

RESEARCH

Open Access



# Measurement properties of cervical joint position error in people with and without neck pain: a systematic review and narrative synthesis

Ahmad AlDahas<sup>1,2</sup> , Nicola R. Heneghan<sup>1</sup> , Shouq Althobaiti<sup>1</sup>, Janet A. Deane<sup>1</sup> , Alison Rushton<sup>3</sup> and Deborah Falla<sup>1\*</sup>

## Abstract

**Introduction** Proprioception can be impaired in people with neck pain. The cervical joint position sense test, which measures joint position error (JPE), is the most common test used to assess neck proprioception. The aim of this systematic review was to assess the measurement properties of this test for the assessment of people with and without neck pain.

**Methods** This systematic review was registered prospectively on Prospero (CRD42020188715). It was designed using the COSMIN guidelines and reported in line with the PRISMA checklist. Two reviewers independently searched Medline, Embase, SportDiscus, and CINAHL Plus databases from inception to the 24th July 2022 with an update of the search conducted until 14th of October 2023. The COSMIN risk of bias checklist was used to assess the risk of bias in each study. The updated criteria for good measurement properties were used to rate individual studies and then the overall pooled results. The level of evidence was rated by two reviewers independently using a modified GRADE approach.

**Results** Fifteen studies were included in this review, 13 reporting absolute JPE and 2 reporting constant JPE. The measurement properties assessed were reliability, measurement error, and validity. The measurement of JPE showed sufficient reliability and validity, however, the level of evidence was low/very low for both measurement properties, apart from convergent validity of the constant JPE, which was high.

**Conclusion** The measure of cervical JPE showed sufficient reliability and validity but with low/very low levels of evidence. Further studies are required to investigate the reliability and validity of this test as well as the responsiveness of the measure.

**Keywords** Neck pain, Proprioception, Sensorimotor, Position sense, Position error, Measurement properties

\*Correspondence:

Deborah Falla

d.falla@bham.ac.uk

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Cervical sensorimotor control relies on the integration of visual, vestibular, and somatosensory information [1]. Afferent information from muscle spindles is known to contribute significantly to cervical proprioception; in particular muscles in the upper cervical region contain a high density of muscle spindles, which implies that they have an important role in neck proprioception [2].

Several outcome measures have been used to assess cervical proprioception with the joint position sense test being the most common test to evaluate joint position error (JPE) [2]. The joint position sense test determines a person's ability to relocate their head back to a target position following active neck movement whilst their vision is occluded. Two commonly used joint position sense tests that measure JPE include the neutral head position (NHP) test, having the participant return to a neutral head position following active movement, or the target head position (THP) test, where target head position is determined by the participant or assessor [3].

Several studies have evaluated cervical proprioception by quantifying JPE in people with neck pain and have shown that cervical proprioception can be impaired in people with neck pain regardless of the aetiology [4–6]. For example, Revel et al. reported a higher repositioning error in people with chronic neck pain (CNP) after returning to neutral from flexion, extension, and right and left rotation when compared to asymptomatic participants [4]. Feipel et al. showed that people with chronic whiplash-associated disorders (WAD) had higher repositioning errors compared to asymptomatic participants [7]. Nevertheless, evidence indicates that cervical proprioception is more likely to be affected and to a greater extent in people that have trauma-induced neck pain [8]. Additionally, people with WAD and dizziness complaints usually have higher repositioning errors when compared to people with WAD but without dizziness [9, 10]. Impaired neck proprioception is thought to be at least partially attributed to a disturbance in cervical afferent activity [2]. Several mechanisms can contribute to this disturbance such as direct trauma to cervical structures, the influence of nociception, the presence of inflammatory mediators [9], and psychological distress [8]. A disturbance in cervical afferent input is also thought to contribute symptoms of dizziness for some patients [8].

Knowledge of the psychometric properties of outcome measures, which includes their reliability, validity, and responsiveness, are important as they reflect data accuracy and precision [11]. Michiels et al. carried out a systematic review investigating the measurement properties of cervical sensorimotor control tests [12]. In their 2012 review, they investigated the reliability and discriminative validity of tests. Although this systematic review did

not use the now recommended Consensus-based Standards for the Selection of health Measurement Instrument (COSMIN) reporting guidelines [13], they reported that the NHP test showed fair to excellent reliability (ICC range: 0.35–0.87) while the THP showed poor to excellent reliability (ICC range: 0.01–0.9). Additionally, the JPE test was able to discriminate between people with and without chronic neck pain.

Given the number of publications since this last review, in this current systematic review, we aimed to build upon this research to synthesise the available evidence in relation to a range of measurement properties (reliability, measurement error, validity, and responsiveness) of the measure of cervical JPE for the assessment of people with and without neck pain.

## Design and methods

This systematic review was designed using the COSMIN risk of bias (RoB) guidelines for reliability and measurement error of outcome measurement instruments as well as the COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) [13, 14] and is reported in accordance with the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) [15]. The protocol was registered with PROSPERO on the 10th of July 2020 (CRD42020188715).

### Deviations from the study protocol

The initial protocol described a systematic review of the measurement properties of proprioception tests for all regions of the spine. However, following an initial review of the literature and appreciation of the number of studies conducted in different spinal regions, the decision was made to focus on the measurement properties of cervical JPE only. Additionally, the original plan was to use the COSMIN RoB checklist for PROMs, however since publishing the protocol, the authors were made aware of the new COSMIN RoB checklist for reliability and measurement error of outcome measurement instruments. Thus, this new tool was used to assess RoB of reliability, measurement error, and criterion validity [14]. The COSMIN RoB checklist for PROMs was used to assess construct and discriminative validity [16] as suggested in the manual for the COSMIN RoB checklist for reliability and measurement error of outcome measurement instruments.

### Eligibility criteria

The following inclusion criteria are based on the Sample, Phenomena of Interest, Design, Evaluation, and Research type (SPIDER) guidelines [17].

- Sample: people with and without neck pain aged  $\geq 18$  years. Those with neck pain included regardless of the stage of their neck pain (e.g., acute, or chronic) or aetiology (e.g., non-specific or attributed to pathology).
- Phenomena of interest: cervical proprioception.
- Design: any study which investigated at least one of the domains (reliability, validity, responsiveness, and their sub-domains) of the COSMIN checklist and reported absolute error (AE) or constant error (CE) in degrees.
- Evaluation: any study that evaluated measurement properties of the measure of cervical JPE.
- Research type: quantitative research.

#### Exclusion criteria

Studies that included patients that had undergone cervical spine surgery and studies not written in English were excluded.

#### Information sources

The following databases were searched as recommended by the COSMIN guidelines for systematic reviews [13], from inception to the 24th July 2022 with an update of the search conducted until 14th of October 2023: MEDLINE, Embase, SportDiscus, and CINAHL plus. Manual searches were carried out for: The Spine Journal, European Spine Journal, Journal of Musculoskeletal Science and Practice, and the Journal of Orthopaedic and Sport Physical Therapy. Grey literature (Open Grey, ProQuest, and EThOS) was hand searched.

#### Search strategy

Following scoping searches and discussions with co-authors, the search strategy was developed, and a librarian was consulted. Search terms are provided in Table 1. Search syntax was translated to meet the requirements of each database.

#### Data management

Endnote software version X9 (Clarivate Analytics) was used to manage citations and bibliographies and store articles found and eliminate duplicates.

#### Study selection

AA carried out the initial search of the databases, after that, two researchers (AA, SA) independently carried out the screening of potentially eligible studies. The screening and selection were carried out in two steps. Step 1: Abstracts and titles using the eligibility criteria. Step 2: Retrieve full text of potentially relevant studies to be screened. Studies were included if both reviewers had agreed on inclusion after screening the full text. In case of any disagreement, a third reviewer (DF) was consulted.

#### Data extraction and data items

Two researchers (AA, SA) independently carried out the data extraction from the included studies. Extracted data items were characteristics of the studies (study design and sample size), characteristics of the participants (age, gender, population), testing instrument, testing protocols, measurement properties (reliability, measurement error, validity, and responsiveness), and results. In case of any disagreement, a third reviewer (DF) was consulted.

#### Risk of bias assessment

Included studies were independently assessed by two reviewers (AA, SA) using the COSMIN RoB checklist for reliability and measurement error of outcome measurement instruments to assess RoB of reliability, measurement error, and criterion validity [14]. The COSMIN RoB checklist for PROMs was used to assess construct and discriminative validity [16]. Both checklists have four scores (very good, adequate, doubtful, and inadequate) [16] that assess measurement properties with regard to design and statistical methods. In case of any disagreement, a third reviewer (DF) was consulted.

#### Data synthesis

Data synthesis of the results was undertaken in accordance with COSMIN guidelines [13]. After assessing the risk of bias, each study was rated using the updated criteria for good measurement properties as sufficient (+),

**Table 1** MEDLINE syntax used in MEDLINE database

<b>Search terms</b>	Neck pain OR neck dysfunction OR cervical pain OR cervical dysfunction AND Propriocept* OR movement sense OR kinesthes* OR repositioning OR repositioning error OR position sense OR motion perception OR active position sense OR passive position sense AND Reliability OR validity OR responsiveness OR reproducibility of results OR reproducib* OR reliab* OR valid* OR stability OR interrater OR interrater OR intrarater OR intrarater OR intra-rater OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intra-observer OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR intraclass correlation OR standard error of measurement OR sensitiv* OR responsive* OR minimal detectable concentration OR interpretab* OR small detectable change OR ceiling effect OR floor effect
---------------------	--

insufficient (-), or indeterminate (?) [13], then, the overall results of each measurement property per outcome measure per population were rated against the criteria of a good measurement property as sufficient (+), insufficient (-), inconsistent ( $\pm$ ), or indeterminate (?) [13]. Table 2 presents the updated criteria for good measurement properties.

The overall level of evidence for each outcome measure and its respective measurement property was then determined independently by two reviewers (AA, SA) using a modified Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach [19]. Table 3 presents the modified GRADE approach used to rate the overall quality of the evidence. More information on how to downgrade the level of evidence can be found in the COSMIN user manual [19].

**Results**

Fifteen studies were included four with CNP, three that did not specify the type of neck pain, one with cervicogenic disc disease, and seven studies that included participants without neck pain. There was a 100% agreement between raters (AA, SA) for the included studies. Search results are summarised in Fig. 1 and Table 4 summarises the extracted data from the included studies.

**Absolute joint position error for people with neck pain**

*Intra-rater reliability*

For the NHP test, six studies investigated intra-rater reliability of absolute JPE. One study included participants with CDD [5] testing right and left rotation using

**Table 3** Modified GRADE approach used to rate the overall level of evidence [13]

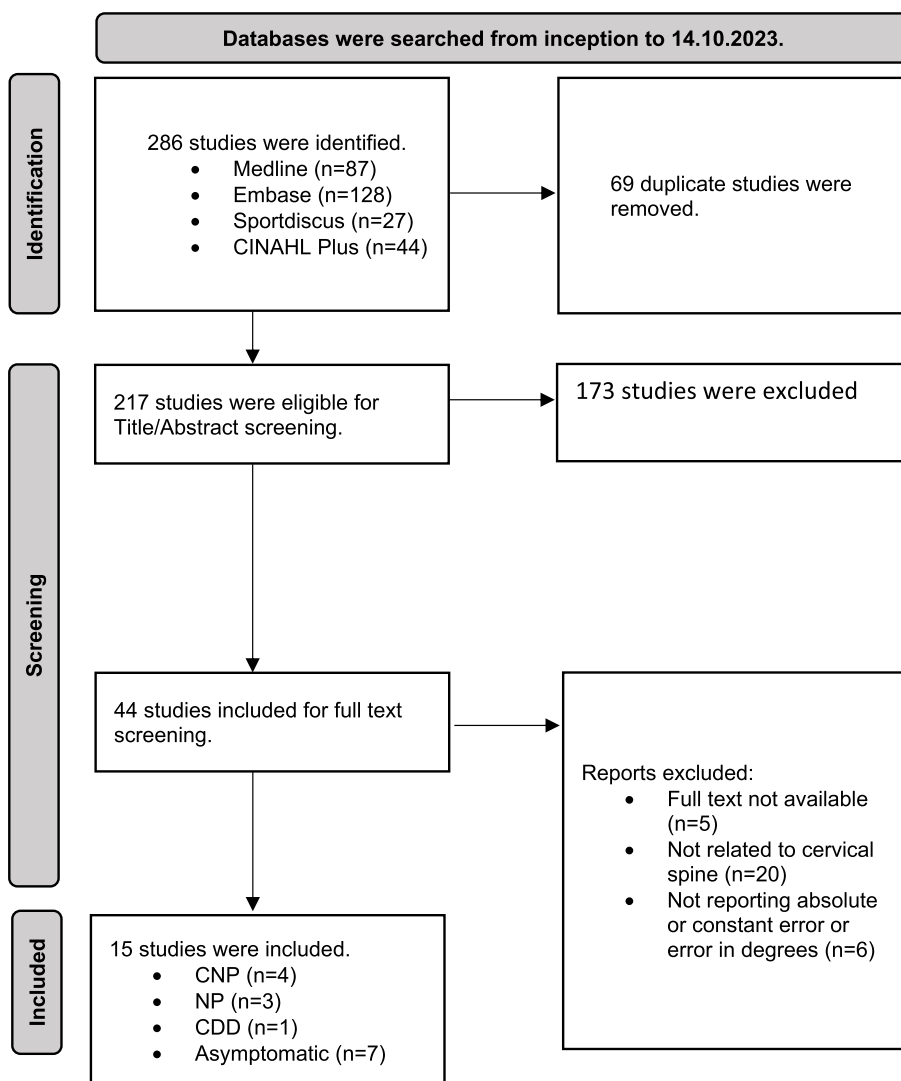
Quality of evidence	Lower if there is
High	<b>Risk of bias</b>
Moderate	-1 Serious
Low	-2 Very serious
Very low	-3 Extremely serious
	<b>Inconsistency</b>
	-1 Serious
	-2 Very serious
	<b>Imprecision</b>
	-1 Sample size (n = 50–100)
	-2 Sample size (n < 50)
	<b>Indirectness</b>
	-1 Serious
	-2 Very serious

a CROM device and 3 trials for their assessment in sitting position, however only the NHP test was reported. This study was rated as inadequate in the RoB checklist and sufficient in the updated criteria for good measurement properties. Three studies mentioned neck pain participants but failed to report type or duration of neck pain [3, 30, 33]. Alahmari et al. [3] carried out their intra-rater reliability assessment for the NHP test, it was rated as inadequate in the RoB checklist and sufficient in the updated criteria for good measurement properties. Burke et al. [30] carried out their intra-rater reliability using two devices, the CROM and laser. Both were rated as inadequate in the RoB checklist and insufficient in the updated criteria for good measurement properties. Cid et al. [33] investigated the intra-rater reliability of the NHP, it was

**Table 2** The updated criteria for good measurement properties [13, 18]

Measurement property	Rating	Criteria
Reliability	Sufficient (+)	ICC or weighted Kappa $\geq 0.7$
	Indeterminate (?)	ICC or weighted Kappa not reported
	Insufficient (-)	ICC or weighted Kappa $< 0.70$
Measurement error	Sufficient (+)	SDC or LoA $< MIC$
	Indeterminate (?)	MIC not defined
	Insufficient (-)	SDC or LoA $> MIC$
Hypothesis testing for construct validity	Sufficient (+)	The result is in accordance with the hypothesis
	Indeterminate (?)	No hypothesis defined (by the review team)
	Insufficient (-)	The result is not in accordance with the hypothesis
Criterion validity	Sufficient (+)	Correlation with gold standard $\geq 0.70$ OR AUC $\geq 0.70$
	Indeterminate (?)	Not all information for ‘+’ reported
	Insufficient (-)	Correlation with gold standard $< 0.70$ OR AUC $< 0.70$
Responsiveness	Sufficient (+)	The result is in accordance with the hypothesis OR AUC $\geq 0.70$
	Indeterminate (?)	No hypothesis defined (by the review team)
	Insufficient (-)	The result is not in accordance with the hypothesis OR AUC $< 0.70$

ICC intraclass correlation coefficient, SDC smallest detectable change, LoA limits of agreement, MIC minimal important change, AUC area under curve



**Fig. 1** Prisma flow diagram of the study selection process [20]. CNP=chronic neck pain. NP=neck pain. CDD=cervicogenic disc disease. n=number of studies

rated as inadequate in the RoB checklist and sufficient in the updated criteria for good measurement properties. Moreover, two studies included CNP participants [27, 31], and tested both right and left rotation in sitting position. Roren et al. [27] included 5 trials in their assessment, and used a laser pointer and US device. Both parts were rated as inadequate in the RoB checklist and insufficient in the updated criteria for good measurement properties. Goncalves and Silva [31] carried out within-day and between-day intra-rater reliability investigations of different types of NHP tests: NHP, figure of 8 (F8T) relocation test, and torsion test (TT). All investigations for were rated as doubtful in RoB checklist and sufficient in the updated criteria for good measurement properties. Nine studies showed sufficient results and four studies

showed insufficient results. Therefore, the overall rating was taken. The overall rating of the intra-rater reliability was rated as sufficient, but the quality of evidence was downgraded to very low due to inconsistency of results and risk of bias (multiple studies with doubtful/inadequate ratings and inconsistency of results) (Table 5).

For the THP test, two studies tested the intra-rater reliability of the THP test [3, 31]. Alahmari et al. [3] was rated as inadequate in the RoB checklist and sufficient in the updated criteria for good measurement properties. Goncalves and Silva [31] carried out a within-day and between-day testing. Both investigations were rated as doubtful in the RoB checklist and sufficient in the updated criteria for good measurement properties. The overall rating of the intra-rater reliability of the

**Table 4** Data extracted from the studies included in this review

Authors	Population and sample size	Age (mean or range)	Testing instrument	Testing position	Testing procedure	Property domain	Statistical test used	Results
Artz et al. [21]	Asymptomatic subjects: $n=40$ , (within day $n=21$ , between day $n=19$ )	29.9 years	3-Space Fastrak	Sitting Standing	<p><b>Trials:</b> 3 trials</p> <p><b>Movements tested:</b> 25%, 50%, 75% in flexion (randomised)</p>	Reliability within-day and between-day intra-rater (at least 1 week apart)	ICC, SEM	<p><b>JPE:</b></p> <p><b>Sitting:</b></p> <p><b>Within-day (ICC range:</b> -0.81–0.66)</p> <p><b>Between-day (ICC range:</b> -0.48–0.77)</p> <p><b>SEM range(sitting):</b></p> <p><b>Within-day</b> (0.71–1.48)</p> <p><b>Between-day</b> (0.72–0.99)</p> <p><b>Standing:</b></p> <p><b>Within-day (ICC range:</b> -0.11–0.68)</p> <p><b>Between-day (ICC range:</b> 0.09–0.58)</p> <p><b>SEM range (standing):</b></p> <p><b>Within-day</b> (0.91–1.48)</p> <p><b>Between-day</b> (0.82–1.22)</p> <p><b>Reliability:</b></p> <p><b>JPE:</b></p> <p><b>NHP (ICC range:</b> 0.35–0.44)</p> <p><b>THP (ICC range:</b> 0.69–0.74)</p> <p><b>Preset trunk rotation (ICC range:</b> 0.52–0.74)</p> <p><b>Figure of 8 relocation test (ICC:</b> 0.67)</p> <p><b>ICC range:</b></p> <p><b>NHP (0.38–0.84)</b></p> <p><b>THP (-0.48–0.83)</b></p> <p><b>SEM: 0.3–4</b></p>
Kristjansson et al. [22]	Asymptomatic Subjects: $n=19$ (12 females)	31.5 years	3-Space Fastrak	Sitting	<p><b>Trials:</b> 3 trials each direction or test</p> <p><b>Movements tested:</b> Right rotation, left rotation</p>	Reliability (Intra-rater between-day in the same week)	ICC (2,1)	
Lee et al. [23]	Asymptomatic Subjects: $n=20$ (11 men)	21.9 years	Motion analysis system, CMS 70P	Sitting	<p><b>Trials:</b> 3 trials each direction</p> <p><b>Movements tested:</b> Flexion/extension, left/right rotation, left/right side-bending (not randomised)</p>	Reliability (Intra-rater within day 10 min in between)	ICC (3, 1) SEM	

**Table 4** (continued)

Authors	Population and sample size	Age (mean or range)	Testing instrument	Testing position	Testing procedure	Property domain	Statistical test used	Results
Strimpakos et al. [24]	Asymptomatic Subjects: n = 35 (17 males)	18–63 years	Zebris CMS20	Intra-rater: sitting, standing Inter-rater (standing)	<b>Trials:</b> 3 trials <b>Movements tested:</b> Flexion, right and left rotation, right and left side bending (not randomised)	Reliability (Intra-rater between day 3 occasion and Inter-rater reliability, 15 min between assessors (10 subjects only)	ICC (1,1), SEM, Bland and Altman	<b>Intra-rater Sitting:</b> (ICC range: -0.01–0.35) <b>Standing:</b> (ICC range: 0.17–0.5) <b>SEM:</b> <b>Standing</b> (1.5–3) <b>Sitting</b> (1.5–3.5) <b>Inter-rater ICC range:</b> -0.2–0.64 <b>SEM:</b> 0.7–2.9
Pinsault et al. [25]	Asymptomatic subjects: n = 44 (22 women)	21.7 years	Laser pointer	Sitting	<b>Trials:</b> 10 trials <b>Movements tested:</b> Right and left rotation	Reliability (intra-rater 1 h apart)	ICC (2, 1), SEM, LoA	<b>ICC range:</b> (0.52–0.81) <b>SEM</b> (0.9) <b>LoA</b> (-2–2.2)
Kramer et al. [26]	Asymptomatic subjects: n = 57 (30 male, 27 females)	18–64 years	Virtual 3D scene via head mounted display + 3-Space Fastrak	Sitting	<b>Trials:</b> 8 trials in total <b>Movements tested:</b> Flexion, Extension, right and left rotation	Reliability (intra-session and inter-session)	ICC (3, 1)	<b>Intra-session ICC:</b> 0.63 <b>Inter-session ICC:</b> 0.48
Roren et al. [27]	Asymptomatic and CNP subjects: n = 82 (41 each group)	Healthy: 30.5 CNP: 54.7 years	Revel visual technique US technique	Sitting	<b>Trials:</b> 10 trials each device (5 trials each movement) <b>Movements tested:</b> Right and left rotation	Reliability: Intra-rater within day 1 h apart Criterion validity for both devices (Revel visual technique vs US technique) Discriminative validity (healthy vs NP group)	Reliability: ICC, Bland and Altman agreement. Criterion validity: Pearson's correlation. Discriminative validity: Kappa agreement	<b>ICC:</b> <b>Revel visual system</b> (0.68) <b>US technique</b> (0.62) <b>Bland and Altman agreement:</b> <b>Revel technique</b> (-3.6–4.2) <b>US technique</b> (-3.8–5.6) <b>Pearson's correlation range:</b> 0.946–0.952 <b>Kappa agreement:</b> 0.65



**Table 4** (continued)

Authors	Population and sample size	Age (mean or range)	Testing instrument	Testing position	Testing procedure	Property domain	Statistical test used	Results
Chen and Treleavan [28]	CNP subjects: n = 25 Asymptomatic subjects: n = 26	18–60 years	Fastrak, Laser pointer	Sitting	<b>Trials:</b> 6 trials each movement <b>Movements tested:</b> Right and left rotation	Criterion validity Discriminative validity Convergent validity	Pearson's correlation (criterion validity) Spearman's correlation (convergent validity) MANOVA (discriminative validity)	<b>Discriminative validity:</b> <b>Conventional Fastrak</b> ( $p=0.28$ ) <b>Conventional Laser</b> ( $p=0.04$ ) <b>Torsion Fastrak</b> ( $p=0.00$ ) <b>Torsion Laser</b> ( $p=0.02$ ) <b>Enbloc Fastrak (head)</b> ( $p=0.43$ ) <b>Enbloc Fastrak (trunk)</b> ( $p=0.42$ ) <b>Criterion validity: Conventional JPE</b> ( $r=0.87$ ) <b>Torsion JPE</b> ( $r=0.67$ ) <b>Convergent validity:</b> No correlation except Conventional Fastrak ( $r=0.51$ ) <b>Reliability:</b> <b>ICC range:</b> 0.79–0.85 <b>SEM range:</b> 1.4–2 <b>Validity:</b> <b>ICC range:</b> 0.43–0.91
Wibault et al. [5]	CDD subjects: n = 24 Asymptomatic subjects: n = 12	CDD: 51 years Healthy: 42 years	Reliability: CROM device Validity: CROM device vs laser pointer	Sitting	<b>Trials:</b> Reliability: 3 trials each direction Validity: 8 trials each direction <b>Movements tested:</b> Reliability and validity: right and left rotation	Reliability: 24 subjects with CDD (Intra-rater within day 1 h in between) Criterion validity: 12 healthy subjects	ICC (2,1) SEM	



**Table 4** (continued)

Authors	Population and sample size	Age (mean or range)	Testing instrument	Testing position	Testing procedure	Property domain	Statistical test used	Results
Dugailly et al. [29]	Asymptomatic and CNP subjects Validity group (n = 17) Reliability group (n = 5)	42 years	Laser + electrogoniometer	Sitting	<b>Trials:</b> 6 trials in each direction <b>Movements tested:</b> Right and left rotation, flexion, extension (no evidence of randomisation)	Criterion validity Intra-rater reliability (1 week apart) Convergent validity	Spearman's correlation ICC Bland Altman (LoA)	<b>ICC range:</b> 90 cm low speed: 0.22–0.47 90 cm high speed: 0.58–0.79 180 cm low speed: 0.52–0.75 180 cm high speed: 0.8–0.86 <b>Convergent validity:</b> JPE vs disability (r = 0.32) JPE vs pain intensity (r = 0.03), JPE vs pain duration (r = 0.14) LoA range: -9 to 9
Burke et al. [30]	Asymptomatic and NP subjects: n = 50	NA	CROM, AL	Sitting	<b>Trials:</b> 3 trials in each direction for two devices (no randomisation) <b>Movements tested:</b> Right and left rotation	Reliability (intra and inter-rater reliability)	ICC (type C)	<b>ICC range:</b> Intra-rater CROM: 0.253–0.386 Intra-rater AL: 0.488–0.556 Inter-rater CROM: 0.717–0.773 Inter-rater AL: 0.589–0.75
Alahmari et al. [3]	NP and Asymptomatic subjects: NP: n = 36 Asymptomatic: n = 33	Healthy: 56 year NP: 36 years	Digital inclinometer	NHP (sitting) THP (sitting and supine)	<b>Trials:</b> 3 trials in all tests and directions <b>Movements tested:</b> NHP (extension) THP (50% of ROM in flexion, extension, right and left side bending, right and left rotation in a randomised order)	Intra-rater reliability, Inter-rater reliability (≤ 3 working days apart)	ICC (2.1.A) SEM	<b>Intra-rater reliability:</b> NHP (ICC range: 0.74–0.78) (SEM: 1.78–1.88) THP: (ICC range: 0.7–0.83) (SEM: 1.78–1.88) <b>Inter-rater reliability:</b> NHP (ICC range: 0.74–0.79) (SEM: 1.79–1.87) THP (ICC range: 0.62–0.84) (SEM: 1.5–2.23)

**Table 4** (continued)

Authors	Population and sample size	Age (mean or range)	Testing instrument	Testing position	Testing procedure	Property domain	Statistical test used	Results
Goncalves and Silva [31]	Asymptomatic vs CNP subjects: n = 66 (33 each group)	Healthy: 43.6 years NP 43.5 years	Laser pointer on a helmet	Sitting	<p><b>Trials:</b> 6 trials each direction</p> <p><b>Movements tested:</b> Right and left rotation</p>	<p>Reliability (Intra-rater within day and between day with 1–2 days in between)</p> <p>Construct validity</p>	<p>Reliability: ICC (2,1) SEM</p> <p>Validity: t-tests or Mann Whitney, Pearson correlation coefficient, Spearman's correlation coefficient</p>	<p><b>Chronic Neck pain: Within day ICC range:</b></p> <p>HRNT: 0.9–0.93 TT: 0.88–0.9 HR30T: 0.73–0.79 F8T: 0.89–0.93</p> <p><b>Between day ICC range:</b></p> <p>HRNT: 0.61–0.85 TT: 0.58–0.71 HR30T: 0.67–0.7 F8T: 0.66–0.85</p> <p><b>Asymptomatic: Within day ICC range:</b></p> <p>HRNT: 0.79–0.89 TT: 0.75–0.87 HR30T: 0.78–0.83 F8T: 0.83–0.93</p> <p><b>Between day ICC range:</b></p> <p>HRNT: 0.75–0.8 TT: 0.57–0.59 HR30T: 0.55–0.76 F8T: 0.8–0.83</p> <p><b>All tests between groups (NP vs healthy) were &lt; 0.05 but the HR30T</b></p> <p><b>Between test correlations ranged between 0.35 and 0.61 and correlations between proprioceptive tests and catastrophizing, fear of movement and disability were, in general, lower than 0.3</b></p>

**Table 4** (continued)

Authors	Population and sample size	Age (mean or range)	Testing instrument	Testing position	Testing procedure	Property domain	Statistical test used	Results
Nikkhoo et al. [32]	Asymptomatic (35 participants)	21.2	US MOCAP (CMS 10, Zebris) + IMU-based mobile devices	Sitting	<p><b>Trials:</b> 5 trials</p> <p><b>Movements tested:</b> Flexion, extension, right and left rotation</p>	<p>Within-day and between-day intra-rater reliability (5–7 days in between)</p> <p>Measurement error</p> <p>Criterion validity</p>	<p><b>Reliability:</b> ICC two-way mixed model,</p> <p><b>Measurement error:</b> SEM</p> <p><b>Validity:</b> Pearson's correlation</p>	<p><b>Reliability:</b> <b>Within-day (US MOCAP):</b> ICC range 0.83–0.93</p> <p><b>Between-day (US MOCAP):</b> ICC range 0.69–0.85</p> <p><b>Within-day (IMU):</b> ICC range 0.66–0.91</p> <p><b>Between-day (IMU):</b> ICC range 0.63–0.76</p> <p><b>Validity:</b> <b>r range:</b> 0.74–0.83</p>
Cid Et al. [33]	Asymptomatic (14 men, 14 women) NP (13 women)	24.4, 23.1, 26.6	Laser pointer	Sitting	<p><b>Trials:</b> 10 trials</p> <p><b>Movements tested:</b> Right and left rotation</p>	<p>Within-day intra-rater reliability (at least 7 days in between)</p>	<p><b>Reliability:</b> ICC (two-way mixed model)</p>	<p><b>Reliability (NHP):</b> <b>NP:</b> ICC range 0.77–0.86</p> <p><b>Asymptomatic:</b> ICC range -0.16–0.5</p>

CNP chronic neck pain, NP neck pain, n number of subjects, NHP neutral head position, THP target head position, WK weighted kappa, ICC intraclass correlation coefficient, SEM standard error of measurement, LoA limits of agreement, JPE joint position error, CROM cervical range of motion, HRNT head repositioning to neutral, TT torsion test, F8T figure-of-eight test, HR30T head repositioning to 30 degrees test, CDD cervical disc disease, IMU inertial measurement unit, r Pearson's or Spearman's correlation

**Table 5** Summary of measurement properties of the measure of absolute JPE

<b>Neutral head position (Neck Pain population)</b>	<b>Summary or pooled results</b>	<b>Overall rating</b>	<b>Quality of evidence</b>
Intra-rater reliability	ICC: 0.58–0.93 Total sample size: 580	Sufficient	Very low evidence for sufficient intra-rater reliability <ul style="list-style-type: none"> <li>• Nine studies showed sufficient results, 4 showed insufficient results (Inconsistent results)</li> <li>• Multiple studies with doubtful/inadequate rating (risk of bias)</li> <li>• No imprecision</li> <li>• No indirectness</li> </ul>
Inter-rater reliability	ICC: 0.58–0.79 Total sample size: 169	Sufficient	Low evidence for sufficient inter-rater reliability <ul style="list-style-type: none"> <li>• Three studies showed sufficient results</li> <li>• Multiple studies with inadequate rating</li> <li>• No inconsistency</li> <li>• No imprecision</li> <li>• No indirectness</li> </ul>
Measurement error	Total sample size: 736	Indeterminate	Not possible to apply GRADE as the minimal important change was not provided
Convergent validity	Correlation ( $r < 0.5$ ) Total sample size: 1890	Insufficient	Low evidence for insufficient convergent validity <ul style="list-style-type: none"> <li>• Thirteen studies showed sufficient results, 17 studies showed insufficient results (Inconsistent results)</li> <li>• Multiple studies with adequate rating (no risk of bias)</li> <li>• No indirectness</li> <li>• No imprecision</li> </ul>
Discriminative validity	Total sample size: 496	Indeterminate	Very Low evidence for indeterminate discriminative validity <ul style="list-style-type: none"> <li>• Seven studies were indeterminate and 1 study was sufficient (inconsistent results)</li> <li>• Multiple studies with inadequate rating</li> <li>• No imprecision</li> <li>• No indirectness</li> </ul>
Criterion validity	$r = 0.87$ – $0.95$ Total sample size: 184	Sufficient	Low evidence for sufficient criterion validity <ul style="list-style-type: none"> <li>• Two studies were sufficient, 1 was insufficient (inconsistent results)</li> <li>• Multiple studies with adequate rating (no risk of bias)</li> <li>• No imprecision</li> <li>• No indirectness</li> </ul>
<b>Target head position (Neck Pain population)</b>	<b>Summary of pooled results</b>	<b>Overall rating</b>	<b>Quality of evidence</b>
Intra-rater reliability	ICC: 0.67–0.83 Total sample size: 135	Sufficient	Low evidence for sufficient intra-rater reliability <ul style="list-style-type: none"> <li>• Three studies showed sufficient results</li> <li>• Multiple studies with doubtful/inadequate rating</li> <li>• No imprecision</li> <li>• No indirectness</li> </ul>
Inter-rater reliability	ICC: 0.58–0.84 Total sample size: 69	Sufficient	Very low evidence of sufficient inter-rater reliability <ul style="list-style-type: none"> <li>• One study showed sufficient results</li> <li>• One study with inadequate rating (risk of bias)</li> <li>• Imprecision</li> <li>• No indirectness</li> </ul>
Measurement error	Total sample size: 204	Indeterminate	Not possible to apply GRADE as the minimal important change was not provided
<b>Neutral head position (asymptomatic population)</b>	<b>Summary or pooled results</b>	<b>Overall rating</b>	<b>Quality of evidence</b>
Intra-rater reliability	ICC: 0.52–0.93 Total sample size: 537	Sufficient	Very low evidence of sufficient intra-rater reliability <ul style="list-style-type: none"> <li>• Eleven studies showed sufficient results, 6 showed insufficient results (inconsistent results)</li> <li>• Multiple studies with doubtful/inadequate rating (risk of bias)</li> <li>• No imprecision</li> <li>• No indirectness</li> </ul>

**Table 5** (continued)

Inter-rater reliability	ICC: -0.2–0.64 Total sample size: 35	Insufficient	Very low evidence of insufficient inter-rater reliability <ul style="list-style-type: none"> <li>• One study showed insufficient results</li> <li>• One study with inadequate rating (risk of bias)</li> <li>• Imprecision (sample size &lt; 100)</li> <li>• No inconsistency</li> <li>• No indirectness</li> </ul>
Measurement error	Total sample size: 509	Indeterminate	Not possible to apply GRADE as the minimal important change was not provided
Intra-session reliability	ICC: 0.63 Total sample size: 57	Insufficient	Very low evidence of insufficient intra-session reliability <ul style="list-style-type: none"> <li>• One study with doubtful rating (risk of bias)</li> <li>• Imprecision</li> <li>• No inconsistency</li> <li>• No indirectness</li> </ul>
Inter-session reliability	ICC: 0.48 Total sample size: 57	Insufficient	Very low evidence of insufficient intra-session reliability <ul style="list-style-type: none"> <li>• One study with doubtful rating (risk of bias)</li> <li>• Imprecision</li> <li>• No inconsistency</li> <li>• No indirectness</li> </ul>
Criterion validity	Total sample size: 71	Inconsistent	Not possible to apply GRADE due to inconsistency of results <ul style="list-style-type: none"> <li>• One study showed indeterminate results and one showed sufficient results</li> </ul>
Target head position (asymptomatic population)	Summary of pooled results	Overall rating	Quality of evidence
Intra-rater reliability	ICC: -0.48–0.83 Total sample size: 165	Sufficient	Very low evidence of sufficient intra-rater reliability <ul style="list-style-type: none"> <li>• Four studies showed sufficient results, three showed insufficient results (inconsistency of results)</li> <li>• Multiple studies with doubtful/inadequate rating (risk of bias)</li> <li>• No indirectness</li> <li>• No imprecision</li> </ul>
Measurement error	Total sample size: 165	Indeterminate	Not possible to apply GRADE as the minimal important change was not provided

THP test was rated as sufficient, but the quality of evidence was downgraded to low due to risk of bias (multiple studies with doubtful/inadequate rating) (Table 5).

#### **Inter-rater reliability**

Only two studies investigated inter-rater reliability of the NHP test in this population, and both did not report type of neck pain. Alahmari et al. [3] was rated as inadequate in the RoB checklist and sufficient in the updated criteria for good measurement properties. Burke et al. [30] carried out their investigation using two devices the laser pointer and the CROM. Both were rated as inadequate in the RoB checklist and sufficient in the updated criteria for good measurement properties. A total of three investigation showing sufficient results. The overall rating was rated as sufficient, but the quality of evidence was downgraded to low due to risk of bias (multiple studies with inadequate ratings) (Table 5).

#### **Measurement error**

For the THP test, five studies investigated measurement error [3, 5, 27, 30, 31]. GRADE was not possible to apply due to minimal important change (MIC) not provided (Table 5). For the THP test, two studies investigated measurement error [3, 31]. GRADE was not possible to apply as the minimal important change was not provided (Table 5).

#### **Convergent validity**

Two studies investigated the convergent validity in this population and were on CNP people. Chen and Treleaven [28] correlated three JPE tests (conventional, TT, Enbloc) with the neck disability index (NDI) and the visual analogue scale (VAS). All parts were rated as adequate in the RoB checklist and insufficient in the updated criteria for good measurement properties, apart from the correlation of JPE conventional with VAS, which showed sufficient results. Goncalves and Silva [31] correlated four JPE tests (NHP, THP, TT, and F8T) against each other and

against disability, pain catastrophising, and fear of movement questionnaires. All parts were rated as adequate in the RoB checklist. Correlation of the tests against the questionnaires were rated as insufficient in the updated criteria for good measurement properties, while correlation of tests against each other were rated as sufficient. Seventeen investigations showed insufficient results and thirteen studies showed sufficient results. The overall rating was taken and rated as insufficient, and the quality of evidence was downgraded to low due to inconsistency of results (Table 5).

#### ***Discriminative validity***

Three studies investigated the discriminative validity in people with CNP. Chen and Treleaven [28] used three tests (JPE conventional, TT, Enbloc), Goncalves and Silva [31] used four tests (NHP, THP, TT, F8T), and Roren et al. [27] used the NHP test. All investigation were rated as inadequate in the RoB checklist. All studies were rate as indeterminate in the updated criteria for good measurement properties due to improper statistical tests used for analysis, apart from the study by Roren et al. [27], which was rated as sufficient. Seven studies showed indeterminate results and one study showed sufficient results. The overall rating of the discriminative validity was rated as indeterminate, and the quality of evidence was downgraded to very low due to inconsistency of results and risk of bias (multiple studies with inadequate rating) (Table 5).

#### ***Criterion validity***

The criterion validity was reported only in CNP population testing for only right and left rotation. Roren et al. [27] correlated the laser pointer against an US device in sitting position for the NHP test only. This study was rated as inadequate in the RoB checklist and sufficient in the updated criteria for good measurement properties. Chen and Treleaven [28] correlated the laser pointer against the 3-Space Fastrak for both the NHP and TT in sitting position. Both parts were rated as adequate in the RoB checklist. The conventional JPE was rated as sufficient, and the TT was rated as insufficient in the updated criteria for good measurement properties. Two investigations showed sufficient results and one showed insufficient results. The overall rating was rated as sufficient, and the quality of evidence was downgraded to low due to inconsistency of results (Table 5).

#### **Absolute joint position error for asymptomatic people**

##### ***Intra-rater reliability***

A total of six studies investigated intra-rater reliability of the NHP test in this population. Kristjansson et al. [22] carried their investigation on four JPE tests: NHP, Preset

trunk rotation, and F8T relocation test. All parts were rated as inadequate in the RoB checklist. The NHP and F8T investigations were rated as insufficient, and Present trunk rotation investigation was rated as sufficient in the updated criteria for good measurement properties. Strimpakos et al. [24] carried out their intra-rater investigation in sitting and standing. Both were rated as inadequate in the RoB checklist and insufficient in the updated criteria for good measurement properties. Pinsault et al. [25] was rated as doubtful in the RoB checklist and sufficient in the updated criteria for good measurement properties. Goncalves and Silva [31] carried out within-day and between day investigations for three NHP tests (NHP, TT, and F8T). All investigations were rater as doubtful in the RoB checklist. The between-day investigation of the TT was rated as insufficient, while the remaining investigations were rated as sufficient in the updated criteria for good measurement properties. Ninkhoo et al. [34] carried out within-day and between-day investigations using US MOCAP and IMU devices. All investigations were rated as doubtful in the RoB checklist and sufficient in the updated criteria for good measurement properties. Cid et al. [33] was rated as doubtful in the RoB checklist and insufficient in the updated criteria for good measurement properties. Eleven studies showed sufficient results and six studies showed insufficient results. The overall rating was sufficient, and the quality of evidence was downgraded to very low due to inconsistency of results and risk of bias (multiple studies with doubtful/inadequate rating) (Table 5).

Regarding the THP test, three studies investigated the intra-rater reliability of this test in this population [21, 22, 31]. Artz et al. [21] carried out within-day and between-day intra-rater reliability of THP test only in sitting and standing. All parts were rated as inadequate in the RoB checklist and insufficient in the updated criteria for good measurement properties, apart from the between-day assessment in sitting, which was rated as sufficient. Kristjansson et al. [22] was rated as inadequate in the RoB checklist and sufficient in the updated criteria for good measurement properties. Goncalves and Silva [31] carried out a within-day and between-day investigations, both investigation were rated as doubtful in the RoB checklist and sufficient in the updated criteria for good measurement properties. Four studies showed sufficient results and three studies showed insufficient results. The overall rating was rated as sufficient, but the quality of evidence was downgraded to very low due to risk of bias and inconsistency of results (Table 5).

##### ***Inter-rater reliability***

Only one study investigated inter-rater reliability of the NHP test [24] in this population. This study was rated as

inadequate in the RoB checklist and insufficient in the updated criteria for good measurement properties. The overall rating was insufficient, and the quality of evidence was downgraded to very low due to risk of bias and low imprecision (sample size < 100) (Table 5).

#### ***Intra-session reliability***

Only one study [26] investigated in intra-session reliability of the NHP test in this population. This study was rated as doubtful in the RoB checklist and insufficient in the updated criteria for good measurement properties. The overall rating was insufficient, and the quality of evidence was very low due to risk of bias and imprecision (sample size < 100) (Table 5).

#### ***Inter-session reliability***

Only one study [26] investigated in inter-session reliability of the NHP test in this population. This study was rated as doubtful in the RoB checklist and insufficient in the updated criteria for good measurement properties. The overall rating was insufficient, and the quality of evidence was very low due to risk of bias and imprecision (sample size < 100) (Table 5).

#### ***Measurement error***

For the NHP test, six studies investigated measurement error [21, 22, 24, 25, 31, 34]. GRADE was not possible to apply due to MIC no provided. For the THP test, three studies investigated measurement error [21, 22, 31]. GRADE was not possible to apply as the minimal important change was not provided.

#### ***Criterion validity***

Two studies investigated criterion validity in this population. Wibault et al. [5] was rated as doubtful in the RoB checklist and indeterminate in the updated criteria for good measurement properties. Nikkhoo et al. [32] was rated as adequate in the RoB checklist and sufficient in the updated criteria for good measurement properties. We were not able to take an overall rating as one study showed sufficient results and the other one showed indeterminate results. Therefore, the overall rating was indeterminate, and no GRADE was applied due to inconsistency of results (Table 5).

#### **Constant joint position error for asymptomatic people**

##### ***Intra-rater reliability***

Two studies investigated the intra-rater reliability of the NHP test. Lee et al. [23] was rated as inadequate in the RoB checklist and sufficient in the updated criteria for good measurement properties. Dugailly et al. [29] carried out four intra-rater reliability investigation of the NHP test; low and fast speeds at 90cm and 180cm from a

target. All four parts were rated as inadequate in the RoB checklist. Only the low speed at 90cm was rated as insufficient, while the remaining three were rated as sufficient in the updated criteria for good measurement properties. Four studies showed sufficient result, one study showed insufficient results. The overall rating was sufficient, and the quality of evidence was downgraded to very low due to inconsistency of results, risk of bias (multiple studies with inadequate ratings), and imprecision (sample size < 100) (Table 6).

For the THP test, only one study investigated the intra-rater reliability of this test [23]. This study was rated as inadequate in the RoB checklist and sufficient in the updated criteria for good measurement properties. The overall rating was sufficient, but the quality of evidence was downgraded to very low due to risk of bias and imprecision (sample size < 100) (Table 6).

##### ***Measurement error***

For the NHP test, two studies investigated the measurement error in this population [23, 29]. GRADE was not possible to apply due to MIC not provided (Table 6). For the THP, only one study investigated measurement error [23]. GRADE was not possible to apply as the minimal important change was not provided (Table 6).

##### ***Convergent validity***

One study by Dugailly et al. [29] correlated the JPE test against disability questionnaire, pain duration, and pain intensity. All parts were rated as adequate in the RoB checklist and insufficient in the updated criteria for good measurement properties. The overall rating was insufficient, and the quality of evidence was high due to multiple studies with adequate ratings (no risk of bias) (Table 6).

##### ***Criterion validity***

The criterion validity was reported only once by Dugailly et al. [29]. This study was rated as doubtful in the RoB checklist and indeterminate in the updated criteria for good measurement properties. The overall rating was indeterminate, and the quality of evidence was downgraded to very low due to risk of bias and imprecision (sample size < 50) (Table 6).

## **Discussion**

This is the first systematic review to synthesise and appraise the measurement properties of cervical JPE in people with and without neck pain using the COSMIN checklist. Our search yielded 8 studies that included neck pain participants and 7 in which asymptomatic participants were included. Absolute and constant errors were reported in this review since they are recommended



**Table 6** Summary of measurement properties of the measure of constant JPE

Neutral head position (asymptomatic population)		Overall rating	Quality of evidence
Intra-rater reliability	Summary or pooled results ICC: 0.38–0.86 Total sample size: 40	Sufficient	Very low evidence of sufficient intra-rater reliability <ul style="list-style-type: none"> <li>• Four studies showed sufficient results and 1 showed insufficient results (inconsistency)</li> <li>• Multiple studies with inadequate rating (risk of bias)</li> <li>• Imprecision (sample size &lt; 100)</li> <li>• No indirectness</li> </ul>
Measurement error	Total sample size: 40	Indeterminate	Not possible to apply GRADE as the minimal important change was not provided
Convergent validity	$r=0.03-0.32$ Total sample size: 213	Insufficient	High evidence for insufficient convergent validity <ul style="list-style-type: none"> <li>• Multiple studies with adequate rating (no risk of bias)</li> <li>• No inconsistency of results</li> <li>• No imprecision</li> <li>• No indirectness</li> </ul>
Criterion validity	Total sample size: 17	Indeterminate	Very Low evidence for indeterminate criterion validity <ul style="list-style-type: none"> <li>• One study with doubtful rating (risk of bias)</li> <li>• Imprecision (sample size &lt; 50)</li> <li>• No indirectness</li> </ul>
Target head position (asymptomatic population)		Overall rating	Quality of evidence
Intra-rater reliability	Summary of pooled results ICC: -0.47–0.83 Total sample size: 20	Sufficient	Very low quality of evidence for sufficient intra-rater reliability <ul style="list-style-type: none"> <li>• One study showed sufficient results</li> <li>• One study with inadequate rating (risk of bias)</li> <li>• Imprecision</li> <li>• No inconsistency</li> <li>• No indirectness</li> </ul>
Measurement error	Total sample size: 20	Indeterminate	Not possible to apply GRADE as the minimal important change was not provided

when assessing JPE [35]. The large range of testing procedures used in the studies reviewed highlight the lack of any consensus in the literature on how best to assess JPE. A key factor contributing to this may be the heterogeneity of neck pain participants recruited for the reviewed studies, each with different clinical features. Given these differences in testing procedures and the vast range in types of neck pain, it is difficult to draw any general conclusions on the gold standard for testing the measurement properties of cervical JPE.

Similar to other systematic reviews, the current systematic review highlighted several issues with the quality of the included studies [36, 37]. Most of the included studies in this review were rated as inadequate or doubtful in the RoB checklist with an overall quality of the evidence being low to very low, apart from the convergent validity of the constant JPE, which was high. This was due to a failure in adhering to COSMIN guidelines when carrying out investigations of measurement properties of outcome measures. For example, according to COSMIN, the time-interval should be long enough to prevent recall bias, and short enough to ensure that the patients have not been changed on the construct to be measured [13]. When assessing the RoB for reliability and measurement error, there are no guidelines for the time-interval between

sessions, therefore, this section was rated as doubtful. Other issues highlighted were statistical tests used for validity investigations. COSMIN recommends Pearson’s or Spearman’s correlation for validity assessment, which the criterion validity in the constant JPE did not use. Therefore, some of the included studies were rated as indeterminate in the updated criteria for good measurement properties. A further limitation in the included studies was when the model of the ICC used for reliability assessment was not stated. When using the RoB checklist for reliability and measurement error [14], if a study used ICC and reported the model used, it should be rated as very good; if the study used ICC but failed to report the model, then it should be rated as inadequate. Three studies failed to report the ICC model used [21, 27, 29], thus, they were rated as inadequate in the RoB checklist. Reporting the ICC model is important because the model used and the type of coefficient will impact on the magnitude of the ICC [38]. Failure to report the ICC model will affect the study’s generalisability and interpretation of the results. Inclusion of a replicable measure of response stability will aid the interpretation of results and comparison between studies.

Another issue in the current review was the inconsistency of results for the criterion validity of absolute JPE

in the asymptomatic population. This inconsistency was probably due to differences in statistical tests used for validity assessment and variations in testing protocols. For example, Wibault et al. [5] correlated the CROM device against a laser pointer after returning from right and left rotation using three trials per movement in their assessment. They used the ICC for their validity assessment, which is not recommended by COSMIN, and thus were rated as indeterminate in the updated criteria for good measurement properties. Nikkhoo et al. [32] correlated the US MOCAP against IMUs after returning from flexion, extension, and bilateral rotation using five trials per movement. This study was rated as sufficient in the updated criteria for good measurement properties. Therefore, it was not possible to draw an overall rating for this measurement property due to inconsistency of the results and it was rated as indeterminate. The convergent validity on the other hand was rated as high. This was due to no risk of bias in the included studies; however, it did not show sufficient results. Sample size was another issue that affected the overall rating of an outcome measure. When applying the modified GRADE approach, sample size should be  $\geq 100$ . However, the total sample size of the inter-rater reliability of absolute JPE in the asymptomatic population was 62 participants; this led to downgrading the overall evidence to one level. Similarly, the criterion validity of constant JPE was downgraded to two levels due to sample size  $< 50$ . In addition, the wording around reliability studies was challenging as several studies did not report the word "reliability" in the title of the study, affecting the quality of the study.

Furthermore, the current systematic review highlighted gaps in the literature when testing the measurement properties of the measure of cervical JPE. First, the testing position. Most of the included studies carried out their investigations in sitting. Only two studies carried out their investigation in sitting as well as standing [21, 24]. However, these two studies did not include any neck pain patients, and only asymptomatic participants were recruited. In addition, they reported only constant JPE, failing to report absolute JPE. A second gap was the lack of investigation of inter-rater reliability of constant JPE in people with neck pain. The third gap we uncovered was regarding the criterion validity of absolute JPE. Although this property was investigated twice, it was limited to right and left rotation. Lastly, the domain of responsiveness was not reported in our systematic review.

### Methodological considerations

This is the first systematic review to summarise and appraise the evidence of measurement properties of the cervical JPE measure using COSMIN guidelines. Two raters carried out the study selection, data extraction, the

risk of bias checklist, and the GRADE approach minimising bias, which is considered a strength of this systematic review. Additionally, we included studies that have reported absolute and constant errors, which is recommended when testing cervical proprioception [35]. Prospective registration with PROSPERO is another strength of this review. A potential limitation is that the principle of lowest rating counts when using the COSMIN risk of bias checklist, thus underestimating the overall quality of the study, and potentially downgrading the overall quality of the evidence.

### Recommendations for future research

Additional research is clearly warranted to assess the measurement properties of the measure of JPE in people with and without neck pain. Another recommendation is to report both absolute and constant errors in future research. Also, assessing the measurement properties of the measure of JPE in standing in addition to sitting is recommended, as well as reporting absolute and constant error for both. Responsiveness of the measure of JPE was not investigated, which we recommend investigating in future research.

### Conclusion

Conclusions about the measurement properties of the measure of cervical JPE were difficult to draw due to lack of consensus on testing procedures and tools used. Further high-quality research to overcome the risk of bias in the included studies is required. Studies are also required to investigate the responsiveness of this measure.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12891-023-07111-4>.

**Additional file 1.**

**Additional file 2.**

**Additional file 3.**

**Additional file 4.**

**Additional file 5.**

**Additional file 6.**

### Acknowledgements

Not applicable

### Authors' contributions

AA is a PhD student and DF is lead supervisor and JD, is co-supervisor. AA drafted the initial version of the manuscript with the supervisory team providing guidance on topic, methodology and analyses. SA was the second reviewer for stage 2 data extraction, risk of bias checklists, updated criteria for good measurement property and the GRADE approach. AA completed the analysis. All authors reviewed and commented on each draft. All authors have approved the final draft for submission.

## Funding

The authors declare that this research received no funding for this research.

## Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

## Declarations

### Ethics approval and consent to participate

No research ethics were required for this systematic review due to no patient data being collected.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

### Author details

<sup>1</sup>Centre of Precision Rehabilitation for Spinal Pain (CPR Spine), School of Sport, Exercise and Rehabilitation Sciences, College of Life and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.

<sup>2</sup>Department of Physical Therapy, Al-Sabah Medical Hospital, Ministry of Public Health, Kuwait City, Kuwait. <sup>3</sup>School of Physical Therapy, Western University, London, ON, Canada.

Received: 19 April 2023 Accepted: 12 December 2023

Published online: 10 January 2024

## References

- Peng B, Yang L, Li Y, Liu T, Liu Y. Cervical proprioception impairment in neck pain-pathophysiology, clinical evaluation, and management: a narrative review. *Pain Ther.* 2021;10:143–64.
- Qu N, Tian HC, De Martino E, Zhang B. Neck pain: do we know enough about the sensorimotor control system? *Front Comput Neurosci.* 2022;16(July):1–10.
- Alahmari K, Reddy RS, Silvian P, Ahmad I, Nagaraj V, Mahtab M. Intra- and inter-rater reliability of neutral head position and target head position tests in patients with and without neck pain. *Brazilian J Phys Ther.* 2017;21(4):259–67.
- Revel M, Andre-Deshays C, Minguet M. Cervicocephalic kinesthetic sensibility in patients with cervical pain. *Arch Phys Med Rehabil.* 1991;72(5):288–91.
- Wibault J, Vaillant J, Vuillerme N, Dederich Å, Peolsson A. Using the cervical range of motion (CROM) device to assess head repositioning accuracy in individuals with cervical radiculopathy in comparison to neck-healthy individuals. *Man Ther.* 2013;18(5):403–9.
- Reddy RS, Tedla JS, Dixit S, Abohashrh M. Cervical proprioception and its relationship with neck pain intensity in subjects with cervical spondylosis. *BMC Musculoskelet Disord.* 2019;20(1):1–7.
- Feipel V, Salvia P, Klein H, Rooze M. Head repositioning accuracy in patients with whiplash-associated disorders. *Spine (Phila Pa 1976).* 2006;31(2):51–8.
- Treleaven J. Dizziness, unsteadiness, visual disturbances, and sensorimotor control in traumatic neck pain. *J Orthop Sports Phys Ther.* 2017;47(7):492–502.
- Treleaven J, Jull G, Sterling M. Dizziness and unsteadiness following whiplash injury: characteristic features and relationship with cervical joint position error. *J Rehabil Med.* 2003;35(1):36–43.
- Mazaheri M, Abichandani D, Kingma I, Treleaven J, Falla D. A meta-analysis and systematic review of changes in joint position sense and static standing balance in patients with whiplash-associated disorder. *PLoS One.* 2021;16(4):e0249659.
- Roach KE. Measurement of health outcomes: reliability, validity and responsiveness. *J Prosthetics Orthot.* 2006;18(6):P8–12.
- Michiels S, De Hertogh W, Truijten 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. <https://doi.org/10.1016/j.gaitpost.2012.10.007>.
- Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res.* 2018;27:1147–57.
- Mokkink LB, Boers M, van der Vleuten CPM, Bouter LM, Alonso J, Patrick DL, et al. 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(1):1–18.
- Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Rev Esp Nutr Humana y Diet.* 2016;20(2):148–60.
- Lidwine B Mokkink, Cecilia AC Prinsen, Donald L Patrick, Jordi Alonso, Lex M Bouter, Henrica CW de Vet, et al. COSMIN Study Design checklist for Patient-reported outcome measurement instruments. *Dep Epidemiol Biostat Amsterdam Public Heal Res Inst Amsterdam Univ Med Centers, Locat VUmc.* 2019;
- Cooke A, Smith D, Booth A. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. *Qual Health Res.* 2012;
- Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, De Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res.* 2012;21(4):651–7.
- Mokkink Cecilia AC Prinsen Donald L Patrick Jordi Alonso Lex M Bouter Henrica CW de Vet Caroline B Terwee Contact LB Mokkink LB. COSMIN manual for systematic reviews of PROMs COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) user manual. 2018. Available from: [www.cosmin.nl](http://www.cosmin.nl). Cited 2020 Feb 11.
- Moher D, Liberati A, Tetzlaff J, Altman DG, Altman D, Antes G, et al. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine.* 2009;8(12):e83138.
- Artz NJ, Adams MA, Dolan P. Sensorimotor function of the cervical spine in healthy volunteers. *Clin Biomech.* 2015;30(3):260–8.
- Kristjansson E, Dall'alba P, Jull G. Cervicocephalic kinaesthesia: Reliability of a new test approach. *Physiother Res Int.* 2001;6(4):224–35.
- Lee HY, Teng CC, Chai HM, Wang SF. Test-retest reliability of cervicocephalic kinesthetic sensibility in three cardinal planes. *Man Ther.* 2006;11(1):61–8.
- Strimpakos N, Sakellari V, Gioftsos G, Kapreli E, Oldham J. Cervical joint position sense: an intra- and inter-examiner reliability study. *Gait Posture.* 2006;23(1):22–31.
- Pinsault N, Fleury A, Virone G, Bouvier B, Vaillant J, Vuillerme N. Test-retest reliability of cervicocephalic relocation test to neutral head position. *Physiother Theory Pract.* 2008;24(5):380–91.
- Kramer M, Honold M, Hohl K, Bockholt U, Rettig A, Elbel M, et al. Reliability of a new virtual reality test to measure cervicocephalic kinaesthesia. *J Electromyogr Kinesiol.* 2009;19(5):e353–61. <https://doi.org/10.1016/j.jelekin.2008.05.005>.
- Roren A, Mayoux-Benhamou MA, Fayad F, Poiraudou 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.
- Chen X, Treleaven J. The effect of neck torsion on joint position error in subjects with chronic neck pain. *Man Ther.* 2013;18(6):562–7.
- Dugaillly PM, De Santis R, Tits M, Sobczak S, Vigne A, Feipel V. Head repositioning accuracy in patients with neck pain and asymptomatic subjects: concurrent validity, influence of motion speed, motion direction and target distance. *Eur Spine J.* 2015;24:2885–91.
- Burke S, Lynch K, Moghul Z, Young C, Saviola K, Schenk R. The reliability of the cervical relocation test on people with and without a history of neck pain. *J Man Manip Ther.* 2016;24(4):210–4.
- 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.
- Nikkhoo M, Niu CC, Fu CJ, Lu ML, Chen WC, Lin YH, et al. Reliability and validity of a mobile device for assessing head control ability. *J Med Biol Eng.* 2021;41(1):45–52. <https://doi.org/10.1007/s40846-020-00577-w>.

33. Cid MM, Calixtre LB, da Silva Grüninger BL, Sousa FS, Oliveira AB. Reliability of the joint position sense error test for women with neck pain and asymptomatic men and women. *J Manipulative Physiol Ther.* 2022;45(5):329–36.
34. Nikkhoo M, Niu CC, Fu CJ, Lu ML, Chen WC, Lin YH, et al. Reliability and validity of a mobile device for assessing head control ability. *J Med Biol Eng.* 2021;41(1):45–52.
35. Hill R, Jensen P, Baardsen T, Kulvik K, Jull G, Treleaven J. Head repositioning accuracy to neutral: A comparative study of error calculation. *Man Ther.* 2009;14(1):110–4.
36. Althobaiti S, Rushton A, Aldahas A, Falla D, Heneghan NR. Practicable performance-based outcome measures of trunk muscle strength and their measurement properties: a systematic review and narrative synthesis. *PLoS One.* 2022;17(6 June):1–30. <https://doi.org/10.1371/journal.pone.0270101>.
37. Middlebrook N, Rushton AB, Abichandani D, Kuithan P, Heneghan NR, Falla D. Measures of central sensitization and their measurement properties in musculoskeletal trauma: a systematic review. *Eur J Pain (United Kingdom).* 2021;25(1):71–87.
38. Harvey ND. A Simple Guide to Inter-rater, Intra-rater and Test-retest Reliability for Animal Behaviour Studies. *OSF Prepr.* 2021;1–13. Available from: <https://osf.io/8stpy>.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

