AUTHOR QUERY FORM

ELSEVIER	**Journal: YMMT** **Article Number: 1577**	**Please e-mail your responses and any corrections to:** **E-mail: d.poulson@elsevier.com**

Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list. Note: if you opt to annotate the file with software other than Adobe Reader then please also highlight the appropriate place in the PDF file. To ensure fast publication of your paper please return your corrections within 48 hours.

For correction or revision of any artwork, please consult http://www.elsevier.com/artworkinstructions.

We were unable to process your file(s) fully electronically and have proceeded by

☐ Scanning (parts of) your article ☐ Rekeying (parts of) your article ☐ Scanning the artwork

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof. Click on the 'Q' link to go to the location in the proof.

Location in article	Query / Remark: click on the Q link to go Please insert your reply or correction at the corresponding line in the proof
Q1	In Tables 1 and 2, change of erected to erect OK?
Q2	The author names have been tagged as given names and surnames (surnames are highlighted in teal color). Please confirm if they have been identified correctly.
Q3	Please note that Table 1B was renumbered as Table 2 and all subsequent tables were renumbered sequentially.
Q4	Please confirm Fabio Gallo's degree "MSc(Biostat)" is correct.
Q5	Please distinguish if "F.G." refers to Fabio Gallo or Federico Giaretta.
Q6	Please distinguish whether "F.G." refers to Fabio Gallo or Federico Giaretta.

Please check this box if you have no corrections to make to the PDF file. ☐

Thank you for your assistance.

Intrarater and Inter-rater Reliability of Active Cervical Range of Motion in Patients With Nonspecific Neck Pain Measured With Technological and Common Use Devices: A Systematic Review With Meta-regression

Angie Rondoni, PT, OMPT,[a] Giacoma Rossettini, PT, OMPT, MSc,[a] Diego Ristori, PT, OMPT, MSc,[a] Fabio Gallo, MSc(Biostat),[b] Marco Strobe, PT, OMPT,[a] Federico Giaretta, PT,[c] Andrea Battistin, PT, OMPT,[d] and Marco Testa, PT, OMPT, PhD[a]

Abstract

Objectives: The purpose of this systematic review was to compare intrarater and inter-rater reliability of active cervical range of motion (ACROM) measures obtained with technological devices to those assessed with low-cost devices in patients with nonspecific neck pain. As a secondary outcome, we investigated if ACROM reliability is influenced by the plane of the assessed movement.

Methods: Medline, Scopus, Embase, the Cochrane Library, CINHAL, PEDro, and gray literature were searched until August 2016. Inclusion criteria were reliability design, population of adults with nonspecific neck pain, examiners of any level of experience, measures repeated at least twice, and statistical indexes on reliability. A device was considered inexpensive if it cost less than €500. The risk of bias of included studies was assessed by Quality Appraisal of Reliability Studies.

Results: The search yielded 35 151 records. Nine studies met all eligibility criteria. Their Quality Appraisal of Reliability Studies mean score was 3.7 of 11. No significant effect of the type of device (inexpensive vs expensive) on intraclass correlation coefficient (ICC) was identified for intrarater (ICC = 0.93 vs 0.91; $P > .99$) and inter-rater reliability (ICC = 0.80 vs 0.87; $P > .99$). The plane of movement did not affect inter-rater reliability ($P = .11$). Significant influences were identified with intrarater reliability ($P = .0001$) of inexpensive devices, where intrarater reliability decreased ($P = .01$) in side bending, compared with flexion-extension.

Conclusions: The use of expensive devices to measure ACROM in adults with nonspecific neck pain does not seem to improve the reliability of the assessment. Side bending had a lower level of intrarater reliability. (J Manipulative Physiol Ther 2017;xx:1-12)

Key Indexing Terms: *Neck Pain; Range of Motion, Articular; Validation Study as Topic; Cost Control*

Introduction

Neck pain (ie, pain in the neck with or without pain referred into 1 or both upper limbs that lasts at least 1 day) is a common complaint in the global population.[1] It was estimated that about the 5% of population at any time suffers from neck pain.[1] Objective evaluation of active range of motion, for neck pain as much as for other condition, is a cornerstone of clinical assessment,[2,3] just like history taking, visual inspection, and passive motion examination.[4]

Changes in active cervical range of motion (ACROM) are considered adequate indicators for treatment effect[3] and

[a] Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genova—Campus of Savona, Savona, Italy.
[b] Department of Health Sciences, University of Genova, Genova, Italy.
[c] Private practice, Verona, Italy.
[d] Private practice, Vicenza, Italy.
Corresponding author: Marco Testa, PT, OMPT, PhD, Campus Universitario di Savona, Palazzina Branca, Via Magliotto 2 – 17100, Savona, Italy. Tel.: +39 3289213515. (e-mail: *marco.testa@unige.it*).
Paper submitted February 19, 2017; in revised form July 7, 2017; accepted July 20, 2017.
0161-4754

https://doi.org/10.1016/j.jmpt.2017.07.002

prognosis for nonspecific neck pain[3] but not useful for diagnosing the condition. However, results from previous systematic reviews on reliability of ACROM measurement indicate conflicting conclusions because of potential selection biases. In former studies, an asymptomatic sample was sometimes used to test ACROM, and reviews pooled data from this population with a symptomatic one.[3,5-7] Moreover, data were pooled from patients with multiple diagnoses, mixing results of measures in specific neck pain (ie, neck pain originating from systemic conditions like rheumatic diseases or identified causes like radiculopathy) with nonspecific ones.[3,5-7] In the end, previous reviews did not perform an adequate assessment of methodological quality of identified studies: In some cases there was a complete lack of assessment[5,6]; in others the assessment was performed by tools that were not validated.[2,7] To avoid such biases, it was decided to create this systematic review with stricter criteria in the selection procedure.

Neck pain is highly disabling and demands direct and indirect costs (eg, public or private health costs, insurance refunds, working days lost). Some studies have estimated that a patient with nonspecific back and neck pain will spend about $5500 per year,[8,9] with a trend of increasing costs,[8,9] mainly because of medical specialty costs, possibly related to innovative contents and collaborative markets with producers of supplies.[10]

To contain costs for spine pain management, it is important to evaluate if devices are reliable and cost-worthy. Therefore, the purpose of this systematic review was to compare intrarater and inter-rater reliability of ACROM measures from technological devices to those assessed with low-cost devices in patients with nonspecific neck pain. As a secondary outcome, we investigated if ACROM reliability is influenced by the plane of the assessed movement.

Methods

Search Strategy

This systematic review was written in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement for reporting systematic reviews and meta-analyses of studies,[10] without registering a protocol review. A systematic search was performed in 6 electronic databases (Medline, Scopus, Embase, the Cochrane Library, CINHAL, and PEDro) from each databases' inception until August 2016; in addition, a gray literature search was performed on articular mobility textbooks, with no limits about year of publication.

A search strategy was built using keywords combined with Boolean operators. Relevant hand-searched articles, found in textbooks, were also included to obtain as complete information as possible. The search strategies are reported in Appendix 1.

A hand search of the reference lists of the articles screened for inclusion was also performed to locate any publications not identified through the electronic database searches.

Eligibility Criteria

Two authors (M.S., F.G.) independently reviewed the articles obtained by the systematic search for eligibility and possible inclusion. Titles and abstracts of all articles were screened for eligibility, based on the criteria listed next. In case of uncertain eligibility, all reviewers screened the full text of the manuscript for inclusion into the systematic review.

Inclusion Criteria. Studies were included only if intrarater or inter-rater reliability design was adopted. Publications in any language as full-text articles and peer review were included. Studies based on participants with nonspecific neck pain[11] were included for the review. Examiners with various levels of experience and education were included for review, and no restrictions were made based on their demographics. Studies were included if the measurement of ACROM was performed at least twice (by the same rater or different raters). The studies were included if they provided statistics about reliability of measurements such as intraclass correlation (ICC), standard error of measurement, and limits of agreement.

Exclusion Criteria. Types of studies excluded were letters, editorials, comments, case studies, protocols, guidelines, conference proceedings, review articles, and those whose full text was not available. Also excluded were studies with asymptomatic participants or with mixed populations (healthy and symptomatic) where data were pooled together without any distinction and those involving participants with other pathologic conditions different from nonspecific neck pain. Students were not included as raters. Studies where the measurement of ACROM was performed only once by a single rater or was not performed were not included. Studies were excluded if they did not provide statistics about reliability of measurements.

Quality Assessment

The Quality Appraisal of Reliability Studies (QAREL) checklist evaluated the risk of bias.[12] It has been reported to have acceptable levels of content validity (good) and inter-rater reliability ($\kappa > 0.60$).[13] It has been used in systematic reviews aimed at reliability of clinical tests in rehabilitation.[14-16] In this systematic review, a QAREL checklist was adopted to assess methodological quality both across studies and within studies.

A QAREL is composed of 11 items and assesses the external validity, internal validity, and statistical methods of reliability studies. Based on guidelines provided, each item is equally weighted and scored as *Yes, No, Unclear,* or *Not Applicable*. Former systematic reviews of inter-rater and intrarater reliability based on QAREL scores have used

50%,[16-18] 60%,[15,17,19] and 70%[17] cutoff points to define high-quality evidence. Because operational definitions of study quality can affect the overall levels of evidence[10] and no consensus on a single cutoff point for defining study quality exists, we interpreted QAREL scores using the mean proposed value of 60%. Studies were therefore defined as high quality if 60% of applicable QAREL checklist items were scored as Yes.

As recommended to improve reliability, criteria by which judgments were made for each item of QAREL were defined before the methodological assessment was done.[13] Two independent researchers (G.R., A.R.) evaluated independently the quality of reliability studies and disagreement was resolved by discussion or use of a third reviewer (M.T.). The agreement between the 2 reviewers for rating studies based on quality scales was calculated by κ coefficient.

Data Extraction

Two independent reviewers (M.S., A.B.) extracted data, including authors' names, year of publication, mean age of patient enrolment, type of instrument used, movement of ACROM measured, ICC, standard error of the mean, and limits of agreement. Discrepancies in interpretation were resolved by discussion seeking consensus and use of a third reviewer (M.T.) if needed. Data are shown in Tables 1 and 2.

Data Analysis and Synthesis

Continuous variables are given as means and standard deviation (SD) and categorical variables as number of participants. The total flexion and extension measure (if not available) was obtained by taking the mean of flexion and extension values in each study. The total rotation and side-bending measure (if not available) was obtained by taking the mean of left and right values for rotation and side bending in each study. To investigate the associations of the ICC with age, ACROM, and tool type, a univariate, and later a multivariate, analysis was performed by using the linear mixed effects model. In addition, the linear mixed effects model, corrected for age effect, was again used to study which direction of movement was more reproducible between inexpensive tools. Studies were considered as a random effect in all mixed effects models.

After fitting each model, residual analyses were performed to check for adequacy of the mixed model assumptions. In particular, the Shapiro-Wilk test and the quantile-quintile plot of the standardized residuals were used to check for normality assumptions of the error terms. The standardized residual vs predicted mean values plot was used to check for the assumption of constant variance. The plot of the fitted values obtained from the model of interest against the observed values was used to assess the linearity of the model.

The likelihood ratio test was used as a test of statistical significance. Differences with a P value < .05 were selected as significant in all the comparisons. Data were acquired and analyzed in R Version 3.2.3 software.[20]

Devices used to measure active cervical range of motion were divided into expensive and inexpensive tools. This division was made to evaluate if complex and technological devices are cost-worthy because of their superior reliability. A former study was found that arbitrarily established a cutoff to determine whether a device was considered affordable or not at maximum €1000.[2] However, in this study a device was considered inexpensive if it cost a maximum of €500, so that it could be bought by self-employed physical therapists without great economic effort, and not only by research laboratories.

Results

Search Results

Searching the databases yielded 35 151 citation postings. Research strategies on each database are detailed in Appendix 1. Two reviewers independently reviewed the titles and abstracts, and 35 114 articles were eliminated. There was no agreement on the selection of 60 studies. Disagreements were resolved by a consensus-based discussion,[21] with a strong level of agreement (percentage of agreement = 99.3%)[21] and high inter-reviewer reliability (κ coefficient k_w = 0.96; 95% CI: 0.95-0.97). A total of 37 articles were regarded as possibly relevant and were retrieved as full articles. Nine studies met all eligibility criteria. Excluded studies and motivation of exclusion are shown in Appendix 2. There was no disagreement in the procedures of application of eligibility criteria and data extraction. The selection process is shown in Figure 1.

Characteristics of the Studies

Two of the selected studies investigated both inter-rater and intrarater reliability.[22,23] The others assessed only 1 value.[24-29]

The number of participants with neck pain ranged from 19[27] to 56 per study, and the sum of them was 288. The mean age of the participants had a minimum value of 33.6 years (SD = 10.3)[26] and a maximum value of 59 years (SD = 5.6).[29]

Six studies used inexpensive devices to measure ACROM: universal inclinometer,[23] standard dual-arm goniometer,[23,28] gravity inclinometer,[27,28,30] and cervical range of motion device.[26,29] Only 3 studies used expensive devices: the Cybex Electronic Digital Inclinometer 320 (EDI-320),[22] the Orthopedic Systems Incorporated (OSI) Computerized Anatometry 6000 Spine Motion Analyzer (SMA),[24] and the Flock-of-Birds system.[25] Only 6 studies reported the mean Neck Disability Index value[22,23,27-30]; only 3 of them reported the mean duration of symptoms,[22,23,28] so no subgroup analysis could be performed to identify differences between patients with

Table 1. *Intrarater Reliability*

Study	Participants	Examiners	Instrument	Type	Position	Direction	ICC	SEM	LoA
Dunleavy et al[29] (2013)	n = 36 Diagnosis = chronic neck pain Age = 59 (SD 5.6) M/F = not declared NDI = 12.6 ± 6	n = 5 examiners (not specified) Experience = not declared Training = yes (15 h)	CROM device	IE	Sit erect posture	Flexion Extension Left side bending Right side bending Left rotation Right rotation	0.93 0.96 0.88 0.92 0.96 0.94	3.5 2.7 2.8 2.1 2.8 2.8	
Fletcher and Bandy[26] (2008)	n = 22 Diagnosis = neck pain Age = 33.6 (SD 10.3) M/F = 7/15 NDI = not applicable (also n = 25 asymptomatic participants, not included in meta-regression)	n = 1 (physical therapist) Experience ≥10 y Training = not declared	CROM device	IE	Sit erect posture	Flexion Extension Left side bending Right side bending Left rotation Right rotation	0.88 (0.73-0.95) 0.92 (0.80-0.97) 0.89 (0.76-0.95) 0.93 (0.83-0.97) 0.96 (0.91-0.98) 0.92 (0.81-0.97)	4.1 3.0 3.9 2.5 2.9 3.3	
Hoving et al[22] (2005)	n = 32 Diagnosis = neck pain Age = 45 (SD 9.2) M/F = 12/20 NDI = 15.2 ± 8.3	n = 2 physical therapists Experience = not declared Training = yes (once a wk for 3 mo, plus a trial on 5 healthy participants before this study)	EDI-320 inclinometer	E	Sit erect posture	Total flexion-extension Total side bending Supine Total rotation	Rater a: 0.96 (0.93-0.98) Rater b: 0.97 (0.93-0.98) Rater a: 0.93 (0.86-0.97) Rater b: 0.93 (0.86-0.96) Rater a: 0.96 (0.91 – 0.98) Rater b: 0.96 (0.92-0.98)		–2.5 ± 11.1 1.0 ± 11.1 –0.1 ± 10.4 –0.6 ± 9.8 –5.9 ± 13.5 –2.7 ± 14.4
Petersen et al[24] (2000)	n = 20 Diagnosis = neck pain Age = 40.2 (SD 8.7) M/F = 7/13 NDI = not applicable (also n = 30 asymptomatic participants, not included in meta-regression)	n = 1 examiner (not specified) Experience ≥14 y Training = yes (4 h)	OSI SMA instrument	E	Sit erect posture	Flexion Extension Left side bending Right side bending Left rotation Right rotation	0.68 0.87 0.96 0.92 0.88 0.94	3.92 2.66 1.85 2.35 2.86 1.90	
Schneider et al[23] (2013)	n = 56 Diagnosis = chronic neck pain Age = 46 (range 21-64) M/F = 19/37 NDI = 19 (range 8-18)	n = 2 physical therapists Experience = 12-16 y Training = yes (1 h)	Universal inclinometer Standard dual-armed goniometer	IE	Sit erect posture	Flexion Extension Left side bending Right side bending Left rotation Right rotation	0.94 (0.90-0.96) 0.95 (0.91-0.97) 0.93 (0.88-0.96) 0.91 (0.85-0.95) 0.97 (0.95-0.98) 0.95 (0.92-0.97)	2.6 2.9 2.4 2.6 2.5 3.2	–11.9 to 9.9 –11.3 to 9.4 –11.7 to 10.0 –9.8 to 7.4 –12.1 to 10.4 –12.1 to 10.4

Study: first author and year of publication; participants: number (n), diagnosis, mean age (with SD or range); examiners: number (n), work experience, training in using the instrument to assess active range of motion; instrument: name of instrument to assess active range of motion; position: position of the patient; direction: direction of measured active movement.
CROM, cervical range of motion; *E,* expensive instrument to assess active range of motion; *EDI,* Electronic Digital Inclinometer; *ICC,* intraclass correlation coefficient, with SD in parentheses, if provided; *IE,* inexpensive instrument to assess active range of motion; *LoA,* limits of agreement; *M/F,* male/female; *NDI,* Neck Disability Index; *OSI SMA,* Orthopedic Systems Incorporated Spine Motion Analyzer; *SD,* standard deviation; *SEM,* standard error of the mean.

Table 2. *Inter-rater Reliability*

Study	Participants	Examiners	Instrument	Type	Position	Direction	ICC	SEM	LoA
Assink et al[25] (2005)	n = 30 Diagnosis: neck pain Age = 51.1 ± 9.5 M/F = 13/17 NDI = not applicable (also n = 30 asymptomatic participants, not included in meta-regression)	n = 2 examiners (not specified) Experience = not declared Training = yes (20 h)	Flock-of-Birds system	E	Sit erect posture	Total flexion-extension Total side bending Total rotation	0.91 (0.83-0.96) 0.76 (0.56-0.88) 0.77 (0.58-0.88)		–12.4 to 9.6 –17.8 to 17.4 –18.6 to 21.8
Cleland et al[28] (2006)	n = 22 Diagnosis = neck pain Age = 41.0 ± 12.9 M/F = 4/18 NDI = 30.9 ± 10.5	n = 4 physical therapists Experience = 3-23 y (12.3 ± 10) Training = yes (manual, videos then 1-h training session)	Gravity inclinometer	IE	Sit erect posture	Flexion Extension Left side bending Right side bending	0.75 (0.50-0.89) 0.74 (0.48-0.88) 0.66 (0.33-0.84) 0.69 (0.40–0.86)	6.8 4.7 3.6 7	0.76 ± 16.6 1.6 ± 11.7 –0.5 ± 9.5 2.2 ± 18.3
			Standard dual-armed goniometer	IE	Sit erect posture	Left rotation Right rotation	0.78 (0.55–0.90) 0.77 (0.52–0.90)	5 5.5	1.6 ± 13.5 –0.3 ± 15.9
Hoving et al[22] (2005)	n = 32 Diagnosis = neck pain Age = 45 (SD 9.2) M/F = 12/20 NDI = 15.2 ± 8.3	n = 2 physical therapists Experience = not declared Training = yes (once a wk for 3 mo, plus a trial on 5 healthy participants before this study)	EDI-320 inclinometer	E	Sit erect posture Supine	Total flexion-extension Total side bending Total rotation	0.95 (0.90-0.98) 0.89 (0.77-0.94) 0.95 (0.90-0.98)		3.3 ± 17.0 0.5 ± 17.0 –1.3 ± 24.6
Piva et al[30] (2006)	n = 30 Diagnosis = neck pain Age = 41.0 ± 12.0 M/F = 12/18 NDI = 24.3 ± 14.8	n = 2 examiners (trained in manual therapy, not further specified) Experience = 2-10 y Training = not declared	Gravity inclinometer	IE	Sit erect posture Supine	Flexion Extension Left side bending Right side bending Left rotation Right rotation	0.86 (0.73-0.93) 0.78 (0.59-0.89) 0.85 (0.70-0.92) 0.87 (0.75 -0.94) 0.91 (0.82-0.96) 0.86 (0.74–0.93)	5.6 5.8 4.2 3.7 4.1 4.8	
Shahidi et al[27] (2012)	n = 19 Diagnosis = chronic neck pain Age = 34.9 ± 9.9 M/F = 10/9 NDI = 14.4 ± 7.3 (also n = 20 asymptomatic participants, not included in meta-regression)	n = 2 physical therapist Experience = 2-23 y Training = yes (3 1-h sessions)	Gravity inclinometer	IE	Sit erect posture Supine	Flexion Extension Left side bending Right side bending Left rotation Right rotation	0.69 (0.36-0.87) 0.78 (0.50-0.91) 0.68 (0.34-0.87); 0.47 (0.06-0.75) 0.70 (0.37-0.87) 0.51 (0.09-0.78)		
Schneider et al[23] (2013)	n = 56 Diagnosis = chronic neck pain Age = 46 (range 21-64) M/F = 19/37 NDI = 19 (range 8-18)	n = 2 physical therapists Experience = 12-16 y Training = yes (1 h)	Universal inclinometer	IE	Sit erect posture	Flexion Extension Left side bending Right side bending	0.93 (0.86-0.96) 0.95 (0.91-0.97) 0.90 (0.68-0.96) 0.90 (0.83-0.94)	3.4 3.1 3.2 3.3	–10.9 to 6.9 –9.4 to 8.7 –10.4 to 4.8 –10.4 to 7.5
			Standard dual-armed goniometer	IE	Sit erect posture	Left rotation Right rotation	0.95 (0.71-0.98) 0.95 (0.91-0.97)	3.4 3.3	–4.2 to 11.1 –7.7 to 10.6

Study: first author and year of publication; participants: number (n), diagnosis, mean age (with SD or range); examiners: number (n), work experience, training in using the instrument to assess active range of motion; instrument: name of instrument to assess active range of motion; position: position of the patient; direction: direction of measured active movement.

E, expensive instrument to assess active range of motion; *EDI,* Electronic Digital Inclinometer; *ICC,* intraclass correlation coefficient, with SD in parentheses, if provided; *IE,* inexpensive instrument to assess active range of motion; *LoA,* limits of agreement; *M/F,* male/female; *NDI,* Neck Disability Index; *SD,* standard deviation; *SEM,* standard error of the mean.

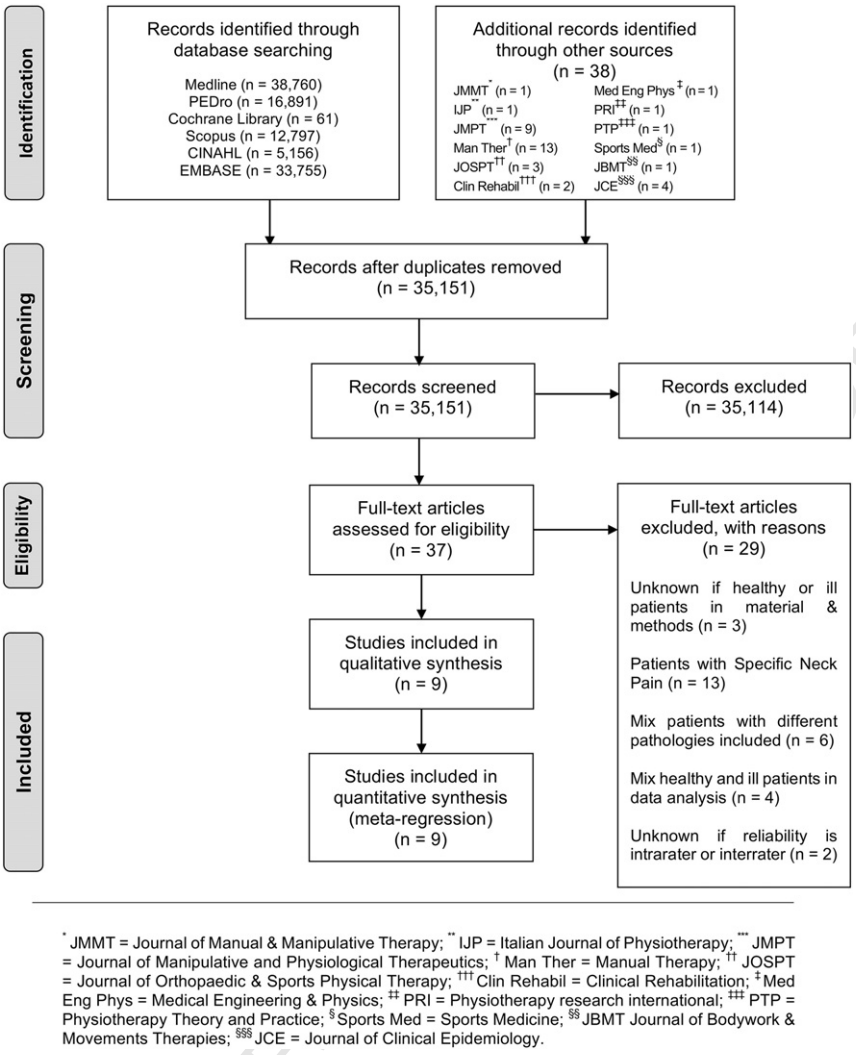

Fig 1. *Flow diagram of the phases of the selection of studies according to PRISMA.*

different disability and between patients with acute, subacute, or chronic neck pain.

All the data are reported in Tables 1 and 2.

Quality Appraisal

Methodological quality of included studies was assessed with the QAREL tool. Across studies, the mean score reached by studies was 3.7 of 11, with a maximum score of 7 of 11[23] and a minimum score of 2 of 11.[24] A complete overview of methodological quality across studies is provided in Table 3.

Regarding risk of bias within studies, internal validity was compromised because populations and raters were largely not sufficiently described or not representative. Only 1 study provided a representative population,[29] and only 1 reported representative raters.[28] All 9 studies instead applied and interpreted appropriately the test.[27] External validity was also lacking, because blinding and practical organization of the test were not clearly reported in most cases. Only a couple of studies stated that raters were blinded to clinical information on patients,[23,27] with only 1 specifically declared blinding to the reference standard for the target disorder,[23] and only 1 did not provide additional cues that were not part of the test. However, blinding of results from 1 rater to other raters, when appropriate, was mainly ensured—6 of 8 studies were scored Yes.[22,23,25,27,28,30] Raters were also blinded to their own findings in most cases, when appropriate, with 3 of 5 studies reporting that.[22,23,26] None reported if the examination order was varied, and only 2 studies ensured an appropriate time interval between 2 measurements.[23,27] Obviously, every study adopted an appropriate statistical measurement of agreement because this item of the QAREL checklist was an inclusion criterion of studies in the present review. A graphical representation of risk of bias within studies is provided in Figure 2.

t3.2 **Table 3.** *Quality Appraisal*

t3.3	Study	External Validity			Internal Validity						Statistical Analysis	
t3.4		Item 1: Representative Sample	Item 2: Representative Raters	Item 10: Appropriate Test	Item 3: Blinding (Other Raters)	Item 4: Blinding (Own Findings)	Item 5: Blinding (Reference/Disease)	Item 6: Blinding (Clinical Information)	Item 7: Blinding (Additional Cues)	Item 8: Examination Order Varied	Item 9: Appropriate Time Interval	Item 11: Appropriate Statistics	Percentage of Yes (%)
t3.5	Assink et al[25] (2005)	N	U	Y	Y	N/A	N	U	U	U	N	Y	30
t3.6	Cleland et al[28] (2006)	U	Y	Y	Y	N/A	N	N	N	U	N	Y	40
t3.7	Dunleavy et al[29] (2013)	Y	U	Y	U	U	U	U	U	U	U	Y	27
t3.8	Fletcher and Bandy[26] (2008)	U	U	Y	U	Y	U	U	U	U	N	Y	27
t3.9	Hoving et al[22] (2005)	U	U	Y	Y	Y	U	U	U	U	N	Y	27
t3.10	Petersen et al[24] (2000)	U	U	Y	N/A	U	U	U	U	U	N	Y	20
t3.11	Piva et al[30] (2006)	U	U	Y	Y	N/A	U	U	U	U	N	Y	30
t3.12	Schneider et al[23] (2013)	U	U	Y	Y	Y	Y	Y	U	U	Y	Y	63
t3.13	Shahidi et al[27] (2012)	U	U	Y	Y	N/A	N	Y	Y	U	Y	Y	60

t3.14 Study: first author and year of publication.
t3.15 *N*, no; *N/A*, not applicable; *U*, unknown; *Y*, yes.

There was no agreement on the scoring of 7 of 110 items valued (10 studies, 11 items each). Disagreements were resolved by a consensus-based discussion, with a high level of agreement (percentage of agreement = 93.6%) and good inter-reviewer reliability (κ coefficient $k_w = 0.83$; 95% CI: 0.71-0.95).

Meta-regression

Regarding the effect of the tool type on ICC (Tables 4 and 5), no significant tool type effect, corrected for age and motion, on ICC was observed for inter-rater and intrarater reliability (Table 4, $P = .99$, and Table 5, $P = .99$, respectively).

No significant motion effect on ICC values, corrected for age, was estimated for inter-rater reliability (Table 6; $P = .11$).

Significant motion effect on ICC values, corrected for age, was estimated for intrarater reliability (Table 7; $P = .0001$). In particular, comparing ICC mean in the side-bending motion group with that in the flexion and extension group, approximately a –0.02 significant decrease ($P < .0001$) was noted.

Regarding diagnostic residual analysis (Appendix 3), the quantile-quintile plots indicated that the circles crossed or were very close to the diagonal line, and there was no significant evidence of rejecting the null hypothesis of normality using the Shapiro-Wilk tests ($P > .05$) for all residual models. The plots of standardized residuals against fitted values indicated no trend, and was randomly scattered around 0, with positive and negative values equally likely at any point. The plots of the fitted values against the observed values indicated that the points were quite close, both to the diagonal line and to the regression line of the observed vs fitted values. According to these results, no violations of the mixed model assumptions were identified.

Discussion

Main Findings

The main purpose of this systematic review was to determine whether technological devices, used to measure active cervical range of motion in adults suffering from nonspecific neck pain, are cost-worthy because of their superior reliability. Meta-regression analysis performed on pooled data indicated that no difference exists in intrarater and inter-rater reliability between expensive and inexpensive devices.

As some recent studies have suggested,[31,32] cost reduction without loss of efficacy in spinal disorder management is a rising concern in clinical practice and in related research. Choosing a measurement instrument that conciliates cost and effectiveness has become particularly relevant because the expenditure determined by the spinal pathologic conditions is growing and new technologies provide appealing, but expensive, tools for clinical practice.

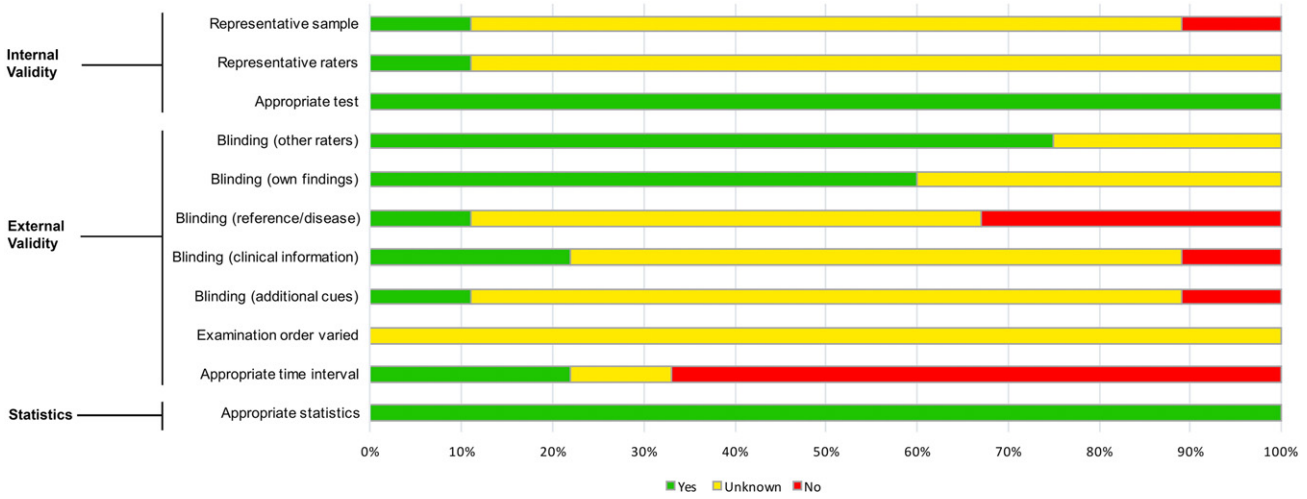

Fig 2. *The figure shows the level of biases within included studies for each Quality Appraisal of Reliability Studies item; the green bar represents the percentage of studies scored "yes" for that item, the yellow bar represents the percentage of studies scored "unknown" for that item, and the red bar represents the percentage of studies scored "no" for that item. When items were not applicable, they were not included because they don't contribute to percentage calculation.*

t4.2 **Table 4.** *Output of the Mixed Effect Model for Interrater Reliability Regarding Tool Types*

	ICC Descriptive Statistics	Mixed Effects Model	
t4.3 Characteristics	Mean (SD)	β (95% CI)	*P*
t4.5 (Intercept)		0.25 (−0.55 to 1.05)	
t4.6			
t4.7 Tool Type			.99
t4.8 Expensive	0.87 (0.09)	0	
t4.9 Inexpensive	0.80 (0.12)	0.03 (−0.15 to 0.21)	
t4.10			
t4.11 Direction			.06
t4.12 Flexion & extension	0.82 (0.09)	0	
t4.13 Rotation	0.86 (0.11)	0.04 (−0.02 to 0.11)	
t4.14 Side bending	0.77 (0.13)	−0.04 (−0.10 to 0.01)	
t4.15			
t4.16 Age	ρ = 0.58	0.01 (−0.01 to 0.03)	.43

Characteristics = variable considered; mean (SD) = ICC mean in the characteristic levels with the standard deviation for categorical variable and ICC mean with standard deviation for continuous variable; β (95% CI) = β regression coefficient with 95% CI estimated using mixed effects model; *P* value = the t4.17 likelihood ratio *P* value adjusted for multiple comparisons by the Bonferroni correction method.
t4.18 *CI,* confidence interval; *ICC,* intraclass correlation coefficient; *SD,* standard deviation.

Common-use and inexpensive measurement instruments that were analyzed for their reliability in the articles included in our review were standard dual-armed goniometer, universal or gravity inclinometer, and CROM device, which cost between €30 and €500.

The expensive instruments analyzed in the papers selected for the present review were EDI-320, OSI SMA, and Flock-of-Birds. The EDI-320 is an electronic, fully portable digital inclinometer.[2,22] OSI SMA consists in a lightweight aluminum linkage with 6 potentiometers that measures regional spinal motion concurrently on the 3 planes of movement by transforming a voltage change in a measure of ACROM degrees.[24] Flock-of-Birds is an electromagnetic tracking system, consisting in a standard range transmitter and 3 receivers: 1 receiver is used to assess natural posture of the participant, and the other 2 are used to measure movements as they are positioned on the forehead and on the sternum.[25]

All these expensive devices, as also ascertained by a previous study,[2] were estimated to cost >€500. A complete overview of their modes of operation and prices could not be performed because those instruments are out of production or have been superseded by new technologies, as declared by producers. The potentially short life of technology-based devices is a relevant feature that should be considered when investing resources for managing spinal disorders. Inclinometers and handheld goniometers

Table 5. *Output of the Mixed Effects Model for Intrarater Reliability Regarding Tool Types*

Characteristics	ICC Descriptive Statistics Mean (SD)	Mixed Effects Model β (95% CI)	*P*
(Intercept)		0.84 (0.71-0.96)	
Tool Type			.99
Expensive	0.91 (0.07)	0	
Inexpensive	0.93 (0.02)	0.01 (–0.04 to 0.06)	
Direction			.56
Flexion & extension	0.91 (0.07)	0	
Rotation	0.95 (0.02)	0.03 (–0.01 to 0.07)	
Side bending	0.92 (0.01)	0.01 (–0.02 to 0.04)	
Age	ρ = 0.29	0.00 (–0 .01 to 0.01)	.79

Characteristics = variable considered; mean (SD) = ICC mean in the characteristic levels with the standard deviation for categorical variable and ICC mean with standard deviation for continuous variable; β (95% CI) = β regression coefficient with 95% CI estimated using mixed effects model; *P* value = the likelihood ratio *P* value adjusted for multiple comparisons by the Bonferroni correction method.
CI, confidence interval; *ICC,* intraclass correlation coefficient; *SD,* standard deviation.

Table 6. *Output of the Mixed Effect Model for Interrater Reliability Regarding Directions*

Characteristics	ICC Descriptive Statistics Mean (SD)	Mixed Effects Model β (95% CI)	*P*
(Intercept)		–0.06 (–0.60 to 0.49)	
Direction			.11
Flexion & extension	0.81 (0.09)	0	
Rotation	0.85 (0.12)	0.04 (–0.03 to 0.10)	
Side bending	0.75 (0.14)	–0.06 (–0.12 to 0.01)	
Age	0.80 (0.12)	0.02 (0.01-0.03)	.04

Characteristics = variable considered; mean (SD) = ICC mean in the characteristic levels with the standard deviation for categorical variable and ICC mean with standard deviation for continuous variable; β (95% CI) = β regression coefficient with 95% CI estimated using mixed effects model; *P* value = the likelihood ratio *P* value adjusted for multiple comparisons by the Bonferroni correction method.
CI, confidence interval; *ICC,* intraclass correlation coefficient; *SD,* standard deviation.

Table 7. *Output of the Mixed Effects Model for Intrarater Reliability Regarding Directions*

Characteristics	ICC Descriptive Statistics Mean (SD)	Mixed Effects Model β (95% CI)	*P*
(Intercept)		0.89 (0.85-0.94)	
Direction			<.0001
Flexion & extension	0.93 (0.02)	0	
Rotation	0.95 (0.01)	0.02 (–0.01 to 0.03)	
Side bending	0.91 (0.01)	–0.02 (–0.03 to –0.01)	
Age	0.91 (0.02)	0.00 (–0.01 to 0.01)	.18

Characteristics = variable considered; mean (SD) = ICC mean in the characteristic levels with the standard deviation for categorical variable and ICC mean with standard deviation for continuous variable; β (95% CI) = β regression coefficient with 95% CI estimated using mixed effects model; *P* value = the likelihood ratio *P* value adjusted for multiple comparisons by the Bonferroni correction method.
CI, confidence interval; *ICC,* intraclass correlation coefficient; *SD,* standard deviation.

are still widespread and constitute common components of the tool box of every practitioner.

Only a couple of previous reviews have been published on this topic[2,7]; they both suggested that the CROM device and single inclinometer have been proven to be reliable,[2,7] and they are more worth using than other devices because of their affordable costs[3] and more solid literature about them.[8]

Unfortunately, their results cannot be directly compared with ours. Both previous reviews summarized data from studies on healthy participants, on mixed population (healthy participants and patients), and on different pathologic conditions (including rheumatic diseases, whiplash, and radicular pathologic conditions), so no firm conclusions can be drawn about reliability in adults with nonspecific neck pain. Moreover, we have conducted a meta-regression analysis on the pooled data that we extracted from the selected papers, whereas in the studies by de Koning et al[2] and Williams et al[7] results were presented only as a summary of findings with quality assessment. The intention of Williams et al was to perform a meta-analysis, but, as they declared, that was impossible because of the heterogeneity of studies.[7]

As a secondary outcome, a multivariate analysis was conducted using a linear mixed effects model, corrected for age, to assess which direction of movement was more reproducible between inexpensive tools.

Inter-rater reliability is not affected by the age or the direction of the movement measured because no statistical difference in ICC values was identified ($P = .11$). The same does not apply to intrarater reliability, which is not affected by age, but is by direction of the movement ($P = .0001$). In particular, the same rater is more likely to obtain the same measure ($P = .01$) when measuring movements on the sagittal plane (ie, flexion and extension), instead of on the frontal plane (ie, side bending). This lack of intrarater reliability in the frontal plane has, however, a small impact on physical therapists' everyday practice. Among cervical movements, side bending is the least tested as pure movement, being physiologically coupled with rotation—that is, more quantitatively represented—and has only a limited involvement in daily functional tasks.[33-35] Moreover, a recent good-quality review with meta-analysis indicated that side bending was also poorly affected by neck complaints, both acute and chronic.[36] Clinicians might be aware that visual estimation of cervical range of motion is not sufficiently reliable, even if they trained in it, as has been reported by previous studies.[2,37-39]

Methodological Consideration

Among the included studies, only 2 were of high quality,[23,27] and they both analyzed inexpensive tools (universal[23] and gravity inclinometer[27] and standard dual-armed goniometer[23]). They both analyzed inter-rater reliability,[23,27] but only 1 also included an intrarater reliability measure.[23] The quality of the other selected studies ranged from 20% to 40% of positive answers to the applicable QAREL items. Even if each of these studies adopted adequate statistics and selection of tests, reporting of blinding, samples, and raters was unclear or incorrect. Hence, because of their high risk of bias, strong conclusions cannot be drawn from the present review, especially about technological devices.

Limitations

A selection bias could have occurred when studies with mixed healthy and unhealthy population were excluded. In these cases, although data of unhealthy participants were requested directly to the authors, we were unable to collect them. Furthermore, studies were included if they investigated intrarater or inter-rater reliability in measuring ACROM as the primary outcome. Studies with this investigation as a secondary outcome may have been missed, and for this reason, a selection bias could have occurred. Reviewer bias is also a possible limitation of this review because reviewers were not blinded to the authors.

No other systematic review was found in the literature about reliability of active cervical range of motion, including only patients with nonspecific neck pain. Extensive research was conducted to identify all the former studies on this topic, without restrictions of language and date. Further information was also requested from authors when necessary. Even so, because the protocol registration of these studies was not compulsory, publication bias cannot be excluded.

Finally, a reporting bias could have occurred because the review was not registered on PROSPERO (an international prospective register of systematic reviews).

Conclusions

Intrarater and inter-rater reliability of ACROM measures recorded by expensive or inexpensive devices was not significantly different. However, our results indicate that a reliable and cost-effective measurement seems possible by means of inexpensive devices in common clinical practice. Nevertheless, the methodological quality of the available studies is quite poor; thus, new studies of better quality would empower the strength of our conclusions about reliability of active cervical range of motion measures.

Funding Sources and Conflicts of Interest

No funding sources or conflicts of interest were reported for this study.

Contributorship Information

Concept development (provided idea for the research): G.R., A.R., M.T.

Design (planned the methods to generate the results): G.R., A.R., M.T.

Supervision (provided oversight, responsible for organization and implementation, writing of the manuscript): G.R., A.R., M.T.

Data collection/processing (responsible for experiments, patient management, organization, or reporting data): M.S., F.G., A.B.

Analysis/interpretation (responsible for statistical analysis, evaluation, and presentation of the results): F.G., D.R.

Literature search (performed the literature search): M.S., F.G., A.B.

Writing (responsible for writing a substantive part of the manuscript): G.R., A.R., F.G., M.S., F.C., A.B.

Critical review (revised manuscript for intellectual content, this does not relate to spelling and grammar checking): M.T., G.R., D.R.

Practical Applications

- Measurement of active cervical range of motion during common clinical practice can be effectively performed with low-cost devices.
- Use of expensive devices should be reserved to specific clinical or research conditions.
- Assessment of active cervical side bending is less reliable.

Appendix A. Supplementary Data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmpt.2017.07.002.

References

1. Hoy D, March L, Woolf A, et al. The global burden of neck pain: estimates from the Global Burden of Disease 2010 study. *Ann Rheum Dis.* 2014;73(7):1309-1315.
2. de Koning CHP, van den Heuvel SP, Staal JB, Smits-Engelsman BCM, Hendriks EJM. Clinimetric evaluation of active range of motion measures in patients with non-specific neck pain: a systematic review. *Eur Spine J.* 2008;17(7):905-921.
3. Snodgrass SJ, Cleland JA, Haskins R, Rivett DA. The clinical utility of cervical range of motion in diagnosis, prognosis, and evaluating the effects of manipulation: a systematic review. *Physiother (UK).* 2014;100(4):290-304.
4. van Trijffel E, Plochg T, van Hartingsveld F, Lucas C, Oostendorp RAB. The role and position of passive intervertebral motion assessment within clinical reasoning and decision-making in manual physical therapy: a qualitative interview study. *J Man Manip Ther.* 2010;18(2):111-118.
5. Antonaci F, Ghirmai S, Bono G, Nappi G. Current methods for cervical spine movement evaluation: a review. *Clin Exp Rheumatol.* 2000;18(2 Suppl 19):S45-S52.
6. Jordan K. Assessment of published reliability studies for cervical spine range-of-motion measurement tools. *J Manipulative Physiol Ther.* 2000;23(3):180-195.
7. Williams MA, McCarthy CJ, Chorti A, Cooke MW, Gates S. A systematic review of reliability and validity studies of methods for measuring active and passive cervical range of motion. *J Manipulative Physiol Ther.* 2010;33(2):138-155.
8. Martin BI, Deyo RA, Mirza SK, et al. Expenditures and health status among adults with back and neck problems. *JAMA.* 2008;299(6):656-664.
9. Kleinman N, Patel AA, Benson C, Macario A, Kim M, Biondi DM. Economic burden of back and neck pain: effect of a neuropathic component. *Popul Health Manag.* 2014;17(4):224-232.
10. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med.* 2009;6(7):1-28.
11. Bogduk N. Neck pain. *Aust Fam Physician.* 1984;13(1):26-30.
12. Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J Clin Epidemiol.* 2010;63(8):854-861.
13. Lucas NP, Macaskill P, Irwig L, et al. The reliability of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J Clin Epidemiol.* 2013;63(8):854-861.
14. Carlsson H, Rasmussen-Barr E. Clinical screening tests for assessing movement control in non-specific low-back pain. A systematic review of intra- and inter-observer reliability studies. *Man Ther.* 2013;18(2):103-110.
15. Barrett E, McCreesh KLJ. Reliability and validity of non-radiographic methods of thoracic kyphosis measurement: a systematic review. *Man Ther.* 2014;19:10-17.
16. McCreesh KM, Crotty JM, Lewis JS. Acromiohumeral distance measurement in rotator cuff tendinopathy: is there a reliable, clinically applicable method? A systematic review. *Sport Med.* 2015;49(5):298-305.
17. Gorgos KS, Wasylyk NT, Van Lunen BL, Hoch MC. Inter-clinician and intra-clinician reliability of force application during joint mobilization: a systematic review. *Man Ther.* 2014;19(2):90-96.
18. Moloney NA, Hall TM, Doody CM. Reliability of thermal quantitative sensory testing: a systematic review. *J Rehabil Res Dev.* 2012;49(2):191.
19. Adhia DB, Bussey MD, Ribeiro DC, Tumilty S, Milosavljevic S. Validity and reliability of palpation-digitization for non-invasive kinematic measurement—a systematic review. *Man Ther.* 2013;18(1):26-34.
20. Team RC. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2014.
21. Furlan JC, Singh J, Hsieh J, Fehlings MG. Methodology of systematic reviews and recommendations. *J Neurotrauma.* 2011;28(8):1335-1339.
22. Hoving JL, Pool JJM, van Mameren H, et al. Reproducibility of cervical range of motion in patients with neck pain. *BMC Musculoskelet Disord.* 2005;6:59.
23. Schneider GM, Jull G, Thomas K, et al. Intrarater and interrater reliability of select clinical tests in patients referred for diagnostic facet joint blocks in the cervical spine. *Arch Phys Med Rehabil.* 2013;94(8):1628-1634.
24. Petersen CM, Johnson RD, Schuit D. Reliability of cervical range of motion using the OSI CA 6000 spine motion analyser on asymptomatic and symptomatic subjects. *Man Ther.* 2000;5(2):82-88.
25. Assink N, Bergman GJD, Knoester B, Winters JC, Dijkstra PU, Postema K. Interobserver reliability of neck-mobility measurement by means of the flock-of-birds electromagnetic tracking system. *J Manipulative Physiol Ther.* 2005;28(6):408-413.

26. Fletcher JP, Bandy WD. Intrarater reliability of CROM measurement of cervical spine active range of motion in persons with and without neck pain. *J Orthop Sports Phys Ther.* 2008;38(10):640-645.
27. Shahidi B, Johnson CL, Curran-Everett D, Maluf KS. Reliability and group differences in quantitative cervicothoracic measures among individuals with and without chronic neck pain. *BMC Musculoskelet Disord.* 2012;13(1):215.
28. Cleland JA, Childs JD, Fritz JM, Whitman JM. Inter-rater reliability of the history and physical examination in patients with mechanical neck pain. *Arch Phys Med Rehabil.* 2006;87(10):1388-1395.
29. Dunleavy K, Goldberg A. Comparison of cervical range of motion in two seated postural conditions in adults 50 or older with cervical pain. *J Man Manip Ther.* 2013;21(1):33-39.
30. Piva SR, Erhard RE, Childs JD, Browder DA. Inter-tester reliability of passive intervertebral and active movements of the cervical spine. *Man Ther.* 2006;11(4):321-330.
31. Paul JC, Petrizzo A, Rizzo JR, et al. Feasibility of a cost-effective, video analysis software-based mobility protocol for objective spine kinematics and gait metrics: a proof of concept study. *PM R.* 2015;7(3):336-339.
32. Brink Y, Louw Q, Grimmer K, Schreve K, van der Westhuizen G, Jordaan E. Development of a cost effective three-dimensional posture analysis tool: validity and reliability. *BMC Musculoskelet Disord.* 2013;14:335.
33. Bennett SE, Schenk RJ, Simmons ED. Active range of motion utilized in the cervical spine to perform daily functional tasks. *J Spinal Disord Tech.* 2002;15(4):307-311.
34. Bible JE, Biswas D, Miller CP, Whang PG, Grauer JN. Normal functional range of motion of the cervical spine during 15 activities of daily living. *J Spinal Disord Tech.* 2010;23(1):15-21.
35. Cobian DG, Daehn NS, Anderson PA, Heiderscheit BC. Active cervical and lumbar range of motion during performance of activities of daily living in healthy young adults. *Spine (Phila Pa 1976).* 2013;38(20):1754-1763.
36. Stenneberg MS, Rood M, de Bie R, Schmitt MA, Cattrysse E, Scholten-Peeters GG. To which degree differs active cervical range of motion between patients with neck pain, whiplash and healthy controls? A systematic review and meta-analysis. *Arch Phys Med Rehabil.* 2017;98(7):1407-1434.
37. Youdas JW, Carey JR, Garrett TR. Reliability of measurements of cervical spine range of motion–comparison of three methods. *Phys Ther.* 1991;71(2):98-104 discussion 105-106.
38. Whitcroft KL, Massouh L, Amirfeyz R, Bannister G. Comparison of methods of measuring active cervical range of motion. *Spine (Phila Pa 1976).* 2010;35(19):E976-E980.
39. Hirsch BP, Webb ML, Bohl DD, et al. Improving visual estimates of cervical spine range of motion. *Am J Orthop (Belle Mead NJ).* 2014;43(11):E261-E265.