Reliability and measurement error of sensorimotor tests in patients with neck pain: a systematic review

Simone Elsig[1,2,3,4*] , Lara Allet[3,4,5], Caroline Henrice Germaine Bastiaenen[2], Rob de Bie[2] and Roger Hilfiker[1]

Abstract

Background Neck pain is one of the leading causes of years lived with disability, and approximately half of people with neck pain experience recurrent episodes. Deficits in the sensorimotor system can persist even after pain relief, which may contribute to the chronic course of neck pain in some patients. Evaluation of sensorimotor capacities in patients with neck pain is therefore important. No consensus exists on how sensorimotor capacities of the neck should be assessed in physiotherapy. The aims of this systematic review are: (a) to provide an overview of tests used in physiotherapy for assessment of sensorimotor capacities in patients with neck pain; and (b) to provide information about reliability and measurement error of these tests, to enable physiotherapists to select appropriate tests.

Methods Medline, CINAHL, Embase and PsycINFO databases were searched for studies reporting data on the reliability and/or measurement error of sensorimotor tests in patients with neck pain. The results for reliability and measurement error were compared against the criteria for good measurement properties. The quality of evidence was assessed according to the modified GRADE method proposed by the COSMIN group.

Results A total of 206 tests for assessment of sensorimotor capacities of the neck were identified and categorized into 18 groups of tests. The included tests did not cover all aspects of the sensorimotor system; tests for the sensory and motor components were identified, but not for the central integration component. Furthermore, no data were found on reliability or measurement error for some tests that are used in practice, such as movement control tests, which apply to the motor component. Approximately half of the tests showed good reliability, and 12 were rated as having good (+) reliability. However, tests that evaluated complex movements, which are more difficult to standardize, were less reliable. Measurement error could not be evaluated because the minimal clinically important change was not available for all tests.

Conclusion Overall, the quality of evidence is not yet high enough to enable clear recommendations about which tests to use to assess the sensorimotor capacities of the neck.

Keywords Neck pain, Sensorimotor tests, Reliability

*Correspondence:
Simone Elsig
simone.elsig@gmail.com
Full list of author information is available at the end of the article

Elsig *et al. Archives of Physiotherapy*   (2023) 13:15

Page 2 of 16

Introduction

Neck pain is the second most common musculoskeletal problem [1]. It is one of the leading causes of years lived with disability worldwide and represents an increasing burden on healthcare systems [2–4]. The economic burden of neck pain, in terms of treatment costs, lost productivity and work-related problems is high [1]. The point prevalence of neck pain in different countries ranges from 2443.9 to 6151.2 cases per 100,000 population, with the highest values in western Europe [1, 5]. The mean percentages for one-year prevalences and lifetime prevalences of adults worldwide are 37.2% and 48.5%, respectively [6]. Although acute neck pain usually resolves within two months, approximately 50% of patients are not completely pain free one year after an episode of neck pain [7–9]. This illustrates the often chronic-episodic course of the condition, with patients experiencing persistent or recurrent episodes of neck pain [10].

Management of patients with neck pain is a major challenge in physiotherapy, mainly because these patients form a very heterogeneous group in terms of the nature of symptoms, symptom distribution, and underlying pain mechanisms [11]. As neck pain is a multidimensional condition, management should consider multiple factors (e.g. pain mechanisms, and psychological, biological, movement and work-related factors). Among the work-related factors, workload, work or study time, sustained postures or body positions during work and computer work are considered as risk factors for the development of neck pain [1, 12]. The different factors can interact, and their expression may be more or less dominant in each patient, thus influencing the clinical approach [1, 12].

Deficits of sensorimotor capacities (SC) may be one of the factors contributing to neck pain, in particular the persistence or recurrence of neck pain [13]. The sensorimotor system is defined as an integrated whole, comprising afferent and efferent information, with central integration and processing components necessary to provide functional joint stability [14]. It is thought to influence, among others, joint position sense, activation of cervical flexor muscles and control of head-eye movement. The SC of the cervical spine are related to neck pain [15] and patients with neck pain often demonstrate reduced SC, e.g. reduced joint position sense [16–18], altered activation patterns of the cervical muscles [19–21], or disturbed head-eye movement control [22]. Furthermore, the persistence of deficits in the sensorimotor system can continue even after pain relief. It is hypothesized that persistence of these deficits may contribute to some patients experiencing recurrent episodes of neck pain [23–25] and the integration of sensorimotor training in the management of patients with neck pain has shown promising results [13]. Therefore, evaluation of SC in patients with neck pain is important [26]. Various tests to evaluate the sensorimotor system have been developed and are widely used in physiotherapy practice and research. However, the terminology used is often confusing, and there is no consensus on how SC of the neck should be assessed [14, 27]. Systematic reviews of tests for SC of the neck have investigated only a limited selection of tests assessing single aspects of SC, such as joint position sense [28] or muscle function [29–31]. A systematic review, providing a comprehensive overview of all available tests to assess all different aspects of SC of the neck, is lacking.

Given that many tests exist for assessment of SC of the neck, the challenge is to choose the most appropriate test for use in a specific situation. From a scientific perspective, knowledge about the quality of a test, i.e. measurement properties, is important when making this decision. The quality of a test depends on three criteria: reliability, validity and responsiveness [32].

This systematic review investigates the domain reliability. Reliability is the degree to which measurements are free from measurement error. The domain reliability includes three measurement properties: reliability, (expressing the proportion of the total variance in the measurements which is due to 'true' differences between patients), measurement error (which is the systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured), and internal consistency [32]. Internal consistency is usually investigated in self-reporting multi-item questionnaires and therefore is not relevant for the single-item tests used to assess SC.

The aim of this systematic review is to include all tests assessing any aspect of SC of the neck. Therefore, since many different tests are described in the literature, this review focusses only on reliability. Of course, when

Elsig *et al. Archives of Physiotherapy*   (2023) 13:15

Page 3 of 16

deciding which test to use, it would also be important to consider the different aspects of validity.

The concepts of reliability and measurement error are related, but focus on different purposes. Reliability focusses on the variability between patients or measurements and is influenced by the variation in the population where the test is used. On the other hand, measurement error is a relevant parameter for measurement of change over time, and it is not affected by population variability [33]. In clinical practice physiotherapists are interested in both concepts. The distinction between patients with and without deficits in the sensorimotor system (diagnostic purpose) is important. But measurement error is also an issue, as change over time, i.e. the evolution of the patient's symptoms, is of interest.

The aims of this systematic review are: (a) to provide an overview of tests used in physiotherapy to assess SC in patients with neck pain; and (b) to provide information about the reliability and measurement error of these tests, to enable physiotherapists to select appropriate tests.

Methods

Design

A meta-analysis of studies investigating the reliability and measurement error of tests assessing SC of patients with neck pain in a physiotherapy setting.

Search strategy

The databases CINAHL, Embase and PsycINFO were searched up to July 2020 and for Medline up to May 2021. Blocks of search terms were developed for: (a) construct of interest (sensorimotor capacities), (b) population (patients with neck pain), (c) the sensitive PubMed filter developed by Terwee et al. [34] for the identification of studies about measurement properties of measurement instruments, and (d) the exclusion filter proposed by Terwee et al. [34] to exclude irrelevant studies. The two filters were adapted to the other databases, adopting the strategy used by Ammann-Reiffer et al. [35]. There was no language restriction. The reference lists of systematic reviews retrieved were hand searched for further eligible studies. The detailed search strategy is shown in Additional file 1.

Selection process

Two reviewers (SE and either RH or MT) screened the titles and abstracts independently, based on the predefined inclusion and exclusion criteria listed in Fig. 1. Disagreements were discussed and, if necessary, a third reviewer (CB) made a decision regarding inclusion. Reviewers in the team were able to read English, German,

Dutch, French, Danish and Norwegian, and no exclusion of relevant papers based on language was noted.

Full-text screening was performed independently by two researchers (SE and RH) using the same predefined criteria (Fig. 1). After each screening step (title/abstract and full text), in the case of any disagreement about inclusion, consensus was reached through discussion with a third reviewer (CB). The screening was carried out using Covidence systematic review software [36].

Data extraction

Data extraction was conducted using REDCap electronic data capture tools hosted at the University of Applied Sciences and Arts Western Switzerland (HES-SO) Valais [37] by SE and RH. The first five studies were checked by a third researcher (CB) to ensure the correct procedure. Data were extracted on study characteristics, reliability, and measurement error of the different tests. Two researchers (SE and RH) assessed methodological quality, applying the COSMIN risk of bias tool in the adapted version for clinician-reported or performance-based outcome measures [38]. Each criterion was rated on a four-point rating system (i.e. very good, adequate, doubtful, or inadequate). The lowest rating determined the overall rating of the study (worst-score-counts method). The detailed tables for risk of bias assessment are shown in Additional file 2 (reliability) and Additional file 3 (measurement error). A third researcher (CB) performed a check of the first studies. Data extraction and synthesis was conducted with all included studies regardless of their methodological quality.

Data synthesis and analysis

The intraclass correlation coefficients of studies that used the same device and similar instructions for the corresponding test were quantitatively pooled. When pooling was not feasible, the study results were qualitatively summarized, by reporting the lowest and highest values. Because of the large number of tests applied for different directions of movement of the neck (left rotation, right rotation, etc.) test directions were summarized with reliability or measurement error values that led to the same conclusion regarding the criteria for good measurement properties, with the lowest and highest value. Tests directions with very different values (i.e. when the conclusion about the appropriateness of the reliability or the measurement error for this direction would be different from that for other directions) were reported separately.

The overall results for the reliability and/or measurement error of single studies or of summarized or pooled studies were compared against the criteria for good measurement properties. In a next step, the quality of evidence was graded according to the modified GRADE

Inclusion criteria
o Study examined test/s evaluating sensorimotor capacities of the neck that can be used in physiotherapy practice
o It is possible to perform the tests in a physiotherapy setting (e.g. tests requiring force plates or motion capture systems are included)
o Study population: persons with neck pain, at least 18 years old. (Neck pain is defined as pain perceived in the anatomical region of the neck with or without radiation to the upper limbs, trunk, and head)
o Include cervicogenic headache and tension type headache only when neck pain is a main complaint and when headache is mainly provoked by a problem in the neck (e.g. "Neck pain with headache")
o The study investigated either reliability or measurement error or both. Regarding reliability, studies were included if they assessed one or more aspects of reliability, i.e. test-retest reliability, inter-rater or intra-rater reliability, expressed as intra-class correlations (ICC), Spearman, Pearson, or kappa coefficients. Studies designed to investigate measurement error should report the standard error of measurement (SEM) and/or limits of agreement.
o Study is available as a full text published in a peer-reviewed journal
Exclusion criteria
o Tests evaluating sensorimotor capacities that are usually not available in physiotherapy (e.g. tests requiring X-ray or magnetic resonance imaging (MRI) are excluded)
o Healthy subjects
o Study population under 18 years of age
o Symptoms primarily in the head or not explicitly related to a problem in the neck
o Questionnaires assessing sensorimotor capacities of the neck
o Diseases of central neural system (e.g. multiple sclerosis, myelopathy, Parkinson's disease, etc.)
o Diseases of peripheral neural system (e.g. peripheral neuropathy, polyneuropathy, cervical radiculopathy, nerve compressions, such as carpal tunnel syndrome or thoracic outlet syndrome)
o Dystonia, spasmodic torticollis, Ménières disease, Huntington's disease
o Vestibular diseases, vestibular dizziness in relation to neck pain is not an exclusion criterion
o Migraine
o Case-reports, single-case studies, abstracts of conference papers

Fig. 1 Criteria for inclusion or exclusion of studies

Elsig *et al. Archives of Physiotherapy*   (2023) 13:15

Page 5 of 16

method proposed by the COSMIN group [38]. The quality of evidence was classified as high, moderate, low, or very low. The score was downgraded for risk of bias (minus one for serious, minus two for very serious and minus three for extremely serious risk of bias), inconsistency (minus one if more than one study per test available $I^2 > 0.5$), and imprecision (minus one if total sample size n = 50–100, minus two if total simple size n < 50). The score was not downgraded for indirectness, due to the restrictive inclusion criteria used in the current study [38]. Detailed tables of the quality of evidence criteria are shown in Additional file 4 (reliability) and Additional file 5 (measurement error).

Results

In total 11,704 studies were found using the search strategy in four databases (Medline, CINAHL, Embase, PsycINFO). First, 3741 duplicates were removed. The remaining 7963 studies were screened for title and abstract, and 7803 were excluded based on the predefined criteria. Of the 160 full-text studies, 118 were excluded. The reasons for exclusion are listed in Fig. 2.

A final total of 42 studies, investigating a total of 206 tests for the assessment of SC in patients with neck pain, were included in the systematic review (Table 1).

Tests were categorized into 18 different groups (e.g. tests for active range of motion in the different movement directions of flexion, extension, lateral flexion,

and rotation with the help of different devices were grouped together as active range of motion tests). Based on the classification of Riemann & Lephart [14], tests for the sensory and the motor components of the sensorimotor system were identified, but no tests for the central integration component were found. Within the sensory component, tests in the subcomponents "tactile" and "conscious proprioceptive senses" were found. As this study did not search for tests assessing pain, the subcomponent "pain" does not contain a test. A list of all groups of tests is shown in Fig. 3.

According to the COSMIN criteria the following 12 tests were rated as good: craniocervical flexion test (test-retest reliability), neck flexor muscle endurance test (inter-rater and test-retest reliability), neck extensor muscle endurance test (inter-rater and test-retest reliability), sternocleidomastoid muscle strength (test-retest reliability), maximal voluntary isometric contraction (test-retest reliability), isometric strength with the help of different devices (test-retest reliability), flexion-relaxation ratio (test-retest reliability), active range of motion test with the help of different devices (inter-rater and test-retest reliability), figure of eight test (inter-rater and test-retest reliability), zigzag test (inter-rater and test-retest reliability), smooth pursuit neck torsion test (test-retest reliability), and rod and frame test (test-retest reliability). An overview of the ratings of all tests is shown Table 2. However, regarding reliability, the quality of evidence was rated as low or very
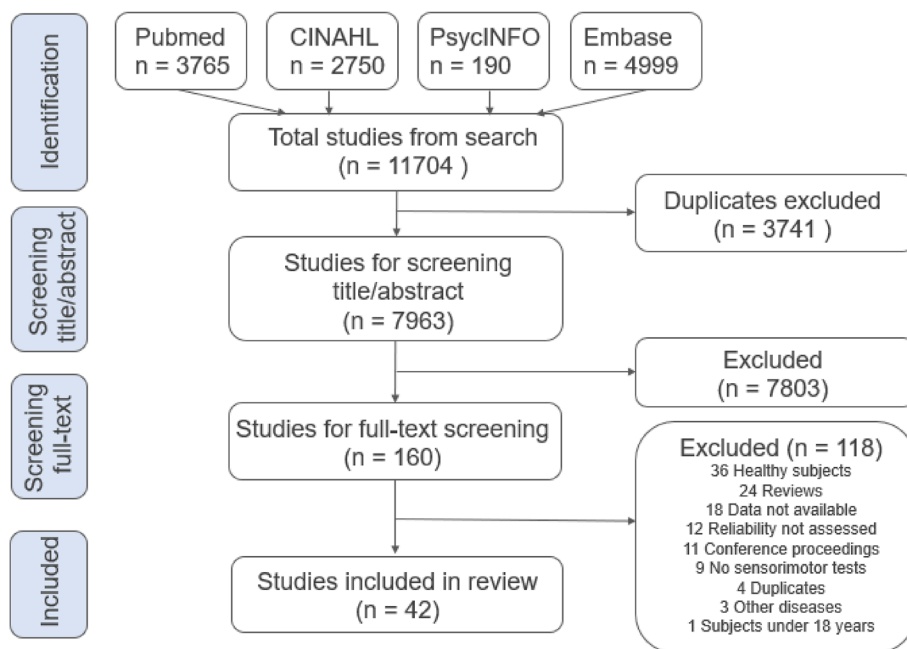

Fig. 2 Flow chart

Elsig *et al. Archives of Physiotherapy* (2023) 13:15

Page 6 of 16

Table 1 Characteristics of the included studies

Study	Population	Age, mean (sd)	Women n (%)
Chiu (2002) [39]; China	Subjects with mechanical neck pain	27.0 (9.5)	12 (57%)
Cibulka (2017) [40]; USA	Healthy adults with mild neck pain gathered through flyers, email, and word of mouth	22.8 (3.5)	23 (62%)
Cleland (2006) [41]; USA	Patients with mechanical neck pain referred to physical therapy at the Rehabilitation Services of a hospital	41 (12.9)	18 (82%)
De Pauw (2020) [42]; Belgium	Patients reporting neck pain, recruitment by advertising on social media and distribution of flyers	28.0 (8.2)	17 (68%)
Dvir (2006) [43]; Israel	Patients injured in whiplash-type accidents	37.1 (9.9)	13 (52%)
Edmondston (2008) [44]; Australia	Subjects with postural neck pain, recruited through poster advertising and through a university physical therapy clinic	36 (11)	14 (67%)
Fletcher (2008) [45]; USA	Subjects with neck pain from a college campus and a community setting	33.6 (10.3)	15 (68%)
Ghorbani (2020) [46]; Iran	Participants with neck pain from the University of Medical Sciences, invited via word of mouth	25.9 (1.04)	13 (65%)
Gonçalves (2019) [47]; Portugal	Individuals with neck pain recruited from a private clinical practice and from the general population	43.6 (13.3)	26 (79%)
Grod (2002) [48]; Canada	Patients with chronic neck pain from two chiropractic offices	38.5 (NA)	11 (58%)
Hanney (2014) [49]; USA	Patients with mechanical neck pain who presented to clinics	48.9 (14.8)	14 (64%)
Harris (2005) [50]; USA	Subjects with neck pain	38 (10)	61%[a]
Hoppenbrouwers (2006) [51]; Netherlands	Patients with neck pain from three physical therapy practices	43.0 (10.9)	15 (60%)
Hoving (2005) [52]; Netherlands	Patients with neck pain, referred by general practitioners for physical therapy	45.5 (9.2)	20 (63%)
Kristjansson (2004) [53]; Iceland	Female patients with chronic whiplash (grades I or II of Quebec Task Force classification) recruited from physiotherapy clinics	30.0 (8.8)[a]	10 (100%)
Kristjansson (2010) [54]; Iceland	Subjects with non-traumatic neck pain	38.0 (8.3)	11 (61%)
Kumbhare (2005) [55]; Canada	Patients with WDA (grade II of Quebec Task Force classification) recruited from a hospital	39.9 (14.9)	49 (69%)
Law (2013) [56]; China	Patients with neck pain from the out-patient Physiotherapy Department of a hospital	44.52 (7.11)	17 (65%)
Lourenço (2016) [57]; Portugal	Students with idiopathic neck pain from a university	20.18 (1.84)	17 (77%)
Majcen Rosker (2021) [58]; Slovenia	Patients with chronic neck pain, referred by an orthopaedic surgeon	46.2 (4.8)	23 (72%)
Martins (2018) [59]; Portugal	Participants with neck pain recruited from the general population	36.8 (2.4)	28 (85%)
Murphy (2010) [60]; New Zealand	Subjects with chronic neck pain recruited through advertisements in local papers and word of mouth	44.8 (8.5)	11 (79%)
O'Leary (2005) [61]; Australia	Subjects with neck pain recruited by printed and electronic advertising within the University	27.9[a]	75%[a]
Pearson (2009) [62]; Canada	Patients with WAD recruited from a rehabilitation and return-to-work program and from advertisements in local newspapers	36.6 (10.8)	6 (43%)
Peolsson (2007) [63]; Sweden	Patients with chronic neck disorders from primary care and from private clinicians	intra-rater: 49 (11); inter-rater: 47 (8)	intra-rater: 9 (90) ; inter-rater: 6 (75)
Petersen (2000) [64]; USA	Subjects with present complaints of local cervical pain	40.2 (8.7)	13 (65%)
Piva (2006) [65]; USA	Patients referred to a University Spine Speciality Centre with a primary complaint of neck pain	41 (12)	18 (60%)
Pourahmadi (2018) [66]; Iran	Subjects with non-traumatic neck pain recruited by purposive and snowball sampling	31.12 (6.38)	20 (50%)
Rheault (1992) [67]; USA	Subjects with a history of cervical spine pathology	37.41 (14.1)	15 (68%)
Röijezon (2010) [68]; Schweden	Women with non-traumatic neck pain recruited by advertising in local papers and by information to job holders	48 (7)	16 (100%)
Roren (2009) [69]; France	Patients with neck pain from a rehabilitation department	54.7 (14.2)	23 (56%)
Schneider (2013) [70]; Canada	Patients with persistent neck pain, referred to a tertiary interventional pain management centre	46 (NA)	37 (66%)
Sebastian (2015) [71]; USA	Patients with a diagnosis of neck pain	Range 30-75y	NA
Shahidi (2012) [72]; USA	Participants with neck pain recruited from a university medical campus and surrounding community	34.9 (9.9)	9 (47%)

Elsig *et al. Archives of Physiotherapy*   (2023) 13:15

Page 7 of 16

Table 1 (continued)

Study	Population	Age, mean (sd)	Women n (%)
Stenneberg (2018) [73]; Netherlands	Patients with neck pain recruited from five primary care physical therapy practices	45.2 (15.3)	19 (73%)
Sterling (2002) [74]; Australia	Patients with chronic neck pain of traumatic or non-traumatic origin, recruited via written advertisement within a university	31.63 (11.5)	13 (68%)
Uddin (2013) [75]; Canada	Patients with mechanical neck disorder	45.43 (11.88)[b]	81%[b]
Vernon (1992) [76]; Canada	Subjects presenting to the problem case clinic of a chiropractic college teaching clinic (mechanical neck pain syndrome, whiplash-type cervical strain injury)	37.5 (8.6)	12 (50%)
Werner (2018) [77]; Switzerland	Subjects with neck pain (non-specific or WAD) attending the physiotherapy department of a hospital	40.1 (12.35)	13 (65%)
Williams (2012) [78]; United Kingdom	Patients following a whiplash injury, attending an Emergency Department	41 (14.8)	13 (68%)
Williams (2012) [78]; United Kingdom	Patients following a whiplash injury, attending an Emergency Department	38 (11.3)	19 (50%)
Ylinen (2004) [79]; Finland	Women with nonspecific chronic neck pain, recruited through local occupational health care services	44 (6)	21 (100%)
Youdas (1991) [80]; USA	Patients referred to a Clinic Department of Physical Medicine and Rehabilitation with orthopaedic disorders	59.1 (15.7)	39 (65%)

WAD Whiplash associated disorders, *NA* Not available

[a] values only for total number of participants (participants with and without neck pain)

[b] values for all 106 patients, reliability was only calculated for 34 patients with unknown age and gender distribution

low for all included studies. The reasons for downgrading are shown in Additional file 4.

Regarding measurement error, the criteria for good measurement error were rated as unknown for all included tests, because the minimal clinically important change was not reported. The quality of evidence was rated very low to high (Table 2). Reasons for downgrading are shown in Additional file 5.

Discussion

This systematic review included 42 studies evaluating 206 tests, with the aim of investigating the reliability and measurement error of tests for SC in patients with neck pain. The main findings are, firstly, that tests for the sensory and motor components of the sensorimotor system were found, but not for the central integration component. Furthermore, no data were found on reliability or measurement error in patients with neck pain for some tests that are used in practice, such as the movement control tests, which would belong to the motor component; secondly, approximately half of the tests, particularly tests that are easier to standardize with regard to test position or movement direction, showed good reliability; and, finally, tests evaluating more complex movements, which are more difficult to standardize, were less reliable.

In general, all included muscle endurance tests, had good (relative) reliability values according to the criteria for good measurement properties proposed by COSMIN, except for the scapula muscle endurance test in standing position. The execution of this test is much more complex and more difficult to standardize than other tests. Furthermore, scapula movements, compensatory movements, muscle recruitment etc. are more difficult to assess compared with neck movements where the movement directions follow the sagittal, frontal, or transversal plane in a more stable way. Similarly, regarding reliability of the isometric muscle strength tests, tests involving the judgement of movements or muscle recruitment around the scapula have lower values for reliability than tests for isometric activity of the head into flexion, extension, lateral flexion, or rotation. Again, this may be because scapula positions are more difficult to standardize, and isometric contractions of the scapula muscles are more difficult to assess regarding compensatory movements than isometric muscle activity of the muscles of the cervical spine.

The test of fast cervical rotations showed very low reliability, possibly due to the very complex characteristics of these movements, which make it difficult to standardize the test. The tests assessing active range of motion (AROM) of the cervical spine showed that assessment of rotation is more difficult compared with the other movement directions. This is particularly evident when the rotation is assessed as a single movement (combined right and left values) and when AROM is assessed with the help of a smartphone. In the current analysis, the values were less reliable for Android phones than for iPhones (see Table 2). This could be due to differences in
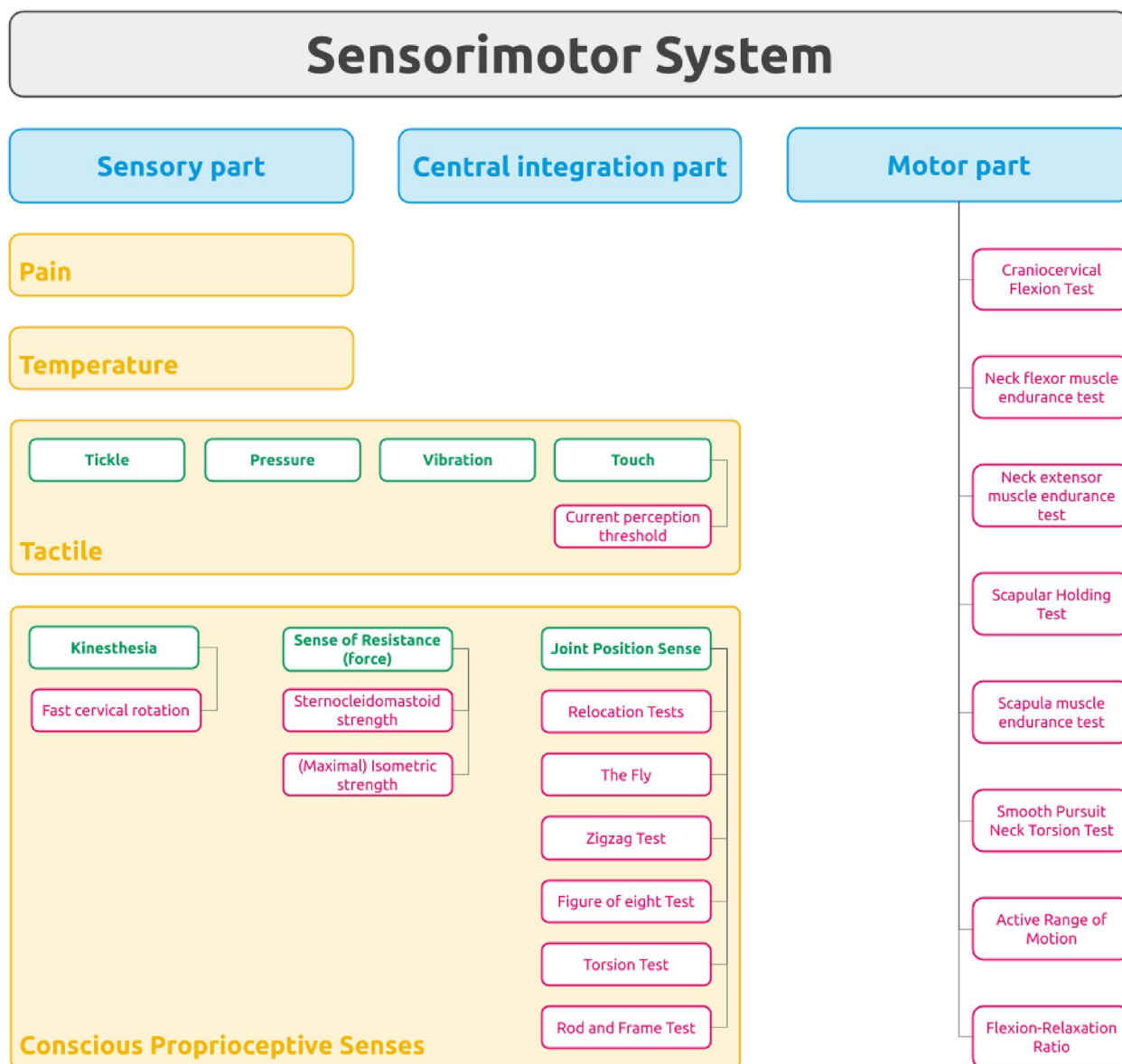
Elsig *et al. Archives of Physiotherapy*   (2023) 13:15

Page 8 of 16

Sensorimotor System

Sensory part

Central integration part

Motor part

Pain

Temperature

Tactile

| Tickle | Pressure | Vibration | Touch |

Current perception threshold

Conscious Proprioceptive Senses

Kinesthesia

Fast cervical rotation

Sense of Resistance (force)

Sternocleidomastoid strength

(Maximal) Isometric strength

Joint Position Sense

Relocation Tests

The Fly

Zigzag Test

Figure of eight Test

Torsion Test

Rod and Frame Test

Craniocervical Flexion Test

Neck flexor muscle endurance test

Neck extensor muscle endurance test

Scapular Holding Test

Scapula muscle endurance test

Smooth Pursuit Neck Torsion Test

Active Range of Motion

Flexion-Relaxation Ratio

Fig. 3 Sensorimotor system definition (according to Riemann & Lephart 2002 (14)) and the 18 groups of tests included in this systematic review *(pink boxes)*

the study protocols. In the study that used an Android phone, it was only held against the head, whereas in the study assessing AROM with an iPhone, the device was fastened securely to the forehead with a rigid strap, which might produce more reliable results. The assessment of AROM with the help of a dynamometer or goniometer showed good test-retest reliability results, but less good values for inter-rater reliability. It is evident that good values for inter-rater reliability are more difficult to achieve, because more sources of variation are included (e.g. different testers). Thus, the standardization of these types of tests is often a problem.

Using the example of the craniocervical flexion test (CCFT), this review shows that tests that require a substantial subjective rating (e.g. judgement of muscle recruitment or movement patterns) lead to lower reliability compared with more objective criteria (e.g. time). The current results are in line with a recent systematic review by Selistre and colleagues [30], investigating clinical tests for measuring strength or endurance of cervical muscles. They found moderate to good intra- and inter-rater reliability for the CCFT, cervical flexor endurance test, cervical extensor endurance test and cervical muscle strength assessed using a handheld dynamometer. The results of

Elsig *et al. Archives of Physiotherapy*   (2023) 13:15

Page 9 of 16

Table 2 Summary of findings

Test	Pooled[1]	Test-retest	Reliability Index (95% CI), (rating of criteria for good reliability)	Quality of the evidence (modified GRADE) Reliability	Quality of the evidence (modified GRADE)	SDC (rating of criteria for good ME)	Quality of the evidence (modified GRADE) ME	Quality of the evidence (modified GRADE) ME
Craniocervical Flexion Test adapted (De Pauw ,2020) [42]		inter-rater	0.64 (0.33-0.82), (-)	+	Very Low	2.72 (?)	++	Low
Craniocervical Flexion Test adapted (De Pauw, 2020) [42]		test-retest	0.65 (0.35-0.83), (-)	+	Very Low	2.58 (?)	++	Low
Craniocervical Flexion Test muscle activation		test-retest	0.7 (0.39-0.86), (+)	+	Very Low	4.96 (?)	++	Low
Craniocervical Flexion Test muscle endurance		test-retest	0.9 (0.8-0.95), (+)	+	Very Low	2.49 (?)	++	Low
Neck flexor muscle endurance Test	x	inter-rater	0.75 (0.29-0.93), (+)	++	Low	6.38 to 31.88 (?)	++	Low
Neck flexor muscle endurance Test	x	test-retest	0.84 (0.68-0.92), (+)	+	Very Low	17.74 to 23.01 (?)	+	Very Low
Neck extensor muscle endurance Test	x	inter-rater	0.84 (0.69-0.92), (+) sic!	++	Low			
Neck extensor muscle endurance Test	x	test-retest	0.84 (0.69-0.92), (+)	++	Low	2.05 to 71.51 (?)	+	Very Low
Scapula muscle endurance Test Standing		test-retest	0.67 (0.31-0.85), (-)	+	Very Low	30.21 (?)	++	Low
Scapular Holding Test adapted (summarized left and right)		inter-rater	0.54-0.63 (0.19-0.82), (-)	+	Very Low	3.02-3.3 (?)	++	Low
Scapular Holding Test adapted (summarized left and right)		test-retest	0.68-0.7 (0.41-0.86), (- to +)	+	Very Low	2.44-3.08 (?)	++	Low
Sterno-cleidomastoid strength (summarized left and right)		test-retest	0.95-0.97 (0.91-0.99), (+)	+	Very Low	4.63-5.04 (?)	++	Low
Isometric Muscle Strength Lower Trapezius (summarized left and right)		inter-rater	0.65-0.78 (0.28-0.91), (- to +)	+	Very Low			
Isometric Muscle Strength Middle Trapezius, Rhomboid (summarized left and right)		inter-rater	0.33-0.59 (-0.07-0.82), (-)	+	Very Low			

Elsig *et al. Archives of Physiotherapy*   (2023) 13:15

Page 10 of 16

Table 2 (continued)

Test	Pooled[1]	Test-retest	Reliability Index (95% CI), (rating of criteria for good reliability)	Quality of the evidence (modified GRADE) Reliability	Quality of the evidence (modified GRADE)	SDC (rating of criteria for good ME)	Quality of the evidence (modified GRADE) ME	Quality of the evidence (modified GRADE) ME
Maximal voluntary isometric contraction (six parameters summarized)		test-retest	0.76-0.91 (NA), (+)	+	Very Low	0.03-26.06 (?)	++	Low
Max isometric strength Flex outer torque Dorsal Head Force		test-retest	0.09 (NA-NA), (-)	+	Very Low	141.36 (?)	++	Low
Isometric strength Dynamometer		inter-rater	0.39-0.72 (-0.1-0.89), (- to +)	+	Very Low			
Isometric strength (Dynamometer, Modified Sphygmomanometer, Multi Cervical Rehabilitation Unit, Neck strength measurement system)	x	test-retest	0.74-0.99 (0.47-1), (+)	+	Very Low	1.94-54.33 (?)	+ to ++	Very Low to Low
Flexion-relaxation ratio		test-retest	0.83 (0.67-0.92), (+)	+	Very Low			
AROM visual estimation (six directions summarized)		inter-rater	0.42-0.82 (NA), (- to +)	+	Very Low			
AROM Universal Goniometer (all directions)	x	inter-rater	0.66-0.82 (0.47-0.93), (- to +)	+ to ++	Very Low to Low	3.41-15.25 (?)	+ to ++	Very Low to Low
AROM Universal Goniometer (all directions)	x	test-retest	0.71-0.89 (0.5-0.97), (+)	+ to ++	Very Low to Low	2.22-12.2 (?)	+ to ++	Very Low to Low
AROM Universal Goniometer Rot (left and right summarized)		test-retest	0.31 (-0.12-0.64)[2], (-)	+	Very Low	27.5 (?)	++	Low
AROM Electronic Goniometer (three directions summarized)		inter-rater	0.81-0.86 (0.62-0.94), (+)	+	Very Low	14.91-22.23 (?)	++	Low
AROM Electronic Goniometer (three directions summarized)		test-retest	0.89-0.92 (0.77-0.96), (+)	+	Very Low	11.28-16.58 (?)	++	Low
AROM Gravity Goniometer (eight movements summarized)		inter-rater	0.74-0.89 (0.26-0.95), (+)	+	Very Low	8.87-15.52 (?)	+ to +++	Very Low to Moderate
AROM Inclinometer (three directions summarized)		test-retest	0.41-0.75 (-0.16-0.89), (- to +)	+	Very Low	14.8-22.06 (?)	++	Low

Elsig *et al. Archives of Physiotherapy*   (2023) 13:15

Page 11 of 16

Table 2 (continued)

Test	Pooled[1]	Test-retest	Reliability Index (95% CI), (rating of criteria for good reliability)	Quality of the evidence (modified GRADE) Reliability	Quality of the evidence (modified GRADE)	SDC (rating of criteria for good ME)	Quality of the evidence (modified GRADE) ME	Quality of the evidence (modified GRADE) ME
AROM Digital inclinometer (six directions summarized)		inter-rater	0.73-0.89 (0.05-0.96), (+)	+	Very Low			
AROM Digital inclinometer (six directions summarized)		test-retest	0.53-0.84 (-0.32-0.94), (- to +)	+	Very Low	7.45-17.52 (?)	++	Low
AROM Digital inclinometer EDI-320 (three directions summarized)		inter-rater	0.89-0.95 (0.77-0.98), (+)	+	Very Low			
AROM Digital inclinometer EDI-320 (three directions summarized)		test-retest	0.93-0.96 (0.86-0.98), (+)	+	Very Low			
AROM Gravity Inclinometer (four pooled directions summarized)	x	inter-rater	0.74-0.86 (0.38-0.95), (+)	+	Very Low	8.59-19.4 (?)	+	Very Low
AROM Gravity Inclinometer (four directions summarized)		test-retest	0.91-0.95 (0.85-0.97), (+)	+	Very Low	6.65-8.04 (?)	++	Low
AROM CROM device (six pooled directions summarized)	x	inter-rater	0.82 - 0.93 (0.73-0.97), (+)	+ to ++	Very Low to Low	10.26-18.02 (?)	+	Very Low
AROM CROM device (six pooled directions summarized)	x	test-retest	0.91-0.95 (0.68-0.98), (+)	+ to ++	Very Low to Low	3.6-11.36 (?)	++	Low
AROM Fastrak (six directions summarized)		test-retest	0.64-0.88 (NA), (- to +)	+	Very Low	6.54-18.63 (?)	++	Low
AROM Multi Cervical Rehabilitation Unit (six directions summarized)		test-retest	0.82-0.96 (0.66-0.98), (+)	+	Very Low			
AROM OSI Spine Motion Analyser (six directions summarized)		test-retest	0.68-0.96 (NA), (- to +)	+	Very Low	5.13-9.12 (?)	+	Very Low
AROM Zebris (six directions summarized)		test-retest	0.81-0.86 (NA), (+)	+	Very Low	15.52-28.27 (?)	++	Low
AROM iPhone (six pooled directions summarized)	x	inter-rater	0.74-0.96 (0.39-0.98), (+)	+ to ++	Very Low to Low	4.1-9.67 (?)	+	Very Low

Table 2 (continued)

Test	Pooled[1]	Test-retest	Reliability Index (95% CI), (rating of criteria for good reliability)	Quality of the evidence (modified GRADE) Reliability	Quality of the evidence (modified GRADE)	SDC (rating of criteria for good ME)	Quality of the evidence (modified GRADE) ME	Quality of the evidence (modified GRADE) ME
AROM iPhone (six pooled directions summarized)	x	test-retest	0.69-0.84 (0.43-0.96), (- to +)	+ to ++	Very Low to Low	2.86-12.5 (?)	+ to ++	Very Low to Low
AROM Android (four directions summarized)		inter-rater	0.9-0.92 (0.72-0.97), (+)	+	Very Low			
AROM Android (four directions summarized)		test-retest	0.85-0.9 (0.61-0.96), (+)	+	Very Low	6.9-12.16 (?)	++	Low
AROM Android Rotation (left and right summarized)		inter-rater	0.17-0.48 (-0.98-0.8), (-)	+	Very Low			
AROM Android Rotation (left and right summarized)		test-retest	0.13-0.52 (-1.42-0.82), (-)	+	Very Low	24.05-32.89 (?)	++	Low
Relocation Test (Rotation pooled, seven movements summarized)	x	test-retest	0.62-0.85 (0.32-0.93), (- to +)	+ to ++	Very Low to Low	3.33-4.16 (?)	+ to ++	Very Low to Low
The Fly (twelve parameters summarized)		test-retest	0.58-0.86 (0.38-0.86), (- to +)	+	Very Low			
The Fly Overshoots (three level summarized)		test-retest	0.14-0.42 (-0.14-0.62), (-)	+	Very Low			
Torsion Test (left and right summarized)		test-retest	0.58.0.71 (0.14-0.85), (- to +)	+	Very Low	2.77-4.43 (?)	++	Low
Figure of eight Test (six parameters summarized)		inter-rater	0.76-1 (0.62-1), (+)	+	Very Low	0.06-5.6 (?)	++++	High
Figure of eight Test (four parameters summarized)		test-retest	0.81-1 (0.64-1), (+)	+	Very Low	0.03-8.09 (?)	++++	High
Zigzag Test (six parameters summarized)		inter-rater	0.8-1 (0.6-1), (+)	+	Very Low	0.08-2.27 (?)	++++	High
Zigzag Test (four parameters summarized)		test-retest	0.95-1 (0.78-1), (+)	+	Very Low	0.03-1.64 (?)	++++	High
Smooth Pursuit Neck Torsion Test (diff) Amplitude 40°, Velocity 20°/s		test-retest	0.75 (0.44-0.81), (+)	+	Very Low	0.11 (?)	++	Low
Rod and Frame Test (three parameters summarized)		test-retest	0.74-0.9 (NA), (+)	+ to ++	Very Low to Low			

Elsig *et al. Archives of Physiotherapy*   (2023) 13:15

Page 13 of 16

Table 2 (continued)

Test	Pooled[1]	Test-retest	Reliability Index (95% CI), (rating of criteria for good reliability)	Quality of the evidence (modified GRADE) Reliability	Quality of the evidence (modified GRADE)	SDC (rating of criteria for good ME)	Quality of the evidence (modified GRADE) ME	Quality of the evidence (modified GRADE) ME
Fast cervical rotation (five directions summarized)		test-retest	0.37-0.86 (-0.16-0.95), (- to +)	+	Very Low	0.3-91.47 (?)	++	Low
Fast cervical rotation Conjunct movements		test-retest	-0.07 (-0.54-0.45), (-)	+	Very Low	18.02 (?)	++	Low
Current perception threshold (nine parameters summarized)		test-retest	0.47-0.86 (0.08-0.93), (- to +)	+	Very Low			

Summarized Smallest and highest ICC from several tests, *CI* Confidence interval, *SDC* Smallest detectable change, *ME* Measurement error, *AROM* Active range of motion

Criteria for good reliability: + (sufficient) if ICC or (weighted) Kappa at least 0.7; ? (indeterminate) if ICC or (weighted) Kappa not reported; - (insufficient) if ICC or (weighted) Kappa < 0.7.

Criteria for good measurement error: + (sufficient) if smallest detectable change (SDC) or Limits of Agreement (LoA) or Coefficient of Variation (CV)*$\sqrt{2}$*1.96 < Minimal Clinically Important Change (MCIC); ? (indeterminate) if MCIC was not defined; - (insufficient) if SDC or LoA or CV*$\sqrt{2}$*1.96 >= MCIC. (37)

Quality of evidence: High: we are very confident that the true measurement property lies close to that of the estimate of the measurement property; Moderate: we are moderately confident in the measurement property estimate: the true measurement property is likely to be close to the estimate of the measurement property, but there is a possibility that it is substantially different; Low: our confidence in the measurement property estimate is limited: the true measurement property may be substantially different from the estimate of the measurement property; Very low: we have very little confidence in the measurement property estimate: the true measurement property is likely to be substantially different from the estimate of the measurement property (38).

the current review are comparable for the CCFT, the cervical flexor endurance test and the cervical extensor endurance test. For the cervical muscle strength tests, the current review performed a more detailed analysis, e.g. Selistre et al. [30] described the cervical strength tests only with the handheld dynamometer and not with other devices. In the current study, the cervical strength tests with dynamometer showed good results for test-retest reliability, but poorer results for intra-rater reliability. The results of the current review are also comparable with those of a recent systematic review of the measurement properties of the CCFT [31]. The authors classified the inter-rater and the intra-rater reliability of the CCFT as positive and the level of evidence as moderate. The measurement error was classified as indeterminate and the level of evidence as unknown. The authors identified the same problems as found in the current review, such as low methodological quality of the included studies and missing data on minimal clinically important change.

The two recent systematic reviews on measurement properties of tests for the SC of the neck included studies with participants with and without neck pain [30, 31]. Both stated that studies on participants with neck pain were lacking, which is in line with the current results. The current review excluded several studies because the

results for participants with neck pain were not reported separately but only together with those for people without neck pain. It was decided to include only studies with data for patients with neck pain, given our interest in the use of the tests in a clinical setting. Because the reliability of a test is influenced by the heterogeneity of the population in which the test is performed, it is important to know the reliability for a comparable population to that in which the test will be administered. It was also surprising that tests such as the CCFT, which is widely used in clinical practice, are so rarely investigated in patients with neck pain.

The major strength of this study is that it included all available tests for assessment of all aspects of sensorimotor control of the neck. However, the study also has a number of limitations. Many tests were performed only on healthy participants or in a mixed group of participants with and without neck pain. Several studies were excluded, including all studies assessing tests for movement control of the neck, as the authors did not report separate data for the patient group. Secondly, the quality of evidence was low to very low regarding reliability for all included studies. It was necessary to downgrade the level of evidence, mainly because of high risk of bias, inconsistency, and low precision. In the assessment of

Elsig *et al. Archives of Physiotherapy*   (2023) 13:15

Page 14 of 16

risk of bias, the item "patient stability" is one of the items that was particularly rated as doubtful in many cases. COSMIN recommends that patient stability should only be rated as very good if the study explicitly describes that the patients' condition did not change between measurements. As this information was often missing, the current review had to rate the patient stability item as doubtful, even though the time interval between measurements was adequate. A further limitation of this review is that the included studies did not report data on interpretability and feasibility of the different tests, which would be important information for the recommendation of specific tests. Finally, this review did not assess aspects of validity, which would certainly also be important for the selection of appropriate tests.

Better studies are needed on reliability, measurement error and validity of tests in patients with neck pain, because the quality of evidence of the existing research is mainly low or very low, and the reliability of some tests (e.g. for movement control) was not evaluated in patients with neck pain at all.

Conclusion
Despite the large number of tests available, the quality of evidence is not yet high enough to conclusively inform clinicians which test to use to assess SC in patients with neck pain.

For clinical practice, this systematic review shows that tests with objective criteria and a thorough standardization should be chosen to ensure higher reliability.

Measurement error could not be evaluated because the minimal clinically important change was not available for all tests.

Abbreviations
AROM	Active range of motion
CCFT	Craniocervical flexion test
CINAHL	Cumulative Index to Nursing and Allied Health Literature
COSMIN	Consensus-based Standards for the selection of health Measurement INstruments
Embase	Excerpta Medica dataBASE
GRADE	Grading of Recommendations Assessment, Development and Evaluation
ICC	Intra-class correlations
Medline	Medical Literature Analysis and Retrieval System Online

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s40945-023-00170-9.

Additional file 1. Search strategy (for Medline). Search strategy for Medline

Additional file 2. Risk of bias Reliability. Items of the assessment of risk of bias for reliability

Additional file 3. Risk of bias Measurement error. Items of the assessment of risk of bias for measurement error

Additional file 4. GRADE Downgrading Reliability. Reasons for downgrading

Additional file 5. GRADE Downgrading Measurement error. Reasons for downgrading

Author details
[1]School of Health Sciences, University of Applied Sciences and Arts Western Switzerland, HES-SO Valais-Wallis, Rathausstrasse 25, 3954, Valais, Leukerbad, Switzerland. [2]Department of Epidemiology, Research Line Functioning, Participation and Rehabilitation, CAPHRI - Care and Public Health Research Institute, Maastricht University, PO Box 616, 6200 Maastricht, the Netherlands. [3]School of Health Sciences, University of Applied Sciences and Arts Western Switzerland, HES-SO Valais-Wallis, Chemin de l'Agasse 5, Valais, Sion, Switzerland. [4]The Sense Innovation & Research Center, Sion and Lausanne, Switzerland. [5]Department of Medicine, Faculty of Medicine, University of Geneva, Geneva, Switzerland.

References
1. Kazeminasab S, Nejadghaderi SA, Amiri P, Pourfathi H, Araj-Khodaei M, Sullman MJ, et al. Neck pain: global epidemiology, trends and risk factors. BMC Musculoskel Dis. 2022;23(26):1–13.
2. Hoy D, March L, Woolf A, Blyth F, Brooks P, Smith E, et al. The global burden of neck pain: estimates from the global burden of disease 2010 study. Ann Rheumat Dis. 2014;73(7):1309–15.
3. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the global burden of Disease Study 2010. Lancet. 2012;380(9859):2163–96.
4. Hurwitz EL, Randhawa K, Yu H, Côté P, Haldeman S. The Global Spine Care Initiative: a summary of the global burden of low back and neck pain studies. Eur Spine J. 2018;27(6):796–801.
5. Safiri S, Kolahi A-A, Hoy D, Buchbinder R, Mansournia MA, Bettampadi D, et al. Bettampadi, et al. Global, regional, and national burden of neck pain

Elsig *et al. Archives of Physiotherapy*   (2023) 13:15

Page 15 of 16

in the general population, 1990-2017: systematic analysis of the Global Burden of Disease Study 2017. BMJ. 2020;368:m791. https://doi.org/10.1136/bmj.m791.

6. Fejer R, Kyvik KO, Hartvigsen J. The prevalence of neck pain in the world population: a systematic critical review of the literature. Eur Spine J. 2006;15(6):834–48.

7. Vos CJ, Verhagen AP, Passchier J, Koes BW. Clinical course and prognostic factors in acute neck pain: an inception cohort study in general practice. Pain Med. 2008;9(5):572–80.

8. Vasseljen O, Woodhouse A, Bjørngaard JH, Leivseth L. Natural course of acute neck and low back pain in the general population: the HUNT study. PAIN®. 2013;154(8):1237–44.

9. Cohen SP, editor. Epidemiology, diagnosis, and treatment of neck pain. Mayo Clinic Proceedings; 2015:90(2):284 – 99.

10. Hogg-Johnson S, van der Velde G, Carroll LJ, Holm LW, Cassidy JD, Guzman J, et al. The burden and determinants of neck pain in the general population: results of the bone and joint decade 2000–2010 Task Force on Neck Pain and its Associated Disorders. J Manipul Physiol Therapeut. 2009;32(2):46–S60.

11. Childs MJD, Fritz JM, Piva SR, Whitman JM. Proposal of a classification system for patients with neck pain. J Orthopaed Sports Phys Ther. 2004;34(11):686–700.

12. Rabey M, Beales D, Slater H, O'Sullivan P. Multidimensional pain profiles in four cases of chronic non-specific axial low back pain: an examination of the limitations of contemporary classification systems. Man Ther. 2015;20(1):138–47.

13. Sremakaew M, Jull G, Treleaven J, Uthaikhup S. Effectiveness of adding rehabilitation of cervical related sensorimotor control to manual therapy and exercise for neck pain: a randomized controlled trial. Musculoskel Sci Pract. 2023;63:102690.

14. Riemann BL, Lephart SM. The sensorimotor system, part I: the physiologic basis of functional joint stability. J Athlet Train. 2002;37(1):71.

15. Beinert K, Preiss S, Huber M, Taube W. Cervical joint position sense in neck pain. Immediate effects of muscle vibration versus mental training interventions: a RCT. Eur J Phys Rehabil Med. 2015;51(6):825–32.

16. Revel M, Andre-Deshays C, Minguet M. Cervicocephalic kinesthetic sensibility in patients with cervical pain. Arch Phys Med Rehabil. 1991;72(5):288–91.

17. Kristjansson E, Dall'Alba P, Jull G. A study of five cervicocephalic relocation tests in three different subject groups. Clin Rehabil. 2003;17(7):768–74.

18. Reddy RS, Meziat-Filho N, Ferreira AS, Tedla JS, Kandakurti PK, Kakaraparthi VN. Comparison of neck extensor muscle endurance and cervical proprioception between asymptomatic individuals and patients with chronic neck pain. J Bodyw Mov Ther. 2021;26:180–6.

19. Falla D, Jull G, Edwards S, Koh K, Rainoldi A. Neuromuscular efficiency of the sternocleidomastoid and anterior scalene muscles in patients with chronic neck pain. Disabil Rehabil. 2004;26(12):712–7.

20. Falla DL, Jull GA, Hodges PW. Patients with neck pain demonstrate reduced electromyographic activity of the deep cervical flexor muscles during performance of the craniocervical flexion test. Spine. 2004;29(19):2108–14.

21. Jull G, Kristjansson E, Dall'Alba P. Impairment in the cervical flexors: a comparison of whiplash and insidious onset neck pain patients. Man Ther. 2004;9(2):89–94.

22. Della Casa E, Affolter Helbling J, Meichtry A, Luomajoki H, Kool J. Head-eye movement control tests in patients with chronic neck pain; inter-observer reliability and discriminative validity. BMC Musculoskel Dis. 2014;15:16.

23. Sterling M, Jull G, Wright A. The effect of musculoskeletal pain on motor activity and control. The J Pain. 2001;2(3):135–45.

24. Jull G, Trott P, Potter H, Zito G, Niere K, Shirley D, et al. A randomized controlled trial of exercise and manipulative therapy for cervicogenic headache. SPINE. 2002;27(17):1835–43.

25. Kristjansson E, Treleaven J. Sensorimotor function and dizziness in neck pain: implications for assessment and management. J Orthopaed sports Phys Ther. 2009;39(5):364–77.

26. Treleaven J. Sensorimotor disturbances in neck disorders affecting postural stability, head and eye movement control. Man Ther. 2008;13(1):2–11.

27. de Zoete RM, Osmotherly PG, Rivett DA, Farrell SF, Snodgrass SJ. Sensorimotor control in individuals with idiopathic neck pain and healthy

individuals: a systematic review and meta-analysis. Arch Phys Med Rehabil. 2017;98(6):1257–71.

28. Michiels S, De Hertogh W, Truijen S, November D, Wuyts F, Van de Heyning P. The assessment of cervical sensory motor control: a systematic review focusing on measuring methods and their clinimetric characteristics. Gait Posture. 2013;38(1):1–7.

29. de Koning CH, van den Heuvel SP, Staal JB, Smits-Engelsman B, Hendriks EJ. Clinimetric evaluation of methods to measure muscle functioning in patients with non-specific neck pain: a systematic review. BMC Musculoskel Dis. 2008;9(1):1–9.

30. Selistre LFA, de Sousa Melo C, de Noronha MA. Reliability and validity of clinical tests for measuring strength or endurance of cervical muscles: a systematic review and meta-analysis. Arch Phys Med Rehabil. 2021;102(6):1210–27.

31. Araujo FXd, Ferreira GE, Scholl Schell M, Castro MPd, Ribeiro DC, Silva MF. Measurement properties of the craniocervical flexion test: a systematic review. Phys Ther. 2020;100(7):1094–117.

32. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737–45.

33. Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable? Injury. 2011;42(3):236–40.

34. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res. 2009;18(8):1115–23.

35. Ammann-Reiffer C, Bastiaenen CH, de Bie RA, van Hedel HJ. Measurement properties of gait-related outcomes in youth with neuromuscular diagnoses: a systematic review. Phys Ther. 2014;94(8):1067–82.

36. Covidence systematic review software Melbourne, Australia. www.covidence.org.

37. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42(2):377–81.

38. Mokkink LB, Boers M, van der Vleuten CPM, Bouter LM, Alonso J, Patrick DL, de Vet HCW, Terwee CB. COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. BMC Med Res Methodol. 2020;20(293).

39. Chiu TTW, Lo SK. Evaluation of cervical range of motion and isometric neck muscle strength: reliability and validity. Clin Rehabil. 2002;16(8):851–8.

40. Cibulka MT, Herren J, Kilian A, Smith S, Mahmutovic F, Dolles C. The reliability of assessing sternocleidomastoid muscle length and strength in adults with and without mild neck pain. Physiother Theory Pract. 2017;33(4):323–30.

41. Cleland JA, Childs JD, Fritz JM, Whitman JM. Interrater reliability of the history and physical examination in patients with mechanical neck pain. Arch Phys Med Rehabil. 2006;87(10):1388–95.

42. De Pauw R, Van Looveren E, Lenoir D, Danneels L, Cagnie B. Reliability and discriminative validity of a screening tool for the assessment of neuromuscular control and movement control in patients with neck pain and healthy individuals. Disabil Rehabil. 2020. https://doi.org/10.1080/09638288.2020.1760948.

43. Dvir Z, Gal-Eshel N, Shamir B, Prushansky T, Pevzner E, Peretz C. Cervical motion in patients with chronic disorders of the cervical spine: a reproducibility study. Spine. 2006;31(13):E394–E9.

44. Edmondston SJ, Wallumrød ME, MacLéid F, Kvamme LS, Joebges S, Brabham GC. Reliability of isometric muscle endurance tests in subjects with postural neck pain. J Manip Physiol Ther. 2008;31(5):348–54.

45. 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):640–5.

46. Ghorbani F, Kamyab M, Azadinia F. Smartphone applications as a suitable alternative to CROM device and inclinometers in assessing the cervical range of motion in patients with nonspecific neck pain. J Chiropr Med. 2020;19(1):38–48.

47. Gonçalves C, Silva AG. Reliability, measurement error and construct validity of four proprioceptive tests in patients with chronic idiopathic neck pain. Musculoskelet Sci Pract. 2019;43:103–9.

48. Grod JP, Diakow PR. Effect of neck pain on verticality perception: a cohort study. Arch Phys Med Rehabil. 2002;83(3):412–5.

49. Hanney WJ, George SZ, Kolber MJ, Young I, Salamh PA, Cleland JA. Inter-rater reliability of select physical examination procedures in patients with neck pain. Physiother Theory Pract. 2014;30(5):345–52.
50. Harris KD, Heer DM, Roy TC, Santos DM, Whitman JM, Wainner RS. Reliability of a measurement of neck flexor muscle endurance. Phys Ther. 2005;85(12):1349–55.
51. Hoppenbrouwers M, Eckhardt MM, Verkerk K, Verhagen A. Reproducibility of the measurement of active and passive cervical range of motion. J Manipulative Physiol Ther. 2006;29(5):363–7.
52. Hoving JL, Pool JJ, Van Mameren H, Devillé WJ, Assendelft WJ, De Vet HC, et al. Reproducibility of cervical range of motion in patients with neck pain. BMC Musculoskel Dis. 2005;6:1–8.
53. Kristjansson E, Hardardottir L, Asmundardottir M, Gudmundsson K. A new clinical test for cervicocephalic kinesthetic sensibility:"the fly. Arch Phys Med Rehabil. 2004;85(3):490–5.
54. Kristjansson E, Oddsdottir GL. The Fly": a new clinical assessment and treatment method for deficits of movement control in the cervical spine: reliability and validity. Spine. 2010;35(23):E1298–E305.
55. Kumbhare DA, Balsor B, Parkinson WL, Harding Bsckin P, Bedard M, Papaioannou A, et al. Measurement of cervical flexor endurance following whiplash. Disabil Rehabil. 2005;27(14):801–7.
56. Law EYH, Chiu TT-W. Measurement of cervical range of motion (CROM) by electronic CROM goniometer: a test of reliability and validity. J Back Musculoskelet Rehabil. 2013;26(2):141–8.
57. Lourenço AS, Lameiras C, Silva AG. Neck flexor and extensor muscle endurance in subclinical neck pain: intrarater reliability, standard error of measurement, minimal detectable change, and comparison with asymptomatic participants in a university student population. J Manipulative Physiol Ther. 2016;39(6):427–33.
58. Majcen Rosker Z, Vodicar M, Kristjansson E. Inter-visit reliability of smooth pursuit neck torsion test in patients with chronic neck pain and healthy individuals. Diagnostics. 2021;11(5):752.
59. Martins F, Bento A, Silva AG. Within-session and between-session reliability, construct validity, and comparison between individuals with and without neck pain of four neck muscle tests. PM&R. 2018;10(2):183–93.
60. Murphy BA, Marshall PW, Taylor HH. The cervical flexion-relaxation ratio: reproducibility and comparison between chronic neck pain patients and controls. Spine. 2010;35(24):2103–8.
61. O'Leary SP, Vicenzino BT, Jull GA. A new method of isometric dynamometry for the craniocervical flexor muscles. Phys Ther. 2005;85(6):556–64.
62. Pearson I, Reichert A, De Serres SJ, Dumas J-P, Côté JN. Maximal voluntary isometric neck strength deficits in adults with whiplash-associated disorders and association with pain and fear of movement. J Orthop Sports Phys Ther. 2009;39(3):179–87.
63. Peolsson A, Hamp C, Albinsson A-K, Engdahl S, Kvist J. Test position and reliability in measurements of dorsal neck muscle endurance. Adv Physiotherapy. 2007;9(4):181–9.
64. Petersen C, Johnson R, 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):82–8.
65. 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):321–30.
66. Pourahmadi MR, Bagheri R, Taghipour M, Takamjani IE, Sarrafzadeh J, Mohseni-Bandpei MA. A new iPhone application for measuring active craniocervical range of motion in patients with non-specific neck pain: a reliability and validity study. Spine J. 2018;18(3):447–57.
67. Rheault W, Albright B, Byers C, Franta M, Johnson A, Skowronek M, et al. Intertester reliability of the cervical range of motion device. J Orthop Sports Phys Ther. 1992;15(3):147–50.
68. Röijezon U, Djupsjöbacka M, Björklund M, Häger-Ross C, Grip H, Liebermann DG. Kinematics of fast cervical rotations in persons with chronic neck pain: a cross-sectional and reliability study. BMC Musculoskel Dis. 2010;11(1):1–10.
69. Roren A, Mayoux-Benhamou M-A, Fayad F, Poiraudeau S, Lantz D, Revel M. Comparison of visual and ultrasound based techniques to measure head repositioning in healthy and neck-pain subjects. Man Ther. 2009;14(3):270–7.
70. Schneider GM, Jull G, Thomas K, Smith A, Emery C, Faris P, 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):1628–34.
71. Sebastian D, Chovvath R, Malladi R. Cervical extensor endurance test: a reliability study. J Bodyw Mov Ther. 2015;19(2):213–6.
72. 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 Musculoskel Dis. 2012;13(1):1–11.
73. Stenneberg MS, Busstra H, Eskes M, van Trijffel E, Cattrysse E, Scholten-Peeters GG, et al. Concurrent validity and interrater reliability of a new smartphone application to assess 3D active cervical range of motion in patients with neck pain. Musculoskelet Sci Pract. 2018;34:59–65.
74. Sterling M, Jull G, Carlsson Y, Crommert L. Are cervical physical outcome measures influenced by the presence of symptomatology? Physiother Res Int. 2002;7(3):113–21.
75. Uddin Z, MacDermid JC, Galea V, Gross AR, Pierrynowski MR. Reliability indices, limits of agreement, and construct validity of the current perception threshold test in mechanical neck disorder. Crit Rev Phys Rehabil Med. 2013;25:3–4.
76. Vernon H, Aker P, Aramenko M, Battershill D, Alepin A, Penner T. Evaluation of neck muscle strength with a modified sphygmomanometer dynamometer: reliability and validity. J Manipulative Physiol Ther. 1992;15(6):343–9.
77. Werner IM, Ernst MJ, Treleaven J, Crawford RJ. Intra and interrater reliability and clinical feasibility of a simple measure of cervical movement sense in patients with neck pain. BMC Musculoskel Dis. 2018;19:1–9.
78. Williams MA, Williamson E, Gates S, Cooke MW. Reproducibility of the cervical range of motion (CROM) device for individuals with sub-acute whiplash associated disorders. Eur Spine J. 2012;21:872–8.
79. Ylinen J, Salo P, Nykänen M, Kautiainen H, Häkkinen A. Decreased isometric neck strength in women with chronic neck pain and the repeatability of neck strength measurements. Arch Phys Med Rehabil. 2004;85(8):1303–8.
80. 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.

Publisher's Note