Manuscript Details

Manuscript number YMATH_2019_508_R2

Title Intra-rater and Inter-rater Reliability of Rehabilitative Ultrasound Imaging of Cervical Multifidus Muscle in Healthy People: Imaging Capturing and Imaging Calculation

Article type Technical & measurement reports

Abstract

Background: Studies have analyzed muscle morphometry of cervical multifidus by using ultrasound imaging, but its reliability is not clearly determined. Objective: To investigate intra- and inter-rater reliability of imaging capturing (probe assessment/patient positioning) and imaging calculation (scan assessment) of cervical multifidus cross sectional area (CSA) by considering the assessor's experience in asymptomatic individuals. Design: Reliability study. Methods: The CSA of C4/C5 cervical multifidus was assessed in 16 asymptomatic subjects. Two examiners performed the imaging capture and also repeated the procedure (probe placement/patient positioning) twice with a 10-min period between each. Other two raters conducted imaging calculations of CSA. Intra-examiner imaging capturing reliability, each rater (experienced and novice) calculated multifidus CSA of both images obtained by each examiner. Inter-examiner imaging capturing reliability, each rater calculated the CSA obtained by each examiner at the first imaging attempt. For imaging calculation reliability, each rater calculated multifidus CSA of all images captured by both examiners. Intra-class correlation coefficients (ICC) and standard error of measurement (SEM) were calculated. Results: Intra- ($ICC_{3,1}$ 0.988-0.996, SEM 0.3%-0.7%) and inter- ($ICC_{3,2}$ 0.958-0.965, SEM 2.6%-3.2%) examiner reliability of imaging capturing was excellent. Intra- and inter-rater reliability of imaging calculation was also excellent for both raters (experienced/novice). No significant differences between experienced or novice examiners or testers were found. Conclusions: This study found that intra- and inter-examiner/rater reliability of imaging capturing (probe assessment/patient positioning) and imaging calculation (scan assessment) of the cervical multifidus CSA at C4/C5 level was excellent in asymptomatic subjects.

Keywords Rehabilitative ultrasound imaging; cervical multifidus; reliability.

Taxonomy Medicine, Medical Specialty

Corresponding Author César Fernández-de-las-Peñas

Corresponding Author's Institution Aalborg University

Order of Authors Juan Antonio Valera-Calero, Sandra Sánchez-Jorge, Javier Alvarez-González, Ricardo Ortega, Josh Cleland, César Fernández-de-las-Peñas, José Luis Arias-Buría

Suggested reviewers Maria Stokes, Emilio Puentedura, Carol A Courtney

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	1
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	1-2
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	3-4
	4	Study objectives and hypotheses	5
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	
Participants	**6**	Eligibility criteria	5
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	5
	8	Where and when potentially eligible participants were identified (setting, location and dates)	5
	9	Whether participants formed a consecutive, random or convenience series	5
Test methods	**10a**	Index test, in sufficient detail to allow replication	6-7
	10b	Reference standard, in sufficient detail to allow replication	6-7
	11	Rationale for choosing the reference standard (if alternatives exist)	6-7
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	6-7
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	6-7
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	6-7
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	6-7
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	8
	15	How indeterminate index test or reference standard results were handled	8
	16	How missing data on the index test and reference standard were handled	8
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	8
	18	Intended sample size and how it was determined	8
RESULTS			
Participants	**19**	Flow of participants, using a diagram	
	20	Baseline demographic and clinical characteristics of participants	8
	21a	Distribution of severity of disease in those with the target condition	8
	21b	Distribution of alternative diagnoses in those without the target condition	8
	22	Time interval and any clinical interventions between index test and reference standard	8
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	8-9
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	8-9
	25	Any adverse events from performing the index test or the reference standard	8-9
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	12
	27	Implications for practice, including the intended use and clinical role of the index test	9-11
OTHER INFORMATION			
	28	Registration number and name of registry	
	29	Where the full study protocol can be accessed	
	30	Sources of funding and other support; role of funders	

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on [http://www.equator-network.org/reporting-guidelines/stard.](http://www.equator-network.org/reporting-guidelines/stard)

Intra-rater and Inter-rater Reliability of Rehabilitative Ultrasound Imaging of Cervical Multifidus Muscle in Healthy People: Imaging Capturing and Imaging Calculation

Abstract

Background: Studies have analyzed muscle morphometry of cervical multifidus by using ultrasound imaging, but its reliability is not clearly determined. **Objective:** To investigate intra- and inter-rater reliability of imaging capturing (probe assessment/patient positioning) and imaging calculation (scan assessment) of cervical multifidus cross sectional area (CSA) by considering the assessor's experience in asymptomatic individuals. **Design:** Reliability study. **Methods:** The CSA of C4/C5 cervical multifidus was assessed in 16 asymptomatic subjects. Two examiners performed the imaging capture and also repeated the procedure (probe placement/patient positioning) twice with a 10-min period between each. Other two raters conducted imaging calculations of CSA. Intra-examiner imaging capturing reliability, each rater (experienced and novice) calculated multifidus CSA of both images obtained by each examiner. Inter-examiner imaging capturing reliability, each rater calculated the CSA obtained by each examiner at the first imaging attempt. For imaging calculation reliability, each rater calculated multifidus CSA of all images captured by both examiners. Intra-class correlation coefficients (ICC) and standard error of measurement (SEM) were calculated. **Results**: Intra- ($ICC_{3,1}$ 0.988-0.996, SEM 0.3%-0.7%) and inter- ($ICC_{3,2}$ 0.958-0.965, SEM 2.6%-3.2%) examiner reliability of imaging capturing was excellent. Intra- and inter-rater reliability of imaging calculation was also excellent for both raters (experienced/novice). No significant differences between experienced or novice examiners or testers were found.

Conclusions: This study found that intra- and inter-examiner/rater reliability of imaging capturing (probe assessment/patient positioning) and imaging calculation (scan assessment) of the cervical multifidus CSA at C4/C5 level was excellent in asymptomatic subjects.

Key words: Rehabilitative ultrasound imaging, cervical multifidus, reliability.

Intra-rater and Inter-rater Reliability of Rehabilitative Ultrasound Imaging of Cervical Multifidus Muscle in Healthy People: Imaging Capturing and Imaging Calculation

Introduction

Mechanical neck pain of insidious onset and whiplash associated disorders (WAD) represent a major health care problem. While the general prognosis of neck pain is positive, up to 50% of individuals continue reporting symptoms after 1-year (Kamper et al, 2008). There is no consensus regarding the potential pathophysiology of neck disorders. One mechanism may be related to the fact that neck pain induces changes in cervical muscle performance. Several studies have reported that people with WAD show increased fatty infiltrate and changes in cross sectional area (CSA) in the posterior cervical muscles, specifically the multifidi (Abbott et al, 2015; Elliot et al, 2014; Snodgrass et al, 2019a). However, most studies have used magnetic resonance imaging (MRI) to assess muscle morphology (Owers et al, 2018), which is not readily available in clinical practice.

A more pragmatic method of measuring muscle morphology is ultrasonography; but it its reliability must first be demonstrated (Whittaker et al, 2017). Some studies have investigated the reliability of rehabilitative ultrasound imaging (RUSI) for examination of the posterior neck muscles (Javanshir et al, 2010). Kristjansson (2004) analyzed the reliability of assessing CSA of the C4 cervical multifidus and reported appropriate intra- and inter-tester agreement in 10 asymptomatic subjects, and acceptable intra-, but questionable inter-tester, agreement in 10 symptomatic subjects. Fernández-de-las-Peñas et al (2008) reported excellent between-scan and good between-day assessment in individuals with mechanical neck pain. Lee et al (2007) found that ultrasound imaging was as reliable and valid as MRI for assessing cervical multifidus thickness after isometric contractions.

An important step before neck muscle composition can routinely be used in research or clinical practice is establishing reliability of imaging calculation (scan assessment) but also reliability of imaging capturing (patient positioning). This is particularly relevant since ultrasonography is operator-dependent and the measurement protocol could influence the imaging calculation. In fact, previous studies had mostly investigated the reliability of imaging calculation (scan assessment) but not the reliability of imaging capturing (probe assessment /patient repositioning). Further, reliability according to the experience of the assessor has not been properly investigated. Our aim was to determine intra- and inter-examiner/rater reliability of imaging calculation (scan/image assessment) and imaging capturing (probe assessment/patient positioning) of cervical multifidus CSA considering assessor experience

Methods

Participants

Asymptomatic volunteers without neck pain symptoms were recruited via local announcements between December 2018 and June 2019. To be eligible to participate, they had to be between 18 and 45 years old and with no history of neck pain the previous year. Exclusion criteria included history of whiplash injury; any pharmacological treatment affecting muscle tone, e.g., muscle relaxants, analgesics; prior history of cervical surgery; cervical radiculopathy or myelopathy; presence of degenerative changes; and any medical condition such as tumor or fracture. The study was approved by the Institutional Review Ethical Committee of XX. All subjects signed the written informed consent prior to their inclusion.

Procedure Assessment - Imaging Capturing

All images were acquired with a Toshiba Xario® 100 ultrasound equipment with a PLU-1005 BT (7-15MHz) linear probe. Gain, frequency, depth, or focus were pragmatically adapted by the examiner for each exam. Participants were placed in a prone position with their arms in 90º abduction and the elbows flexed to 90º. The head/neck were stabilized using the plinth's facial hole. A passive cranio-cervical flexion movement was performed by the examiner to achieve a neutral position of the neck/head.

It has been estimated that measurements of C4 would exhibit less error (Lee et al, 2007); therefore, we assessed C4/C5 multifidus. To identify the cervical multifidus level, the C2 spinous process was identified by palpation. At that point, the US probe was moved caudally two segments until the posterior arch of C4 vertebra was visually identified. Then, the transducer was moved lateral over the articular pillar (**Fig. 1A**). The image (scan) was captured when the most superficial point of the spinous tubercle cortical surface and the most superficial point of C4/C5 joint were visualized simultaneously (**Fig. 1B**).

Measurement Assessment - Imaging Calculation

Once the US image was captured, it was transferred to offline Oxyiri® Software for calculating the CSA of the cervical multifidus by using on-screen calipers traced around the following contours: 1, inferior limit: internal echogenic fascia between cervical multifidus and rotator muscle (deep to cervical multifidus); 2, superior limit: echogenic fascia between cervical multifidus and semispinalis; 3, medial limit: echogenic spinous process (**Fig. 2**).

Examiners

A total of 4 examiners participated in the study. For imaging capturing, one experienced (10 years of practice) and one novice (1 year of practice) examiner performed the patient positioning/procedure placement and captured two images of the posterior neck muscles as described. Each examiner repeated the assessment, twice with a 10-min period between. Participants were repositioned on each assessment. On each assessment, an image of the posterior neck muscles was obtained (Fig. 1B).

Another two examiners, identified in our study as raters, again one experienced and the second novice (same experience as examiners) participated in the imaging calculation of all images. Every image was coded to blind raters using alphanumerical codes. The order of assessment and raters was numerically randomized between participants.

Reliability Calculations

We assessed the reliability of both imaging capturing and imaging calculation by considering the examiner/rater experience. For imaging capturing intra-examiner reliability, each rater (experienced/novice) calculated CSA of both images obtained by each examiner (experienced/novice) at both assessments. For imaging capturing inter-examiner reliability, each rater (experienced/novice) calculated the CSA of the image captured by each examiner (experienced/novice) at the first positioning assessment.

For imaging calculation intra-rater reliability, each rater (experienced/novice) determined the CSA of all images obtained by both examiners twice, one-week apart. For imaging calculation inter-rater reliability, each rater (experienced/novice) calculated the CSA of all images obtained by both examiners once.

Statistical Analysis

Statistical analysis was performed using the SPSS package, Version 21 software for Mac OS. Normal distribution of the data was verified by using the Shapiro-Wilk test. Intra- and inter- examiner/rater reliability was estimated using 2-way mixed-model, consistency-type intra class correlation coefficients (ICC). Reliability was classified as fair ($ICC<0.50$), moderate ($0.50<ICC<0.75$), good ($0.75<ICC<0.90$) or excellent ($ICC>0.90$) (Koo and Li 2016). Standard error of measurement (SEM) was calculated using the resulting ICC values and standard deviation (SD): *SEM (%) = (SDx√1-ICC) x 100* to assess measurement precision. All tests were two-tailed with p-values <0.05 considered significant.

Results

From a total of 20 subjects responding to the announcement, 4 were excluded due to previous whiplash injury (n=2) and history of neck pain the previous year (n=2). Sixteen asymptomatic subjects (50% male) were finally included (total 64 images, n=4 per subject). **Table 1** provides demographic features. Male exhibited higher anthropometric outcomes and larger CSA than female ($P<0.001$). A positive correlation between multifidus CSA and weight (r:0.473, P=0.006), height (r:0.385; P=0.03), and BMI (r:0.481, P=0.005) was found.

Table 2 shows reliability data of imaging capturing. In general, imaging capturing intra-examiner reliability ($ICC_{3,1}$) was excellent ranging from 0.988 to 0.996 with a SEM from 0.3% to 0.7%. No difference between experienced/novice examiner/testers was found. Inter-examiner reliability was also excellent for both experienced ($ICC_{3,2}$ 0.965) and novice ($ICC_{3,2}$ 0.958) raters, with SEM of 2.6% and 3.2%, respectively.

Reliability data of imaging calculation of both raters is shown on **Table 3**. Again, intra- and inter-rater reliability was excellent for both raters but with smaller SEM for intra-rater reliability.

Discussion

This is the first study assessing reliability of both imaging capturing (positioning) and imaging calculation (assessment) of the neck muscles considering the experience of the assessor. In general, intra- and inter- rater reliability of imaging capturing and calculation of cervical multifidus CSA was excellent when applied by an experienced or novice assessor in asymptomatic individuals. Our findings are similar to those previously found in patients with mechanical neck pain (Fernández-de-las-Peñas et al, 2008) and slightly better than those previously reported by Kristjansson (2004). In addition, reliability data of ultrasound imaging assessment was slightly superior to reliability data obtained for MRI. Snodgrass et al (2019b) reported good to excellent intra- (ICC 0.78-0.96), but fair to good inter- (ICC 0.44-0.88) rater reliability of cervical multifidus assessment using MRI in 5 asymptomatic individuals. Interestingly, no differences between an experienced and novice examiner/rater were observed. This is a relevant topic since ultrasound imaging assessment is operator-dependent. Two points are relevant for ultrasound assessment, probe angulation and pressure. Whittaker et al (2009) showed that probe angulation of less than approx. 10 degrees, as it is commonly done in clinical practice, do not distort measurements of tissue thickness. No data about probe pressure is available. It is possible that strict positioning and measurement protocols followed in our study could explain high reliability values.

This study can help for developing specific protocols for image capturing and calculation for both research and clinical practice. For instance, morphometry assessment, i.e., CSA, is based on anatomical surrounding muscle contours. Previous studies recognized that lack of proper visualization of the fascial layers dividing the cervical multifidus from surrounding muscles (Fernández-de-las-Peñas et al, 2008; Kristjansson, 2004; Lee et al, 2007). One potential reason for this lack of clarity of muscular fascial layers could be the presence of fatty infiltration, a sign potentially associated with WAD (Owers et al, 2018). In the current study, we assessed cervical multifidus, without including the rotators, as it was conducted in previous studies (Fernández-de-las-Peñas et al, 2008; Kristjansson, 2004; Lee et al, 2007). The main reason was that the cervical multifidus attaches to the posterior aspect of the facet capsules (Anderson et al, 2005) and play a relevant role in proprioception. In such a scenario, fascial layers as border contours for determining muscle morphology may have highly relevance. Since the current study included asymptomatic individuals, fascial layers surrounding the multifidus were properly identified (Fig. 2). Rankin et al (2005) were not able to consistently identify fascial divisions between multifidus/semispinalis and between multifidus/rotators; therefore, they provided normative data of the whole muscle group and not from individual muscles. It is also possible that technical improvements of ultrasound imaging equipment, offering much better-definition and quality images, can also influence proper visualization of fascial layers between muscles.

We found positive correlations between CSA and height, weight, and BMI supporting the fact that muscle morphometry is associated to anthropometric features (Rezasoltani et al 1998). This could explain the CSA variability found in previous studies. In fact, the only study investigating normative data of posterior cervical muscles reported that weight, rather than gender, was a relevant cofounder factor for CSA (Rankin et al, 2005). Previous studies investigating differences in CSA between neck pain patients and healthy subjects did not control for the anthropometric variables (Fernández-de-las-Peñas et al, 2008; Kristjansson, 2004; Lee et al, 2007). De Pauw et al (2016) concluded that although there is evidence of reduced CSA in the cervical multifidus in people with mechanical neck pain, more studies are needed due to the inconsistency on the results. Future studies investigating differences between pain populations and healthy subjects should include these considerations.

Finally, this study has some limitations. First, we included asymptomatic subjects. We do not know if similar results would be observed in patients with mechanical neck pain or WAD. Second, our sample was small. Therefore, our results should not be considered as potential normative data, future studies are needed to determine muscle morphology data of posterior neck muscles separately, e.g., semispinalis, splenius, multifidus, and rotators. In addition, reliability of imaging capturing (probe assessment/patient positioning) was tested within 10 minutes on the same day which may not be clinically relevant. Imaging capturing reliability should ideally be assessed on different days in future studies. Finally, it is also important to consider that fatty infiltration quantification is not possible with ultrasound imaging and is currently possible with MRI, since MRI calculates the amount of fat by differentiating fat and soft-aqueous tissue signal intensities (Elliott et al, 2013). Therefore, RUSI should be only used for determining muscle morphology, e.g., CSA, size, or muscle function, but not intramuscular quality, e.g., fat, or fibers.

Conclusions

We found that ultrasound assessment of cervical multifidus at C4/C5 level is highly reliable for evaluating CSA in asymptomatic people since imaging capturing and imaging calculation exhibited excellent intra- and inter- examiner reliability. Reliability was similar independently of the assessor experience. This paper proposes technical considerations for future studies assessing muscle morphometry in neck pain populations.

Legend of Figures

Figure 1: (A) Ultrasound probe placement over the cervical multifidus at C4 level; (B) Ultrasound image showing the superficial posterior neck muscles.

Figure 2: Cross sectional area (CSA) assessment of the cervical multifidus. Borders were marked as follows: 1, inferior limit: internal echogenic fascia between cervical multifidus and rotator muscle; 2, superior limit: echogenic fascia between cervical multifidus and semispinalis; 3, medial limit: echogenic spinous process

References

- Abbott R, Pedler A, Sterling M, Hides J, Murphey T, Hoggarth M, Elliott J. The geography of fatty infiltrates within the cervical multifidus and semispinalis cervicis in individuals with chronic whiplash-associated disorders. J Orthop Sports Phys Ther 2015; 45: 281-8.
- Anderson JS, Hsu AW, Vasavada AN. Morphology, architecture, and biomechanics of human cervical multifidus. Spine 2005; 30: E86-91.
- De Pauw R, Coppieters I, Kregel J, De Meulemeester K, Danneels L, Cagnie B. Does muscle morphology change in chronic neck pain patients? A systematic review. Man Ther 2016; 22: 42-9.
- Elliott JM, Walton DM, Rademaker A, Parrish TB. Quantification of cervical spine muscle fat: a comparison between T1-weighted and multi-echo gradient echo imaging using a variable projection algorithm (VARPRO). BMC Med Imaging 2013; 13: 30.
- Elliott JM, Pedler AR, Jull GA, Van Wyk L, Galloway GG, O□Leary SP. Differential changes in muscle composition exist in traumatic and nontraumatic neck pain. Spine 2014; 39: 39-47.
- Fernández-de-las-Peñas C, Albert-Sanchís JC, Buil M, Benitez JC, Alburquerque-Sendín F. Cross-sectional area of cervical multifidus in females with chronic bilateral neck pain compared to controls. J Orthop Sports Phys Ther 2008; 38: 175-80.

- Kamper SJ, Rebbeck TJ, Maher CG, McAuley JH, Sterling M. Course and prognostic factors of whiplash: a systematic review and meta-analysis. Pain 2008; 138: 617-629.
- Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016; 15: 155-63.
- Javanshir K, Amiri M, Mohseni-Bandpei MA, Rezasoltani A, Fernández-de-las-Peñas C. Ultrasonography of the cervical muscles: a critical review of the literature. J Manipul Physiol Ther 2010; 33: 630-7.
- Kristjansson E. Reliability of ultrasonography for the cervical multifidus muscle in asymptomatic and symptomatic subjects. Man Ther 2004; 9: 83-8.
- Lee JP, Tseng WY, Shau YW, Wang CL, Wang HK, Wang SF. Measurement of segmental cervical multifidus contraction by ultrasonography in asymptomatic adults. Man Ther 2007; 12: 286-294.
- Lin YJ, Chai HM, Wang SF. Reliability of thickness measurements of the dorsal muscles of the upper cervical spine: an ultrasonographic study. J Orthop Sports Phys Ther 2009; 39: 850-7.
- Owers DS, Perriman DM, Smith PN, Neeman T, Webb AL. Evidence for cervical muscle morphometric changes on magnetic resonance images after whiplash: A systematic review and meta-analysis. Injury 2018; 49: 165-176
- Rankin G, Stokes M, Newham D. Size and shape of the posterior neck muscles measured by ultrasound imaging: normal values in males and females of different ages. Man Ther 2005; 10: 108-115.

- Rezasoltani A, Kallinen M, Mälkiä E, Vihko V. Neck semispinalis capitis muscle size in sitting and prone positions measured by real time ultrasonography. Clin Rehabil 1998; 12: 36-44.
- Snodgrass SJ, Croker C, Yerrapothu M, Shepherd S, Stanwell P, Holder C, Oldmeadow C, Elliott J. Cervical muscle volume in individuals with idiopathic neck pain compared to asymptomatic control: A cross-sectional magnetic resonance imaging study. Musculosket Sci Pract 2019a Aug 13:102050.
- Snodgrass SJ, de Zoete RMJ, Croker C, Yerrapothu M, Elliott JM. Reliability of cervical muscle volume quantification using magnetic resonance imaging. Musculosket Sci Pract 2019b; 44: 102056.
- Whittaker JL, Teyhen DS, Elliott JM, Cook K, Langevin HM, Dahl HH, Stokes M. Rehabilitative ultrasound imaging: understanding the technology and its applications. J Orthop Sports Phys Ther 2007; 37: 434-49.
- Whittaker JL, Warner MB, Stokes MJ. Induced transducer orientation during ultrasound imaging: effects on abdominal muscle thickness and bladder position. Ultrasound Med Biol 2009; 35: 1803-1811

Intra-rater and Inter-rater Reliability of Rehabilitative Ultrasound Imaging of Cervical Multifidus Muscle in Healthy People: Imaging Capturing and Calculation Reliability Study

Highlights

1. Intra- and inter- examiner reliability of cervical multifidus imaging capturing and imaging calculation was excellent.
2. No significant differences between experienced and novice examiners and testers were found.
3. These values were obtained in asymptomatic individuals.

Intra-rater and Inter-rater Reliability of Rehabilitative Ultrasound Imaging of Cervical Multifidus Muscle in Healthy People: Imaging Capturing and Imaging Calculation

Background: Studies have analyzed muscle morphometry of cervical multifidus by using ultrasound imaging, but its reliability is not clearly determined. **Objective:** To investigate intra- and inter-rater reliability of imaging capturing (probe assessment/patient positioning) and imaging calculation (scan assessment) of cervical multifidus cross sectional area (CSA) by considering the assessor's experience in asymptomatic individuals. **Design:** Reliability study. **Methods:** The CSA of C4/C5 cervical multifidus was assessed in 16 asymptomatic subjects. Two examiners performed the imaging capture and also repeated the procedure (probe placement/patient positioning) twice with a 10-min period between each. Other two raters conducted imaging calculations of CSA. Intra-examiner imaging capturing reliability, each rater (experienced and novice) calculated multifidus CSA of both images obtained by each examiner. Inter-examiner imaging capturing reliability, each rater calculated the CSA obtained by each examiner at the first imaging attempt. For imaging calculation reliability, each rater calculated multifidus CSA of all images captured by both examiners. Intra-class correlation coefficients (ICC) and standard error of measurement (SEM) were calculated. **Results**: Intra- ($ICC_{3,1}$ 0.988-0.996, SEM 0.3%-0.7%) and inter- ($ICC_{3,2}$ 0.958-0.965, SEM 2.6%-3.2%) examiner reliability of imaging capturing was excellent. Intra- and inter-rater reliability of imaging calculation was also excellent for both raters (experienced/novice). No differences between experienced or novice examiners or testers were found. **Conclusions:** This study found that intra- and inter-examiner/rater reliability of imaging capturing (probe assessment/ patient positioning) and imaging calculation (scan assessment) of the cervical multifidus CSA at C4/C5 level was excellent in asymptomatic subjects.

Key words: Rehabilitative ultrasound imaging, cervical multifidus, reliability.

Title Page

Title

Intra-rater and Inter-rater Reliability of Rehabilitative Ultrasound Imaging of Cervical Multifidus Muscle in Healthy People: Imaging Capturing and Imaging Calculation

Authors

Juan A. Valera-Calero[1] PT, PhD candidate; Sandra Sánchez-Jorge[2] PT, PhD; Javier Álvarez-González[3] PT, MSc; Ricardo Ortega-Santiago[4,5] PT, PhD; Joshua A. Cleland[6,7,8] PT, MSc, PhD; César Fernández-de-las-Peñas[4,5] PT, PhD, DMSc; José L. Arias-Buría[4,5] PT, MSc, PhD.

Affiliations

(1) Escuela Internacional de Doctorado, Universidad Rey Juan Carlos, Alcorcón, Spain

(2) Professional Practice, Madrid, Spain.

(3) Servicio de Radiodiagnóstico. Hospital General Universitario Gregorio Marañón, Madrid, Spain.

(4) Department of Physical Therapy, Occupational Therapy, Rehabilitation and Physical Medicine, Universidad Rey Juan Carlos, Alcorcón, Spain

(5) Cátedra Institucional en Docencia, Clínica e Investigación en Fisioterapia: Terapia Manual, Punción Seca y Ejercicio Terapéutico, Universidad Rey Juan Carlos, Alcorcón, Madrid, Spain.

(6) Department of Physical Therapy, Franklin Pierce University, Manchester, New Hampshire.

(7) Rehabilitation Services, Concord Hospital, Concord, New Hampshire;

(8) Manual Therapy Fellowship Program, Regis University, Denver, Colorado.

Address for reprint requests / corresponding author.

César Fernández de las Peñas
Facultad de Ciencias de la Salud
Universidad Rey Juan Carlos
Avenida de Atenas s/n
28922 Alcorcón, Madrid, SPAIN
E-mail address: cesar.fernandez@urjc.es

Telephone number: + 34 91 488 88 84

Fax number: + 34 91 488 89 57

Intra-rater and Inter-rater Reliability of Rehabilitative Ultrasound Imaging of Cervical Multifidus Muscle in Healthy People: Imaging Capturing and Imaging Calculation

Abstract

Background: Studies have analyzed muscle morphometry of cervical multifidus by using ultrasound imaging, but its reliability is not clearly determined. **Objective:** To investigate intra- and inter-rater reliability of imaging capturing (probe assessment/patient positioning) and imaging calculation (scan assessment) of cervical multifidus cross sectional area (CSA) by considering the assessor's experience in asymptomatic individuals. **Design:** Reliability study. **Methods:** The CSA of C4/C5 cervical multifidus was assessed in 16 asymptomatic subjects. Two examiners performed the imaging capture and also repeated the procedure (probe placement/patient positioning) twice with a 10-min period between each. Other two raters conducted imaging calculations of CSA. Intra-examiner imaging capturing reliability, each rater (experienced and novice) calculated multifidus CSA of both images obtained by each examiner. Inter-examiner imaging capturing reliability, each rater calculated the CSA obtained by each examiner at the first imaging attempt. For imaging calculation reliability, each rater calculated multifidus CSA of all images captured by both examiners. Intra-class correlation coefficients (ICC) and standard error of measurement (SEM) were calculated. **Results**: Intra- ($ICC_{3,1}$ 0.988-0.996, SEM 0.3%-0.7%) and inter- ($ICC_{3,2}$ 0.958-0.965, SEM 2.6%-3.2%) examiner reliability of imaging capturing was excellent. Intra- and inter-rater reliability of imaging calculation was also excellent for both raters (experienced/novice). No significant differences between experienced or novice examiners or testers were found.

Conclusions: This study found that intra- and inter-examiner/rater reliability of imaging capturing (probe assessment/patient positioning) and imaging calculation (scan assessment) of the cervical multifidus CSA at C4/C5 level was excellent in asymptomatic subjects.

Key words: Rehabilitative ultrasound imaging, cervical multifidus, reliability.

Intra-rater and Inter-rater Reliability of Rehabilitative Ultrasound Imaging of Cervical Multifidus Muscle in Healthy People: Imaging Capturing and Imaging Calculation

Introduction

Mechanical neck pain of insidious onset and whiplash associated disorders (WAD) represent a major health care problem. While the general prognosis of neck pain is positive, up to 50% of individuals continue reporting symptoms after 1-year (Kamper et al, 2008). There is no consensus regarding the potential pathophysiology of neck disorders. One mechanism may be related to the fact that neck pain induces changes in cervical muscle performance. Several studies have reported that people with WAD show increased fatty infiltrate and changes in cross sectional area (CSA) in the posterior cervical muscles, specifically the multifidi (Abbott et al, 2015; Elliot et al, 2014; Snodgrass et al, 2019a). However, most studies have used magnetic resonance imaging (MRI) to assess muscle morphology (Owers et al, 2018), which is not readily available in clinical practice.

A more pragmatic method of measuring muscle morphology is ultrasonography; but it its reliability must first be demonstrated (Whittaker et al, 2017). Some studies have investigated the reliability of rehabilitative ultrasound imaging (RUSI) for examination of the posterior neck muscles (Javanshir et al, 2010). Kristjansson (2004) analyzed the reliability of assessing CSA of the C4 cervical multifidus and reported appropriate intra- and inter-tester agreement in 10 asymptomatic subjects, and acceptable intra-, but questionable inter-tester, agreement in 10 symptomatic subjects. Fernández-de-las-Peñas et al (2008) reported excellent between-scan and good between-day assessment in individuals with mechanical neck pain. Lee et al (2007) found that ultrasound imaging was as reliable and valid as MRI for assessing cervical multifidus thickness after isometric contractions.

An important step before neck muscle composition can routinely be used in research or clinical practice is establishing reliability of imaging calculation (scan assessment) but also reliability of imaging capturing (patient positioning). This is particularly relevant since ultrasonography is operator-dependent and the measurement protocol could influence the imaging calculation. In fact, previous studies had mostly investigated the reliability of imaging calculation (scan assessment) but not the reliability of imaging capturing (probe assessment /patient repositioning). Further, reliability according to the experience of the assessor has not been properly investigated. Our aim was to determine intra- and inter-examiner/rater reliability of imaging calculation (scan/image assessment) and imaging capturing (probe assessment/patient positioning) of cervical multifidus CSA considering assessor experience

Methods

Participants

Asymptomatic volunteers without neck pain symptoms were recruited via local announcements between December 2018 and June 2019. To be eligible to participate, they had to be between 18 and 45 years old and with no history of neck pain the previous year. Exclusion criteria included history of whiplash injury; any pharmacological treatment affecting muscle tone, e.g., muscle relaxants, analgesics; prior history of cervical surgery; cervical radiculopathy or myelopathy; presence of degenerative changes; and any medical condition such as tumor or fracture. The study was approved by the Institutional Review Ethical Committee of XX. All subjects signed the written informed consent prior to their inclusion.

Procedure Assessment - Imaging Capturing

All images were acquired with a Toshiba Xario® 100 ultrasound equipment with a PLU-1005 BT (7-15MHz) linear probe. Gain, frequency, depth, or focus were pragmatically adapted by the examiner for each exam. Participants were placed in a prone position with their arms in 90º abduction and the elbows flexed to 90º. The head/neck were stabilized using the plinth's facial hole. A passive cranio-cervical flexion movement was performed by the examiner to achieve a neutral position of the neck/head.

It has been estimated that measurements of C4 would exhibit less error (Lee et al, 2007); therefore, we assessed C4/C5 multifidus. To identify the cervical multifidus level, the C2 spinous process was identified by palpation. At that point, the US probe was moved caudally two segments until the posterior arch of C4 vertebra was visually identified. Then, the transducer was moved lateral over the articular pillar (**Fig. 1A**). The image (scan) was captured when the most superficial point of the spinous tubercle cortical surface and the most superficial point of C4/C5 joint were visualized simultaneously (**Fig. 1B**).

Measurement Assessment - Imaging Calculation

Once the US image was captured, it was transferred to offline Oxyiri® Software for calculating the CSA of the cervical multifidus by using on-screen calipers traced around the following contours: 1, inferior limit: internal echogenic fascia between cervical multifidus and rotator muscle (deep to cervical multifidus); 2, superior limit: echogenic fascia between cervical multifidus and semispinalis; 3, medial limit: echogenic spinous process (**Fig. 2**).

Examiners

A total of 4 examiners participated in the study. For imaging capturing, one experienced (10 years of practice) and one novice (1 year of practice) examiner performed the patient positioning/procedure placement and captured two images of the posterior neck muscles as described. Each examiner repeated the assessment, twice with a 10-min period between. Participants were repositioned on each assessment. On each assessment, an image of the posterior neck muscles was obtained (Fig. 1B).

Another two examiners, identified in our study as raters, again one experienced and the second novice (same experience as examiners) participated in the imaging calculation of all images. Every image was coded to blind raters using alphanumerical codes. The order of assessment and raters was numerically randomized between participants.

Reliability Calculations

We assessed the reliability of both imaging capturing and imaging calculation by considering the examiner/rater experience. For imaging capturing intra-examiner reliability, each rater (experienced/novice) calculated CSA of both images obtained by each examiner (experienced/novice) at both assessments. For imaging capturing inter-examiner reliability, each rater (experienced/novice) calculated the CSA of the image captured by each examiner (experienced/novice) at the first positioning assessment.

For imaging calculation intra-rater reliability, each rater (experienced/novice) determined the CSA of all images obtained by both examiners twice, one-week apart. For imaging calculation inter-rater reliability, each rater (experienced/novice) calculated the CSA of all images obtained by both examiners once.

Statistical Analysis

Statistical analysis was performed using the SPSS package, Version 21 software for Mac OS. Normal distribution of the data was verified by using the Shapiro-Wilk test. Intra- and inter- examiner/rater reliability was estimated using 2-way mixed-model, consistency-type intra class correlation coefficients (ICC). Reliability was classified as fair ($ICC<0.50$), moderate ($0.50<ICC<0.75$), good ($0.75<ICC<0.90$) or excellent ($ICC>0.90$) (Koo and Li 2016). Standard error of measurement (SEM) was calculated using the resulting ICC values and standard deviation (SD): $SEM\ (\%) = (SD x \sqrt{1-ICC})\ x\ 100$ to assess measurement precision. All tests were two-tailed with p-values <0.05 considered significant.

Results

From a total of 20 subjects responding to the announcement, 4 were excluded due to previous whiplash injury (n=2) and history of neck pain the previous year (n=2). Sixteen asymptomatic subjects (50% male) were finally included (total 64 images, n=4 per subject). **Table 1** provides demographic features. Male exhibited higher anthropometric outcomes and larger CSA than female ($P<0.001$). A positive correlation between multifidus CSA and weight (r:0.473, P=0.006), height (r:0.385; P=0.03), and BMI (r:0.481, P=0.005) was found.

Table 2 shows reliability data of imaging capturing. In general, imaging capturing intra-examiner reliability ($ICC_{3,1}$) was excellent ranging from 0.988 to 0.996 with a SEM from 0.3% to 0.7%. No difference between experienced/novice examiner/testers was found. Inter-examiner reliability was also excellent for both experienced ($ICC_{3,2}$ 0.965) and novice ($ICC_{3,2}$ 0.958) raters, with SEM of 2.6% and 3.2%, respectively.

Reliability data of imaging calculation of both raters is shown on **Table 3**. Again, intra- and inter-rater reliability was excellent for both raters but with smaller SEM for intra-rater reliability.

Discussion

This is the first study assessing reliability of both imaging capturing (positioning) and imaging calculation (assessment) of the neck muscles considering the experience of the assessor. In general, intra- and inter- rater reliability of imaging capturing and calculation of cervical multifidus CSA was excellent when applied by an experienced or novice assessor in asymptomatic individuals. Our findings are similar to those previously found in patients with mechanical neck pain (Fernández-de-las-Peñas et al, 2008) and slightly better than those previously reported by Kristjansson (2004). In addition, reliability data of ultrasound imaging assessment was slightly superior to reliability data obtained for MRI. Snodgrass et al (2019b) reported good to excellent intra- (ICC 0.78-0.96), but fair to good inter- (ICC 0.44-0.88) rater reliability of cervical multifidus assessment using MRI in 5 asymptomatic individuals. Interestingly, no differences between an experienced and novice examiner/rater were observed. This is a relevant topic since ultrasound imaging assessment is operator-dependent. Two points are relevant for ultrasound assessment, probe angulation and pressure. Whittaker et al (2009) showed that probe angulation of less than approx. 10 degrees, as it is commonly done in clinical practice, do not distort measurements of tissue thickness. No data about probe pressure is available. It is possible that strict positioning and measurement protocols followed in our study could explain high reliability values.

This study can help for developing specific protocols for image capturing and calculation for both research and clinical practice. For instance, morphometry assessment, i.e., CSA, is based on anatomical surrounding muscle contours. Previous studies recognized that lack of proper visualization of the fascial layers dividing the cervical multifidus from surrounding muscles (Fernández-de-las-Peñas et al, 2008; Kristjansson, 2004; Lee et al, 2007). One potential reason for this lack of clarity of muscular fascial layers could be the presence of fatty infiltration, a sign potentially associated with WAD (Owers et al, 2018). In the current study, we assessed cervical multifidus, without including the rotators, as it was conducted in previous studies (Fernández-de-las-Peñas et al, 2008; Kristjansson, 2004; Lee et al, 2007). The main reason was that the cervical multifidus attaches to the posterior aspect of the facet capsules (Anderson et al, 2005) and play a relevant role in proprioception. In such a scenario, fascial layers as border contours for determining muscle morphology may have highly relevance. Since the current study included asymptomatic individuals, fascial layers surrounding the multifidus were properly identified (Fig. 2). Rankin et al (2005) were not able to consistently identify fascial divisions between multifidus/semispinalis and between multifidus/rotators; therefore, they provided normative data of the whole muscle group and not from individual muscles. It is also possible that technical improvements of ultrasound imaging equipment, offering much better-definition and quality images, can also influence proper visualization of fascial layers between muscles.

We found positive correlations between CSA and height, weight, and BMI supporting the fact that muscle morphometry is associated to anthropometric features (Rezasoltani et al 1998). This could explain the CSA variability found in previous studies. In fact, the only study investigating normative data of posterior cervical muscles reported that weight, rather than gender, was a relevant cofounder factor for CSA (Rankin et al, 2005). Previous studies investigating differences in CSA between neck pain patients and healthy subjects did not control for the anthropometric variables (Fernández-de-las-Peñas et al, 2008; Kristjansson, 2004; Lee et al, 2007). De Pauw et al (2016) concluded that although there is evidence of reduced CSA in the cervical multifidus in people with mechanical neck pain, more studies are needed due to the inconsistency on the results. Future studies investigating differences between pain populations and healthy subjects should include these considerations.

Finally, this study has some limitations. First, we included asymptomatic subjects. We do not know if similar results would be observed in patients with mechanical neck pain or WAD. Second, our sample was small. Therefore, our results should not be considered as potential normative data, future studies are needed to determine muscle morphology data of posterior neck muscles separately, e.g., semispinalis, splenius, multifidus, and rotators. In addition, reliability of imaging capturing (probe assessment/patient positioning) was tested within 10 minutes on the same day which may not be clinically relevant. Imaging capturing reliability should ideally be assessed on different days in future studies. Finally, it is also important to consider that fatty infiltration quantification is not possible with ultrasound imaging and is currently possible with MRI, since MRI calculates the amount of fat by differentiating fat and soft-aqueous tissue signal intensities (Elliott et al, 2013). Therefore, RUSI should be only used for determining muscle morphology, e.g., CSA, size, or muscle function, but not intramuscular quality, e.g., fat, or fibers.

Conclusions

We found that ultrasound assessment of cervical multifidus at C4/C5 level is highly reliable for evaluating CSA in asymptomatic people since imaging capturing and imaging calculation exhibited excellent intra- and inter- examiner reliability. Reliability was similar independently of the assessor experience. This paper proposes technical considerations for future studies assessing muscle morphometry in neck pain populations.

Legend of Figures

Figure 1: (A) Ultrasound probe placement over the cervical multifidus at C4 level; (B) Ultrasound image showing the superficial posterior neck muscles.

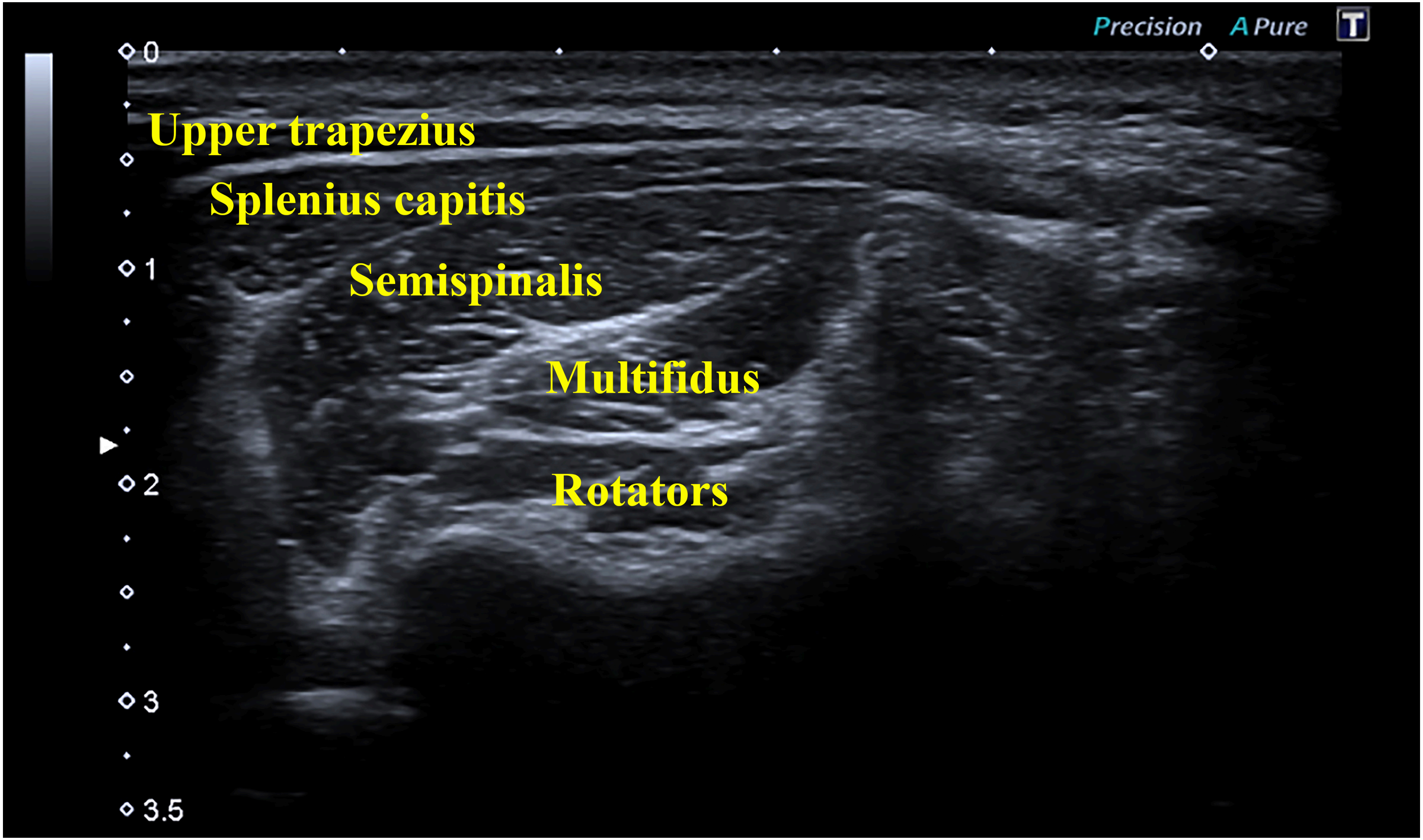
Figure 2: Cross sectional area (CSA) assessment of the cervical multifidus. Borders were marked as follows: 1, inferior limit: internal echogenic fascia between cervical multifidus and rotator muscle; 2, superior limit: echogenic fascia between cervical multifidus and semispinalis; 3, medial limit: echogenic spinous process

References

- Abbott R, Pedler A, Sterling M, Hides J, Murphey T, Hoggarth M, Elliott J. The geography of fatty infiltrates within the cervical multifidus and semispinalis cervicis in individuals with chronic whiplash-associated disorders. J Orthop Sports Phys Ther 2015; 45: 281-8.
- Anderson JS, Hsu AW, Vasavada AN. Morphology, architecture, and biomechanics of human cervical multifidus. Spine 2005; 30: E86-91.
- De Pauw R, Coppieters I, Kregel J, De Meulemeester K, Danneels L, Cagnie B. Does muscle morphology change in chronic neck pain patients? A systematic review. Man Ther 2016; 22: 42-9.
- Elliott JM, Walton DM, Rademaker A, Parrish TB. Quantification of cervical spine muscle fat: a comparison between T1-weighted and multi-echo gradient echo imaging using a variable projection algorithm (VARPRO). BMC Med Imaging 2013; 13: 30.
- Elliott JM, Pedler AR, Jull GA, Van Wyk L, Galloway GG, O□Leary SP. Differential changes in muscle composition exist in traumatic and nontraumatic neck pain. Spine 2014; 39: 39-47.
- Fernández-de-las-Peñas C, Albert-Sanchís JC, Buil M, Benitez JC, Alburquerque-Sendín F. Cross-sectional area of cervical multifidus in females with chronic bilateral neck pain compared to controls. J Orthop Sports Phys Ther 2008; 38: 175-80.

- Kamper SJ, Rebbeck TJ, Maher CG, McAuley JH, Sterling M. Course and prognostic factors of whiplash: a systematic review and meta-analysis. Pain 2008; 138: 617-629.
- Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016; 15: 155-63.
- Javanshir K, Amiri M, Mohseni-Bandpei MA, Rezasoltani A, Fernández-de-las-Peñas C. Ultrasonography of the cervical muscles: a critical review of the literature. J Manipul Physiol Ther 2010; 33: 630-7.
- Kristjansson E. Reliability of ultrasonography for the cervical multifidus muscle in asymptomatic and symptomatic subjects. Man Ther 2004; 9: 83-8.
- Lee JP, Tseng WY, Shau YW, Wang CL, Wang HK, Wang SF. Measurement of segmental cervical multifidus contraction by ultrasonography in asymptomatic adults. Man Ther 2007; 12: 286-294.
- Lin YJ, Chai HM, Wang SF. Reliability of thickness measurements of the dorsal muscles of the upper cervical spine: an ultrasonographic study. J Orthop Sports Phys Ther 2009; 39: 850-7.
- Owers DS, Perriman DM, Smith PN, Neeman T, Webb AL. Evidence for cervical muscle morphometric changes on magnetic resonance images after whiplash: A systematic review and meta-analysis. Injury 2018; 49: 165-176
- Rankin G, Stokes M, Newham D. Size and shape of the posterior neck muscles measured by ultrasound imaging: normal values in males and females of different ages. Man Ther 2005; 10: 108-115.

- Rezasoltani A, Kallinen M, Mälkiä E, Vihko V. Neck semispinalis capitis muscle size in sitting and prone positions measured by real time ultrasonography. Clin Rehabil 1998; 12: 36-44.
- Snodgrass SJ, Croker C, Yerrapothu M, Shepherd S, Stanwell P, Holder C, Oldmeadow C, Elliott J. Cervical muscle volume in individuals with idiopathic neck pain compared to asymptomatic control: A cross-sectional magnetic resonance imaging study. Musculosket Sci Pract 2019a Aug 13:102050.
- Snodgrass SJ, de Zoete RMJ, Croker C, Yerrapothu M, Elliott JM. Reliability of cervical muscle volume quantification using magnetic resonance imaging. Musculosket Sci Pract 2019b; 44: 102056.
- Whittaker JL, Teyhen DS, Elliott JM, Cook K, Langevin HM, Dahl HH, Stokes M. Rehabilitative ultrasound imaging: understanding the technology and its applications. J Orthop Sports Phys Ther 2007; 37: 434-49.
- Whittaker JL, Warner MB, Stokes MJ. Induced transducer orientation during ultrasound imaging: effects on abdominal muscle thickness and bladder position. Ultrasound Med Biol 2009; 35:1803–1811


Precision
A Pure
0
1
2
3
3.5
Upper trapezius
Splenius capitis
Semispinalis
Multifidus
Rotators

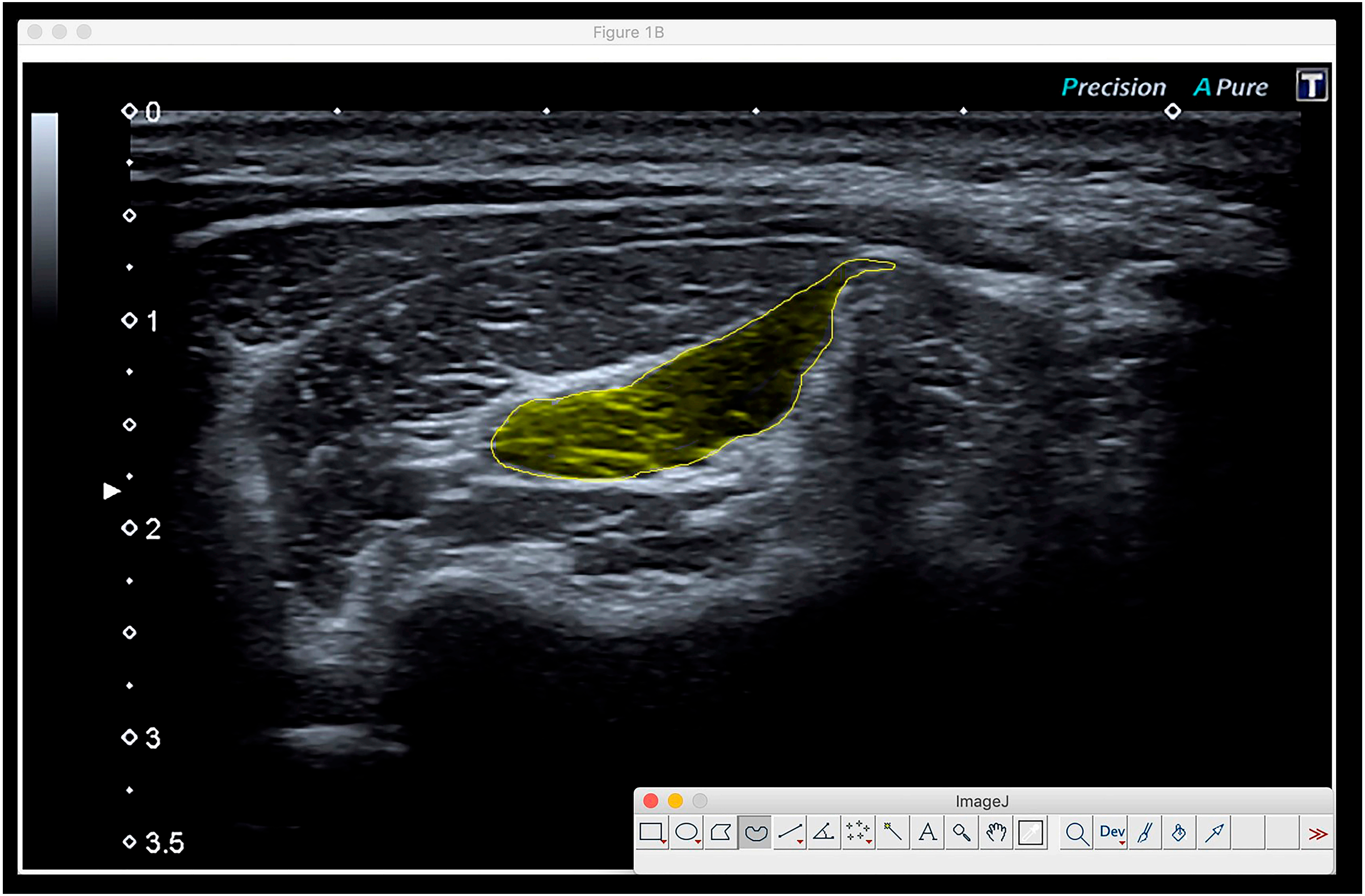


Figure 1B
Precision
A Pure
0
1
2
3
3.5
ImageJ
Dev

Table 1: Demographic Features of Participants by Gender

	Total (n=16)	Male (n=8)	Female (n=8)
Age (y)	28.5 (9.5)	31.0 (10.0)	25.5 (8.0)
Height (m)	1.70 (0.1)	1.80 (0.1)*	1.60 (0.1)
Weight (kg)	65.3 (15.0)	77.75 (9.0)*	53.0 (6.5)
BMI (kg/m^2)	21.9 (3.0)	23.78 (1.9)*	20.0 (1.9)
CSA right (cm^2)	1.30 (0.40)	1.50 (0.5)*	1.10 (0.3)
CSA left (cm^2)	1.20 (0.3)	1.40 (0.3)*	1.05 (0.15)

Values are expressed as means (SD)

* Significant differences between male and female ($P<0.05$)

Table 2: Intra- and Inter-Examiner Reliability of Imaging Capturing - Probe Assessment/Patient Positioning

	ICC (95% CI)	SD (cm^2)	SEM
Intra-Examiner ($ICC_{3,1}$) Reliability (experienced)			
Experienced Rater	0.995 (0.989 - 0.997)	0.05	0.3%
Novice Rater	0.988 (0.976 - 0.994)	0.07	0.7%
Intra-Examiner ($ICC_{3,1}$) Reliability (novice)			
Experienced Rater	0.996 (0.992 - 0.998)	0.05	0.3%
Novice Rater	0.993 (0.985 - 0.996)	0.07	0.5%
Inter-Examiner Reliability ($ICC_{3,2}$)			
Experienced Rater	0.965 (0.929 - 0.983)	0.14	2.6%
Novice Rater	0.958 (0.913 - 0.979)	0.16	3.2%

ICC: Intraclass Correlation Coefficient; SEM: Standard Error of Measurement
SD: Standard Deviation

Table 3: Intra- and Inter-Rater Reliability of Imaging Calculation – Scan Assessment

	ICC (95% CI)	SD (cm^2)	SEM
Intra-rater Experienced Reliability ($ICC_{3,1}$)	0.996 (0.994 - 0.997)	0.05	0.3%
Intra-rater Novice Reliability ($ICC_{3,1}$)	0.938 (0.912 - 0.956)	0.18	4.4%
Inter-rater Reliability ($ICC_{3,2}$)	0.922 (0.890 - 0.945)	0.21	5.8%

ICC: Intraclass Correlation Coefficient; SEM: Standard Error of Measurement
SD: Standard Deviation