee, Eugene, et al.[28]	34 patients with cervical spondylotic myelopathy (CSM)	ICC	Between C1-T1 at each disk level	Whole cord	Intraobserver **Per cord level: FA ICC [95% CI] for mean ROI method for observer1, 2,3 and 4, respectively**. **C1/2**: 0.460 [0.149–0.688], 0.558 [0.275–0.752], 0.986 [0.973–0.993] and 0.995 [0.990–0.997]**; C2/3**: 0.892 [0.794–0.944], 0.919 [0.845–0.959], 0.977 [0.955–0.987], and 0.992 [0.985–0.996]**; C3/4**: 0.948 [0.898–0.974], 0.843 [0.709–0.919], 0.991 [0.982–0.996], and 0.992 [0.983–0.996] **; C4/5**: 0.910 [0.828–0.954], 0.959 [0.920–0.979], 0.945 [0.894–0.972], and 0.994 [0.989–0.997] **; C5/6**: 0.635 [0.383–0.800], 0.807 [0.647–0.899], 0.964 [0.930–0.982], and 0.980 [0.961–0.990] **C6/7**: 0.582 [0.308–0.767], 0.813 [0.657–0.902], 0.946 [0.895–0.973], and 0.990 [0.980–0.995]**; C7/T1**: 0.834 [0.692–0.913], 0.735 [0.532–0.858], 0.980 [0.961–0.990], and 0.995 [0.989–0.997] Intraobserver **Per cord level: FA ICC [95% CI] for manual ROI method** **C1/2**: 0.541 [0.253–0.741], 0.524 [0.230–0.730], 0.969 [0.940–0.985], and 0.997 [0.9932–0.998]**; C2/3**: 0.798 [0.632–0.894], 0.888 [0.788–0.943], 0.944 [0.891–0.972], and 0.993 [0.986–0.996]**; C3/4**: 0.914 [0.834–0.956], 0.868 [0.753–0.932], 0.989 [0.978–0.995], and 0.984 [0.968–0.992] **C4/5**: 0.923 [0.851–0.961], 0.917 [0.840–0.957], 0.984 [0.968–0.992], and 0.990 [0.980–0.995]**; C5/6**: 0.888 [0.788–0.943], 0.783 [0.607–0.885], 0.981 [0.963–0.991], and 0.994 [0.988–0.997]**; C6/7**: 0.660 [0.418–0.814], 0.683 [0.452–0.828], 0.966 [0.932–0.983], and 0.992 [0.984–0.996] **; C7/T1**: 0.814 [0.659–0.903], 0.711 (0.494–0.844], 0.972 [0.944–0.986], and 0.983 [0.966–0.992] Intraobserver **Per cord level: FA ICC [95% CI] for Sagittal ROI method** **C1/2**: 0.810 [0.653–0.901], 0.684 [0.453–0.828], 0.889 [0.790–0.943], and 0.645 [0.396–0.805]**; C2/3**: 0.646 [0.398–0.806], 0.816 [0.663–0.904], 0.926 [0.856–0.962], and 0.973 [0.948–0.987]**; C3/4**: 0.891 [0.794–0.944], 0.942 [0.886–0.970], 0.913 [0.834–0.956], and 0.965 [0.931–0.982]**; C4/5**: 0.797 [0.631–0.893], 0.808 [0.650–0.899], 0.865 [0.746–0.930], and 0.974 [0.949–0.987]**; C5/6**: 0.659 [0.418–0.814], 0.840 [0.703–0.917], 0.934 [0.872–0.966], and 0.887 [0.785–0.942] **C6/7**: 0.518 [0.223–0.727], 0.813 [0.658–0.902], 0.924 [0.854–0.961], and 0.977 [0.955–0.989]**; C7/T1**: 0.798 [0.632–0.894], 0.815 [0.661–0.903], 0.919 [0.844–0.959], and 0.924 [0.854–0.961]	Interobserver **Per cord level: FA ICC [95% CI] for mean ROI method first and second measurements of observer 1 and 2 &first and second measurements of observer 3 and 4** **C1/2**: 0.775 [0.595–0.881], 0.510 [0.213–0.721] & 0.690 [0.462–0.832],0.647 [0.400–0.807] **C2/3**: 0.925 [0.855–0.962], 0.906 [0.819–0.952] & 0.652 [0.406–0.810], 0.700 [0.477–0.838] **C3/4**: 0.954 [0.909–0.977], 0.868 [0.752–0.932] & 0.682 [0.451–0.827], 0.646 [0.398–0.806] **C4/5**: 0.948 [0.898–0.974], 0.932 [0.869–0.966] & 0.663 [0.423–0.816], 0.757 [0.566–0.871] **C5/6**: 0.783 [0.608–0.885], 0.580 [0.305–0.765] & 0.641 [0.391–0.803], 0.634 [0.380–0.799] **C6/7**: 0.596 [0.327–0.776], 0.854 [0.727–0.924] & 0.461 [0.150–0.689], 0.555 [0.271–0.750] **C7/T1**: 0.892 [0.795–0.945], 0.675 [0.440–0.823] & 0.678 [0.445–0.825], 0.694 [0.469–0.834] Interobserver **Per cord level: FA ICC [95% CI] for manual ROI method first and second measurements of observer 1 and 2 &first and second measurements of observer 3 and 4** **C1/2**: 0.744 [0.546–0.864], 0.635 [0.382–0.799] & 0.551 [0.266–0.747], 0.579 [0.304–0.765] **; C2/3**: 0.930 [0.864–0.964], 0.789 [0.618–0.889] & 0.747 [0.551–0.865], 0.709 [0.491–0.843] **; C3/4**: 0.899 [0.807–0.948], 0.923 [0.851–0.961] & 0.792 [0.622–0.890], 0.800 [0.636–0.895] **; C4/5**: 0.967 [0.936–0.984], 0.913 [0.834–0.956] & 0.756 [0.565–0.870], 0.773 [0.591–0.880] **; C5/6**: 0.843 [0.708–0.918], 0.838 [0.700–0.916] & 0.594 [0.325–0.774], 0.607 [0.342–0.782] **; C6/7**: 0.728 [0.520–0.854], 0.617 [0.357–0.789] & 0.478 [0.1716–0.700], 0.570 [0.291–0.759]; **C7/T1**: 0.861 [0.740–0.928], 0.799 [0.635–0.894] & 0.715 [0.501–0.847], 0.679 [0.447–0.826] Interobserver **Per cord level: FA ICC [95% CI] for Sagittal ROI method first and second measurements of observer 1 and 2 &first and second measurements of observer 3 and 4** **C1/2**: 0.809 [0.651–0.900], 0.781 [0.604–0.884] & 0.479 [0.173–0.701], 0.430 [0.113–0.668] **C2/3**: 0.667 [0.428–0.818], 0.814 [0.660–0.903] & 0.275 [-0.065–0.558], 0.302 [-0.036–0.578] **C3/4**: 0.924 [0.853–0.961], 0.907 [0.823–0.953] & 0.532 [0.241–0.735], 0.624 [0.366–0.793] **C4/5**: 0.795 [0.628–0.892], 0.830 [0.689–0.912] & 0.422 [0.102–0.662], 0.410 [0.088–0.654] **C5/6**: 0.676 [0.442–0.824], 0.771 [0.589–0.879] & 0.222 [-0.121–0.517], 0.404 [0.081–0.650]**; C6/7**: 0.649 [0.402–0.808], 0.780 [0.604–0.884] & 0.595 [0.326–0.775], 0.628 [0.372–0.795]**; C7/T1**: 0.762 [0.575–0.874], 0.764 [0.578–0.875] & 0.084 [-0.257–0.406], 0.157 [-0.186–0.467]

AD, axial diffusivity; CI, Confidence Interval; CV, coefficient of variation; D_{BA}, Bland-Altman differences; FA, fractional anisotropy; ICC, intra-class correlation; Lt, left; MD, mean diffusivity; RD, radial diffusivity; Rt, right.

Table 3. Quality assessment with diverse studies (QuADS) scores for each paper

Paper	Smith et al. 2009[18]	Mohamed, Feroze B., et al. 2010[19]	Barakat, N., et al. 2011[20]	Mulcahey et al. (2011)[21]	Barakat et al.[22]	Taso, Manuel, et al.[23]	By, et al.[24]	Peterson, D. J., et al. (2017)	Martin, A. R., et al. (2017)	Guan, Li, et al.[27]	Lee, Eugene, et al.[28]
Criteria	**Score (0–3)**										
1.Theoretical or conceptual underpinning to the research	3	2	3	3	2	2	3	3	3	1	2
2.Statement of research aim/s	3	3	3	3	3	2	3	3	3	1	3
3.Clear description of research setting and target population	3	3	3	3	1	2	2	3	3	1	3
4.The study design is appropriate to address the stated research aim/s	3	3	3	3	3	3	3	3	3	2	3
5.Appropriate sampling to address the research aim/s	0	1	2	2	2	1	0	1	2	1	0
6.Rationale for choice of data collection tool/s	2	2	2	2	2	3	3	3	3	1	1
7.The format and content of data collection tool is appropriate to address the stated research aim/s	3	3	3	3	3	3	3	3	3	1	2
8.Description of data collection procedure	3	2	2	3	2	3	3	3	3	1	2
9.Recruitment data provided	1	2	3	3	2	1	1	1	3	1	3
10.Justification for analytic method selected	2	2	2	2	2	3	1	3	3	1	1
11.The method of analysis was appropriate to answer the research aim/s	3	3	3	3	3	3	3	3	3	2	3
12.Evidence that the research stakeholders have been considered in research design or conduct	0	0	0	0	0	0	0	0	0	0	0
13.Strengths and limitations critically discussed	3	3	2	2	2	1	2	3	2	1	3
Total scores	29	29	31	31	27	27	27	32	34	14	26
Score as Percentage	74%	74%	79%	79%	69%	69%	69%	82%	87%	35%	66%

Table 4. Mean and standard deviation (SD) normative value of mean diffusivity (MD, $\times10^{-3}$ mm^2/s), axial diffusivity (AD, $\times10^{-3}$ mm^2/s), radial diffusivity (RD, $\times10^{-3}$ mm^2/s), and fractional anisotropy (FA, unitless) for white matter (WM) and grey matter (GM) in healthy and non-healthy CSC

Author (population)	Gender (mean age)	C-spinal level	DTI metrics (mean ± SD)
Mohamed, Feroze B., et al.[19](paediatrics)	Not mentioned (healthy: 15.2 years and SCI: 11.6 years)	Between C1-C7	**Averaged for entire WM and GM ROI** **Healthy:** MD = 0.72±0.17; AD = 1.23±0.29; RD = 0.44±0.24, and FA = 0.62±0.11. **SCI:** MD = 1.27±0.67; AD = 1.65±0.65; RD = 1.06±0.69, and FA = 0.39±0.22.
Barakat, N., et al.[22] (paediatrics)	6 males and 19 females (healthy: 13.28 years, age ranged 7–21 years)	Between C1-T1	**Averaged for entire WM and GM ROI** **Healthy:** MD = 0.59±0.15; AD = 0.97±0.20; RD = 0.41±0.13; and FA = 0.50±0.11.
Barakat et al.[22] (paediatrics)	Not mentioned (healthy: 12.10 years, age ranged 9–15)	Between C1-T1	Averaged for entire WM and GM **Healthy:** **Freehand ROI:** **Rater1:** **Trial 1:** MD = 0.92±0.53; AD = 1.35±0.52; RD = 0.70±0.54, and FA = 0.50±0.13. **Trial 2:**MD = 0.72±0.18; AD = 1.16±0.18; RD = 0.52±0.17, and FA = 0.47±0.13. **Rater2:** **Trial 1:** MD = 1.10±0.65; AD = 1.53±0.65; RD = 0.88±0.64; and FA = 0.48±0.12. **Trial 2:** MD = 0.72±0.14; AD = 1.20±0.16; RD = 0.48±0.17; and FA = 0.52±0.10. **Fixed ROI:** **Rater1:** **Trial 1:** MD = 0.70±0.16; AD = 1.18±0.19; RD = 0.46±0.14; and FA = 0.54±0.10. **Trial 2:** MD = 0.71±0.14; AD = 1.18±0.16; RD = 0.47±0.13, and FA = 0.54±0.10. **Rater2,** **Trial 1:** MD = 1.16±0.56; AD = 1.64±0.58; RD = 0.91±0.54, and FA = 0.53±0.10. **Trial 2:** MD = 0.70±0.12; AD = 1.19±0.15; RD = 0.45±0.11, and FA = 0.54±0.10.
Taso, Manuel, et al.[23] (adults)	26 male and 22 females (43 ± 15 years, age ranged 21–68 years)	At C2 and C5	**WM:** **For age group<35:** **C2**: MD = 1.03±0.09; AD = 2.05±0.07; RD = 0.52±0.13; and FA = 0.72±0.06. **C5:** MD = 0.99±0.07; AD = 1.91±0.06; RD = 0.53±0.11; and FA = 0.68±0.07. **For age group between 35 and 50:** **C2:** MD = 1.01±0.08; AD = 2.04±0.09; RD = 0.50±0.13; and FA = 0.73±0.06. **C5:** MD = 1.02±0.07; AD = 1.95±0.07; RD = 0.55±0.10; and FA = 0.68±0.05. **For age group between above 50:** **C2:** MD = 0.99±0.07; AD = 1.06±0.16; RD = 0.50±0.08; and FA = 0.71±0.04. **C5:** MD = 1.06±0.16; AD = 1.06±0.16; RD = 0.63±0.18; and FA = 0.65±0.06. **GM:** **For age group<35:** **C2**: MD = 0.97±0.07, and FA = 0.55±0.08. **C5:** MD = 0.91±0.09, and FA = 0.46±0.07. **For age group between 35 and 50:** **C2:** MD = 0.95±0.08, and FA = 0.57±0.07. **C5:** MD = 0.89±0.10, and FA = 0.48±0.08. **For age group above 50:** **C2:** MD = 0.95±0.20, and FA = 0.52±0.06. **C5:** MD = 0.96±0.06, and FA = 0.44±0.07.
By, Samantha, et al.[24] (adults)	three male and 2 females (age ranged 25–36 years)	Between C2-C5	**Healthy:** **WM with six directions: scan1**: MD = 1.03±0.12; AD = 2.11±0.27; RD = 0.49±0.08; FA = 0.73±0.05; **scan2**: MD = 1.00±0.08; AD = 2.06±0.17; RD = 0.47±0.10; FA = 0.74±0.07. **WM with 15 directions: scan1**: MD = 1.00±0.07; AD = 1.92±0.13; RD = 0.53±0.08; FA = 0.68±0.05; **scan2**: MD = 0.97±0.06; AD = 1.89±0.13; RD = 0.52±0.10; FA = 0.68±0.07. **WM with 32 directions: scan1**: MD = 0.96±0.06; AD = 1.89±0.12; RD = 0.50±0.07; FA = 0.69±0.04. **scan2**: MD = 1.02±0.13; AD = 1.91±0.10; RD = 0.57±0.18; FA = 0.66±0.10. **GM with six directions: scan1**: MD = 0.99±0.12; AD = 1.79±0.23; RD = 0.59±0.11; FA = 0.61±0.07. **scan2**: MD = 0.99±0.09; AD = 1.79±0.20; RD = 0.58±0.08; FA = 0.62±0.06. **GM with 15 directions: scan1**: MD = 0.94±0.06; AD = 1.62±0.13; RD = 0.61±0.10; FA = 0.55±0.08. **scan2**: MD = 0.94±0.07; AD = 1.59±0.12; RD = 0.62±0.10; FA = 0.54±0.08. **GM with 32 directions: scan1**: MD = 0.92±0.07; AD = 1.60±0.19; RD = 0.57±0.08; FA = 0.57±0.09. **scan2**: MD = 0.94±0.11; AD = 1.59±0.13; RD = 0.61±0.15; FA = 0.56±0.10.

Reproducibility evaluation of DTI metrics

Different outcome measures were quantified for analysing the reproducibility of DTI studies. The coefficient of variation (CV) of DTI metrics was reported in four studies[23,25–27] that used different ROIs in the analyses (Table 2). The intrarater reproducibility of FA was high for the entire WM with CV of 3.8% and individual WM tracts with CV ranging from 8.0 to 18.2% across WM tracts. Similarly, MD also showed high test-retest reproducibility (WM: CV, 5.6%) with CV ranging from 11.6% in the fasciculus cuneatus to 18.3% in the ventral reticulospinal tract. Similar to FA, the MD of the ventral reticulospinal tract had the lowest reproducibility of the tracts assessed, likely due to its small

size.The interrater reproducibility of FA and MD was higher than the corresponding test-retest with CV less than 10 and 11.2%, respectively, except for ventral reticulospinal tract and rubrospinal tract regions with a CV for FA of 13.0 and 11.2%, respectively, and the CV of MD was less than 15% for all individual WM pathways except for the rubrospinal tract region. These different findings might be attributed to the fact that the Peterson, Rutman et al.[25] used different ROIs, which were automatically selected and reported. When larger volumes of interest were selected, the variation between repeated measurements decreased by 25.5%. Smaller WM tracts have a poorer degree of reproducibility than bigger WM tracts when it comes to DTI metrics in the cervical spinal cord and this was attributed to the poor in-plane image resolution, as with improved resolution, DTI measures of smaller WM tracts are more likely be more reproducible.[25]

The ICC of DTI metrics was reported in six studies[19–22,25,28] that used different ROIs in the analyses (Table 2) and ranged from poor (ICC<0.37) to excellent reproducibility (ICC 0.91–0.99). In 2011, Mohamed, Feroze B., et al demonstrated fair to moderate ICC in all DTI metrics with test-retest ICC≥0.72. In 2012, Mulcahey et al demonstrated good to strong reproducibility in the MD, AD, and RD per cervical cord level. The ICC for MD ranged from 0.80 to 0.95, AD ranged from 0.82 to 0.94, and RD ranged from 0.82 to 0.94 per cervical cord level. FA demonstrated moderate to good ICC at C1,mid-C4 and between C5-C6 and C7-T1 levels with ICC ranging from 0.50 to 0.89. In the same year, Barakat, N., et al reported a moderate-to-strong reproducibility of all DTI metrics with ICC of 0.87, 0.95, 0.97 and 0.91 for FA, MD, AD, and RD, respectively. Barakat et al. in 2015 reported a moderate-(ICC = 0.5)-to-strong (ICC = 0.84) between-rater and within-rater agreement using two different ROI methods (free-hand and fixed-size ROIs). FA values revealed the highest variability among DTI metrics in the ICC values (ranging from 0.10 to 0.87). The upper spinal cord levels between C1 and mid-C3 revealed the lower agreement value. RD revealed slightly higher agreement values than FA (0.26–0.83). Further, comparing to free hand drawn ROIs, the fixed-size ROIs for RD and FA reported lower agreement. Recently, Lee et al.[28] reported an overall variation in the ICC of interobserver reproducibility of FA ranging from poor-(ICC = 0.37)-to-excellent (ICC = 0.82) agreement among three different ROI methods (mean, manual, and sagital), with relatively less agreement for the sagittal ROI method. The test-retest reproducibility showed an excellent agreement with ICC ranging between 0.88 and 0.99 at almost every cervical cord level. In 2017, Peterson, et al reported moderate-to-excellent reproducibility of DTI metrics. The intrarater reproducibility of FA was high for the entire WM (ICC = 0.93) and to a lesser extent in all WM individual tracts (ICC = 0.81) with ICC ranging from 0.47 in the ventral reticulospinal tract to 0.80 in the lateral corticospinal tract. The ICC was shown to be improved with larger volumes of interest. Similarly, MD metrics had also high test-retest reproducibility measurements with ICC of 0.86 in WM and 0.75 in all individual WM tracts. The interrater reproducibility of FA and MD was higher than the coresponding test-retest with ICC larger than 0.80 in all individual WM tracts, except for the ventral reticulospinal tract and ventral corticospinal tract regions where the ICC of FA was 0.73 and 0.76, respectively, and the ICC of MD was larger than 0.70 for all individual WM pathways except for the rubrospinal tract region (ICC = 0.65).

The normalised Bland-Altman differences of DTI metrics was reported in three studies[18,24,27] which used multiple ROIs in the analyses (Table 2). The 95% CI overlaps zero for the mean differences between repeated scans with acceptable level of agreement for all the ROIs drawn. In all three studies, the normalised Bland–Altman (BA) difference for inter-readers analysis ranged from 1.89 to 2.06% and the intrareaders assessment was between 2.38 and 4.54% for all DTI metrics with the exception of RD for Smith et al. (2016), which showed a higher percentage (D_{BA} = 8.44%), and was attributed to a strong dependency on image resolution, as the sequence parameters resulted in reduced resolution and an enhanced signal-to-noise ratio (SNR).

DISCUSSION

Reproducible measurements of DTI would greatly facilitate the evaluation of progressive diseases that involve the cervical spinal cord, assist in treatment strategies as well as being able to be used for monitoring when disease is stable. The main purpose of this systematic review paper was to assess previous DTI studies that specially evaluated the repeatability and reproducibility of DTI metrics in assessments of the cervical spinal cord. Our findings show overall that DTI metrics reveal fairly good agreement between repeated measurements.

Factors that impact the reproducibility of DTI metrics

One of factors that might impact the reproducibility of DTI metrics is the segmentation methods applied for the cervical spinal cord. A total of seven studies[18–22,27,28] used manual ROIs, which are more bias-related, inconvenient, and labour-intensive and do not provide atlas-based segmentation assessments and tract-specific information. Moreover, as the cord volume decreases between C1 and T1, manual ROIs, at each disk and mid-level of cervical vertebral body, are subject to contamination from cerebrospinal fluid (CSF) or nearby tissue structures as discussed by Barakat, Shah et al.[22] ROIs which cover the entire cord, both WM and grey matter (GM) are frequently employed which provide limited details particularly about the disease influence on the WM regions, specifically when considering the differences in DTI indices between WM and GM, as well as the possible microstructural changes in disease impact between these tissues.[25] Moreover, in some spinal cord diseases such as cervical spondylotic myelopathy (CSM) disease, low spatial resolution DTI images and an atrophied spinal cord make it more challenging to draw an accurate ROI that involves only the spinal cord.[28] Conversely, four studies in this review[23–26] applied an automatic segmentation method. This approach provides tract-based indices and robust readouts from different ROIs, with the outcomes validated by suitable reproducibility analysis that can provide increased specificity with respect to clinical damage values when compared to the entire cord regions.[25] Another factor that might impact the reproducibility of DTI metrics is the cervical spinal cord level examined, as observed in Barakat et al,[20 22] and [28] where DTI characteristics were acquired at

different levels of spinal cord and reported different ICC values. One possible explanation for the variation of ICC of DTI parametric values in these studies was signal drop-off near the neck coil's edge. The lower cervical and higher thoracic levels were positioned at the neck coil's extremities, where SNRs declined, and CSF was accidentally included in areas with prominent image artifacts. Furthermore, the cervical levels closest to the heart (C4C7) are those most susceptible to cardiac motion.[32,33] As a result, some cardiac motion artifacts may have influenced the ROI selection or placement and when individual DTI parameters were measured, FA and RD demonstrated the greatest variability in ICC values. Moreover, cervical spine measurements may also be impacted by some movement artifacts related to breathing, CSF pulsation, blood flow and/or swallowing resulting in ghosting artifacts, particularly if the phase encoding direction was applied in the anterior to posterior (AP) direction. The reproducibility agreement was found to be low particularly at lower (C6, C7) cervical spinal cord levels in the Barakat, Shah et al.[22] study. cardiac motion-related artifacts can be minimised using cardiac triggering during the scanning. Although cardiac gating is beneficial for controlling spinal cord movement and reducing flow-related artifacts from CSF, it generally increases the scan time, potentially resulting in more motion artifacts from other sources. The most frequently utilised diffusion-weighted MRI pulse sequence is echo-planar imaging (EPI) ,which may help in reducing motion-related artifacts due to its very rapid acquisition time.[34,35] In Martin et al.[26], DTI sequences either with or without cardiac gating were used on 10 subjects and no differences were found for the CV of FA under the two conditions. Different acquisition parameters could be another impacting factor for the reproducibility of DTI measures. It was observed that the reproducibility of equivalent DTI technique varied greatly between included studies (Table 2). The clinical nature of DTI techniques necessitates optimising the SNR and contrast-to-noise ratio (CNR) by altering the DTI parameters to acquire the best images with the most diagnostic information per subject, which intrinsically influences reproducibility as well. Changes in, for example, echo time (TE), repetition time (TR) or voxel size have a significant impact on the signal intensity of voxels. These findings imply that the wide range of values for DTI parameters' reproducibility is suggested to be due to the wide range of settings for TE, TR, b-values, number of signal averages, voxel size and sequences used to generate the images. Standardized methods employing the same TR, TE, matrix sizes, and other parameters would improve DTI reproducibility and would help in quantitative analysis of DTI technique. Furthermore, reduced FOV (rFOV) EPI sequences such as ZOOM (zonally magnified oblique multislice) are also possible in DTI protocols, which provides adequate SNR and lower data distortions.[36] In addition, the number of averages was different in most of the included studies (ranging from 1 to 6 acquisition averages). It was observed that more signal averages produce lower relative noise and reduces the variation resulted from image corruption sources such as inconsistence in hardware in MRI systems and motion-related artifacts. However, as more signal averages lead to longer scan time, this consequently provides more time for the subject to move. In addition, the CNR was also seen to be higher with increased numbers of signal averages.[24] The selection of the number of diffusion gradient directions may also play a role in the reproducibility of DTI metrics, as reported by By et al.[24] where they investigated the effect of the number of diffusion gradient directions on the reproducibility of the quantitative DTI using normalised D_{BA} as well as limits of agreement (LoA). They compared results obtained 6, 15, and 32 directions with automatically segmented WM and GM ROIs between C2 and C5 and found that the D_{BA} of the two repeated scans were below 14% for all metrics, with the largest D_{BA} being observed from the 32-direction data. The 15-directions, on the other hand, reported the lowest D_{BA} for most of the metrics (MD: WM D_{BA} = 2.28%, GM D_{BA} = 0.07%; RD WM D_{BA} = 3.1%, FA GM D_{BA} = 0.09%, FA GM D_{BA} = 2.43%) showing a small variation overall across different time points with the 15-direction option as well as producing high and reproducible contrast. This might be due to the higher number of averages used in the 15-direction scheme (four times) comparing to the 32-direction scheme where only two excitations were applied; however, with more gradient directions, the boundaries of the cervical spinal cord can be readily detected, as small directions could result in larger variation in all DTI metrics, especially in the principle eigenvector (PEV) direction, which is an essential concern for tractography and might have an impact on the reproducibility of DTI metrics. Another possible factor that influences the reproducibility of DTI metrics is image artifacts.Although, DTI has been widely applied in brain studies, for example, in MS patients,[37] its application in the spinal cord remains challenging due to its small structural size, artifacts related to cerebrospinal fluid (CSF) contamination, as CSF DTI parameters are very different to GM and WM values,[22] magnetic field inhomogeneities and respiratory and cardiac motion.[27,38–41] In addition, partial volume contamination between WM and GM or between WM and CSF, distorted anatomy, inaccurate registration to the spinal cord toolbox (SCT) templates[42] may also influence the reproducibility of DTI metrics. These artifacts were found have an impact on the reproducibility of DTI indices as reported by[26] where a diminished reproducibility were observed specifically in mid cervical levels (MCLs) as well as caudal levels, but the overall variation was acceptable (CV<5%).

None of included studies have discussed the effect of age-related changes to the ICC and CV calculations, and thus, further research is required to examine the impact of this on the reliability and reproducibility of DTI metrics in CSC.

Normative values for DTI metrics for WM and GM

Imaging of the spinal cord is challenging and there is possibility of increased motion in a less-than-cooperative patients. Therefore, obtaining accurate DTI values may be more difficult in the paediatric population, especially with those who have had SCI. In injured cord, the FA map reveals reduced values, while MD, AD, and RD show increased values relative to a healthy population. This supports the usefulness of DTI technique in evaluating the tissue microstructure changes. In healthy paediatric population studies,[19,20,22] there was a gradual increase in AD and decrease in MD, and FA along the length of CSC in the superior-to-inferior direction (C1-T1), while RD stayed relatively constant. In addition, in Taso, et al[23] study, the differences between WM and GM was clearly seen where lower MD and FA

values observed in GM than WM. This is due to the fact that water diffusivity in GM tissues is substantially less restricted than in WM tissues. Moreover, there was also reduction in the AD and FA values of WM and FA of GM in C5 in comparison to that in C2 (fewer fibres in C5 than C2), while RD in WM was found to be higher at C5 comparing to C2 (higher extracellular space at C5). This was in line with previous reports,[43,44] which examined the morphological structure of WM and GM as the cord volume decreases between C1 and T1 (superior-to-inferior). Using a fixed ROI method such as that in the Barakat et al[22] study is difficult to maintain at levels where cord volume become much smaller compared with a free-hand ROI method, and some artefacts might arise due to tissue contamination with cerebrospinal fluid (CSF) impacting the DTI values. The use of the automated image processing software such as spinal cord toolbox (SCT)[42] might help to reduce the partial volume effects and provide better measurements of the DTI metrics. Further, DTI metrics could be effected depending on the cord levels examined as at the upper and lower end of neck coil (where SNR decreases) the DTI indices might be influenced by signal drop-off.[22]

DTI metrics have previously been reported for brain WM with differences in the values of DTI indices related to gender[45] as well as age.[46,47] In the current review, only the study by Taso, et al[23] investigated the DTI metrics over three different age groups (<35 years, between 35 and 50 years, and >50 years) at C2 and C5 found that at C5 people < 50 years had reduced AD and FA for WM metrics, and lower MD and FA values for GM in comparison with middle-age subjects (between 35 and 50 years). The same study also reported that none of the DTI metrics showed any differences between males and females in any ROI or age group.

The effect of different gradient directions on DTI measures was investigated in one study by By, et al,[24] where they applied three different gradient schemes (6,15, and 32 directions taking three different scanning times 4.5, 9, and 18 min) and generally found that the CNR was higher for a larger number of gradient directions and the CNR reduces as the scanning time decreases. In addition, they qualitatively assessed the images of DTI metrics and reported that the difference between GM and WM in the spinal cord for FA, AD, and RD was observed in all schemes, with less separation between different tissues at the 6-direction scheme. Additionally, the contrast between CSF and CSC was also less for the lower gradient direction scheme. The MD value was not sensitive due to its less dependent on directions number.[24]

Limitations

To the best of our knowledge, this is the first systematic review to evaluate the reproducibility of DTI metrics within the cervical spinal cord. However, it is important to note that there are some considerable barriers that limit this study. Firstly, a meta-analysis was not conducted due to a variety of reasons, including the heterogeneous nature of data collected, the fact that all included studies did not have the similar outcome measurements, that different regions of interest and vertebral levels were chosen among the included studies and that not all DTI metrics were evaluated. This review covered only the cervical spinal cord area as this area of the spinal cord is specifically affected in certain diseases; however, including the thoracic and lumber vertebral regions could improve the repeatability and reproducibility values for spinal cord DTI metrics generally.

CONCLUSION

This study has shown that the use of DTI and its related measures in the clinical setting, specifically in the evaluation of cervical spinal cord pathologies, is feasible and reproducible. However, cervical spinal cord DTI suffers from limitations that prevent it from being used routinely in research and clinical settings. More research into the techniques used is required to significantly overcome and improve these limitations, as well as to detect reliable time-specific changes in different DTI parameters, so that this technique may be utilised reliably and accurately in the clinical environment. The normative values of DTI metrics suggest that level-dependent evaluations should be considered carefully when assessing different pathologies of the spinal cord.

ACKNOWLEDGEMENTS

Authors would like to thank Mr. Rajeh Assiri for reviewing the protocol of this review.

CONTRIBUTORS

HA: Conceptualisation, Methodology, Writing –protocol and original draft. JF, CH: Investigation, Supervision, Re- sources, Writing – review & editing, Validation. RM: Writing – review & editing, Validation.

FUNDING

Financial disclosures: This study was financially funded 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.

REFERENCES

1. Beaulieu C. The basis of Anisotropic water diffusion in the nervous system–a technical review. *NMR Biomed* 2002; **15**: 435–55. https://doi.org/10.1002/nbm.782
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**: 247–54. https://doi.org/10.1006/jmrb.1994.1037
3. Schulte T, Sullivan EV, Müller-Oehring EM, Adalsteinsson E, Pfefferbaum A. Corpus Callosal Microstructural integrity influences Interhemispheric processing: a diffusion Tensor imaging study. *Cereb Cortex* 2005;

15: 1384–92. https://doi.org/10.1093/cercor/bhi020
4. Mukherjee P, Berman JI, Chung SW, Hess CP, Henry RG. Diffusion Tensor MR imaging and fiber Tractography: theoretic underpinnings. *AJNR Am J Neuroradiol* 2008; **29**: 632–41. https://doi.org/10.3174/ajnr.A1051
5. Bartlett JW, Frost C. Reliability, Repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound Obstet Gynecol* 2008; **31**: 466–75. https://doi.org/10.1002/uog.5256
6. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021; **372**: n71. https://doi.org/10.1136/bmj.n71
7. Haynes RB, McKibbon KA, Wilczynski NL, Walter SD, Werre SR, Hedges Team. Optimal search strategies for retrieving scientifically strong studies of treatment from MEDLINE: Analytical survey. *BMJ* 2005; **330**: 1179. https://doi.org/10.1136/bmj.38446.498542.8F
8. Ingui BJ, Rogers MA. Searching for clinical prediction rules in MEDLINE. *J Am Med Inform Assoc* 2001; **8**: 391–97. https://doi.org/10.1136/jamia.2001.0080391
9. Geersing G-J, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons KGM. Search filters for finding Prognostic and diagnostic prediction studies in MEDLINE to enhance systematic reviews. *PLoS One* 2012; **7**(2): e32844. https://doi.org/10.1371/journal.pone.0032844
10. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme version. 2006; **1**(1): b92.
11. Marenco S, Rawlings R, Rohde GK, Barnett AS, Honea RA, Pierpaoli C, et al. Regional distribution of measurement error in diffusion Tensor imaging. *Psychiatry Research: Neuroimaging* 2006; **147**: 69–78. https://doi.org/10.1016/j.pscychresns.2006.01.008
12. Carlson HL, Laliberté C, Brooks BL, Hodge J, Kirton A, Bello-Espinosa L, et al. Reliability and variability of diffusion Tensor imaging (DTI) Tractography in pediatric epilepsy. *Epilepsy & Behavior* 2014; **37**: 116–22. https://doi.org/10.1016/j.yebeh.2014.06.020
13. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing Rater reliability. *Psychol Bull* 1979; **86**: 420–28. https://doi.org/10.1037//0033-2909.86.2.420
14. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating Normed and standardized assessment instruments in psychology. *Psychological Assessment* 1994; **6**: 284–90. https://doi.org/10.1037/1040-3590.6.4.284
15. Harrison R, Jones B, Gardner P, Lawton R. Quality assessment with diverse studies (Quads): an appraisal tool for methodological and reporting quality in systematic reviews of Mixed- or multi-method studies. *BMC Health Serv Res* 2021; **21**: 144. https://doi.org/10.1186/s12913-021-06122-y
16. Sirriyeh R, Lawton R, Gardner P, Armitage G. Reviewing studies with diverse designs: the development and evaluation of a new tool. *J Eval Clin Pract* 2012; **18**: 746–52. https://doi.org/10.1111/j.1365-2753.2011.01662.x
17. Page M, McKenzie J, Bossuyt P, Boutron I, Hoffmann T, Mulrow C. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. [Internet]. 2021; **372**.
18. Smith SA, Jones CK, Gifford A, Belegu V, Chodkowski B, Farrell JAD, et al. Reproducibility of Tract-Specific Magnetization transfer and diffusion Tensor imaging in the Cervical spinal cord at 3 Tesla. *NMR Biomed* 2009; **23**: n. https://doi.org/10.1002/nbm.1447
19. Mohamed FB, Hunter LN, Barakat N, Liu C-SJ, Sair H, Samdani AF, et al. Diffusion Tensor imaging of the pediatric spinal cord at 1.5 T: preliminary results. *AJNR Am J Neuroradiol* 2011; **32**: 339–45. https://doi.org/10.3174/ajnr.A2334
20. Barakat N, Mohamed FB, Hunter LN, Shah P, Faro SH, Samdani AF, et al. Diffusion Tensor imaging of the normal pediatric spinal cord using an inner field of view echo-planar imaging sequence. *AJNR Am J Neuroradiol* 2012; **33**: 1127–33. https://doi.org/10.3174/ajnr.A2924
21. Mulcahey MJ, Samdani A, Gaughan J, Barakat N, Faro S, Betz RR, et al. Diffusion Tensor imaging in pediatric spinal cord injury: preliminary examination of Reliability and clinical correlation. *Spine (Phila Pa 1976)* 2012; **37**: E797–803. https://doi.org/10.1097/BRS.0b013e3182470a08
22. Barakat N, Shah P, Faro SH, Gaughan JP, Middleton D, Mulcahey MJ, et al. Inter-and intra-Rater reliability of diffusion Tensor imaging parameters in the normal pediatric spinal cord. *World J Radiol* 2015; **7**: 279–85. https://doi.org/10.4329/wjr.v7.i9.279
23. Taso M, Girard OM, Duhamel G, Le Troter A, Feiweier T, Guye 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**: 817–32. https://doi.org/10.1002/nbm.3530
24. By S, Smith AK, Dethrage LM, Lyttle BD, Landman BA, Creasy JL, et al. Quantifying the impact of underlying measurement error on Cervical spinal cord diffusion Tensor imaging at 3t. *J Magn Reson Imaging* 2016; **44**: 1608–18. https://doi.org/10.1002/jmri.25308
25. Peterson DJ, Rutman AM, Hippe DS, Jarvik JG, Chokshi FH, Reyes MR, 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. *AJNR Am J Neuroradiol* 2017; **38**: 2015–20. https://doi.org/10.3174/ajnr.A5334
26. Martin AR, De Leener B, Cohen-Adad J, Cadotte DW, Kalsi-Ryan S, Lange SF, et al. Clinically feasible Microstructural MRI to quantify Cervical spinal cord tissue injury using DTI, MT, and T2*-Weighted imaging: assessment of normative data and reliability. *AJNR Am J Neuroradiol* 2017; **38**: 1257–65. https://doi.org/10.3174/ajnr.A5163
27. Guan L, Chen X, Hai Y, Ma X, He L, Wang G, et al. High-Resolution diffusion Tensor imaging in Cervical Spondylotic Myelopathy: a preliminary Follow-Up study. *NMR Biomed* 2017; **30**(10): 10. https://doi.org/10.1002/nbm.3769
28. Lee E, Lee JW, Bae YJ, Kim HJ, Kang Y, Ahn JM. Reliability of pre-operative diffusion Tensor imaging parameter measurements of the Cervical spine in patients with Cervical Spondylotic Myelopathy. *Sci Rep* 2020; **10**: 1–9. https://doi.org/10.1038/s41598-020-74624-6
29. Wei X, Guttmann CRG, Warfield SK, Eliasziw M, Mitchell JR. Has your patient's multiple sclerosis lesion burden or brain atrophy actually changed *Mult Scler* 2004; **10**: 402–6. https://doi.org/10.1191/1352458504ms1061oa
30. Velasco-Annis C, Akhondi-Asl A, Stamm A, Warfield SK. Reproducibility of brain MRI Segmentation Algorithms: empirical comparison of local MAP PSTAPLE, Freesurfer, and FSL-FIRST. *J Neuroimaging* 2018; **28**: 162–72. https://doi.org/10.1111/jon.12483
31. Brander A, Koskinen E, Luoto TM, Hakulinen U, Helminen M, Savilahti S, et al. Diffusion Tensor imaging of the Cervical spinal cord in healthy adult population: normative values and measurement reproducibility at 3t MRI. *Acta Radiol* 2014; **55**: 478–85. https://doi.org/10.1177/0284185113499752
32. Kharbanda HS, Alsop DC, Anderson AW, Filardo G, Hackney DB. Effects of cord

motion on diffusion imaging of the spinal cord. *Magn Reson Med* 2006; **56**: 334–39. https://doi.org/10.1002/mrm.20959
33. Figley CR, Stroman PW. Investigation of human Cervical and upper Thoracic spinal cord motion: implications for imaging spinal cord structure and function. *Magnetic Resonance in Med* 2007; **58**: 185–89. https://doi.org/10.1002/mrm.21260
34. Poustchi-Amin M, Mirowitz SA, Brown JJ, McKinstry RC, Li T. Principles and applications of echo-planar imaging: a review for the general Radiologist. *Radiographics* 2001; **21**: 767–79. https://doi.org/10.1148/radiographics.21.3.g01ma23767
35. Stroman PW, Wheeler-Kingshott C, Bacon M, Schwab JM, Bosma R, Brooks J, et al. The current state-of-the-art of spinal cord imaging: methods. *Neuroimage* 2014; **84**: 1070–81. https://doi.org/10.1016/j.neuroimage.2013.04.124
36. Samson RS, Lévy S, Schneider T, Smith AK, Smith SA, Cohen-Adad J, et al. ZOOM or non-ZOOM? assessing spinal cord diffusion Tensor imaging protocols for multi-centre studies. *PLoS One* 2016; **11**(5): e0155557. https://doi.org/10.1371/journal.pone.0155557
37. Filippi M, Rocca MA, Comi G. The use of quantitative magnetic-resonance-based techniques to monitor the evolution of multiple sclerosis. *Lancet Neurol* 2003; **2**: 337–46. https://doi.org/10.1016/s1474-4422(03)00408-3
38. Clark CA, Werring DJ. Diffusion Tensor imaging in spinal cord: methods and applications–a review. *NMR Biomed* 2002; **15**: 578–86. https://doi.org/10.1002/nbm.788
39. Clark CA, Werring DJ, Miller DH. Diffusion imaging of the spinal cord in vivo: estimation of the principal Diffusivities and application to multiple sclerosis. *Magn Reson Med* 2000; **43**: 133–38. https://doi.org/10.1002/(SICI)1522-2594(200001)43:1<133::AID-MRM16>3.0.CO;2-X
40. Mottershead JP, Schmierer K, Clemence M, Thornton JS, Scaravilli F, Barker GJ, et al. High field MRI correlates of myelin content and axonal density in multiple sclerosis. *J Neurol* 2003; **250**: 1293–1301. https://doi.org/10.1007/s00415-003-0192-3
41. Wheeler-Kingshott CAM, Hickman SJ, Parker GJM, Ciccarelli O, Symms MR, Miller DH, et al. Investigating Cervical spinal cord structure using axial diffusion Tensor imaging. *Neuroimage* 2002; **16**: 93–102. https://doi.org/10.1006/nimg.2001.1022
42. De Leener B, Lévy S, Dupont SM, Fonov VS, Stikov N, Louis Collins D, et al. SCT: spinal cord Toolbox, an open-source software for processing spinal cord MRI data. *Neuroimage* 2017; **145**: 24–43. https://doi.org/10.1016/j.neuroimage.2016.10.009
43. Taso M, Le Troter A, Sdika M, Ranjeva J-P, Guye M, Bernard M, et al. Construction of an in vivo human spinal cord Atlas based on high-resolution MR images at Cervical and Thoracic levels: preliminary results. *Magn Reson Mater Phy* 2014; **27**: 257–67. https://doi.org/10.1007/s10334-013-0403-6
44. Fradet L, Arnoux P-J, Ranjeva J-P, Petit Y, Callot V. Morphometrics of the entire human spinal cord and spinal canal measured from in vivo high-resolution anatomical magnetic resonance imaging. *Spine (Phila Pa 1976)* 2014; **39**: E262–9. https://doi.org/10.1097/BRS.0000000000000125
45. Hsu J-L, Leemans A, Bai C-H, Lee C-H, Tsai Y-F, Chiu H-C, et al. Gender differences and age-related white matter changes of the human brain: a diffusion Tensor imaging study. *Neuroimage* 2008; **39**: 566–77. https://doi.org/10.1016/j.neuroimage.2007.09.017
46. Barnea-Goraly N, Menon V, Eckert M, Tamm L, Bammer R, Karchemskiy A, et al. White matter development during childhood and adolescence: a cross-sectional diffusion Tensor imaging study. *Cereb Cortex* 2005; **15**: 1848–54. https://doi.org/10.1093/cercor/bhi062
47. Bonekamp D, Nagae LM, Degaonkar M, Matson M, Abdalla WMA, Barker PB, et al. Diffusion Tensor imaging in children and adolescents: reproducibility, Hemispheric, and age-related differences. *Neuroimage* 2007; **34**: 733–42. https://doi.org/10.1016/j.neuroimage.2006.09.020