

Clinimetric evaluation of active range of motion measures in patients with non-specific neck pain: a systematic review

Chantal H. P. de Koning · Sylvia P. van den Heuvel ·
J. Bart Staal · Bouwien C. M. Smits-Engelsman ·
Erik J. M. Hendriks

Received: 12 November 2007 / Revised: 11 February 2008 / Accepted: 9 March 2008 / Published online: 22 April 2008
© The Author(s) 2008

Abstract The study is to provide a critical analysis of the research literature on clinimetric properties of instruments that can be used in daily practice to measure active cervical range of motion (ACROM) in patients with non-specific neck pain. A computerized literature search was performed in Medline, Cinahl and Embase from 1982 to January 2007. Two reviewers independently assessed the clinimetric properties of identified instruments using a criteria list. The search identified a total of 33 studies, investigating three different types of measurement instruments to determine ACROM. These instruments were: (1) different types of goniometers/inclinometers, (2) visual estimation, and (3) tape measurements. Intra- and inter-observer reliability was demonstrated for the cervical range of motion instrument (CROM), Cybex electronic digital instrument (EDI-320) and a single inclinometer. The presence of agreement was assessed for the EDI-320 and a single inclinometer. The CROM received a positive rating for construct validity. When clinical acceptability is taken into

account both the CROM and the single inclinometer can be considered appropriate instruments for measuring the active range of motion in patients with non-specific neck pain in daily practice. Reliability is the aspect most frequently evaluated. Agreement, validity and responsiveness are documented less frequently.

Keywords Non-specific neck pain · Active range of motion · Reproducibility of results · Systematic review

Introduction

Neck pain is a common musculoskeletal disorder. The incidence of neck pain in the Netherlands has been estimated as 23.1 per 1,000 person years [12]. In general, women have more neck pain than men [12, 27]. In the Netherlands, 51% of patients with acute non-specific neck pain who consulted their general practitioners were referred to physiotherapists for treatment [79].

Neck pain may result from many causes (trauma, infections or inflammatory conditions, rheumatic disorders and congenital diseases), but most often no specific cause can be found and the condition is labelled as non-specific neck pain [10]. In their clinical examination, physiotherapists and other healthcare providers may routinely perform an assessment of the active cervical range of motion (ACROM) to assess the level of impairment associated with neck pain as well as the results of treatment. Typical ACROM assessments of the cervical spine include flexion and extension in the sagittal plane, lateral flexion in the frontal plane and rotation in the transverse plane. Tests or instruments which are used to examine ACROM in patients with non-specific neck pain should meet several clinimetric

C. H. P. de Koning (✉) · B. C. M. Smits-Engelsman
Avans+, University for Professionals, Breda,
The Netherlands
e-mail: cdekoning@zeelandnet.nl

S. P. van den Heuvel · E. J. M. Hendriks
Dutch Institute for Allied Health Care (NPi),
Amersfoort, The Netherlands

J. B. Staal · E. J. M. Hendriks
Department of Epidemiology, Centre for Evidence Based
Physiotherapy and Caphri Research Institute, Maastricht
University, Maastricht, The Netherlands

B. C. M. Smits-Engelsman
Motor Control Lab, Department of Kinesiology,
K.U. Leuven, Leuven, Belgium

prerequisites, such as an acceptable level of reproducibility, validity and responsiveness [29].

Recently, several systematic reviews have been published on the assessment of passive cervical range of motion (PCROM) and palpation procedures of the cervical spine [62, 64, 75]. Among the palpation procedures, pain provocation tests were found to be the most reliable [62, 64] and the assessment of regional PCROM was found to be more reliable than segmental PCROM [62]. However, in these reviews it was also concluded that most studies were of a poor methodological quality [62, 64, 75].

In 2000, a review was published that assessed the reliability of tools used to measure ACROM. This review concluded that the cervical range of motion device has shown promising reliability [38]. Our overview takes account not only of reliability, but also of validity and responsiveness.

The objective of the present systematic review, therefore, is to provide an overview of the current knowledge on clinimetric properties of instruments that are practical to use when evaluating ACROM in patients with non-specific neck pain.

Methods

Study selection

An extensive search was conducted in the MEDLINE (1982 to January 2007), CINAHL (1982 to January 2007) and EMBASE (1996 to January 2007) databases. The following search terms were used: neck, cervical, reproducibility of results, reliability, reproducibility, validation studies, validity, responsiveness, range of motion, active motion, movement. In addition, after the selection of relevant studies, the specific names of identified assessment instruments were used for an additional computerized search to identify supplementary relevant studies. References from retrieved papers were searched for additional studies. The principle investigator (CK) screened the potentially relevant papers retrieved for eligibility according to the following inclusion criteria:

- The paper had to be written in English or Dutch;
- Studies had to pertain to the cervical or upper thoracic spine;
- Studies had to investigate the reproducibility, validity or responsiveness of instruments or tests for measuring ACROM;
- The instrument or test used had to be described clearly, enabling eventual replication of the test,
- And the instrument or test had to be portable, affordable (maximum 1,000 Euros) and easy to use (time to test

maximum 5 min) by healthcare professionals in daily practice.

Studies were excluded if they were non-published papers (thesis studies).

Data abstraction and quality assessment

We investigated the following clinimetric properties: intra-observer reliability, interobserver reliability, agreement, construct validity, responsiveness and interpretability. To interpret the data a checklist was composed that was partly based on criteria developed by the Scientific Advisory Committee of the Medical Outcome Trust [46] and a checklist developed by Bot et al. [11] (Table 1).

Description of the instruments for the assessment of ACROM

Descriptive data extracted from the publications included the target population and the examiners, description of test/instrument and protocol used, description of test–retest interval, blinding of examiners for participants and each other's or reference test result, and explanation of withdrawals.

Reproducibility

Reproducibility is the extent to which an instrument yields stable scores over time among respondents who are assumed not to have changed [67]. Reproducibility was assessed by rating reliability and agreement. Reliability represents the extent to which individuals can be distinguished from each other, despite measurement errors. Agreement represents a lack of measurement error [67].

The weighted kappa was considered adequate for calculating the reliability of ordinal data, and calculation of the intra-class correlation coefficient (ICC) was considered an adequate measure for ordinal or parametric data [31]. Intraobserver reliability and interobserver reliability were rated positive if the ICC was >0.85 and >0.70 respectively [29, 66]. A kappa coefficient above 0.60 was rated positively for intra- and inter-observer reliability. This is based on the Landis and Koch scale [43] 0.41–0.60 moderate correlation, 0.61–0.80 substantial correlation and 0.81–1.00 almost perfect correlation. Application of the Pearson reliability coefficient was rated as doubtful, as it neglects systematic observer bias [31].

Agreement is the ability to achieve the same value with repeated measurements. For this review, calculations of the 95% limits of agreement (LoA), standard error of measurement (SEM), smallest detectable change (SDC) or

Table 1 Checklist used for the assessment of clinimetric properties of the studies included

Clinimetric property	Definition	Criteria
Reproducibility	Degree to which repeated measurements in stable persons provide similar answers	K: nominal/ordinal data ICC: ordinal/parametric data
Reliability	The extent to which patients can be distinguished from each other, despite measurement error	+ Adequate design, method, intraobserver ICC > 0.85 or $K > 0.41$ interobserver ICC > 0.70 or $K > 0.41$
Agreement	The ability to achieve the same value with repeated measurements	± Adequate design, method, intraobserver ICC < 0.85 or $K < 0.40$ interobserver ICC < 0.70 or $K < 0.40$? Doubtful method used o No information found Limits of agreement, SEM or SDC are presented + Adequate design method and result ? Doubtful method used – No information found
Construct validity	The extent to which a test identifies the concept or trait of which is being measured	Pearson R or Spearman Rho + Adequate design, method, $r > 0.65$? Doubtful method used – Inadequate construct validity o No information found
Responsiveness	Ability of an instrument to detect important change over time in the concept being measured	Hypotheses were formulated and results are in agreement + Adequate design method and result ? Doubtful method used – Inadequate responsiveness o No information found
Interpretability	The degree to which one can assign qualitative meaning to quantitative scores	Authors provided information on the interpretation of scores, MIC defined Mean and SD scores before and after treatment

K Kappa statistics, *ICC* intraclass correlation coefficient, *SEM* standard error of measurement, *SDC* smallest detectable change, *MIC* minimal important change, *SD* standard deviation

minimal detectable change (MDC) were considered sufficient. The SDC or MDC reflect the smallest within-person change in score that can be interpreted as a real change, above measurement error [11, 67]. It is not possible to define adequate cut-off points for the result of an agreement study. For that reason, a positive rating was given when an adequate method for agreement was used.

Validity

Validity is the degree to which an instrument measures what it is supposed to measure. Construct validity is the extent to which scores on a particular instrument relate to other measures in a manner that is consistent with theoretically derived hypotheses concerning the concept being measured [67]. A Pearson's correlation coefficient or Spearman correlation coefficient above 0.65 was rated positively for construct validity [29, 66].

Responsiveness

Responsiveness refers to the ability of an instrument to detect important change over time in the concept being measured and is therefore considered to be a measure of longitudinal validity. There is no single agreed method of assessing or expressing an instrument's responsiveness [29, 67]. Responsiveness is considered to have been studied adequately if hypotheses have been specified and the results corresponded to these hypotheses [11]. It was not possible to define adequate cut-off points for the result of a responsiveness study. For that reason, a positive rating was given when an adequate method for responsiveness was used.

Interpretability

Interpretability is defined as the degree to which scores and change scores can be interpreted and qualitative meaning

can be assigned to quantitative scores. The investigators should provide information about what difference in score would be clinically meaningful. We rated this on the basis of whether the authors had presented a minimal important change (MIC) or if information was presented that could aid in interpreting scores—for instance, presentation of means and standard deviations (SD) of patient scores before and after treatment, data on distribution of scores in relevant subgroups and relating changes in the instrument score to patients' global perceived change [11, 67].

Overall quality

To obtain an overall score for the quality of the instruments, the number of positive ratings on the above-mentioned points was summed for each instrument.

Two investigators (C.K. & S.H.) independently assessed the studies included according to the criteria list. Disagreements between the reviewers were resolved by discussion. If disagreement persisted about the assignment of a score to an item, a third person (E.H.) was consulted to decide on the final rating.

Results

Selection of the studies

The search generated 549 hits. After screening titles and abstracts, 481 studies were excluded. Of the remaining 68 studies, 33 publications were included after reading the full article [1–3, 8, 15, 16, 20, 32–37, 41, 48, 51, 53–56, 59, 61, 68–71, 73, 77, 78, 80, 84–86].

Reasons for exclusion were the cost of the instruments used ($n = 24$), (these instruments were mainly tested in university laboratories and we estimated that the instruments cost more than 1,000 Euros) [4, 5, 7, 14, 17, 19, 22–24, 28, 39, 40, 42, 44, 45, 49, 50, 52, 57, 58, 63, 65, 81, 83]; no clinimetric evaluation of the instrument ($n = 9$) [6, 9, 18, 24, 25, 30, 60, 72, 82], and only assessment of PCROM [47].

The instruments that were included based on the criteria of clinical acceptability were: goniometers/inclinometers (30 articles); visual estimation (two articles) and tape measurements (four articles). Relevant data on study population, examiners, study protocol and results from these studies are displayed in Tables 2, 3 and 4. Initially, there was 94% agreement on the items that were rated. All disagreements between the two reviewers were solved with discussion.

Goniometers/inclinometers

Goniometers are versatile devices that measure range of motion in grades and depend on landmarks. Inclinometers

are fluid-filled goniometric instruments that depend on gravity. In the literature we included, the terms goniometer and inclinometer are used interchangeably. The instruments included are: variations of universal goniometers [20, 41, 48, 56, 73, 86], a Myrin goniometer [8, 78], a Spin-T goniometer [1, 2, 32], the cervical range of motion instrument (CROM) [16, 33, 54, 55, 61, 69–71, 84, 85], single inclinometers [15, 20, 33, 59, 80], an electronic digital inclinometer EDI-320 [35, 68], a simple inclinometer [51], a gravity action inclinometer [37], a liquid inclinometer [3] and a spirit inclinometer [53] (Fig. 1).

Information on reliability was found for all instruments. In general, the ICC was reported except for most articles published before 1995. The reliability for the CROM, a single inclinometer, and the EDI-320 were rated positively (ICC intraobserver > 0.85 , ICC interobserver > 0.70).

Information on agreement was found for the CROM [54], a universal goniometer [20], spin-T goniometer [32], a single inclinometer [20, 59, 80] and the EDI-320 [35].

The EDI-320 (LoA intraobserver F/E $-2.5 \pm 11.1^\circ$ LFL $-0.1 \pm 10.4^\circ$ ROT $-5.9 \pm 13.5^\circ$; LoA interobserver F/E $3.3 \pm 17^\circ$, LFL $0.5 \pm 17^\circ$, ROT $-1.3 \pm 24.6^\circ$) and a single inclinometer (SEM ranged from 3.6° to 7° and MDC ranged from 10° to 19°) were rated positively (Fig. 2).

Construct validity of the CROM was determined by comparing CROM with radiographics, a single inclinometer and opto-electronic systems [33, 55, 69–71]. The Spin-T goniometer was compared with motion star equipment [2]. The rangiometer and universal goniometer were compared with each other [86] and with the age and disease duration of patients with ankylosing spondylitis [48]. A single inclinometer was compared with results on the neck disability index (NDI) [59] and radiographics [15].

Construct validity of the spin-T goniometer, universal goniometer, rangiometer and single inclinometer were rated as doubtful, because different statistics were used as the predefined criterion for the Pearson correlation coefficient (r) or $r < 0.65$. The CROM gets a positive rating; $r > 0.65$.

Information on responsiveness was only found for the universal goniometer, measured on patients with ankylosing spondylitis [48] and therefore gets a doubtful rating for the present patient group.

Visual estimation

When performing a visual estimation the patient sits and the examiner visually estimates the ACROM. Two articles described the interobserver reliability of visual estimation [34, 77]. Both articles presented Kappa values that were below the predefined criteria and therefore the reliability of visual estimation was rated as inadequate. Information on agreement, validity and responsiveness was not found.

Table 2 Characteristics of included studies as regards goniometric/inclinometric measurements

Author and tool	Purpose	Patient	Examiner	Procedure/protocol	Results
Agarwal [1] Spin-T goniometer	Intraobserver Reliability	Thirty healthy subjects	?	Patient seated near front wall for reference subject seated fl/ext/lfl/rot Warming-up? All movement measured three times test-retest interval no, blinding ex-pt? Reference test Motion star recording simultaneously foam head model and four men. Readings fl/ext/lfl/rot Warming-up?	ICC intra > 0.96
Agarwal [2] Spin-T goniometer	Validity	Four healthy subjects and foam head model	?	Reference test Motion star recording simultaneously foam head model and four men. Readings fl/ext/lfl/rot Warming-up?	Coefficient determination (R^2) higher than 0.99
Alantara [3] Liquid inclinometer Gravitation inclinometer	Intra and interobserver Reliability	Twenty-four healthy subjects	two physical therapists	Intraobserver one-year interval Interobserver one week interval Blinding ex-pt?	Reliability factor Intra Fl/EXT 0.68 LFL 0.61 ROT 0.37 systematic difference (P) Fl/EXT 0.58 LFL 0.38 ROT 0.00 Inter Fl/EXT 0.69 LFL 0.79 ROT 0.86 systematic difference (P)Fl/EXT 0.42 LFL 0.31 ROT 0.24
Balogun [8] Myrin goniometer	Intra and interobserver Reliability	Twenty-one healthy subjects	Three physical therapist	Subject seated Myrin location: Fl-ext above earlobe LFL forehead ROT vertex Warming-up: ? Number of measurements? test-retest interval 4 days blinding ex-pt?	Pearson r Gonio Fl 0.31–0.34 EXT 0.71–0.86 ROT R 0.49–0.67 ROT L 0.53–0.56 LFL R 0.049–0.78 LFL L 0.61–0.67
Bush [15] inclinometer	Reliability Validity	Three healthy subjects	Thirty-four physical therapists	Patient seated FL/EXT/LFL. Patient supine ROT three methods used reliability single method double method, stabilization method Validation RX flexion/extension/lfl CT scan rotation test-retest interval?, blinding ex-pt?	ICC inter Single 0.91–0.93 Double 0.89–0.93 Stab 0.89–0.94 VAL Criteria within $\pm 5^\circ$ RX FL: single/double EXT: single/stab LFL/ROT not valid
Capuano-Pucci [16] CROM	Intra and interobserver Reliability	Twenty healthy subjects	?	Subject seated fl/ext/lfl/rot Warming-up three repetitions half cycle ROM test-retest interval 2 days, blinding ex-pt?	Pearson S Intra tester one 0.63–0.90 tester two 0.62–0.91 Inter session one 0.80–0.87 session two 0.74–0.85

Table 2 continued

Author and tool	Purpose	Patient	Examiner	Procedure/protocol	Results
Cleland [20] Inclinometer Universal goniometer	Intra and interobserver Reliability Agreement	Twenty-two patients mechanical neck pain	Four physical therapists	Inclinometer patient seated FL/ EXT/LFL UG Rotation Patient seated Test-retest interval one directly after the other, blinding examiner-patient?	<i>Inclino</i> ICC inter FL 0.75 (0.50– 0.89) EXT 0.74 (0.48–0.88) LFL R 0.66 (0.33–0.84) LFL L 0.69 (0.40–0.86) SEM FL 6.8°, EXT 4.7°, LFL R 3.6°, LFL L 7° MDC FL 18.8°, EXT 13.0°, LFL R 10.0°, LFL L 19° LoA FL 0.76 ± 16.6, EXT 1.6 ± 11.7, LFL R-0.5 ± 9.5, LFL L 2.2 ± 18.3 <i>UG</i> ICC inter ROT R 0.77(0.52– 0.90) ROT L 0.57(0.14–0.81) SEM ROT R 5°, ROT L 5.5° MDC ROT R 13.9° ROT L 13.9° LoA ROT R –0.3 ± 15.9 ROT L –1.5 ± 6.4
Haynes [32] Spin-T goniometer	Intra and interobserver Reliability Agreement	Twenty-three healthy subjects	Chiropractor & physical therapist	Patient seated corner of room for reference subject seated FL/EXT/LFL/ROT warming-up: two all movement measured: two times Test-retest interval one directly after the other, blinding ex-pt yes	ICC Intra I/II FI 0.91/0.95 Ext 0.96/0.1 LFL R 0.87/0.98 LFL L 0.87/0.98 ROT R 0.94/0.96 ROT L 0.97/0.97 Inter FI 0.85 (0.68–0.93) Ext 0.91 (0.80– 0.96) LFL R 0.75 (0.50–0.89) LFL L 0.82 (0.62–0.92) Rot R 0.96 (0.91–0.98) Rot L 0.96 (0.89–0.98) SEM FI 2.7° EXT 4.1° LFL L 3.9° LFL R 4.3° ROT L 2.7° ROT R 2.6°
Hole [33] CROM Inclinometer	Intra and interobserver Reliability Validity	Thirty healthy subjects	Final year chiropractic student	Subject seated FL/EXT/LFL/ ROT, inclinometer pt supine for rotation Warming up? 2 measurements each movement, half cycle ROM Test-retest interval?, blinding ex-pt yes	<i>CROM</i> ICC Intra FI/EXT 0.96 LFL R 0.96 LFL L 0.96 ROT R 0.92 ROT L 0.92 Inter FI/EXT 0.88 LFL R 0.82 LFL L 0.86 ROT R 0.94 ROT L 0.90 <i>Inclino</i> ICC Intra FI/Ext 0.94 LFL R 0.94 LFL L 0.88 ROT R 0.93 ROT L 0.84 Inter FI/Ext 0.84 LFL R 0.82 LFL L 0.81 ROT R 0.76 ROT L 0.86 Validity ICC FL/EXT 0.80, LFL R 0.80, LFL L 0.78, ROT R –0.23 ROT L –0.12

Table 2 continued

Author and tool	Purpose	Patient	Examiner	Procedure/protocol	Results
Hoving [35] EDI-320	Intra and interobserver Reliability Agreement	Thirty-two patients median 13.5 weeks a-specific neck pain	Two physical therapists	Patient seated FL/EXT/LFL patient supine ROT full cycle ROM Warming-up 2 × Test-retest interval one directly after the other, blinding ex-pt yes	ICC intra rater 1/rater 2 (95%CI) Fl/Ext 0.96 (0.93–0.98)/0.97 (0.86–0.97) LFL 0.93(0.86–0.97)/0.93 (0.86–0.96) Rot 0.91 (0.91–0.98)/0.96 (0.92/0.98) Inter (95%CI) Fl/Ext 0.95 (0.90–0.98) LFL 0.89 (0.77–0.94) Rot 0.95 (0.90–0.98) LoA intra F/E -2.5±11.1° LFL - 0.1 ± 10.4° ROT -5.9 ± 13.5° LoA inter F/E 3.3 ± 17°, LFL 0.5 ± 17°, ROT -1.3 ± 24.6° SDD F/E (11.1°) rot (13.5°) equal to predefined criteria LFL (10.4°) predefined (7.2°)
Jenkinson [37] inclinometer	Intra and interobserver Reliability	Forty subjects with Ankylosing spondylitis	Three physical therapists	PT seated FL/EXT pt supine ROT Warming-up? Test-retest interval measured on consecutive days, blinding ex-pt?	Pearson <i>r</i> intra 0.99 <i>P</i> < 0.001 inter 0.98 <i>P</i> < 0.001
Kadir [41] Goniometer head-gear based	Intra and interobserver Reliability	Thirty subjects both healthy subjects and subjects with neck pain	?	Patient seated Warming-up? Full cycle ROM measured Test-retest interval one directly after the other, blinding ex-pt?	Intra 95% CI of differences, experienced observer fl/ext ± 13.5°, lfl ± 9.4°, rot ± 7.6°, inexperienced observer fl/ext ± 15°, lfl ± 11.4°, rot ± 10.4° Interobserver >90% total variation because of variation between volunteers (2-way ANOVA). Variation between observers accounted for 1.6% or less of variation, only rotation was significantly different between observers (<i>P</i> < 0.5), range of differences fl/ext 2°–21°, lfl 2°–18°, rot 5°–44°

Table 2 continued

Author and tool	Purpose	Patient	Examiner	Procedure/protocol	Results
Klüber-Moffett [51] Inclinometer	Intra and interobserver Reliability	Twelve healthy females for intraobserver 14 healthy females for interobserver	?	Patient seated FL/EXT/LFL/ROT Warming-up? Test-retest interval 3× within 1 h, blinding ex-pt?	Intra: repeated measures anova: no sign difference between 3 measurements; 9 of 216 differences >10°, individual measurements rarely fluctuate by more than 15°; 95% prediction interval for difference between observation: fl ± 3.2°, ext ± 13.6°, rot r ± 9.4°, rot l ± 13.2°, lfl r ± 7.2°, lfl l ± 10.7° Inter: paired student <i>t</i> -test no significant difference between observers; 95% prediction interval for difference between observation: fl ± 15.1°, ext ± 19.8°, rot r ± 11.3°, rot l ± 11.6°, lfl r ± 8.9°, lfl l ± 7.8°
Maksymowich [48] Goniometer	Intra and interobserver Reliability Agreement Validity Responsiveness	Patients with ankylosing spondylitis 44 reliability, 33 responsiveness	Clinician nurse, rheumatologist, trained medical student	pt seated rotation goniometer centre of the head, one arm moved over the plane of the nose as the pt rotates. Score 0: >70, 1: 20–70, 2: <20. Test-retest interval? blinding ex-pt?	ICC Intra 0.98 & 0.97 Inter 0.95 LoA –10.1 to 12.0 Spearman Rho gonio Age –0.41 (<0.001) disease duration –0.42 (<0.001) BASDAI –0.27 (0.001) Responsiveness ES 0.34 SRM 0.86 <i>P</i> < 0.001
O' Driscoll [53] Spirit inclinometer (hydrogoniometer)	Intraobserver Reliability	Ten subjects intraobserver study Ten subjects interobserver study	Metrologists and clinician with several years experience	pt seated FL/EXT/LFL pt supine ROT Warming-up? Intraobserver one-month interval, blinding ex-pt?	Intraobserver SD of differences Fl ± ext 4.1°, lfl 4.7°, rot 3.7° Mean difference ± standard error fl/ext 1.9 ± 3.1° lfl 1.5 ± 3.2° rot 1.9 ± 3.8°
Olson [54] CROM	Intra and interobserver Reliability Agreement	Twelve healthy subjects	?	Subject seated FL/EXT/LFL/ROT Warming-up 5 times, 2 measurements each movement, half cycle ROM Test-retest interval? blinding ex-pt?	ICC Intra Fl 0.88 EXT 0.99 LFL R 0.98 LFL L 0.98 ROT R 0.99 ROT L 0.97 Inter Fl 0.58 Ext 0.97 LFL R 0.96 LFL L 0.94 ROT R 0.96 ROT L 0.98 SEM Fl 4° EXT 3° LFL L 2° LFL R 2° ROT L 3° ROT R 2°

Table 2 continued

Author and tool	Purpose	Patient	Examiner	Procedure/protocol	Results
Ordway [55] CROM	Validity	Twenty healthy subjects	?	Reference test Radiographic/computerised tracking system 3 space Patient seated, flexion extension Warming-up three times, number measurements? simultaneously, blinding ex-pt?	Test Newman Keuls post hoc test, no difference CROM-RX, CROM-3 space significant difference ($P < 0.05$)
Pellechia [56] Plastic goniometer with paperclip placed through axis of rotation	Intra and interobserver Reliability	Hundred healthy subjects, intratester, 35 healthy subjects intertester	Two physical therapists, 13 and 2 years experience	Subject seated cervical lateral flexion Warming-up? 2 repetitions for each measurement, Test-retest interval?, blinding ex-pt?	ICC intraobserver LFL L 0.91 LFL R 0.94 ICC interobserver LFL L 0.65 LFLF R 0.86
Piva [59] inclinometer	Interobserver Reliability, Agreement Validity	Thirty neck pain patients	Manual therapists	Patient seated, extension/flexion/IFI Patient supine rotation Warming-up? Test-retest interval one directly after the other, blinding ex-pt? Validity correlation AROM-NDI	ICC Inter (95%CI) FI 0.78 (0.59–0.89) EXT 0.86 (0.73–0.93) LFL R 0.87 (0.75–0.94) LFL L 0.85 (0.70–0.92) ROT R 0.86 (0.74–0.93) ROT L 0.91 (0.82–0.96) SEM; FI 5.8° EXT 5.6° LFL L 4.2° LFL R 3.7° ROT L 4.1° ROT R 4.8° MDC; FI 16° EXT 16° LFL L 12° LFL R 10° ROT L 11° ROT R 13° Pearson r AROM FL/EXT r 0.43, ROT r 0.40
Rheault [61] CROM	Interobserver Reliability	Twenty-two subjects with cervical sprain/spondylosis/degenerative disc/herniated disc/undiagnosed neck pain	?	Subject seated FL/EXT/LFL/ROT Warming-up ? 2 measurements each movement, half cycle ROM Test-retest interval ?, blinding ex-pt?	ICC Inter FI 0.76 EXT 0.98 LFL R 0.87 LFL L 0.86 ROT R 0.81 ROT L 0.82
Toussignant [69] CROM	Validity	Thirty-one healthy subjects	Three physical therapists	Reference test Radiographic Flexion/extension, warming-up yes, simultaneous measurements, blinding ex-pt?	Pearson r Flexion 0.97 Extension 0.98

Table 2 continued

Author and tool	Purpose	Patient	Examiner	Procedure/protocol	Results
Tousignant [68] EDI 320	Intra and interobserver Reliability	Forty-four healthy subjects	Two testers 2 recorders 4 year physical therapy students	Patient seated FL/EXT half cycle ROM test-retest interval one directly after the other, blinding ex-pt?	ICC intra 0.77–0.83 inter 0.66–0.80
Tousignant [70] CROM	Validity	Twenty-four subjects, minimum 18 years old, who received physiotherapy past 3 months for neck pain	Three physical therapists	Reference test Radiographic Lateral flexion left/right measurements simultaneously, blinding ex-pt?	Pearson r LFL L 0.82 (95% CI 0.62–0.92) LFL R 0.84 (95% CI 0.66–0.93)
Tousignant [71] CROM	Validity	Forty-five healthy subjects	Research assistant with rehabilitation background	Reference test Optitrack Rot/lfl/ fl/ex Warming-up: 1 measurement 1 number of measurements simultaneously, blinding ex-pt?	Pearson r FL 0.98 (95% CI 0.97–0.99) EXT 0.99 (95% CI 0.98–0.99) LFL R 0.91 (95% CI 0.85–0.95) LFL L 0.89 (95% CI 0.82–0.93) ROT R 0.89 (95% CI 0.81–0.94) ROT L 0.94 (95% CI 0.90–0.97)
Tucci [73] Universal goniometer gravity goniometer with head adapter	Interobserver Reliability	Ten healthy subjects	?	pt seated FL/EXT/LFL pt supine ROT Warming-up several trails test-retest interval within one hour, blinding ex-pt?	ICC Inter, UG 2 experienced examiners Fl –0.08 (NS) Ext 0.82 ($P < 0.05$) ROT r 0.52 ($P < 0.05$) ROT L 0.604 ($P < 0.05$) LFL r 0.21 (NS) LFL L 0.33(NS) Interobserver, gravity goniometer and head adapter, 1 experienced and 1 novice examiner Fl 0.84 ($P < 0.01$) Ext 0.86 ($P < 0.01$) ROT R 0.80 ($P < 0.01$) ROT L 0.91 ($P < 0.01$) LFL R 0.87 ($P < 0.01$) LFL L 0.82 ($P < 0.01$)
Viitanen [78] Myrin goniometer	Reliability	Fifty-two male patients with ankylosing spondylitis	Physiotherapists	fl/ext/lfl/rot Warming-up no, number measurements 2 test-retest interval on successive days, blinding ex-pt?	ICC ranged 0.89–0.98, no other details
Youdas [84] CROM	Intra and interobserver Reliability	Sixty subjects orthopaedic disorders, mainly cervical muscle pain	Eleven physical therapists	Subject seated fl/ext/lfl/rot Warming up three repetitions two measurements each movement, half cycle ROM Test-retest interval one directly after the other, blinding ex-pt?	ICC Intra Fl 0.95 EXT 0.90 LFL R 0.92 LFL L 0.84 ROT R 0.93 ROT L 0.90 Inter Fl 0.86 EXT 0.86 LFL R 0.88 LFL L 0.73 ROT R 0.92 ROT L 0.82

Table 2 continued

Author and tool	Purpose	Patient	Examiner	Procedure/protocol	Results
Youdas [85] CROM	Intra and interobserver Reliability	Healthy subjects intraobserver reliability 6 subjects, interobserver reliability 20 subjects	5 physical therapists, faculty members school for physical therapy	Subject seated FL/EXT/LFL/ROT Warming up three repetitions 2 measurements each movement until AROM stopped by muscle tightness/pain/substitution movement half cycle ROM Test-retest interval one directly after the other, blinding ex-pt?	ICC (range) Intra FI 0.23–0.88 EXT 0.89–0.96 LFL R 0.81–0.95 LFL L 0.67–0.90 ROT R 0.58–0.99 ROT L 0.81–0.95 Inter FI 0.83 EXT 0.90 LFL R 0.87 LFL L 0.89 ROT R 0.82 ROT L 0.66
Wainner [80] inclinometer	Interobserver Reliability Agreement	Eighty-two subjects, 18 cervico radiculopathy, 28 carpal tunnel syndrome, 36 healthy subjects	Two physical therapists	Warming-up two repetitions. Patient seated FL/EXT/LFL Test-retest interval one directly after the other, blinding ex-pt yes	ICC Inter (95%CI) FI 0.79(0.65–0.88) EXT 0.84 (0.70–0.95) LFL R 0.68 (0.62–0.87)LF L L 0.63 (0.40–0.78) SEM FL 4.6° EXT 4.8° LFL L 5.3° LFL R 5.4°
Zachman [86] Goniometer Rangiometer	Interobserver Reliability and Validity	Twenty-four healthy subjects	Two chiropractors	Subject seated FL/EXT/LFL/ROT Warming-up three repetitions Test-retest interval? Blinding ex-pt?	Pearson r inter rangiometer FI 0.64 EXT 0.89 ROT R 0.62 ROT L 0.69 LFL R 0.84 LFL L 0.79 Goniometer FI 0.54 EXT 0.85 ROT R 0.52 ROT L 0.47 LFL R 0.43 LFL L 0.61 Concurrent validity: r 0.31–0.83

ICC Intraclass correlation coefficient, FI flexion, EXT extension, ROT R rotation right, ROT L rotation left, LFL R lateral flexion right, LFL L lateral flexion left, Intra intraobserver, Inter interobserver, ROM range of motion, Pearson r Pearson correlation coefficient, SEM standard error of measurement, CROM cervical range of motion instrument, LoA limits of agreement, 95%CI 95% confidence interval, EX examiner, Pt patient

Table 3 Characteristics of included studies as regards visual estimation

Author and tool	Purpose	Patient	Examiner	Procedure/protocol	Results
Hoppenbrouwers [34] Visual estimation	Interobserver Reliability	Sixty-nine subjects, healthy and non-specific neck pain	Twenty-four-year physiotherapy students	Active ROM, as described by Dos Winkel, warming up one repetition, 1 measurement classification restricted or not test-retest interval? blinding ex-pt yes	K inter Fl 0.57 Ext 0.88 ROT (l/r) 0.49 (L: 0.43/R: 0.54) LFL (l/r) 0.35 (L: 0.33/R: 0.36) All movements Together 0.52
Viikari-Juntura [77] Visual estimation	Interobserver Reliability	Fifty-two patients neck and radicular pain	Physician in physical medicine and rehabilitation and physical therapist	Warming-up active ROM Fl/ext normal >45° limited 30–45° markedly limited <30° ROT normal >80° limited 60–80° markedly limited <60° LFL normal >30° limited 20°–30° markedly limited <20° test-retest interval 1 h, blinding ex-pt yes	Kw inter Fl 0.43 Ext 0.56 ROT R 0.56 ROT L 0.40 LFL R 0.51 LFL L 0.41

ROM Range of motion, FL flexion, EXT extension, ROT rotation, LFL lateral flexion, K Kappa statistics, Kw weighted Kappa

Tape measurement

Neck mobility was measured with tape in centimetres with different landmarks as reference marks. Different measurement protocols were found. Two studies described the Pearson correlation coefficient [8, 36] as a statistic measure for intra- and inter-observer reliability. Information on reliability, described by the ICC, agreement and responsiveness were found for patients with ankylosing spondylitis [48, 78]. Tape measurement was rated as doubtful for reproducibility and responsiveness.

Overall quality

The rating of the clinimetric qualities of the instruments included is presented in Table 5, summarizing each aspect as positive, inadequate, doubtful or insufficient quality. Only a few studies gave an adequate description of the study design and population characteristics. Eight studies included only patients with non-specific neck pain; twenty publications provided insufficient information on the methodological aspects to enable a good appraisal of the study design. Furthermore, information about non-response and subjects' loss to follow-up was often lacking.

Discussion

An extensive search strategy led to the identification of three different types of instruments for the evaluation of active cervical ROM which are easily applicable in daily clinical practice and for which clinimetric properties have been investigated. Overall, the CROM, a single inclinometer and EDI-320 had the best ratings on such clinimetric aspects as reproducibility, validity and responsiveness. When clinical acceptability is taken into account the CROM or a single inclinometer can be considered to be the

most appropriate instruments for the assessment of ACROM in patients with non-specific neck pain. In other patients groups, as for example in patients with cervical disc prostheses the CROM can possibly be used to measure follow-up of range of motion. The authors believe that the CROM may be a cheap and safe alternative instrument instead of radiography. Radiography then only has to be used in those patients in which ACROM does not improve as was expected.

The CROM had been studied most extensively. None of the instruments received positive ratings for all items of the methodological quality checklist. It has been advocated that agreement parameters are required for instruments that are used for evaluative purposes and reliability parameters are required for instruments that are used for discriminative purposes [21]. Instruments used to measure ACROM are mainly used for evaluative purposes, but reliability parameters have been studied more extensively than agreement parameters. Agreement, interpretability and responsiveness are clinimetric properties, which in general have not been tested. For evaluative purposes, these clinimetric properties are important because measurement error should be smaller than the minimal change that is considered to be important [67]. Parameters were only found for the single inclinometer and the EDI-320 agreement in studies that have an adequate study design. MDC values of a single inclinometer ranged from 10° (lateral flexion) to 16° (flexion/extension) [20, 59, 80].

The methodological quality of the studies included varied. In total we included 33 publications. In general, the methodological quality of the older literature (published before 1995) is lower than the more recent literature. Older articles mainly describe reliability and eight of those articles did not use the ICC as a statistical measurement for reliability [3, 8, 16, 36, 37, 41, 51, 86].

In order to ensure external validity it is necessary to include patients with neck pain who are likely to undergo

Table 4 Characteristics of included studies as regards tape measurement

Author plus tool	Purpose	Patient	Examiner	Procedure/protocol	Results
Balogun [8] Tape	Intra and Interobserver Reliability	Twenty-one healthy subjects	Three physical therapist	Subject seated FL/EXT/LFL/ROT Protocol tape as in Hsieh test-retest interval 4 days blinding ex-pt?	Intra Pearson r FL 0.26–0.49 EXT 0.72–0.88 ROT R 0.59–0.86 ROT L 0.58–0.83 LFL R 0.53–0.86 LFL L 0.65–0.77 Inter Pearson r FL 0.30–0.80 EXT 0.86–0.92 ROT R 0.74–0.87 ROT L 0.68–0.85 LFL R 0.51–0.79 LFL L 0.47–0.82
Hsieh [36] Tape	Intraobserver Reliability and Agreement	Thirty-four healthy subjects	Physical therapist and physical therapist assistant	Subject seated warming-up? Two repetitions FI-EXT distance sternal notch-chin ROT distance acromion process-chin LFL distance acromion process-lowest point ear lobe Test-retest interval one direct after each other, blinding ex-pt no	Intra Pearson r FL 0.95 0.86 Ext 0.94 0.79 ROT R 0.88 0.78 ROT L 0.80 0.81 LFL R 0.88 0.91 LFL L 0.87 0.86 SEM (tester1 & 2) FI 1.4 & 1.91 EXT 1.18 & 3.30 ROT R 1.97 & 2.95 ROT L 2.55 & 2.75 LFL R 1.77 & 2.25 LFL L 2.05 2.69
Maksymowich [48] Tape	Intra and Interobserver Reliability, agreement, validity, responsiveness	Patients with ankylosing spondylitis 44 for reliability, 33 for responsiveness	Clinician nurse, rheumatologist, trained medical student	Pen mark on suprasternal notch, pt rotate head, distance pen mark and tragus of the ear is recorded. goniometer centre of the head, one arm moved over the plane of the nose as the pt rotates. Score 0,1,2: rotation>70, 20–70, <20. Test-retest interval? blinding ex-pt?	ICC Intra tape 0.80 & 0.89 Inter tape 0.82 LoA tape – 1.13 to 1.48 Spearman Rho tape Age –0.18 (0.004) Disease duration – 0.28(<0.001) BASDAI 0.04 (NS) Responsiveness ES 0.15 SRM 0.26 P 0.14
Viitanen [78] Tape	Reliability	Fifty-two male patients with ankylosing spondylitis	Physiotherapists	Warming-up no LFL distance ear lobe-tuberculum coronoideus clavicular ROT distance chin top- tuberculum coronoideus clavicular test-retest interval on successive days, blinding ex-pt?	ICC ranged 0.89–0.98, no other details

Pearson R Pearson correlation coefficient, FL flexion, EXT extension, ROT R rotation right, ROT L rotation left, LFL L lateral flexion left, PT patient, SEM standard error of measurement, ICC intraclass correlation coefficient, ES effect size, SRM standardized response mean



Fig. 1 Inclinometer

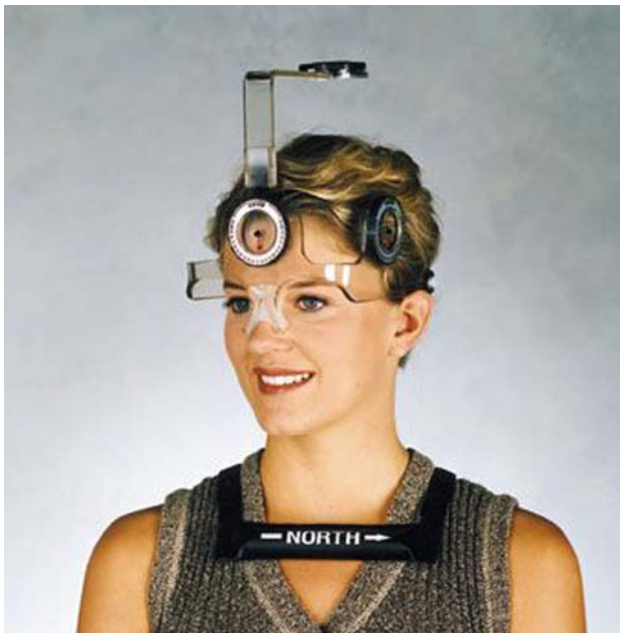


Fig. 2 Cervical range of motion instrument

the same measurement procedure in daily practice [13]. Nineteen articles used healthy subjects [2, 3, 8, 15, 16, 32, 33, 36, 51, 53–56, 68, 69, 71, 73, 85, 86], 4 articles used patients other than non-specific neck pain patients [37, 48, 78, 80] and 9 articles included patients with non-specific neck pain [20, 34, 35, 41, 59, 61, 70, 77, 84].

Thirteen publications did not describe the training and the results of the training undergone by those doing the rating prior to the test, which is however an important aspect of external validity [74]. Blinding is an important aspect of the internal validity of a study. It can be divided

into examiner–participant blinding and examiner–examiner blinding. Blinding of examiner–participant is often not described in the publications included in this review. The aspect of blinding examiner–examiner is given more consideration in the publications.

This review has several limitations. Although much effort was made to find all the published studies, selection bias may have occurred because only Dutch or English-language articles were included. Effort was put into reference tracking but it is possible that studies were missed. Furthermore, unpublished studies were not included. Reviewer bias is also a possible limitation of this review because reviewers were not blinded to the authors.

There is no gold standard for evaluating clinimetric qualities. The checklist used in our review was based on the checklist made by Bot et al. [11]. This list has been used previously for patient-assessed questionnaires instead of instruments to evaluate the functional status of the patient [11, 26, 76]. This checklist, however, was chosen for its quality and international consensus on terminology. Assigning value labels for ranges of Kappa and ICC statistics and correlation coefficients was done in accordance with other authors [29, 43, 66, 67].

Jordan [38] concluded in his review that the CROM has shown promise as regards reliability but its practicality for clinical use is questionable because of the costs involved and its dimensions. The author also concluded that evidence on the single inclinometer is lacking. Since the publication of this review new data have been published on the validity of the CROM [69–71] and on the reproducibility of the inclinometer [15, 59, 80]. These studies generally have a high methodological quality and show good reproducibility of the inclinometer and good construct validity of the CROM.

The findings of this systematic review have implications for research and clinical practice. Researchers should give careful consideration to the study design and presentation of the results. The construct validity of a single inclinometer should be investigated by making comparisons with other instruments to measure ACROM. Future research should also report agreement parameters. Clinicians need to be cognizant that ACROM should be measured with the CROM or a single inclinometer and that visual estimation

Table 5 Summary of the quality assessment of clinimetric properties of the instruments included

+ Positive rating, – inadequate rating, ? doubtful rating, O insufficient information

	Reliability	Agreement	Validity	Responsiveness	Interpretability
CROM (ten studies)	+	?	+	O	O
Universal goniometer (five studies)	?	?	?	?	O
Visual estimation (two studies)	–	O	O	O	O
Tape measure (four studies)	?	?	?	?	O
Single inclinometer (four studies)	+	+	?	O	O
EDI-320 (two studies)	+	+	O	+	O

is not reliable. Furthermore, in future research different patient groups have to be studied, including for example also patients with cervical disc prostheses, in order to validate CROM against radiography.

Conclusion

The present review provides information for researchers and clinicians to facilitate choice amongst existing instruments for measuring ACROM. A systematic computerized literature search of databases revealed three different types of instruments that are practical to use when measuring ACROM in patients with non-specific neck pain: visual estimation, tape measurements, different types of goniometers/inclinometers. When a healthcare professional decides that measuring ACROM on a patient with non-specific neck pain is necessary, a single inclinometer and CROM are to be recommended based on their best ratings for clinimetric properties and practicality. Visual estimation should not be used to measure ACROM.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Agarwal S, Allison GT, Singer KP (2005a) Reliability of the spin-T cervical goniometer in measuring cervical range of motion in an asymptomatic Indian population. *J Manipulative Physiol Ther* 28(7):487–492
- Agarwal S, Allison GT, Singer KP (2005b) Validation of the spin-T goniometer, a cervical range of motion device. *J Manipulative Physiol Ther* 28(8):604–609
- Alantara H, Hurri H, Heliövaara M et al (1994) Flexibility of the spine: normative values of goniometric and tape measurements. *Scan J Rehabil Med* 14:147–154
- Alund M, Larsson SE (1990) Three-dimensional analysis of neck motion; a clinical method. *Spine* 15:87–91
- Amiri M, Jull G, Bullock-Saxton J (2003) Measuring range of active cervical rotation in a position of full head flexion using the 3D Fastrak measurement system: an intra-tester reliability study. *Man Ther* 8(3):176–179
- Antonaci F, Bulgheroni M, Ghirmai S, Lanfranchi S, Dalla Toffola E, Sandrini G, Nappi G (2002) 3D kinematic analysis and clinical evaluation of neck movements in patients with whiplash injury. *Cephalalgia* 22(7):533–542
- Assink N, Bergman GJ, Knoester B, Winters JC, Dijkstra PU, Postema K (2005) Interobserver reliability of neck-mobility measurement by means of the flock-of-birds electromagnetic tracking system. *J Manipulative Physiol Ther* 28(6):408–413
- Balogun JA, Abereje OK, Olaogun MO, Obajuluwa VA (1989) Inter and intratester reliability of measuring neck motions with tape measure and Myrin gravity-reference goniometer. *J Orthop Sports Phys Ther* 10:248–253
- Bergman GJ, Knoester B, Assink N, Dijkstra PU, Winters JC (2005) Variation in the cervical range of motion over time measured by the “flock of birds” electromagnetic tracking system. *Spine* 30(6):650–654
- Bogduk N (1984) Neck pain. *Aust Fam Physician* 13(1):26–30
- Bot SD, Terwee CB, van der Windt DA, Bouter LM, Dekker J, de Vet HC (2004) Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. *Ann Rheum Dis* 63(4):335–341
- Bot SD, van der Waal JM, Terwee CB, van der Windt DA, Schellevis FG, Bouter LM, Dekker J (2005) Incidence and prevalence of complaints of the neck and upper extremity in general practice. *Ann Rheum Dis* 64(1):118–123
- Bruton A, Conway JH, Holgate ST (2000) Reliability: What is it, and how is it measured? *Physiotherapy* 86:94–99
- Bulgheroni MV, Antonaci F, Ghirmai S, Sandrini G, Nappi G, Pedotti A (1998) A 3D kinematic method for evaluating voluntary movements of the cervical spine in humans. *Funct Neurol* 13(3):239–245
- Bush KW, Collins N, Portman L, Tillet N (2000) Validity and intertester reliability of cervical range of motion using inclinometer measurements. *J Man Manipulative Ther* 8(2):52–61
- Capuano-Pucci D, Rheault W, Aukai J, Bracke M, Day R, Patrick M (1991) Intratester and intertester reliability of cervical range of motion device. *Arch Phys Med Rehabil* 18:338–340
- Chiu TT, Sing KL (2002) Evaluation of cervical range of motion and isometric neck muscle strength: reliability and validity. *Clin Rehabil* 16(8):851–858
- Christensen HW, Nilsson N (1998a) Natural variation of cervical range of motion: a one-way repeated-measures design. *J Manipulative Physiol Ther* 21(6):383–387
- Christensen HW, Nilsson N (1998b) The reliability of measuring active and passive cervical range of motion: an observer-blinded and randomized repeated-measures design. *J Manipulative Physiol Ther* 21(5):341–347
- Cleland JA, Childs JD, Fritz JM, Whitman JM (2006) Interrater reliability of the history and physical examination in patients with mechanical neck pain. *Arch Phys Med Rehabil* 87(10):1388–1395
- de Vet HC, Terwee CB, Knol DL, Bouter LM (2006) When to use agreement versus reliability measures. *J Clin Epidemiol* 59(10):1033–1039
- Dvir Z, Gal-Eshel N, Shamir B, Prushansky T, Pevzner E, Peretz C (2006) Cervical motion in patients with chronic disorders of the cervical spine: a reproducibility study. *Spine* 31(13):E394–E399
- Dvir Z, Prushansky T (2000) Reproducibility and instrument validity of a new ultrasonography-based system for measuring cervical spine kinematics. *Clin Biomech (Bristol, Avon)* 15(9):658–664
- Dvorak J, Antinnes JA, Panjabi M, Loustalot D, Bonomo M (1992) Age and gender related normal motion of the cervical spine. *Spine* 17(10 Suppl):S393–S398
- Dvorak J, Panjabi MM, Grob D, Novotny JE, Antinnes JA (1993) Clinical validation of functional flexion/extension radiographs of the cervical spine. *Spine* 18(1):120–127
- Eecheute C, Vaes P, Van Aerschot L, Asman S, Duquet W (2007) The clinimetric qualities of patient-assessed instruments for measuring chronic ankle instability: a systematic review. *BMC Musculoskelet Disord* 8:6
- Fejer R, Kyvik KO, Hartvigsen J (2006) The prevalence of neck pain in the world population: a systematic critical review of the literature. *Eur Spine J* 15(6):834–848
- Ferrario VF, Sforza C, Serrao G, Grassi G, Mossi E (2002) Active range of motion of the head and cervical spine: a three-dimensional investigation in healthy young adults. *J Orthop Res* 20(1):122–129
- Fitzpatrick R, Davey C, Buxton MJ, Jones DR (1998) Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess* 2(14):i–iv (1–74)

30. Frobin W, Leivseth G, Biggemann M, Brinckmann P (2002) Sagittal plane segmental motion of the cervical spine. A new precision measurement protocol and normal motion data of healthy adults. *Clin Biomech (Bristol, Avon)* 17(1):21–31
31. Haas M (1991) Statistical methodology for reliability studies. *J Manipulative Physiol Ther* 14(2):119–132
32. Haynes MJ, Edmondston S (2002) Accuracy and reliability of a new, protractor-based neck goniometer. *J Manipulative Physiol Ther* 25(9):579–586
33. Hole DE, Cook JM, Bolton JE (1995) Reliability and concurrent validity of two instruments for measuring cervical range of motion: effects of age and gender. *Man Ther* 1(1):36–42
34. Hoppenbrouwers M, Eckhardt MM, Verkerk K, Verhagen A (2006) Reproducibility of the measurement of active and passive cervical range of motion. *J Manipulative Physiol Ther* 29(5):363–367
35. Hoving JL, Pool JJ, van Mameren H, Deville WJ, Assendelft WJ, de Vet HC, de Winter AF, Koes BW, Bouter LM (2005) Reproducibility of cervical range of motion in patients with neck pain. *BMC Musculoskelet Disord* 6:59
36. Hsieh CY, Yeung BW (1986) Active neck motion measurements with a tape measure. *J Orthop Sports Phys Ther* 8:88–92
37. Jenkinson TR, Mallorie PA, Whitelock HC, Kennedy LG, Garrett SL, Calin A (1994) Defining spinal mobility in ankylosing spondylitis (AS). The Bath AS Metrology Index. *J Rheumatol* 21(9):1694–1698
38. Jordan K (2000) Assessment of published reliability studies for cervical spine range-of-motion measurement tools. *J Manipulative Physiol Ther* 23(3):180–195
39. Jordan K, Dziedzic K, Jones PW, Ong BN, Dawes PT (2000) The reliability of the three-dimensional FASTRAK measurement system in measuring cervical spine and shoulder range of motion in healthy subjects. *Rheumatology (Oxford)* 39(4):382–388
40. Jordan K, Haywood KL, Dziedzic K (2004) Assessment of the 3-dimensional Fastrak measurement system in measuring range of motion in ankylosing spondylitis. *J Rheumatol* 31(11):2207–2215
41. Kadir N, Grayson MF, Goldberg AA, Swain MC (1981) A new goniometer. *Rheumatol Rehabil* 20(4):219–226
42. Koerhuis CL, Winters JC, van der Helm FC, Hof AL (2003) Neck mobility measurement by means of the 'Flock of Birds' electromagnetic tracking system. *Clin Biomech (Bristol, Avon)* 18(1):14–18
43. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33(1):159–174
44. Lantz CA, Chen J, Buch D (1999) Clinical validity and stability of active and passive cervical range of motion with regard to total and unilateral uniplanar motion. *Spine* 24(11):1082–1089
45. Lantz CA, Klein G, Chen J, Mannion A, Solinger AB, Dvorak J (2003) A reassessment of normal cervical range of motion. *Spine* 28(12):1249–1257
46. Lohr KN, Aaronson NK, Alonso J, Burnam MA, Patrick DL, Perrin EB, Roberts JS (1996) Evaluating quality-of-life and health status instruments: development of scientific review criteria. *Clin Ther* 18(5):979–992
47. Love S, Gringmuth RH, Kazemi M, Cornacchia P, Schmolke M (1998) Interexaminer and intraexaminer reliability of cervical passive range of motion using the CROM and Cybex 320 EDI. *J Can Chiropract Assoc* 42(4):222–228
48. Maksymowich WP, Mallon C, Richardson R, Conner-Spady B, Jauregui E, Chung C, Zappala L, Pile K, Russell AS (2006). Development and validation of a simple tape-based measurement tool for recording cervical rotation in patients with ankylosing spondylitis: comparison with a goniometer-based approach. *J Rheumatol* 33(11):2242–2249
49. Malmstrom EM, Karlberg M, Melander A, Magnusson M (2003) Zebris versus Myrin: a comparative study between a three-dimensional ultrasound movement analysis and an inclinometer/compass method: intratester reliability, concurrent validity, intertester comparison, intratester reliability, and intraindividual variability. *Spine* 28(21):E433–E440
50. Mannion AF, Klein GN, Dvorak J, Lanz C (2000) Range of global motion of the cervical spine: intraindividual reliability and the influence of measurement device. *Eur Spine J* 9(5):379–385
51. Moffett JAK, Hughes I, Griffiths P (1989) Measurement of cervical spine movements using a simple inclinometer. *Physiotherapy* 19:302–305
52. Morphett AL, Crawford CM, Lee D (2003) The use of electromagnetic tracking technology for measurement of passive cervical range of motion: a pilot study. *J Manipulative Physiol Ther* 26(3):152–159
53. O'Driscoll SL, Tomenson J (1982) The cervical spine. *Clin Rheum Dis* 8(3):617–630
54. Olson SL, O'Connor DP, Birmingham G, Broman P, Herrera L (2000) Tender point sensitivity, range of motion, and perceived disability in subjects with neck pain. *J Orthop Sports Phys Ther* 30(1):13–20
55. Ordway NR, Seymour R, Donelson RG, Hojnowski L, Lee E, Edwards WT (1997) Cervical sagittal range-of-motion analysis using three methods. Cervical range-of-motion device, 3space, and radiography. *Spine* 22(5):501–508
56. Pellechia GL, Bohannon RW (1998) Active lateral flexion range of motion measurements obtained with a modified goniometer, reliability and estimates of normal. *J Manipulative Physiol Ther* 21:443–447
57. Peollson A, Hedlund R, Ertzgaard S, Oberg B (2000) Intra- and intertester reliability and range of motion of the neck. *Physiother Can* 52(3):233–242
58. Petersen CM, Johnson RD, Schuit D (2000) Reliability of cervical range of motion using the OSI CA 6000 spine motion analyser on asymptomatic and symptomatic subjects. *Man Ther* 5(2):82–88
59. Piva SR, Erhard RE, Childs JD, Browder DA (2006) Inter-tester reliability of passive intervertebral and active movements of the cervical spine. *Man Ther* 11(4):321–330
60. Pringle RK (2003) Intra-instrument reliability of 4 goniometers. *J Chiropract Med* 3(2):91–95
61. Rheault W, Albright B, Byers C, Franta M, Johnson A, Skowronek M (1992) Intertester reliability of the cervical range of motion device. *J Orthop Sports Phys Ther* 15(3):147–150
62. Seffinger MA, Najm WI, Mishra SI, Adams A, Dickerson VM, Murphy LS, Reinsch S (2004) Reliability of spinal palpation for diagnosis of back and neck pain: a systematic review of the literature. *Spine* 29(19):E413–E425
63. Solinger AB, Chen J, Lantz CA (2000) Standardized initial head position in cervical range-of-motion assessment: reliability and error analysis. *J Manipulative Physiol Ther* 23(1):20–26
64. Stochkendahl MJ, Christensen HW, Hartvigsen J, Vach W, Haas M, Hestbaek L, Adams A, Bronfort G (2006) Manual examination of the spine: a systematic critical literature review of reproducibility. *J Manipulative Physiol Ther* 29(6):475–485 (485 e471–e410)
65. Strimpakos N, Sakellari V, Gioufios G, Papathanasiou M, Broutzos E, Kelekis D, Kapreli E, Oldham J (2005) Cervical spine ROM measurements: optimizing the testing protocol by using a 3D ultrasound-based motion analysis system. *Cephalalgia* 25(12):1133–1145
66. Swinkels RA, Bouter LM, Oostendorp RA, van den Ende CH (2005) Impairment measures in rheumatic disorders for rehabilitation medicine and allied health care: a systematic review. *Rheumatol Int* 25(7):501–512
67. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC (2007) Quality criteria were

- proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 60(1):34–42
68. Tousignant M, Boucher N, Bourbonnais J, Gravelle T, Quesnel M, Brosseau L (2001) Intratester and intertester reliability of the Cybex electronic digital inclinometer (EDI-320) for measurement of active neck flexion and extension in healthy subjects. *Man Ther* 6(4):235–241
 69. Tousignant M, de Bellefeuille L, O'Donoghue S, Grahovac S (2000) Criterion validity of the cervical range of motion (CROM) goniometer for cervical flexion and extension. *Spine* 25(3):324–330
 70. Tousignant M, Duclos E, Lafleche S, Mayer A, Tousignant-Lafamme Y, Brosseau L, O'Sullivan JP (2002) Validity study for the cervical range of motion device used for lateral flexion in patients with neck pain. *Spine* 27(8):812–817
 71. Tousignant M, Smeesters C, Breton AM, Breton E, Corriveau H (2006) Criterion validity study of the cervical range of motion (CROM) device for rotational range of motion on healthy adults. *J Orthop Sports Phys Ther* 36(4):242–248
 72. Trott PH, Percy MJ, Ruston SA, Fulton I, Brien C (1996) Three-dimensional analysis of active cervical motion: the effect of age and gender. *Clin Biomech (Bristol, Avon)* 11(4):201–206
 73. Tucci SM, Hicks JE, Gross EG, Campbell W, Danoff J (1986) Cervical motion assessment: a new, simple and accurate method. *Arch Phys Med Rehabil* 67(4):225–230
 74. Van Genderen FR, de Bie RA, Helders PJM, van Meeteren NLU (2003) Reliability research: towards a more clinically relevant approach. *Phys Ther Rev* 8:169–176
 75. van Trijffel E, Anderegg Q, Bossuyt PM, Lucas C (2005) Inter-examiner reliability of passive assessment of intervertebral motion in the cervical and lumbar spine: a systematic review. *Man Ther* 10(4):256–269
 76. Veenhof C, Bijlsma JW, van den Ende CH, van Dijk GM, Pisters MF, Dekker J (2006) Psychometric evaluation of osteoarthritis questionnaires: a systematic review of the literature. *Arthritis Rheum* 55(3):480–492
 77. Viikari-Juntura E (1987) Interexaminer reliability of observations in physical examinations of the neck. *Phys Ther* 67(10):1526–1532
 78. Viitanen JV, Kokko ML, Heikkila S, Kautiainen H (1998) Neck mobility assessment in ankylosing spondylitis: a clinical study of nine measurements including new tape methods for cervical rotation and lateral flexion. *Br J Rheumatol* 37(4):377–381
 79. Vos C, Verhagen A, Passchier J, Koes B (2007) Management of acute neck pain in general practice: a prospective study. *Br J Gen Pract* 57(534):23–28
 80. Wainner RS, Fritz JM, Irrgang JJ, Boninger ML, Delitto A, Allison S (2003) Reliability and diagnostic accuracy of the clinical examination and patient self-report measures for cervical radiculopathy. *Spine* 28(1):52–62
 81. Wang SF, Teng CC, Lin KH (2005) Measurement of cervical range of motion pattern during cyclic neck movement by an ultrasound-based motion system. *Man Ther* 10(1):68–72
 82. Wolfenberger VA, Bui Q, Batenchuk GB (2002) A comparison of methods of evaluating cervical range of motion. *J Manipulative Physiol Ther* 25(3):154–160
 83. Wu SK, Lan HH, Kuo LC, Tsai SW, Chen CL, Su FC (2007) The feasibility of a video-based motion analysis system in measuring the segmental movements between upper and lower cervical spine. *Gait Posture* 26(1):161–166
 84. Youdas JW, Carey JR, Garrett TR (1991) Reliability of measurements of cervical spine range of motion—comparison of three methods. *Phys Ther* 71(2):98–104 (discussion 105–106)
 85. Youdas JW, Garrett TR, Suman VJ, Bogard CL, Hallman HO, Carey JR (1992) Normal range of motion of the cervical spine: an initial goniometric study. *Phys Ther* 72(11):770–780
 86. Zachman ZJ, Traina AD, Keating JC, Bolles ST, Braun-Porter L (1989) Interexaminer reliability and concurrent validity for the measurement of cervical ranges of motion. *J Manipulative Physiol Ther* 12:205–210