Musculoskeletal Science and Practice 44 (2019) 102056

Contents lists available at ScienceDirect

Musculoskeletal Science and Practice

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

Original article

Reliability of cervical muscle volume quantification using magnetic resonance imaging

Suzanne J. Snodgrass[a,b,*], de Zoete Rutger M.J.[a,b,c], Croker Christopher[a,b], Meghana Yerrapothu[d], James M. Elliott[d,e]

[a] Faculty of Health and Medicine, The University of Newcastle, Australia
[b] Centre for Brain and Mental Health Research, The University of Newcastle, Australia
[c] Recover Injury Research Centre, NHMRC Centre of Research Excellence in Recovery Following Road Traffic Injuries, The University of Queensland, Herston, Australia
[d] Neuromuscular Imaging Research Laboratory, Northwestern University, Feinberg School of Medicine, Department of Physical Therapy and Human Movement Sciences, Chicago, USA
[e] Northern Sydney Local Health District, The Kolling Institute, and Faculty of Health Sciences, The University of Sydney, Australia

ARTICLE INFO

Keywords:
Cervical muscle volume
Magnetic resonance imaging
Reliability

ABSTRACT

Background: Magnetic resonance imaging (MRI) is used to quantify the size and structure of the architecturally complex cervical spine musculature of individuals with traumatic and idiopathic neck pain. However, to our knowledge, no scan-rescan reliability data is available on neck muscle volumes.

Objectives: This study investigates the intra- and inter-rater reliability and scan-rescan reliability in cervical muscle volume investigations.

Design: Clinical Measurement, Reliability study.

Methods: MRI scans were performed and repeated (within 1 h) for five asymptomatic individuals. Two raters manually traced levator scapulae, multifidus including semispinalis cervicis, semispinalis capitis, splenius capitis including splenius cervicis, and sternocleidomastoid using Analyze software (v12.0). Reliability was determined using intra-class correlation coefficients, Lin's concordance coefficient and Bland-Altman plots, with interpretation of reliability coefficients using the criteria from Fleiss.

Results: Intra-rater reliability of muscle quantification was excellent (ICCs ranging from 0.78 to 0.96). Inter-rater reliability was excellent for sternocleidomastoid(ICC 0.92, 95% CI 0.80, 0.97) and splenius capitis (ICC 0.77, 0.51, 0.90), and ranged from fair to good for levator scapulae (0.63, 0.18, 0.85), multifidus (0.73, 0.44, 0.88), and semispinalis capitis (0.50, 0.08, 0.77). The scan-rescan reliability was excellent for all muscles (ICCs ranging from 0.94 to 0.98).

Conclusion: Threats to reliability appear to be more related to manual quantification of muscles on images rather than protocols related to re-positioning a participant in the scanner and repeating the same protocol.

The current findings suggest that the proposed methods can be used in establishing normative data for cervical muscle volume and comparing individuals with and without neck pain.

1. Introduction

Neck pain is a prevalent problem in the general population: globally approximately 70% of adults experience neck pain at least once in their lives (Brontfort et al., 2012; Hoy et al., 2014). Despite a wide variety of available treatments, many neck pain conditions are still poorly understood, leading to a high proportion of cases of recurrent neck pain (Haldeman et al., 2008) and providing rationale for further research into potential pathophysiological explanations.

Magnetic resonance imaging (MRI) of the cervical spine is, when clinically warranted, commonly used in individuals with neck pain to evaluate potential pathology that may be related to clinical symptoms, and thus, potentially amenable to targeted treatments (Kaiser and Holland, 1998). Structures of interest include vertebrae, joints, discs, and also cervical musculature (Matsumoto et al., 1998; Siivola et al., 2002). Cervical spine muscle quality has been defined by the size of muscle tissue quantified by muscle volume or muscle cross-sectional area on MRI, as well as the amount or proportion of fatty infiltrate in the muscle (Elliott et al., 2018). Cervical

* Corresponding author. Discipline of Physiotherapy, School of Health Sciences, The University of Newcastle, Callaghan, NSW, 2308, Australia.
E-mail addresses: Suzanne.Snodgrass@newcastle.edu.au (S.J. Snodgrass), rutger.dezoete@uon.edu.au (R.M.J. de Zoete), christopher.croker@uon.edu.au (C. Croker), myerr95@gmail.com (M. Yerrapothu), jim.elliott@sydney.edu.au (J.M. Elliott).

https://doi.org/10.1016/j.msksp.2019.102056
Received 17 March 2019; Received in revised form 6 June 2019; Accepted 4 September 2019
Available online 05 September 2019
2468-7812/

spine muscle quality has been shown to be affected (increased muscle fat infiltration) in individuals with traumatic neck pain (Abbott et al., 2015; Elliott et al., 2006, 2012, 2014, 2015; Karlsson et al., 2016), but to a lesser extent in individuals with idiopathic neck pain (reduced size but less muscle fat infiltration) (Elliott et al., 2014; Elliott et al., 2008a,b). While cause and effect is difficult to ascribe, changes in muscle composition have been suggested to be associated with the presence and potentially maintenance of neck pain(Abbott et al., 2015; Elliott et al., 2006, 2011; Karlsson et al., 2016), however, inconsistent associations exist (Anderson et al., 2012; Matsumoto et al., 2010, 2012; Ulbrich et al., 2012). We believe such discrepancies across studies, may in part be due to limitations in and the different approaches used for assessment and analysis of imaging findings (Elliott et al., 2018).

The use of MRI to investigate the composition of cervical and lumbar spine musculature has been widely reported in symptomatic populations but compositional muscle changes in asymptomatic individuals, to our knowledge, appear most commonly across the lumbar spine (Amabile et al., 2017; Valentin et al., 2015). Normative data of cervical muscle composition is lacking and needed. Recently, recommendations for building consensus around the imaging parameters, sequences, and methodologies for investigations of cervical muscle composition were published (Elliott et al., 2018). It is possible these recommendations will help to build consensus around consistent measurements of cervical muscle volume and muscle fatty infiltration in patients with and without neck pain.

As the cervical spine consists of an architecturally complex, multilayered musculature, manual measures of muscle quantification can be challenging and assessor-dependent. Our own experiences, and in discussing with colleagues familiar with, or interested in, measuring cervical muscle composition, have revealed that identifying the muscle borders on the available transverse (axial) slices can prove challenging, even for the most experienced. An important step before investigations of muscle composition can routinely be utilised in research or clinical practice is establishing reliability and agreement of the measure (Elliott et al., 2018). This study investigates the intra- and inter-rater reliability of quantification of cervical muscle composition, as well as the reliability of the scanning procedure (scan-rescan reliability) used to quantify cervical muscle compositional volume on transverse MRI images using a fat/water Dixon sequence.

2. Materials and methods

2.1. Procedures

A convenience sample of five asymptomatic participants were scanned twice on the same day within 1 h. Participants were asked to exit the scanning room and be repositioned for the second scan, at least 20 min later (the amount of time it took to scan a different participant). A single experienced and registered radiographer positioned the head in approximately neutral, using the same coil for every participant to standardise alignment. A foam pad was placed under the head for participant comfort and their head was secured on either side with additional padding as appropriate to head size to minimise head movement. The radiographer ensured the participant remained stationary by observing them on a monitor. No additional or specific instructions were used for positioning and the radiographer used their standard practice. Data for this reliability study were extracted from MR scans that were collected for a larger study Snodgrass et al., 2019, approved by the University of Newcastle Research Ethics Committee (H-2015-0235). All participants provided written informed consent.

2.2. MRI parameters

MRI was undertaken on a Siemens Magnetom Prisma 3 T scanner with a 64-channel head coil. An axial, volumetric interpolated breath-hold examination (VIBE) using two-point Dixon fat-water separation (Dixon-VIBE) (TR/TE1/TE2 7.05/2.46/3.69 ms) was undertaken with a 320 × 320 mm field of view (FoV) and 448 × 448 acquisition matrix (0.7 mm in-plane resolution) with a slice thickness of 3 mm. A single slab with 52 slices was acquired covering the cephalad portion of C3 through the caudal portion of the T2 vertebral end plate in a scan duration of 6:23 min. Axial slices were aligned parallel to the C2/3 intervertebral disc allowing MR slices to perpendicularly intersect muscles, a necessity for measuring muscle cross-section.

2.3. Muscle volume quantification

Prior to quantifying muscle volumes on individual MRI images, two raters identified the location of each slice in relation to the cervical vertebrae by assigning each slice to a specific spinal level using visualization of its location on a sagittal localizer view in OsiriX Lite Version 8 (Pixmeo, Bernex, Switzerland). Individual slices were assigned to vertebral levels by first identifying the slice closest to the midsection of each intervertebral disc. Slices between these were assigned to their corresponding vertebral levels using the same sagittal view.

Muscle volume was quantified by a researcher ("Rater 1 or 2") manually tracing the fascial boundary of selected neck muscles using a computer mouse on every second slice collected, with interpolation to the remaining slices performed in Analyze Direct (v. 12.0 – www.analyzedirect.com). Raters were an undergraduate student in the final year of a physiotherapy degree and a research assistant with a neuroscience degree. See Fig. 1 for an example of neck muscle identification at level C4. To extract data from slices that were not manually identified, Analyze software performed automated interpolation, with the software detecting and outlining fascial borders identified as differences in pixel intensity using also the data from manually traced slices to estimate muscle borders. Muscle border outlines on interpolated slices were checked for accuracy by the researcher (Rater 1 or Rater 2) and edited as needed. Three-dimensional models of each muscle generated by the software that included interpolated slices were also visually examined to augment the manual identification and interpolation processes (Fig. 2), and if needed, the identified muscle borders were edited before data extraction. Raters were blinded to all participant information and did not participate in data collection.

Muscles were differentiated with reference to an MRI anatomical atlas outlining the muscles at each level (Au et al., 2016). Muscles were followed in caudal and cephalad directions to their origin and insertions where relevant to confirm identification of muscles. Muscles quantified bilaterally were multifidus (with semispinalis cervicis) [Mult], semispinalis capitis [Semisp], splenius capitis (with splenius cervicis) [Splenius], levator

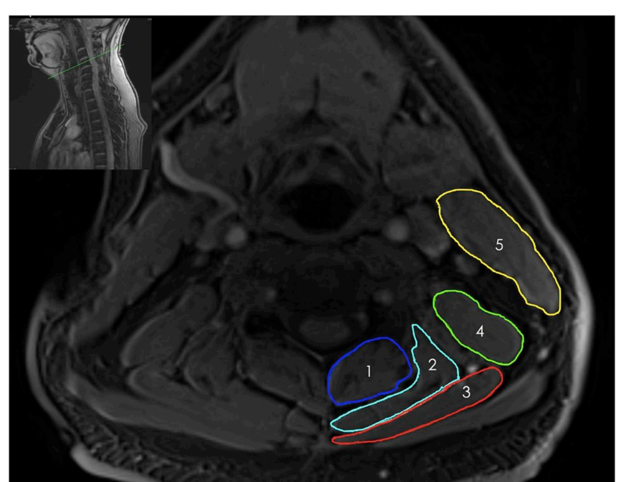

Fig. 1. Cross-sectional aspect at spinal level C4 used for manual identification of cervical muscle volumes (1 = multifidus with semispinalis cervicis; 2 = semispinalis capitis; 3 = splenius capitis with splenius cervicis; 4 = levator scapulae; 5 = sternocleidomastoid).

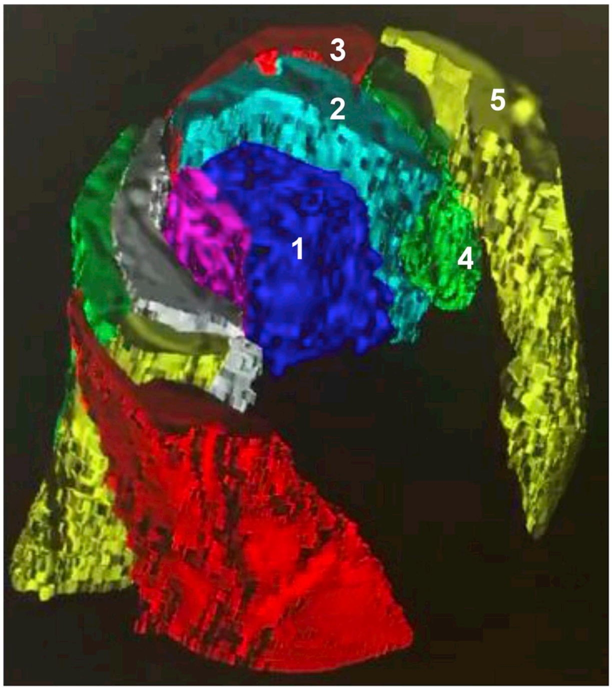

Fig. 2. Three-dimensional model created in Analyze 12.0 and used by raters to check their identification of the lower cervical extensor muscles (1 = multifidus with semispinalis cervicis; 2 = semispinalis capitis; 3 = splenius capitis with splenius cervicis; 4 = levator scapulae; 5 = sternocleidomastoid) as quantified by MRI at spinal levels C3–C6.

scapulae [Lev Scap], and sternocleidomastoid [SCM]. These muscles were selected as they encompass all major deep and superficial lower cervical extensor muscles, as well as one large flexor/rotator (sternocleidomastoid). Multifidus was traced with semispinalis cervicis and splenius capitis was traced with splenius cervicis because there were poorly visualized fascial borders between the muscles comprising these muscle pairs that risked introducing measurement error. These muscle group pairs will be referred to as 'multifidus' and 'splenius' further in this manuscript. From cephalad to caudad, muscles were quantified on slices corresponding to the intervertebral disc of C2-3 to the intervertebral disc of C6-7 (i.e., spinal levels C3–C6). The exception to this was the multifidus which was quantified from slices corresponding to the intervertebral disc of C3-4 to the intervertebral disc of C6-7 (spinal levels C4-6) due to its anatomical location.

2.4. Data analysis

Five participants were considered adequate for assessing reliability of these methods, as each participant had a large number of observations, resulting from examing the left and right-sided muscles separately and multiple MR images (or slices) per scan. To determine the reliability of the identification of spinal levels from the MR images, two raters identified spinal levels on 10 scans from 5 participants; slices where there was disagreement were noted and percent agreement was calculated. To determine the intra-rater reliability of identification of volumes from the MR images, one rater manually identified muscles on two scans from two participants. To determine the inter-rater reliability of muscle volume identification, two raters traced muscles on 10 scans from five participants. Intra- and inter-rater reliability were calculated for each muscle using data for each slice. To provide a more intuitive representation of overall muscle volume reliability, inter-rater reliability was also calculated using the overall volume data for each muscle (summed data for individual participants for all slices where a muscle was traced) and including the left and right as separate muscles (n = 10 for each muscle, i.e., left and right muscles for five participants). This was not done for intra-rater reliability as there would be only four observations per muscle for comparison, as only two participants' muscles were manually identified twice by the same rater.

To determine the reliability of the scanning protocol, the muscle volumes for each muscle of the five participants that were rescanned were compared; this was evaluated using data from Rater 1.

For intra- and inter-rater reliability of muscle volume identification and for the scan to scan reliability, the following statistics are presented: intra-class correlation coefficients ($ICC_{2,1}$) interpreted using the guidelines of Fleiss et al. where < 0.4 is poor, 0.4 to 0.75 is fair to good, and > 0.75 is excellent (Fleiss, 1986), standard error of measurement (SEM), Lin's concordance coefficient and Bland-Altman plots.

3. Results

Participants were a mean age of 35.8 years (SD 15.3, range 24–59) and all were male. Two of the five participants had previously sought treatment for idiopathic neck pain, but neither had received treatment in the previous 12 months.

3.1. Identification of spinal levels

On three occasions, raters disagreed on the identification of the slice corresponding to the intervertebral disc (three different spinal levels on three different participants). This resulted in disparate labelling by the two raters on six slices out of a total of 256 across the five participants, resulting in 97.7% agreement in the identification of spinal levels. The number of slices varied between participants because of the varying length of different participants' cervical spines and the use of a standardised slice thickness.

3.2. Intra- and inter-rater reliability of muscle volume quantification

3.2.1. Data from each MR slice

Intra-rater reliability of muscle volume identification was excellent, with most ICCs above 0.9 and only one muscle with an ICC of 0.779 (Table 1). Examining ICCs for inter-rater reliability of muscle volume

Table 1
Intra-rater reliability of manual identification of left and right cervical muscle volumes of asymptomatic participants (n = 2) using repeated measures from a single rater and data measured from each individual MRI image (or slice).

Muscle	Number of observations	Time 1 mean(SD)	Time 2 mean (SD)	ICC	95% CI	SEM	Lin's	95% CI
Levator scapulae	98	999.8 (473.7)	981.31 (763.7)	.961	.943, .974	122.19	.961	.942, .974
Multifidus[a]	74	1201.3 (286.1)	1167.9 (321.5)	.954	.919, .973	65.17	.953	.930, .969
Semispinalis	98	933.6 (331.8)	923.5 (328.0)	.937	.907, .957	82.80	.936	.907, .957
Splenius capitis[b]	98	839.0 (209.4)	786.6 (290.5)	.779	.646, .859	93.96	.777	.690, .843
Sternocleidomastoid	98	1080.2 (244.7)	1111.1 (290.5)	.903	.854, .935	83.35	.902	.862, .931
Total	466	999.3 (344.4)	985.1 (355.6)	.940	.928, .950	85.73	.931	.929, .950

SEM: standard error of measurement; Lin's: Lin's concordance coeficient.
[a] Includes semispinalis cervicis.
[b] Includes splenius cervicis.

Table 2
Inter-rater reliability of manual identification of left and right cervical muscle volumes of asymptomatic participants (n = 5) using repeated measures from two raters and data measured from each individual MRI image (or slice).

Muscle	Number of observations	Rater 1 mean(SD)	Rater 2 mean (SD)	ICC	95% CI	SEM	Lin's	95% CI
Levator scapulae	510	1046.5 (563.1)	1194.1 (636.7)	.903	.757, .950	186.85	.903	.888, .917
Multifidus[a]	418	1256.1 (301.2)	1282.8 (395.9)	.727	.678, .769	182.11	.727	.682, .766
Semispinalis	512	936.8 (305.7)	1091.9 (472.5)	.577	.428, .681	253.06	.576	.527, .621
Splenius capitis[b]	512	859.3 (178.5)	882.7 (277.2)	.476	.407, .541	164.93	.476	.415, .533
Sternocleidomastoid	512	1111.3 (296.6)	1120.4 (328.6)	.837	.809, .861	126.22	.837	.809, .860
Total	2464	1033.8 (378.3)	1107.9 (460.8)	.793	.752, .825	190.88	.793	.778, .806

SEM: standard error of measurement; Lin's: Lin's concordance coeficient.
[a] Includes semispinalis cervicis.
[b] Includes splenius cervicis.

identification, analysed by comparing the tracings of two raters for each MR slice/image, reliability was excellent for Lev Scap and SCM, and fair to good for Splenius, Semisp and Mult (Table 2).

3.2.2. Data from each muscle

Table 3 reports inter-rater reliability for the quantification of volume for each muscle. Reliability was excellent for Splenius and SCM, and fair to good for Lev Scap, Mult and Semisp (Table 3). Fig. 3 shows the Bland-Altman and Lin's concordance coefficient plots for inter-rater reliability of manual identification of muscle volumes.

3.3. Scan-rescan reliability

Scan-rescan reliability is reported in Table 4. The scan-rescan reliability was excellent for all muscles, with all ICCs being ≥ 0.935. Fig. 4 shows the Bland-Altman and Lin's concordance coefficient plots for the scan-rescan reliability.

4. Discussion

This preliminary study demonstrates that the intra- and inter-rater reliability, as well as the scan-rescan reliability, of manual identification of cervical muscle volumes from MRI images shows good reliability when compared to muscle volume quantifications in other areas (Fischmann et al., 2013; Kivle et al., 2018). This suggests that the proposed MRI methods (Elliott et al., 2018) and manual muscle identification technique can be reliably used in investigations of cervical spine muscle composition. However, whether the same reliability metrics can be demonstrated in a population of symptomatic participants where muscle compositional changes are expected, remains unknown and needs to be established.

The inter-rater reliability ranged between excellent (for the Splenius and SCM) and fair to good (for the Lev Scap, Mult, and Semisp). The $ICC_{2,1}$ value for the inter-rater reliability for the Semisp muscle volume was lower (0.502) than those for the other muscles (ranging from 0.628 to 0.915). Potentially this is due to its anatomical location between the Mult (including the semispinalis cervicis) and the Splenius (including the splenius cervicis), making it difficult to differentiate as it rapidly changes shape in the two-dimensional image view from C7 to C3 (Elliott et al., 2018). Though volume quantification of the multifidus demonstrated a good inter-rater reliability with an $ICC_{2,1}$ value of 0.728, the Bland-Altman plot (Fig. 1), suggests reliability may not be as good when the values for Mult are larger. For six of the 20 muscle volumes quantified on the images, the difference between raters sat outside 2 SD from the mean difference. In four of these six instances, the mean Mult muscle volume was greater than any other Mult volume measures, and the other two instances were also large values for Mult. This suggests that there is more potential for error when manually identifying Mult when it is larger. In the current study, the Mult was traced together with the semispinalis cervices and splenius capitis and it is difficult to delineate the fascial borders of the Mult from these other two muscles. Combining these additional muscles, each that may be visible at different spinal levels, may contribute to the reliability of quantifying Mult. These findings support the use of software such as Analyze, where the fat and water images can be viewed simultaneously, facilitating the ability to visually check muscle borders on each image while performing the manual identification generating a region of interest (ROI). Furthermore, using a semi-automated segmentation algorithm from higher resolution images may prove useful for quantifying volume in these challenging muscle groups. However, this is not without an increase in scanning time, which could increase the risk movement artefact. Nevertheless, future research may be able to determine the optimal imaging parameters for a semi-automated (or someday, automated) segmentation algorithm.

The intra-rater reliability of manually identified cervical muscle volumes was excellent (ICCs 0.903-0.961) for the Lev Scap, Mult, Semisp, and SCM muscles, and good (ICC 0.779) for the Splenius. The lower $ICC_{2,1}$ value for the Splenius is possibly due to the position of this muscle in relation to the Semisp. Similar to the lower inter-rater reliability for the Semisp, differentiation between the Splenius and Semisp can be difficult, which may lead to lower intra-rater reliability.

The scan-rescan reliability analyses were performed using two scans

Table 3
Inter-rater reliability of manual identification of left and right cervical muscle volumes of asymptomatic participants (n = 5) who were scanned twice (ie, two observations per person for both left and right muscles) using repeated measures from two raters and summed data from an individual's MRI images (slices) for each of five muscles.

Muscle	# observations	Rater 1 mean(SD)	Rater 2 mean (SD)	ICC	95% CI	SEM	Lin's	95% CI
Levator scapulae	18	26686.4 (10526.1)	34041.0 (20446.5)	.628	.182, .850	9445.37	.568	.396, .701
Multifidus[a]	20	24178.2 (3367.1)	24859.8 (4330.0)	.728	.439, .882	2007.15	.571	.341, .737
Semispinalis	20	23982.4 (4154.4)	27952.5 (8749.5)	.502	.082, .770	4553.09	.490	.252, .672
Splenius capitis[b]	20	21997.8 (3504.4)	22596.8 (4199.2)	.770	.513, .902	1847.26	.761	.508, .893
Sternocleidomastoid	20	28450.0 (6443.9)	28682.0 (6154.5)	.915	.800, .966	1836.52	.911	.792, .964
Total	98	24841.4 (6241.2)	27495.5 (10838.5)	.662	.504, .771	4964.87	.660	.570, .733

SEM: standard error of measurement; Lin's: Lin's concordance coeficient.
[a] Includes semispinalis cervicis.
[b] Includes splenius cervicis.

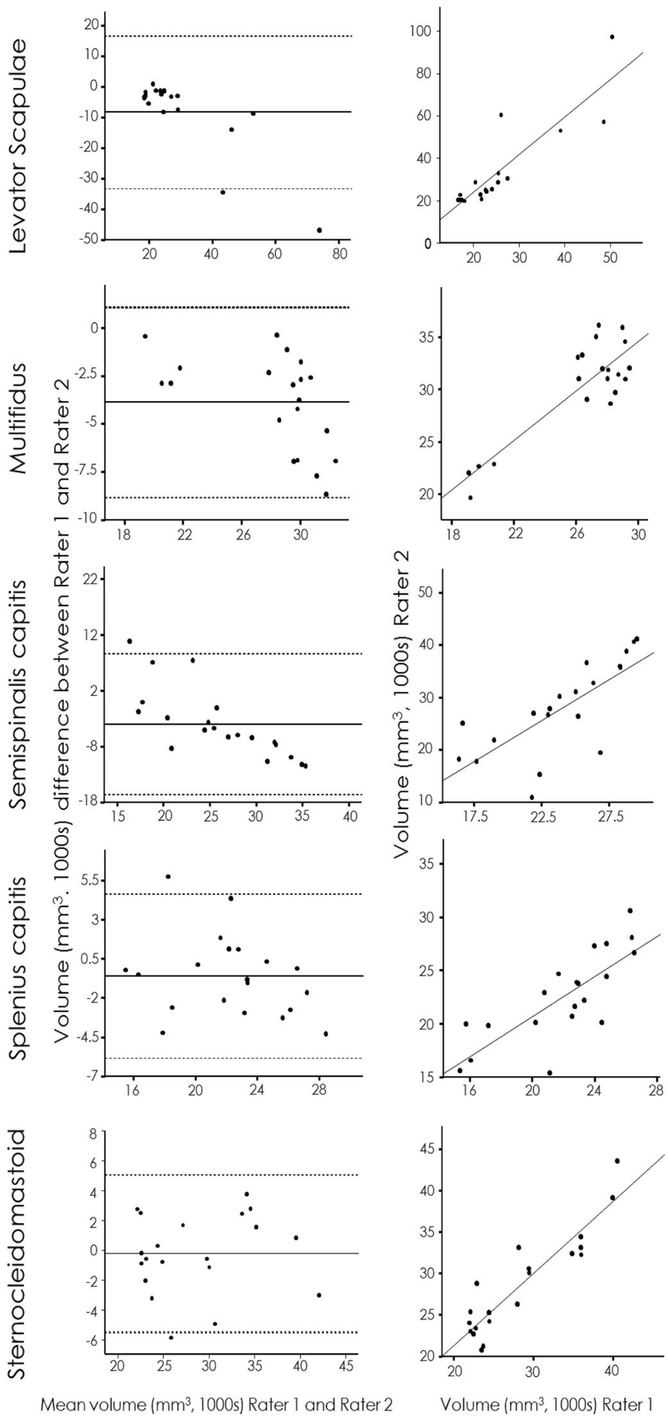

Fig. 3. Bland-Altman plots (left) and Lin's concordance (right) for inter-rater reliability of muscle volumes for five cervical muscles from the intervertebral discs of C2-3 through C6-7 quantified by two raters. All axes present muscle volumes (mm^3) in 1000s. Multifidus includes semispinalis cervicis; splenius capitis includes splenius cervicis.

from each of the five participants, both manually identified by Rater 1. The $ICC_{2,1}$ values ranged from 0.935 to 0.984, indicating an excellent scan-rescan reliability. The lower level of agreement reported for the inter-rater reliability analysis, i.e., inconsistent manual identification of combined muscles, was not present in the scan-rescan analysis, suggesting that future studies performing repeated measurements from MRI might benefit from having a single person performing the manual identification. The data also indicate that consistency in patient positioning resulting in reliable MR images for cervical muscle volume can be achieved using simple standardised methods and instructions from a radiographer.

To our knowledge, few studies have investigated the reliability of MRI scanning procedures, or scan-rescan reliability, of the cervical muscles or other soft-tissues. Fischmann et al. (2013) investigated the reproducibility of determining the cross-sectional area of lower-limb muscles from MRI images by measuring the between-scan differences in the identified ROIs, by quantifying the percent of the identified ROIs from each scan that overlapped. In two positions, the ROI overlap was 78% and 71% for the thigh, and 72% and 78% for the calf, suggesting repeated muscle CSA identification over two scans was approximately 75% similar. Scan-rescan ICC values for the thigh muscles ranged from 0.930 to 0.996, and for the calf 0.904-0.978. Furthermore, in a recent study investigating inter-rater reliability of gluteus muscle CSA quantification, Kivle et al. (2018) reported ICC values of 0.98 (gluteus medius) and 0.95 (gluteus minimus). Notably, manually identifying the cervical spine musculature is likely more difficult than the lower limb, however the $ICC_{2,1}$ values reported in the current study are only slightly lower to those reported for inter-rater reliability and similar for scan-rescan reliability of lower-limb musculature (Fischmann et al., 2013; Kivle et al., 2018). Giedd et al. (1995) investigated the reliability of quantitative volumetric measures of the cerebrum, caudate nucleus, and lateral ventricles using a scan-rescan procedure in asymptomatic individuals and found ICCs ranging from 0.88 to 0.99.

The present study investigated the scan-rescan and rater-reliability of cervical muscle volume quantification. When muscle volume quantifications are performed over time periods of varying duration (when compositional changes in muscle would not be expected, e.g. minutes to hours), good scan-rescan reliability is needed. If a change is found, it should be considered whether this is a variation due to the scanning/analysis procedure or an actual physiological change. Based on the data from the current study, inter-scan standard errors of measurement and variations between repeated measures of MRI data for research or clinical use would be expected to be minimal if the scans are performed soon after one another.

In recent investigations of cervical muscle volume, several studies (Amabile et al., 2017; Elliott et al., 2008a,b; Elliott et al., 2012; Matsumoto et al., 2012; Ulbrich et al., 2012) have examined the cross-sectional muscle area (CSA) in mm^2. For this study, however, three-dimensional (3D) muscle volumes in mm^3 were calculated from measurements made from the two-dimensional slices. Software used to quantify volume allowed the researchers to view the muscles in three dimensions (3D), providing a cross-check for possible errors in muscle border identification, which could be corrected in either the 2D or 3D view. For this reason, volume data exported in mm^3 was used for analysis rather than CSA. The CSA values were approximately one third of the volume values so it is unlikely using the CSA values would have changed the reliability outcomes reported.

4.1. Relevance of muscle quality investigations

Research on muscle quality, including investigations on morphology, fatty infiltration, and muscle volume, has found several clinically relevant differences between individuals with neck pain and asymptomatic controls. In both individuals with whiplash associated disorder and idiopathic neck pain, changes in volumetric measurements of certain muscles (e.g., Mult, Splenius, Semispinalis) were identified (Elliott et al., 2008a,b), while it was suggested that levels of fatty infiltrates in these muscles were heightened in those exposed to trauma (Elliott et al., 2014). Furthermore, the proposal that values of muscle fatty infiltrates are associated with the severity of traumatic neck disorders (specifically, whiplash injury from a motor vehicle collision) has been reported in several studies (Abbott et al., 2015; Elliott et al., 2011; Karlsson et al., 2016). Emerging evidence suggests such changes are also present in degenerative cervical myelopathy and spinal cord injury, both of which have consistent radiological findings of pathology (Fortin et al., 2017, 2018; Smith et al., 2015).

Table 4
Scan-rescan reliability of left and right cervical muscle volume quantification of asymptomatic participants (n = 5) using summed data from an individual's MRI images (slices) for each of five muscles.

Muscle	# observations	Scan 1 mean (SD)	Scan 2 mean (SD)	ICC	95% CI	SEM	Lin's	95% CI
Levator scapulae	10	27205.5 (11388.7)	26167.4 (10178.9)	.973	.899, .993	1771.95	.970	.903, .991
Multifidus[a]	10	26180.7 (3707.0)	24305.5 (3254.6)	.984	.938, .996	440.29	.982	.933, .995
Semispinalis	10	23999.8 (4657.0)	23965.0 (3840.3)	.948	.805, .997	968.83	.943	.831, .981
Splenius capitis[b]	10	21855.9 (3785.6)	22139.7 (3398.6)	.935	.771, .983	915.82	.929	.756, .980
Sternocleidomastoid	10	27985.7 (6805.8)	28914.2 (6392.5)	.979	.832,.996	956.31	.977	.924, .993
Total	50	25019.59 (6826.91)	25098.35 (6220.42)	.970	.948, .983	1129.93	.970	.949, .982

SEM: standard error of measurement; Lin's: Lin's concordance coeficient.
[a] Includes semispinalis cervicis.
[b] Includes splenius cervicis.

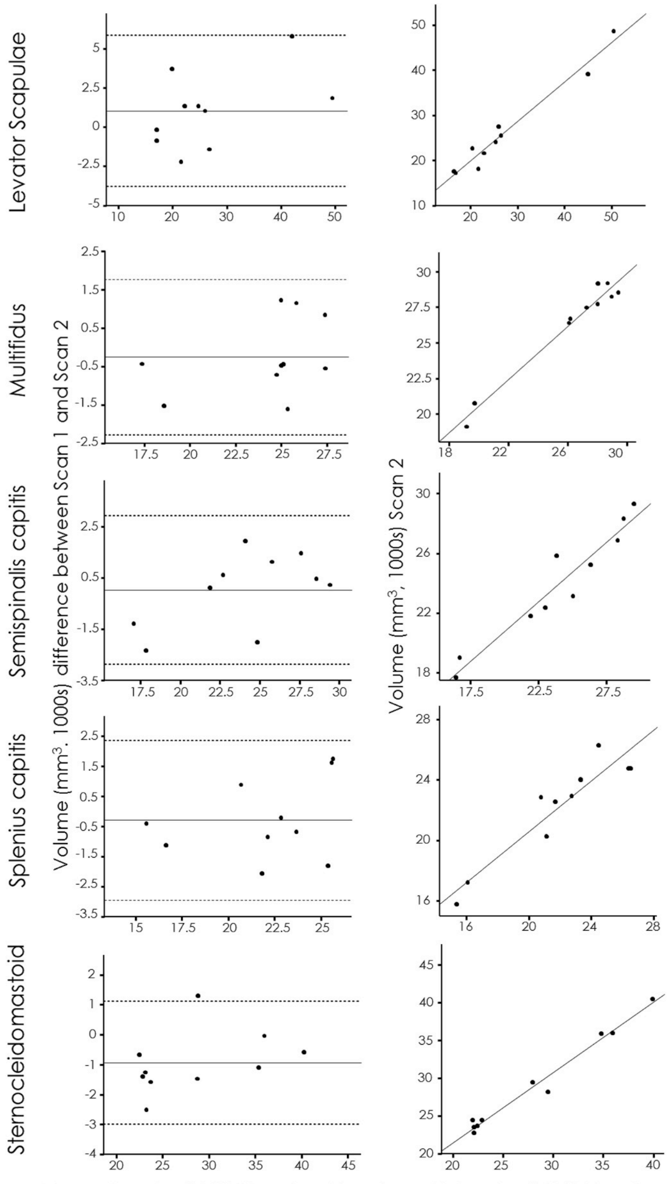

Fig. 4. Bland-Altman plots (left) and Lin's concordance (right) for inter-rater reliability of muscle volumes for five cervical muscles from the intervertebral discs of C2-3 through C6-7 quantified from two separate MRI scans by a single rater. All axes present muscle volumes (mm^3) in 1000s. Multifidus includes semispinalis cervicis; splenius capitis includes splenius cervicis.

4.2. Limitations

In this investigation, quantifications of muscle volume were performed by manually identifying the individual cervical muscles of asymptomatic individuals. Others have reported there might be considerable differences between muscle volumes with and without fatty infiltrates and this may affect the reliability of manual identification in symptomatic individuals. Additionally, individuals with varying levels of traumatic and non-traumatic neck pain may behave differently throughout the scanning procedure, i.e. they may have difficulties getting into the scanner, and they may demonstrate more unwanted pain-dependent movement affecting image quality. These, and likely other yet to be measured factors, should be taken into consideration in the assessment of the scan-rescan reliability of individuals with neck pain.

Researchers in the current study used a standard computer mouse to manually identify muscle borders, which is less efficient and potentially less accurate compared to manual tracing using a high definition touch tablet (Taslakian et al., 2018). However, using a computer mouse has been shown to be comparable to the tablet when segmenting a different body region (Taslakian et al., 2018), and using a computer mouse improves generalisability to clinical practice where only standard computers are likely to be available. Future workspaces will benefit from various machine learning algorithms to accurately and efficiently measure muscle volumes; in a fraction of the time it takes using semi-automated and manual segmentation applications.

5. Conclusions

This study found inter-rater reliability of cervical muscle quantification using MRI ranged from fair to good (Semisp, Lev Scap, Mult) to excellent (Splenius and SCM). The scan-rescan reliability, assessed by a single rater, was excellent for all muscles, suggesting issues of reliability are primarily related to manually contouring the borders of the muscles on the images. These findings suggest that the proposed methods for MRI scanning for the purpose of cervical muscle volume quantification can be used in further research and may provide directions for clinical assessment of individuals with neck pain. Future work may benefit from higher resolution scans and machine learning algorithms that are currently underway.

Declaration of interest

None.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.msksp.2019.102056.

References

Abbott, R., Pedler, A.R., Sterling, M., Hides, J., Murphey, T., Hoggarth, M., Elliott, J., 2015. The geography of fatty infiltrates within the cervial multifidus and semispinalis cervicis in individuals with chronic whiplash-associated disorders. J. Orthop. Sport. Phys. Ther. 45 (4) 10.2519/.

Amabile, C., Moal, B., Chtara, O.A., Pillet, H., Raya, J.G., Iannessi, A., et al., 2017. Estimation of spinopelvic muscles' volumes in young asymptomatic subjects: a quantitative analysis. Surg. Radiol. Anat. 39 (4), 393–403. https://doi.org/10.1007/s00276-016-1742-6.

Anderson, S.E., Boesch, C., Zimmermann, H., Busato, A., Hodler, J., Bingisser, R., et al., 2012. Are there cervical spine findings at MR imaging that are specific to acute symptomatic whiplash injury? A prospective controlled study with four experienced blinded readers. Radiology 262 (2), 567–575.

Au, J., Perriman, D.M., Pickering, M.R., Buirski, G., Smith, P.N., Webb, A.L., 2016. Magnetic resonance imaging atlas of the cervical spine musculature. Clin. Anat. 1–44. https://doi.org/10.1002/ca.22731.

Brontfort, G., Evans, R., Anderson, A.V., Svendsen, K.H., Bracha, Y., Grimm, R.H., 2012. Spinal manipulation, medication, or home exercise with advice for acute and subacute neck pain. Ann. Intern. Med. 156, 1–10.

Elliott, J.M., Cornwall, J., Kennedy, E., Abbott, R., Crawford, R.J., 2018. Towards defining muscular regions of interest from axial magnetic resonance imaging with anatomical cross-reference: part II - cervical spine musculature. BMC Muscoskelet. Disord. 19 (1), 171. https://doi.org/10.1186/s12891-018-2074-y.

Elliott, J.M., Courtney, D.M., Rademaker, A., Pinto, D., Sterling, M.M., Parrish, T.B., 2015. The rapid and progressive degeneration of the cervical multifidus in whiplash: an MRI study of fatty infiltration. Spine (Phila Pa 1976) 40 (12), E694–E700. https://doi.org/10.1097/BRS.0000000000000891.

Elliott, J.M., Jull, G., Noteboom, J., Galloway, G., 2008a. MRI study of the cross-sectional area for the cervical extensor musculature in patients with persistent whiplash associated disorders (WAD). Man. Ther. 13 (3), 258–265. https://doi.org/10.1016/j.math.2007.01.012.

Elliott, J.M., Jull, G., Noteboom, J.T., Darnell, R., Galloway, G., Gibbon, W.W., 2006. Fatty infiltration in the cervical extensor muscles in persistent whiplash-associated disorders. Spine 31 (11), E847–E855.

Elliott, J.M., Pedler, A., Kenardy, J., Galloway, G., Jull, G., Sterling, M., 2011. The temporal development of fatty infiltrates in the neck muscles following whiplash injury: an association with pain and posttraumatic stress. PLoS One 6 (6), e21194. https://doi.org/10.1371/journal.pone.0021194.

Elliott, J.M., Pedler, A.R., Jull, G.A., Van Wyk, L., Galloway, G.G., O'Leary, S.P., 2014. Differential changes in muscle composition exist in traumatic and nontraumatic neck pain. Spine 39 (1), 39–47. https://doi.org/10.1097/brs.0000000000000033.

Elliott, J.M., Pedler, A.R., Theodoros, D., Jull, G.A., 2012. Magnetic resonance imaging changes in the size and shape of the oropharynx following acute whiplash injury. J. Orthop. Sport. Phys. Ther. 42 (11), 912–918. https://doi.org/10.2519/jospt.2012.4280.

Elliott, J.M., Sterling, M., Noteboom, J.T., Darnell, R., Galloway, G., Jull, G., 2008b. Fatty infiltrate in the cervical extensor muscles is not a feature of chronic, insidious-onset neck pain. Clin. Radiol. 63 (6), 681–687. https://doi.org/10.1016/j.crad.2007.11.011.

Fischmann, A., Morrow Jasper, M., Sinclair Christopher, D.J., Reilly Mary, M., Hanna Michael, G., Yousry, T., Thornton John, S., 2013. Improved anatomical reproducibility in quantitative lower-limb muscle MRI. J. Magn. Reson. Imaging 39 (4), 1033–1038. https://doi.org/10.1002/jmri.24220.

Fleiss, J.L., 1986. The Design and Analysis of Clinical Experiments. John Wiley & Sons, Inc, New York.

Fortin, M., Dobrescu, O., Courtemanche, M., Sparrey, C.J., Santaguida, C., Fehlings, M.G., Weber, M.H., 2017. Association between paraspinal muscle morphology, clinical symptoms, and functional status in patients with degenerative cervical myelopathy. Spine 42 (4), 232–239. https://doi.org/10.1097/brs.0000000000001704.

Fortin, M., Wilk, N., Dobrescu, O., Martel, P., Santaguida, C., Weber, M.H., 2018. Relationship between cervical muscle morphology evaluated by MRI, cervical muscle strength and functional outcomes in patients with degenerative cervical myelopathy. Musculoskelet Sci. Pract. 38, 1–7. https://doi.org/10.1016/j.msksp.2018.07.003.

Giedd, J.N., Kozuch, P., Kaysen, D., Vaituzis, A.C., Hamburger, S.D., Bartko, J.J., Rapoport, J.L., 1995. Reliability of cerebral measures in repeated examinations with magnetic resonance imaging. Psychiatry Res. 61 (2), 113–119. https://doi.org/10.1016/0925-4927(95)02593-M.

Haldeman, S., Carroll, L.J., Cassidy, J.D., Schubert, J., Nygren, Å., 2008. The bone and joint decade 2000–2010 task force on neck pain and its associated disorders. Eur. Spine J. 17 (S1), 5–7. https://doi.org/10.1007/s00586-008-0619-8.

Hoy, D.G., March, L., Woolf, A., Blyth, F., Brooks, P., Smith, E., et al., 2014. The global burden of neck pain: estimates from the global burden of disease 2010 study. Ann. Rheum. Dis. 73 (7), 1309–1315. https://doi.org/10.1136/annrheumdis-2013-204431.

Kaiser, J.A., Holland, B.A., 1998. Imaging of the cervical spine. Spine 23 (24), 2701–2712.

Karlsson, A., Leinhard, O.D., Aslund, U., West, J., Romu, T., Smedby, O., et al., Peolsson, A., 2016. An investigation of fat infiltration of the multifidus muscle in patients with severe neck symptoms associated with chronic whiplash-associated disorder. J. Orthop. Sport. Phys. Ther. 46 (10), 886–893. https://doi.org/10.2519/jospt.2016.6553.

Kivle, K., Lindland, E., Mjaaland, K.E., Pripp, A.H., Svenningsen, S., Nordsletten, L., 2018. The gluteal muscles in end-stage osteoarthritis of the hip: intra- and interobserver reliability and agreement of MRI assessments of muscle atrophy and fatty degeneration. Clin. Radiol. 73 (7). https://doi.org/10.1016/j.crad.2018.02.014. 675 e617-675 e624.

Matsumoto, M., Fujimura, Y., Suzuki, N., Nishi, Y., Nakamura, M., Yabe, Y., Shiga, H., 1998. MRI of cervical intervertebral discs in asymptomatic subjects. J. Bone Jt. Surg. 80-B (1), 19–24.

Matsumoto, M., Ichihara, D., Okada, E., Chiba, K., Toyama, Y., Fujiwara, H., et al., 2012. Cross-sectional area of the posterior extensor muscles of the cervical spine in whiplash injury patients versus healthy volunteers–10 year follow-up MR study. Injury 43 (6), 912–916. https://doi.org/10.1016/j.injury.2012.01.017.

Matsumoto, M., Okada, E., Ichihara, D., Chiba, K., Toyama, Y., Fujiwara, H., et al., 2010. Prospective ten-year follow-up study comparing patients with whiplash-associated disorders and asymptomatic subjects using magnetic resonance imaging. Spine 35 (18), 1684–1690.

Siivola, S.M., Levoska, S., Tervonen, O., Ilkko, E., Vanharanta, H., Keinanen-Kiukaanniemi, S., 2002. MRI changes of cervical spine in asymptomatic and symptomatic young adults. Eur. Spine J. 11 (4), 358–363. https://doi.org/10.1007/s00586-001-0370-x.

Smith, A.C., Parrish, T.B., Hoggarth, M.A., McPherson, J.G., Tysseling, V.M., Wasielewski, M., et al., 2015. Potential associations between chronic whiplash and incomplete spinal cord injury. Spinal Cord Ser. Cases 1. https://doi.org/10.1038/scsandc.2015.24.

Snodgrass, S., Croker, C., Yerrapothu, M., Shepherd, S., Stanwell, P., Holder, C., Oldmeadow, C., Elliott, J., 2019. Cervical muscle volume in individuals with idiopathic neck pain compared to asymptomatic controls: a cross-sectional magnetic resonance imaging study. Musculoskeletal Science and Practice in press.

Taslakian, B., Pires, A., Halpern, D., Babb, J.S., Axel, L., 2018. Stylus/tablet user input device for MRI heart wall segmentation: efficiency and ease of use. Eur. Radiol. https://doi.org/10.1007/s00330-018-5435-x.

Ulbrich, E.J., Aeberhard, R., Wetli, S., Busato, A., Boesch, C., Zimmermann, H., et al., 2012. Cervical muscle area measurements in whiplash patients: acute, 3, and 6 months of follow-up. J. Magn. Reson. Imaging 36 (6), 1413–1420. https://doi.org/10.1002/jmri.23769.

Valentin, S., Licka, T., Elliott, J., 2015. Age and side-related morphometric MRI evaluation of trunk muscles in people without back pain. Man. Ther. 20 (1), 90–95. https://doi.org/10.1016/j.math.2014.07.007.