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. Intraclass correlation coefficients (ICC) and standard error of measurement (SEM) were calculated. Results: Intra- (ICC3,1 0.988-0.996, SEM 0.3%-0.7%) and inter- (ICC3,2 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

Number of space Number of space TITLE OR ABSTRACT I Identification as a study of diagnostic accuracy using at least one measure of accuracy furth as sensitivity, specificity, predictive values, or AUC I ABSTRACT I Identification as a study of diagnostic accuracy using at least one measure of accuracy furth as sensitivity, specificity, predictive values, or AUC I ABSTRACT I Identification as a study of diagnostic accuracy using at least one measure of accuracy furth as sensitivity, specificity, predictive values, or AUC I ABSTRACT I Identification as a study of diagnostic accuracy using at least one measure of accuracy furth as sensitivity, specificity, predictive values, or AUC I INTREODUCTION I Identification as a study of accine the index test and references standard were performed (prospective study) or after (persopective study) I Participants I Identification as a study of a field participants were identified (setting, leadion and dates) 5 Participants I Identification as a study of inform previous test, inclusion in registry) I I I Identification as a study of inform previous test, inclusion in registry I I I I I Identification and antionale for test, positivity	Section & Tonic	No	Item	Reported on page
TTLE OR ABSTRACT Image: sensitivity, specificity, predictive values, or AUC) Image: sensitivity, specificity, predictive values, or AUC) ABSTRACT Image: sensitivity, specificity, predictive values, or AUC) Image: sensitivity, specificity, predictive values, or AUC) ABSTRACT Image: sensitivity, specificity, predictive values, or AUC) Image: sensitivity, specificity, predictive values, or AUC) Image: sensitivity, specificity, or Autor and Status, and conclusions Image: sensitivity, specificity, and sensitivity, and sensitivity, specificity, and sensitivity, and sensitivity, specificity, and sensitivity, and sensitity, and sensitivity, and sensit sensitivity, and sensititi	Section & Topic			#
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 aummary of study design, methods, results, and conclusions (brown of the second	TITLE OR ABSTRACT			
Image: sec: sec: sec: sec: sec: sec: sec: se		1	Identification as a study of diagnostic accuracy using at least one measure of accuracy	1
ABSTRACT Image: constraint of the specific guidance, see STARD for Abstracts) Image: constraint of the specific guidance, see STARD for Abstracts) INTRODUCTION Image: constraint of the specific guidance, see STARD for Abstracts) 3 INTRODUCTION Image: constraint of the specific guidance, see STARD for Abstracts) 3 INTRODUCTION Image: constraint of the specific guidance, see STARD for Abstracts) 3 INTRODUCTION Image: constraint of the specific guidance, see STARD for Abstracts) 3 INTRODUCTION Image: constraint of the specific guidance, see STARD for Abstracts of the index test and reference standard were proformed (prospecific study) or after (retrospecific study) 3 Participants Image: constraint of the index test and reference standard were identified (setting, location and dates) 5 Participants Image: constraint of the index test in sufficient detail to allow reglication 6 Image: constraint of the index test in sufficient detail to allow reglication 6 7 Image: constraint of the index test in sufficient detail to allow reglication 6 7 Image: constraint of the reference standard (if alternations entils are exolubate) 6 7 Image: constraint of the reference standard (if start patibility cut offs or result categories of the reference standard, distinguishing pre-specified from exploratory			(such as sensitivity, specificity, predictive values, or AUC)	
92 Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts) 1-2 INTRODUCTION 7 7 93 Scientific and clinical background, including the intended use and clinical role of the index test study design 5 METHODS 7 7 Study design 5 5 METHODS 7 7 Study design 6 Eligibility criteria 5 Participants 6 Eligibility criteria 5 7 0 Nuchter participants formed a consecutive, random or convenience series 5 7 0 Nuchter participants formed a consecutive, random or convenience series 5 7 10 Reference standard, in sufficient detail to allow replication 6-7 7 11 Rationale for test pastificient detail to allow replication 6-7 7 12 Definition of and rationale for test pastificient detail to allow replication 6-7 7 12 Reference standard, distiguishing pre-specified from exploratory 6-7 7 14 Reference standard, distiguishing pre-specified from exploratory 6-7 7 14 Reference standard, distiguishing pre-specified from exploratory 6-7 7 14 Reference standard, di	ABSTRACT			
INTRODUCTOR i i INTRODUCTOR i i <td< td=""><td></td><td>2</td><td>Structured summary of study design, methods, results, and conclusions</td><td>1-2</td></td<>		2	Structured summary of study design, methods, results, and conclusions	1-2
INTRODUCTION V Scientific and clinical background, including the intended use and clinical role of the index test 34 METHODS V V V Study design 5 Study design 5 Study design 5 V V Participants 6 Eligibility criteria 5 On what basis potentially cligble participants were identified 5 5 Participants 6 Eligibility criteria 5 On what basis potentially cligble participants were identified (setting, location and dates) 5 Test methods 100 Reference standard, in sufficient detail to allow replication 6-7 100 Index test, in sufficient detail to allow replication 6-7 -7 111 Rationale for test positivity cut-for or sult categories 6-7 112 Definition of and rationale for test positivity cut-for or sult categories 6-7 113 Netther clinical information and reference standard for sults were available 6-7 114 the index test, distinguishing pre-specified from exploratory 8 115 berliniton of anarizonale for test p			(for specific guidance, see STARD for Abstracts)	
9494Scientific and clinical background, including the intended use and clinical role of the index test94METHODS45METHODS77Study degian8Whether data collection was planned before the index test and reference standard7Porticipants6Eligibility criteria58Where and when potentially eligible participants were identified59Whether participants formal a consecutive, random or convenience series510index test, in sufficient detail to allow replication6711Rationals for choosing the reference standard (if alternatives colst)6712Politicitor of and rationale for test positivity cut offs or result categories6713Rationals for choosing the reference standard (if alternatives colst)6714Whether clinical information and reference standard (if alternatives colst)6715Politition of and rationale for test positivity cut offs or result categories6714Whether clinical information and reference standard results were available to the reference standard, distinguishing pre-specified from exploratory6715How indistribuicity and reference standard results were available to the assessors of the reference standard results were available to the assessors of the reference standard results were available to the assessors of the reference standard results were handled816How indistributin of severified form golymatic participants8817May adves of washing or comparing measures of diagnostic accuracy to the as	INTRODUCTION			
InterfactSet of the set of th		3	Scientific and clinical background, including the intended use and clinical role of the index test	3-4
METHODS Fm Mether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study) Methods Participants 6 Eligibility criteria 5 R On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry) 5 R Where and when potentially eligible participants were identified (setting, location and dates) 5 Test methods 10a Index test, in sufficient detail to allow replication 6-7 Test methods 10a Index test, in sufficient detail to allow replication 6-7 R Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory 6-7 R Definition of and rationale for test positivity cut-offs or result categories of the reference standard distinguishing pre-specified from exploratory 6-7 R Definition for estimation and midex test results were available 6-7 R Definition for estimating or comparing measures of diagnostic accuracy 8 R Definition of setting information and reference standard 8 Analysis 14 Methode store stindard		4	Study objectives and hypotheses	5
Study design S Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study) Participants 6 Eligibility critera 5 Participants 6 Eligibility critera 5 Participants 6 Eligibility critera 5 Participants 8 Where and when potentially eligible participants were identified (setting, location and dates) 5 Participants 8 Where and when potentially eligible participants were identified (setting, location and dates) 5 Test methods 10 Index test, in sufficient detail to allow replication 6-7 11 Rationale for choosing the reference standard (if alternative soits) 6-7 12 Definition of and rationale for test positivity cut-offs or result categories 6-7 12 Definition of and rationale for test positivity cut-offs or result categories 6-7 13 Whether clinical information and reference standard results were available 6-7 14 Methor stinical information and reference standard results were available 6-7 15 How index test results were available 6-7 <tr< td=""><td>METHODS</td><td></td><td></td><td></td></tr<>	METHODS			
Participants 6 Eligibility criteria 5 Participants 6 Eligibility criteria 5 On what basis potentially eligible participants were identified 5 (such as symptoms, results from previous tests, inclusion in registry 5 10 Reference standard, in sufficient details to allow replication 6-7 11 Rationale for choosing the reference standard (if alternatives exist) 6-7 12 Definition of and rationale for test positivity cut-offs or result categories 6-7 13 Rationale for these positivity cut-offs or result categories 6-7 14 Rationale for test positivity cut-offs or result categories 6-7 15 Definition of and rationale for test positivity cut-offs or result categories 6-7 16 Whether clinical information and reference standard for sult saver available 6-7 17 Other reference standard, distinguishing pre-specified from exploratory 6-7 18 Whether clinical information and reference standard were handled 8 19 Nethest or estimating index test results were available 6-7 14 Methods for estimating index test results were available 6-7 15 How indetermination comparing measures of diagnostic accuracy 8 16 How indetermination dor coligaram 8	Study design	5	Whether data collection was planned before the index test and reference standard	
Participants 6 Eligibility criteria 5 2 On what basis potentially eligible participants were identified (setting, location and dates) 5 2 Where and when potentially eligible participants were identified (setting, location and dates) 5 3 Where and when potentially eligible participants were identified (setting, location and dates) 5 3 Where and when potentially eligible participants were identified (setting, location and dates) 5 3 Where and when potentially eligible participants were identified (setting, location and dates) 6-7 3 Whether participants formed a consecutive, random or convenience series 6-7 4 Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory 6-7 5 Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory 6-7 6 How horted trainal for criterion standard results were available 6-7 6 How indeterminate index test or reference standard results were available 6-7 7 How indeterminate index test or reference standard results were handled 8 8 How indeterminate index test or reference standard results were handled 8 8 How indeterminate index test or reference standard results were handled			were performed (prospective study) or after (retrospective study)	
9 0 what basis potentially eligible participants were identified (setting, location and dates) 5 10 8 Where and when potentially eligible participants were identified (setting, location and dates) 5 11 Reference standard, in sufficient detail to allow replication 6-7 100 Reference standard, in sufficient detail to allow replication 6-7 11 Rationale for choosing the reference standard (if alternatives exist) 6-7 120 Definition of and rationale for test positivity cut-offs or result categories of the reference standard, disting pre-specified from exploratory 6-7 130 Whether clinical information and reference standard results were available to the performers/readers of the index test results were available 6-7 141 Methods for estimating or comparing measures of diagnostic accuracy 8 151 How indeterminate index test and reference standard results were available 8 152 How indeterminate index test and reference standard results were handled 8 153 How indeterminate index test and reference standard results were handled 8 154 How indeterminate index test and reference standard results were handled 8 155 How indeterminate index test and reference standard results were handled 8 154 How indeterminate index test and reference standard from exploratom 8 <t< td=""><td>Participants</td><td>6</td><td>Eligibility criteria</td><td>5</td></t<>	Participants	6	Eligibility criteria	5
Such a symptoms, results from previous tests, incluision in registry)Image: Such a symptoms, results from previous tests, incluision in registry)Such a symptoms, results from previous tests, incluision in registry)9Whether participants formed a consecutive, random or convenience series57100Index test, in sufficient detail to allow replication6-7101Rationale for choosing the reference standard (if alternatives exist)6-711Rationale for choosing the reference standard (if alternatives exist)6-7120Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory6-7121Definition of and rationale for test positivity cut-offs or result categories of the reference standard results were available to the performers/readers of the index test results were available to the assessors of the reference standard results were handled6-7131Muehter clinical information and index test results were handled8141Methods for estimating or comparing measures of diagnostic accuracy8151How indicetrminate index test or reference standard results were handled8162How mising data on the index test and reference standard results were handled8163Index estand results and handled test and reference standard results were handled8164How mising data on the index test and reference standard results were handled8165How mising data on the index test and reference standard results were handled8166How mising data on the index te		7	On what basis potentially eligible participants were identified	5
8Where and when potentially eligible participants were identified (setting, location and dates)57 est methods10aIndex test, in sufficient detail to allow replication6-710bReference standard, in sufficient detail to allow replication6-711Rationale for choosing the reference standard (if atternatives exist)6-712aDefinition of and rationale for test positivity cut-offs or result catepories of the index test, distinguishing pre-specified from exploratory6-713aWhether clinical information and reference standard distinguishing pre-specified from exploratory6-713aWhether clinical information and reference standard results were available to the performers/readers of the index test.6-714Methods for estimating or comparing measures of diagnostic accuracy815How indestriminate index test results were available to the assessors of the reference standard distuly were handled817Any anayses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory818Intended sample size and how it was determined819Flow of participants, using a diagram919Flow of servity of disease in those withou the target condition819Distribution of servity of diseas in those withou the target condition8101Distribution of servity of disease in those withou the target condition819Flow of participants, using a diagram9101Distribution of servity of disease in those withou the target condition8102 <td></td> <td></td> <td>(such as symptoms, results from previous tests, inclusion in registry)</td> <td></td>			(such as symptoms, results from previous tests, inclusion in registry)	
9Whether participants formed a consecutive, random or convenience series5Test methods10aIndex test, in sufficient detail to allow replication6-710bReference standard, in sufficient detail to allow replication6-711Rationale for choosing the reference standard (if alternatives exist)6-712aDefinition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory6-712aDefinition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory6-713aWhether clinical information and reference standard results were available to the performers/readers of the index test6-713bWhether clinical information and index test results were available to the assessors of the reference standard results were available to the assessors of the reference standard results were handled813bMethods for estimating or comparing measures of diagnostic accuracy814Methods for estimating or comparing measures of diagnostic accuracy815How missing data on the index test and reference standard were handled816How missing data on the index test and reference standard817Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory818Intended sample size and how it was determined819Flow of participants, using a diagram920Baseline demographic and clinical characteristics of participants		8	Where and when potentially eligible participants were identified (setting, location and dates)	5
Test methods 10a Index test, 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 13a Whether clinical information and reference standard, distinguishing pre-specified from exploratory 6-7 13b Whether clinical information and reference standard results were available to the performere/readers of the index test is to the performere/readers of the index test is 6-7 13b Whether clinical information and index test results were available to the assessors of the reference standard results were handled 6-7 14b Methods for reference standard results were handled 8 15 How indeterminate index test or reference standard 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 19 Flow of participants, using a diagram 8 10a Distribution of severity of disease in those with the target condition		9	Whether participants formed a consecutive, random or convenience series	5
100 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 assessors of the reference standard faignostic accuracy 8 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 sing data on the index test and reference standard results were handled 8 16 How indeterminate index test or reference standard results 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 19 Flow of participants, using a diagram	Test methods	10a	Index test, 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 6-7 12b Definition of and rationale for test positivity cut-offs or result categories 6-7 12b Definition of and rationale for test positivity cut-offs or result categories 6-7 of the reference standard, distinguishing pre-specified from exploratory 6-7 13b Whether clinical information and reference standard results were available to the assessors of the reference standard results were available to the assessors of the reference standard results were handled 6-7 Analysis 14 Methods for estimating or comparing measures of diagnostic accuracy 8 Analysis 14 Method state on the index test and reference standard results were handled 8 15 How indeterminate index test and reference standard were handled 8 16 How missing data on the index test and reference standard for test positivity of test or servity of disease in those with the target condition 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 16 How insting data on the index test re		10b	Reference standard, in sufficient detail to allow replication	6-7
12a Definition of and rationale for test positivity cut-offs or result categories 6-7 12b Definition of and rationale for test positivity cut-offs or result categories 6-7 12b Definition of and rationale for test positivity cut-offs or result categories 6-7 12b Definition of and rationale for test positivity cut-offs or result categories 6-7 12b Whether clinical information and reference standard results were available 6-7 12b Whether clinical information and index test results were available 6-7 12b Whether clinical information and index test results were available 6-7 12b Whether clinical information and index test results were available 6-7 12b Methods for estimating or comparing measures of diagnostic accuracy 8 12c How index test or reference standard results were handled 8 12 How insing data on the index test and reference standard were handled 8 12 Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory 8 12b How index dample size and how it was determined 8 12c Porticipants, using a diagram 16 12a Distribution of alternative diagnoses in those with the target condition 8 12a Distribution of alternative diagnoses in those with pot the targ		11	Rationale for choosing the reference standard (if alternatives exist)	6-7
Image: Section of the index test, distinguishing pre-specified from exploratory 6-7 Image: Section of the reference standard, distinguishing pre-specified from exploratory 6-7 Image: Section of the reference standard, distinguishing pre-specified from exploratory 6-7 Image: Section of the reference standard, distinguishing pre-specified from exploratory 6-7 Image: Section of the reference standard 6-7 Image: Section of the reference standard 6-7 Image: Section of the reference standard 8 Image: Section of the reference standard 8 Image: Section of the reference standard results were handled 8 Image: Section of the reference standard of the reference standard were handled 8 Image: Section of the reference standard of the reference standard results were handled 8 Image: Section of the reference standard of the reference standard results were handled 8 Image: Section of the reference standard of the reference standard results were handled 8 Image: Section of the reference standard of the reference standard were handled 8 Image: Section of the reference standard of the reference standard section were handled 8 Image: Section of the reference standard 8 Image: Section of the reference standard section secti		12a	Definition of and rationale for test positivity cut-offs or result categories	6-7
12b Definition of and rationale for test positivity cut-offs or result categories 6-7 of the reference standard, distinguishing pre-specified from exploratory 6-7 13b Whether clinical information and reference standard results were available to the performers/readers of the index test results were available to the performers/readers of the index test results were available 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 Analysis 14 Methods for estimating or comparing measures of diagnostic accuracy 8 15 How indeterminate index test or reference standard 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 19 Flow of participants, using a diagram 20 20 Baseline demographic and clinical characteristics of participants 8 21 Distribution of severity of disease in those with the target condition 8 22 Time interval and any clinical interventions between index test and reference standard 8 23 <td< td=""><td></td><td></td><td>of the index test, distinguishing pre-specified from exploratory</td><td></td></td<>			of the index test, distinguishing pre-specified from exploratory	
interm 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 Analysis 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 19 Flow of participants, using a diagram 7 20 Baseline demographic and clinical characteristics of participants 8 21a Distribution of alternative diagnoses in those with the target condition 8 22b Distribution of the index test results (or their distribution) by the results of the reference standard 8-9 22c Cross tabulation of the index test results (or their distribution) by the results of the reference standard 8-9 23 Any adverse events from performing the index test or the reference standard egeneralisability 8-9		12b	Definition of and rationale for test positivity cut-offs or result categories	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 or reference standard were handled 8 17 Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory 8 RESULTS Intended sample size and how it was determined 8 20 Baseline demographic and clinical characteristics of participants 8 21a Distribution of severity of disease in those with the target condition 8 21a Distribution of the index test results (or their distribution) 8-9 21a Cross tabulation of the index test results (or their distribution) 8-9 21a Cross tabulation of the reference standard 8-9 21a Cross tabulation of the reference standard 8-9 22 Time interval and any clinical interventions between index test and reference standard 8-9 22 Start performing the index test or the reference standard 8-9 23			of the reference standard, distinguishing pre-specified from exploratory	
IndexIndexIndexIndexIndex13bWhether clinical information and index test results were available to the assessors of the reference standard6-7Analysis14Methods for estimating or comparing measures of diagnostic accuracy815How indeterminate index test or reference standard results were handled816How missing data on the index test and reference standard were handled817Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory818Intended sample size and how it was determined820Baseline demographic and clinical characteristics of participants821Distribution of severity of disease in those with the target condition822Time interval and any clinical interventions between index test and reference standard823Cross tabulation of the index test results (or their distribution) by the results of the reference standard8-924Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)8-924Study limitations, including sources of potential bias, statistical uncertainty, and generalisability120THER INFORMATION29Registration number and name of registry9-1102Bestivation number and name of registry9-1103Sources of funding and other support; role of funders9-11		13a	Whether clinical information and reference standard results were available	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 RESULTS 18 Intended sample size and how it was determined 8 RESULTS 19 Flow of participants, using a diagram 8 20 Baseline demographic and clinical characteristics of participants 8 21 Distribution of severity of disease in those with the target condition 8 22 Time interval and any clinical interventions between index test and reference standard 8 22 Time interval and any clinical interventions between index test and reference standard 8 23 Cross tabulation of the index test results (or their distribution) 8-9 24 Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals) 8-9 24 Estimates of diagnostic accuracy of potential bias, statistical uncertainty, and generalisability 9-11 DISCUSSION 24 Study limitations, in			to the performers/readers of the index test	· _
Analysis14Methods for estimating or comparing measures of diagnostic accuracy8Analysis15How indeterminate index test or reference standard results were handled816How missing data on the index test or reference standard were handled817Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory818Intended sample size and how it was determined8RESULTS20Baseline demographic and clinical characteristics of participants821aDistribution of severity of disease in those with the target condition822Time interval and any clinical interventions between index test and reference standard823Cross tabulation of the reference standard824Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)8-925Any adverse events from performing the index test or the reference standard8-9DISCUSSION27Implications, including sources of potential bias, statistical uncertainty, and generalisability12OTHER INFORMATION28Registration number and name of registry5-30Sources of funding and other support: role of funders-		13b	Whether clinical information and index test results were available	6-7
Analysis 14 Methods for estimating or comparing measures of algobic 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 Participants 19 Flow of participants, using a diagram 10 20 Baseline demographic and clinical characteristics of participants 8 21a Distribution of severity of disease in those with the target condition 8 21a Distribution of alternative diagnoses in those without the target condition 8 22a Time interval and any clinical interventions between index test and reference standard 8 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 24 Estimates of participants, including sources of potential bias, statistical uncertainty, and generalisability 12 DISCUSSION 23 Ruy adverse even	A		to the assessors of the reference standard	0
15How indeterminate index test or reference standard were handled816How missing data on the index test and reference standard were handled817Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory818Intended sample size and how it was determined8RESULTS19Flow of participants, using a diagram1020Baseline demographic and clinical characteristics of participants821aDistribution of severity of disease in those with the target condition822Time interval and any clinical interventions between index test and reference standard823Cross tabulation of the index test results (or their distribution) by the results of the reference standard8-924Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals) ay nadverse events from performing the index test or the reference standard8-9DISCUSSION26Study limitations, including sources of potential bias, statistical uncertainty, and generalisability12OTHER INFORMATION28Registration number and name of registry2429Where the full study protocol can be accessed9-11	Analysis	14	Methods for estimating or comparing measures of diagnostic accuracy	8
10How missing data on the index test and reference standard were handled317Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory818Intended sample size and how it was determined8RESULTS19Flow of participants, using a diagram1020Baseline demographic and clinical characteristics of participants821aDistribution of severity of disease in those with the target condition821bDistribution of alternative diagnoses in those without the target condition822Time interval and any clinical interventions between index test and reference standard823Cross tabulation of the index test results (or their distribution) by the results of the reference standard8-924Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)8-925Any daverse events from performing the index test or the reference standard generalisability8-926Study limitations, including sources of potential bias, statistical uncertainty, and generalisability12CTHER INFORMATION28Registration number and name of registry9-11CHER INFORMATION29Where the full study protocol can be accessed530Sources of funding and other support: role of funders5		15	How indeterminate index test or reference standard results were handled	8 0
17 Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory of a several presspecified from exploratory of a several presspecification from exploratory of a several presspecified from explory of a several presspecified from exploratory of a se		17	Any analysis of variability in diagnostic accuracy, distinguishing the specified from evaluations	0 0
Interded sample size and now it was determined a RESULTS Image: constraint of the interded sample size and now it was determined a Participants 19 Flow of participants, using a diagram 8 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 22 Time interval and any clinical interventions between index test and reference standard 8 7est 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 OTHER 27 Implications for practice, including the intended use and clinical role of the index test 9-11 OTHER <td></td> <td>1/</td> <td>Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory</td> <td>0</td>		1/	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	0
ResolutionFlow of participants, using a diagramImage: Constraint of the index set of participants and clinical characteristics of participantsImage: Constraint of the index set of participants20Baseline demographic and clinical characteristics of participants821aDistribution of severity of disease in those with the target condition821bDistribution of alternative diagnoses in those without the target condition822Time interval and any clinical interventions between index test and reference standard822Time interval and any clinical interventions between index test and reference standard823Cross tabulation of the index test results (or their distribution) by the results of the reference standard8-924Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)8-925Any adverse events from performing the index test or the reference standard826Study limitations, including sources of potential bias, statistical uncertainty, and generalisability9-110THER INFORMATION28Registration number and name of registry9-1129Where the full study protocol can be accessed30Sources of funding and other support; role of funders		19	Intended sample size and now it was determined	ŏ
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 22 Time interval and any clinical interventions between index test and reference standard 8 23 Cross tabulation of the index test results (or their distribution) 8-9 24 Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals) 8-9 24 Estimates of from performing the index test or the reference standard 8-9 25 Any adverse events from performing the index test or the reference standard 8-9 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 0THER 28 Registration number and name of registry 29 29 Where the full study protocol can be accessed 30 Sources of funding and other support; role of funders	RESULIS	10	Clow of participants using a diagram	
20Baseline defining aprile and chinical characteristics of participants321aDistribution of severity of disease in those with the target condition821bDistribution of alternative diagnoses in those without the target condition822Time interval and any clinical interventions between index test and reference standard822Time interval and any clinical interventions between index test and reference standard823Cross tabulation of the index test results (or their distribution) by the results of the reference standard8-924Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)8-925Any adverse events from performing the index test or the reference standard8-9DISCUSSION26Study limitations, including sources of potential bias, statistical uncertainty, and generalisability12OTHER INFORMATION28Registration number and name of registry9-1129Where the full study protocol can be accessed29Where the full study protocol can be accessed30Sources of funding and other support; role of funders5050	Participants	17	Provior participante, using a diagram	0
21aDistribution of severity of disease in those with the target conditiono21bDistribution of alternative diagnoses in those without the target condition822Time interval and any clinical interventions between index test and reference standard87est results23Cross tabulation of the index test results (or their distribution) by the results of the reference standard8-924Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)8-925Any adverse events from performing the index test or the reference standard8-9DISCUSSION26Study limitations, including sources of potential bias, statistical uncertainty, and generalisability12OTHER INFORMATION28Registration number and name of registry9-1129Where the full study protocol can be accessed29Where the full study protocol can be accessed30Sources of funding and other support; role of funders30Sources of funding and other support; role of funders		20 01 a	Distribution of severity of disease in these with the target condition	U Q
210Distribution of alternative diagnoses in those without the target collitiono22Time interval and any clinical interventions between index test and reference standard8Test results23Cross tabulation of the index test results (or their distribution) by the results of the reference standard8-924Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)8-925Any adverse events from performing the index test or the reference standard8-9DISCUSSION26Study limitations, including sources of potential bias, statistical uncertainty, and generalisability12OTHER INFORMATION28Registration number and name of registry9-1129Where the full study protocol can be accessed30Sources of funding and other support; role of funders		21d 21h	Distribution of alternative diagnoses in those without the target condition	U Q
Test results23Cross tabulation of the index test results (or their distribution) by the results of the reference standard8-924Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)8-925Any adverse events from performing the index test or the reference standard8-9DISCUSSION26Study limitations, including sources of potential bias, statistical uncertainty, and generalisability12OTHER INFORMATION28Registration number and name of registry9-1129Where the full study protocol can be accessed20Sources of funding and other support; role of funders		22 710	Time interval and any clinical interventions between index test and reference standard	8
Instruction Instruction Organization of the index test results (of their distribution) Organization of the index test results (of their distribution) by the results of the reference standard 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 0 27 Implications for practice, including the intended use and clinical role of the index test 9-11 OTHER 28 Registration number and name of registry 29 Where the full study protocol can be accessed 29 30 Sources of funding and other support; role of funders 30 Sources of funding and other support; role of funders	Test results	22 22	Cross tabulation of the index test results (or their distribution)	9_0
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 6 5 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 28 Registration number and name of registry 29 29 Where the full study protocol can be accessed 30 Sources of funding and other support; role of funders	10311030113	23	by the results of the reference standard	0 /
21 Example of diagnostic decaded, and their processin (out it or role of the index test) 0 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 0THER 17 Implications for practice, including the intended use and clinical role of the index test 9-11 OTHER 28 Registration number and name of registry 29 29 Where the full study protocol can be accessed 30 30 Sources of funding and other support; role of funders 50		24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	8-9
DISCUSSION 26 Study limitations, including sources of potential bias, statistical uncertainty, and generalisability 12 OTHER 27 Implications for practice, including the intended use and clinical role of the index test 9-11 OTHER 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 30 Sources of funding and other support; role of funders 57		25	Any adverse events from performing the index test or the reference standard	8-9
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 P 9-11 INFORMATION 28 Registration number and name of registry 29 29 Where the full study protocol can be accessed 30 Sources of funding and other support; role of funders	DISCUSSION			
Image: Second		26	Study limitations, including sources of potential bias, statistical uncertainty, and	12
27 Implications for practice, including the intended use and clinical role of the index test 9-11 OTHER NFORMATION 28 Registration number and name of registry 29 Where the full study protocol can be accessed 9 30 Sources of funding and other support; role of funders			generalisability	
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		27	Implications for practice, including the intended use and clinical role of the index test	9-11
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	OTHER			
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	INFORMATION			
29 Where the full study protocol can be accessed 30 Sources of funding and other support; role of funders		28	Registration number and name of registry	
30 Sources of funding and other support; role of funders		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.



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

5 Abstract

4

Background: Studies have analyzed muscle morphometry of cervical multifidus by using 6 ultrasound imaging, but its reliability is not clearly determined. Objective: To investigate 7 intra- and inter-rater reliability of imaging capturing (probe assessment/patient positioning) 8 and imaging calculation (scan assessment) of cervical multifidus cross sectional area (CSA) 9 by considering the assessor's experience in asymptomatic individuals. **Design:** Reliability 10 study. Methods: The CSA of C4/C5 cervical multifidus was assessed in 16 asymptomatic 11 subjects. Two examiners performed the imaging capture and also repeated the procedure 12 (probe placement/patient positioning) twice with a 10-min period between each. Other two 13 raters conducted imaging calculations of CSA. Intra-examiner imaging capturing reliability, 14 each rater (experienced and novice) calculated multifidus CSA of both images obtained by 15 each examiner. Inter-examiner imaging capturing reliability, each rater calculated the CSA 16 17 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 18 correlation coefficients (ICC) and standard error of measurement (SEM) were calculated. 19 **Results**: Intra- (ICC_{3.1} 0.988-0.996, SEM 0.3%-0.7%) and inter- (ICC_{3.2} 0.958-0.965, SEM 20 2.6%-3.2%) examiner reliability of imaging capturing was excellent. Intra- and inter-rater 21 reliability of imaging calculation was also excellent for both raters (experienced/novice). 22 No significant differences between experienced or novice examiners or testers were found. 23

24	Conclusions: This study found that intra- and inter-examiner/rater reliability of imaging
25	capturing (probe assessment/patient positioning) and imaging calculation (scan assessment)
26	of the cervical multifidus CSA at C4/C5 level was excellent in asymptomatic subjects.
27	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

33 Introduction

34 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, 35 up to 50% of individuals continue reporting symptoms after 1-year (Kamper et al, 2008). 36 37 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 38 performance. Several studies have reported that people with WAD show increased fatty 39 infiltrate and changes in cross sectional area (CSA) in the posterior cervical muscles, 40 specifically the multifidi (Abbott et al, 2015; Elliot et al, 2014; Snodgrass et al, 2019a). 41 However, most studies have used magnetic resonance imaging (MRI) to assess muscle 42 morphology (Owers et al, 2018), which is not readily available in clinical practice. 43

A more pragmatic method of measuring muscle morphology is ultrasonography; but 44 45 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 46 the posterior neck muscles (Javanshir et al, 2010). Kristjansson (2004) analyzed the 47 reliability of assessing CSA of the C4 cervical multifidus and reported appropriate intra-48 and inter-tester agreement in 10 asymptomatic subjects, and acceptable intra-, but 49 questionable inter-tester, agreement in 10 symptomatic subjects. Fernández-de-las-Peñas et 50 al (2008) reported excellent between-scan and good between-day assessment in individuals 51 with mechanical neck pain. Lee et al (2007) found that ultrasound imaging was as reliable 52 and valid as MRI for assessing cervical multifidus thickness after isometric contractions. 53

An important step before neck muscle composition can routinely be used in research 54 55 or clinical practice is establishing reliability of imaging calculation (scan assessment) but also reliability of imaging capturing (patient positioning). This is particularly relevant since 56 ultrasonography is operator-dependent and the measurement protocol could influence the 57 imaging calculation. In fact, previous studies had mostly investigated the reliability of 58 imaging calculation (scan assessment) but not the reliability of imaging capturing (probe 59 60 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-61 examiner/rater reliability of imaging calculation (scan/image assessment) and imaging 62 63 capturing (probe assessment/patient positioning) of cervical multifidus CSA considering assessor experience 64

65

66 Methods

67 **Participants**

68 Asymptomatic volunteers without neck pain symptoms were recruited via local announcements between December 2018 and June 2019. To be eligible to participate, they 69 had to be between 18 and 45 years old and with no history of neck pain the previous year. 70 71 Exclusion criteria included history of whiplash injury; any pharmacological treatment affecting muscle tone, e.g., muscle relaxants, analgesics; prior history of cervical surgery; 72 cervical radiculopathy or myelopathy; presence of degenerative changes; and any medical 73 condition such as tumor or fracture. The study was approved by the Institutional Review 74 Ethical Committee of XX. All subjects signed the written informed consent prior to their 75 76 inclusion.

77 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 90° abduction and the elbows flexed to 90°. 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**).

91 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**).

97

- 98
- 99
- 100

101 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.

112 **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.

- 123
- 124

125 Statistical Analysis

Statistical analysis was performed using the SPSS package, Version 21 software for 126 Mac OS. Normal distribution of the data was verified by using the Shapiro-Wilk test. Intra-127 and inter- examiner/rater reliability was estimated using 2-way mixed-model, consistency-128 type intra class correlation coefficients (ICC). Reliability was classified as fair (ICC<0.50), 129 moderate (0.50<ICC<0.75), good (0.75<ICC<0.90) or excellent (ICC>0.90) (Koo and Li 130 2016). Standard error of measurement (SEM) was calculated using the resulting ICC values 131 and standard deviation (SD): SEM (%) = (SDx $\sqrt{1-ICC}$) x 100 to assess measurement precision. 132 All tests were two-tailed with p-values <0.05 considered significant. 133

134

135 **Results**

From a total of 20 subjects responding to the announcement, 4 were excluded due to 136 previous whiplash injury (n=2) and history of neck pain the previous year (n=2). Sixteen 137 asymptomatic subjects (50% male) were finally included (total 64 images, n=4 per subject). 138 139
 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 140 weight (r:0.473, P=0.006), height (r:0.385; P=0.03), and BMI (r:0.481, P=0.005) was found. 141 142 **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 143 from 0.3% to 0.7%. No difference between experienced/novice examiner/testers was found. 144 Inter-examiner reliability was also excellent for both experienced (ICC_{3,2} 0.965) and novice 145 (ICC_{3.2} 0.958) raters, with SEM of 2.6% and 3.2%, respectively. 146

147 Reliability data of imaging calculation of both raters is shown on Table 3. Again,
148 intra- and inter-rater reliability was excellent for both raters but with smaller SEM for intra149 rater reliability.

150

151 **Discussion**

This is the first study assessing reliability of both imaging capturing (positioning) 152 and imaging calculation (assessment) of the neck muscles considering the experience of the 153 assessor. In general, intra- and inter- rater reliability of imaging capturing and calculation 154 of cervical multifidus CSA was excellent when applied by an experienced or novice 155 assessor in asymptomatic individuals. Our findings are similar to those previously found in 156 patients with mechanical neck pain (Fernández-de-las-Peñas et al. 2008) and slightly better 157 than those previously reported by Kristiansson (2004). In addition, reliability data of 158 ultrasound imaging assessment was slightly superior to reliability data obtained for MRI. 159 Snodgrass et al (2019b) reported good to excellent intra- (ICC 0.78-0.96), but fair to good 160 161 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 162 examiner/rater were observed. This is a relevant topic since ultrasound imaging assessment 163 is operator-dependent. Two points are relevant for ultrasound assessment, probe angulation 164 and pressure. Whittaker et al (2009) showed that probe angulation of less than approx. 10 165 degrees, as it is commonly done in clinical practice, do not distort measurements of tissue 166 thickness. No data about probe pressure is available. It is possible that strict positioning and 167 measurement protocols followed in our study could explain high reliability values. 168

8

This study can help for developing specific protocols for image capturing and calculation 169 170 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 171 proper visualization of the fascial layers dividing the cervical multifidus from surrounding 172 muscles (Fernández-de-las-Peñas et al, 2008; Kristjansson, 2004; Lee et al, 2007). One 173 potential reason for this lack of clarity of muscular fascial layers could be the presence of 174 175 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 176 177 in previous studies (Fernández-de-las-Peñas et al, 2008; Kristjansson, 2004; Lee et al, 178 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 179 a scenario, fascial layers as border contours for determining muscle morphology may have 180 highly relevance. Since the current study included asymptomatic individuals, fascial layers 181 surrounding the multifidus were properly identified (Fig. 2). Rankin et al (2005) were not 182 able to consistently identify fascial divisions between multifidus/semispinalis and between 183 multifidus/rotators; therefore, they provided normative data of the whole muscle group and 184 not from individual muscles. It is also possible that technical improvements of ultrasound 185 186 imaging equipment, offering much better-definition and quality images, can also influence proper visualization of fascial layers between muscles. 187

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 190 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.

199 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 200 or WAD. Second, our sample was small. Therefore, our results should not be considered as 201 202 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 203 204 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 205 reliability should ideally be assessed on different days in future studies. Finally, it is also 206 important to consider that fatty infiltration quantification is not possible with ultrasound 207 imaging and is currently possible with MRI, since MRI calculates the amount of fat by 208 differentiating fat and soft-aqueous tissue signal intensities (Elliott et al, 2013). Therefore, 209 RUSI should be only used for determining muscle morphology, e.g., CSA, size, or muscle 210 211 function, but not intramuscular quality, e.g., fat, or fibers.

- 212
- 213
- 214
- 215

216

10

218 Conclusions

219	We found that ultrasound assessment of cervical multifidus at C4/C5 level is highly
220	reliable for evaluating CSA in asymptomatic people since imaging capturing and imaging
221	calculation exhibited excellent intra- and inter- examiner reliability. Reliability was similar
222	independently of the assessor experience. This paper proposes technical considerations for
223	future studies assessing muscle morphometry in neck pain populations.
224	
225	
226	
227	
228	Legend of Figures
229	Figure 1: (A) Ultrasound probe placement over the cervical multifidus at C4 level; (B)
230	Ultrasound image showing the superficial posterior neck muscles.
231	Figure 2: Cross sectional area (CSA) assessment of the cervical multifidus. Borders were
232	marked as follows: 1, inferior limit: internal echogenic fascia between cervical multifidus
233	and rotator muscle; 2, superior limit: echogenic fascia between cervical multifidus and
234	semispinalis; 3, medial limit: echogenic spinous process
235	
236	
237	
238	
239	

References

241	•	Abbott R, Pedler A, Sterling M, Hides J, Murphey T, Hoggarth M, Elliott J. The
242		geography of fatty infiltrates within the cervical multifidus and semispinalis cervicis
243		in individuals with chronic whiplash-associated disorders. J Orthop Sports Phys
244		Ther 2015; 45: 281-8.
245	•	Anderson JS, Hsu AW, Vasavada AN. Morphology, architecture, and biomechanics
246		of human cervical multifidus. Spine 2005; 30: E86-91.
247	•	De Pauw R, Coppieters I, Kregel J, De Meulemeester K, Danneels L, Cagnie B.
248		Does muscle morphology change in chronic neck pain patients? A systematic
249		review. Man Ther 2016; 22: 42-9.
250	•	Elliott JM, Walton DM, Rademaker A, Parrish TB. Quantification of cervical spine
251		muscle fat: a comparison between T1-weighted and multi-echo gradient echo
252		imaging using a variable projection algorithm (VARPRO). BMC Med Imaging
253		2013; 13: 30.
254	•	Elliott JM, Pedler AR, Jull GA, Van Wyk L, Galloway GG, O Leary SP.
255		Differential changes in muscle composition exist in traumatic and nontraumatic
256		neck pain. Spine 2014; 39: 39-47.
257	•	Fernández-de-las-Peñas C, Albert-Sanchís JC, Buil M, Benitez JC, Alburquerque-
258		Sendín F. Cross-sectional area of cervical multifidus in females with chronic
259		bilateral neck pain compared to controls. J Orthop Sports Phys Ther 2008; 38: 175-
260		80.

261	•	Kamper SJ, Rebbeck TJ, Maher CG, McAuley JH, Sterling M. Course and
262		prognostic factors of whiplash: a systematic review and meta-analysis. Pain 2008;
263		138: 617-629.
264	•	Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation
265		coefficients for reliability research. J Chiropr Med. 2016; 15: 155-63.
266	•	Javanshir K, Amiri M, Mohseni-Bandpei MA, Rezasoltani A, Fernández-de-las-
267		Peñas C. Ultrasonography of the cervical muscles: a critical review of the literature.
268		J Manipul Physiol Ther 2010; 33: 630-7.
269	•	Kristjansson E. Reliability of ultrasonography for the cervical multifidus muscle in
270		asymptomatic and symptomatic subjects. Man Ther 2004; 9: 83-8.
271	•	Lee JP, Tseng WY, Shau YW, Wang CL, Wang HK, Wang SF. Measurement of
272		segmental cervical multifidus contraction by ultrasonography in asymptomatic
273		adults. Man Ther 2007; 12: 286-294.
274	•	Lin YJ, Chai HM, Wang SF. Reliability of thickness measurements of the dorsal
275		muscles of the upper cervical spine: an ultrasonographic study. J Orthop Sports
276		Phys Ther 2009; 39: 850-7.
277	•	Owers DS, Perriman DM, Smith PN, Neeman T, Webb AL. Evidence for cervical
278		muscle morphometric changes on magnetic resonance images after whiplash: A
279		systematic review and meta-analysis. Injury 2018; 49: 165-176
280	•	Rankin G, Stokes M, Newham D. Size and shape of the posterior neck muscles
281		measured by ultrasound imaging: normal values in males and females of different
282		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

- Intra- and inter- examiner reliability of cervical multifidus imaging capturing and imaging calculation was excellent.
- 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-Calero¹ PT, PhD candidate; Sandra Sánchez-Jorge² PT, PhD; Javier Álvarez-González³ 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 PeñasTelephone number: + 34 91 488 88 84Facultad de Ciencias de la SaludFax number: + 34 91 488 89 57Universidad Rey Juan CarlosFax number: + 34 91 488 89 57Avenida de Atenas s/n28922 Alcorcón, Madrid, SPAINE-mail address: cesar.fernandez@urjc.esE-mail address: cesar.fernandez@urjc.es

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

5 Abstract

4

Background: Studies have analyzed muscle morphometry of cervical multifidus by using 6 ultrasound imaging, but its reliability is not clearly determined. Objective: To investigate 7 intra- and inter-rater reliability of imaging capturing (probe assessment/patient positioning) 8 and imaging calculation (scan assessment) of cervical multifidus cross sectional area (CSA) 9 by considering the assessor's experience in asymptomatic individuals. **Design:** Reliability 10 study. Methods: The CSA of C4/C5 cervical multifidus was assessed in 16 asymptomatic 11 subjects. Two examiners performed the imaging capture and also repeated the procedure 12 (probe placement/patient positioning) twice with a 10-min period between each. Other two 13 raters conducted imaging calculations of CSA. Intra-examiner imaging capturing reliability, 14 each rater (experienced and novice) calculated multifidus CSA of both images obtained by 15 each examiner. Inter-examiner imaging capturing reliability, each rater calculated the CSA 16 17 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 18 correlation coefficients (ICC) and standard error of measurement (SEM) were calculated. 19 **Results**: Intra- (ICC_{3.1} 0.988-0.996, SEM 0.3%-0.7%) and inter- (ICC_{3.2} 0.958-0.965, SEM 20 2.6%-3.2%) examiner reliability of imaging capturing was excellent. Intra- and inter-rater 21 reliability of imaging calculation was also excellent for both raters (experienced/novice). 22 No significant differences between experienced or novice examiners or testers were found. 23

24	Conclusions: This study found that intra- and inter-examiner/rater reliability of imaging
25	capturing (probe assessment/patient positioning) and imaging calculation (scan assessment)
26	of the cervical multifidus CSA at C4/C5 level was excellent in asymptomatic subjects.
27	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

33 Introduction

34 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, 35 up to 50% of individuals continue reporting symptoms after 1-year (Kamper et al, 2008). 36 37 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 38 performance. Several studies have reported that people with WAD show increased fatty 39 infiltrate and changes in cross sectional area (CSA) in the posterior cervical muscles, 40 specifically the multifidi (Abbott et al, 2015; Elliot et al, 2014; Snodgrass et al, 2019a). 41 However, most studies have used magnetic resonance imaging (MRI) to assess muscle 42 morphology (Owers et al, 2018), which is not readily available in clinical practice. 43

A more pragmatic method of measuring muscle morphology is ultrasonography; but 44 45 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 46 the posterior neck muscles (Javanshir et al, 2010). Kristjansson (2004) analyzed the 47 reliability of assessing CSA of the C4 cervical multifidus and reported appropriate intra-48 and inter-tester agreement in 10 asymptomatic subjects, and acceptable intra-, but 49 questionable inter-tester, agreement in 10 symptomatic subjects. Fernández-de-las-Peñas et 50 al (2008) reported excellent between-scan and good between-day assessment in individuals 51 with mechanical neck pain. Lee et al (2007) found that ultrasound imaging was as reliable 52 and valid as MRI for assessing cervical multifidus thickness after isometric contractions. 53

An important step before neck muscle composition can routinely be used in research 54 55 or clinical practice is establishing reliability of imaging calculation (scan assessment) but also reliability of imaging capturing (patient positioning). This is particularly relevant since 56 ultrasonography is operator-dependent and the measurement protocol could influence the 57 imaging calculation. In fact, previous studies had mostly investigated the reliability of 58 imaging calculation (scan assessment) but not the reliability of imaging capturing (probe 59 60 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-61 examiner/rater reliability of imaging calculation (scan/image assessment) and imaging 62 63 capturing (probe assessment/patient positioning) of cervical multifidus CSA considering assessor experience 64

65

66 Methods

67 **Participants**

68 Asymptomatic volunteers without neck pain symptoms were recruited via local announcements between December 2018 and June 2019. To be eligible to participate, they 69 had to be between 18 and 45 years old and with no history of neck pain the previous year. 70 71 Exclusion criteria included history of whiplash injury; any pharmacological treatment affecting muscle tone, e.g., muscle relaxants, analgesics; prior history of cervical surgery; 72 cervical radiculopathy or myelopathy; presence of degenerative changes; and any medical 73 condition such as tumor or fracture. The study was approved by the Institutional Review 74 Ethical Committee of XX. All subjects signed the written informed consent prior to their 75 76 inclusion.

77 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 90° abduction and the elbows flexed to 90°. 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**).

91 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**).

97

- 98
- 99
- 100

101 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.

112 **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.

- 123
- 124

125 Statistical Analysis

Statistical analysis was performed using the SPSS package, Version 21 software for 126 Mac OS. Normal distribution of the data was verified by using the Shapiro-Wilk test. Intra-127 and inter- examiner/rater reliability was estimated using 2-way mixed-model, consistency-128 type intra class correlation coefficients (ICC). Reliability was classified as fair (ICC<0.50), 129 moderate (0.50<ICC<0.75), good (0.75<ICC<0.90) or excellent (ICC>0.90) (Koo and Li 130 2016). Standard error of measurement (SEM) was calculated using the resulting ICC values 131 and standard deviation (SD): SEM (%) = (SDx $\sqrt{1-ICC}$) x 100 to assess measurement precision. 132 All tests were two-tailed with p-values <0.05 considered significant. 133

134

135 **Results**

From a total of 20 subjects responding to the announcement, 4 were excluded due to 136 previous whiplash injury (n=2) and history of neck pain the previous year (n=2). Sixteen 137 asymptomatic subjects (50% male) were finally included (total 64 images, n=4 per subject). 138 139
 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 140 weight (r:0.473, P=0.006), height (r:0.385; P=0.03), and BMI (r:0.481, P=0.005) was found. 141 142 **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 143 from 0.3% to 0.7%. No difference between experienced/novice examiner/testers was found. 144 Inter-examiner reliability was also excellent for both experienced (ICC_{3,2} 0.965) and novice 145 (ICC_{3.2} 0.958) raters, with SEM of 2.6% and 3.2%, respectively. 146

147 Reliability data of imaging calculation of both raters is shown on Table 3. Again,
148 intra- and inter-rater reliability was excellent for both raters but with smaller SEM for intra149 rater reliability.

150

151 **Discussion**

This is the first study assessing reliability of both imaging capturing (positioning) 152 and imaging calculation (assessment) of the neck muscles considering the experience of the 153 assessor. In general, intra- and inter- rater reliability of imaging capturing and calculation 154 of cervical multifidus CSA was excellent when applied by an experienced or novice 155 assessor in asymptomatic individuals. Our findings are similar to those previously found in 156 patients with mechanical neck pain (Fernández-de-las-Peñas et al. 2008) and slightly better 157 than those previously reported by Kristiansson (2004). In addition, reliability data of 158 ultrasound imaging assessment was slightly superior to reliability data obtained for MRI. 159 Snodgrass et al (2019b) reported good to excellent intra- (ICC 0.78-0.96), but fair to good 160 161 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 162 examiner/rater were observed. This is a relevant topic since ultrasound imaging assessment 163 is operator-dependent. Two points are relevant for ultrasound assessment, probe angulation 164 and pressure. Whittaker et al (2009) showed that probe angulation of less than approx. 10 165 degrees, as it is commonly done in clinical practice, do not distort measurements of tissue 166 thickness. No data about probe pressure is available. It is possible that strict positioning and 167 measurement protocols followed in our study could explain high reliability values. 168

This study can help for developing specific protocols for image capturing and calculation 169 170 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 171 proper visualization of the fascial layers dividing the cervical multifidus from surrounding 172 muscles (Fernández-de-las-Peñas et al, 2008; Kristjansson, 2004; Lee et al, 2007). One 173 potential reason for this lack of clarity of muscular fascial layers could be the presence of 174 175 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 176 177 in previous studies (Fernández-de-las-Peñas et al, 2008; Kristjansson, 2004; Lee et al, 178 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 179 a scenario, fascial layers as border contours for determining muscle morphology may have 180 highly relevance. Since the current study included asymptomatic individuals, fascial layers 181 surrounding the multifidus were properly identified (Fig. 2). Rankin et al (2005) were not 182 able to consistently identify fascial divisions between multifidus/semispinalis and between 183 multifidus/rotators; therefore, they provided normative data of the whole muscle group and 184 not from individual muscles. It is also possible that technical improvements of ultrasound 185 186 imaging equipment, offering much better-definition and quality images, can also influence proper visualization of fascial layers between muscles. 187

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 190 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.

199 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 200 or WAD. Second, our sample was small. Therefore, our results should not be considered as 201 202 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 203 204 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 205 reliability should ideally be assessed on different days in future studies. Finally, it is also 206 important to consider that fatty infiltration quantification is not possible with ultrasound 207 imaging and is currently possible with MRI, since MRI calculates the amount of fat by 208 differentiating fat and soft-aqueous tissue signal intensities (Elliott et al, 2013). Therefore, 209 RUSI should be only used for determining muscle morphology, e.g., CSA, size, or muscle 210 211 function, but not intramuscular quality, e.g., fat, or fibers.

- 212
- 213
- 214
- 215

216

10

218 Conclusions

219	We found that ultrasound assessment of cervical multifidus at C4/C5 level is highly
220	reliable for evaluating CSA in asymptomatic people since imaging capturing and imaging
221	calculation exhibited excellent intra- and inter- examiner reliability. Reliability was similar
222	independently of the assessor experience. This paper proposes technical considerations for
223	future studies assessing muscle morphometry in neck pain populations.
224	
225	
226	
227	
228	Legend of Figures
229	Figure 1: (A) Ultrasound probe placement over the cervical multifidus at C4 level; (B)
230	Ultrasound image showing the superficial posterior neck muscles.
231	Figure 2: Cross sectional area (CSA) assessment of the cervical multifidus. Borders were
232	marked as follows: 1, inferior limit: internal echogenic fascia between cervical multifidus
233	and rotator muscle; 2, superior limit: echogenic fascia between cervical multifidus and
234	semispinalis; 3, medial limit: echogenic spinous process
235	
236	
237	
238	
239	

References

241	•	Abbott R, Pedler A, Sterling M, Hides J, Murphey T, Hoggarth M, Elliott J. The
242		geography of fatty infiltrates within the cervical multifidus and semispinalis cervicis
243		in individuals with chronic whiplash-associated disorders. J Orthop Sports Phys
244		Ther 2015; 45: 281-8.
245	•	Anderson JS, Hsu AW, Vasavada AN. Morphology, architecture, and biomechanics
246		of human cervical multifidus. Spine 2005; 30: E86-91.
247	•	De Pauw R, Coppieters I, Kregel J, De Meulemeester K, Danneels L, Cagnie B.
248		Does muscle morphology change in chronic neck pain patients? A systematic
249		review. Man Ther 2016; 22: 42-9.
250	•	Elliott JM, Walton DM, Rademaker A, Parrish TB. Quantification of cervical spine
251		muscle fat: a comparison between T1-weighted and multi-echo gradient echo
252		imaging using a variable projection algorithm (VARPRO). BMC Med Imaging
253		2013; 13: 30.
254	•	Elliott JM, Pedler AR, Jull GA, Van Wyk L, Galloway GG, O Leary SP.
255		Differential changes in muscle composition exist in traumatic and nontraumatic
256		neck pain. Spine 2014; 39: 39-47.
257	•	Fernández-de-las-Peñas C, Albert-Sanchís JC, Buil M, Benitez JC, Alburquerque-
258		Sendín F. Cross-sectional area of cervical multifidus in females with chronic
259		bilateral neck pain compared to controls. J Orthop Sports Phys Ther 2008; 38: 175-
260		80.

261	•	Kamper SJ, Rebbeck TJ, Maher CG, McAuley JH, Sterling M. Course and
262		prognostic factors of whiplash: a systematic review and meta-analysis. Pain 2008;
263		138: 617-629.
264	•	Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation
265		coefficients for reliability research. J Chiropr Med. 2016; 15: 155-63.
266	•	Javanshir K, Amiri M, Mohseni-Bandpei MA, Rezasoltani A, Fernández-de-las-
267		Peñas C. Ultrasonography of the cervical muscles: a critical review of the literature.
268		J Manipul Physiol Ther 2010; 33: 630-7.
269	•	Kristjansson E. Reliability of ultrasonography for the cervical multifidus muscle in
270		asymptomatic and symptomatic subjects. Man Ther 2004; 9: 83-8.
271	•	Lee JP, Tseng WY, Shau YW, Wang CL, Wang HK, Wang SF. Measurement of
272		segmental cervical multifidus contraction by ultrasonography in asymptomatic
273		adults. Man Ther 2007; 12: 286-294.
274	•	Lin YJ, Chai HM, Wang SF. Reliability of thickness measurements of the dorsal
275		muscles of the upper cervical spine: an ultrasonographic study. J Orthop Sports
276		Phys Ther 2009; 39: 850-7.
277	•	Owers DS, Perriman DM, Smith PN, Neeman T, Webb AL. Evidence for cervical
278		muscle morphometric changes on magnetic resonance images after whiplash: A
279		systematic review and meta-analysis. Injury 2018; 49: 165-176
280	•	Rankin G, Stokes M, Newham D. Size and shape of the posterior neck muscles
281		measured by ultrasound imaging: normal values in males and females of different
282		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







	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)

Table 1: Demographic Features of Participants by Gender

Values are expressed as means (SD)

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

	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 (<mark>ICC_{3,1}) Reliability (novice)</mark>					
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%		

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

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)	<mark>0.18</mark>	<mark>4.4%</mark>
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