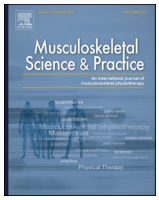
Musculoskeletal Science and Practice 64 (2023) 102738

Contents lists available at ScienceDirect

Musculoskeletal Science and Practice

journal homepage: www.elsevier.com/locate/msksp



Original article

Physical testing in patients with acute whiplash-associated disorders: A within session test-retest reliability study

E. Anarte-Lazo [a,b], C. Rodriguez-Blanco [c,*], D. Falla [b], C. Bernal-Utrera [c]

[a] Doctoral Program in Health Sciences, University of Seville, Seville, Spain
[b] Centre of Precision Rehabilitation for Spinal Pain (CPR Spine), School of Sport, Exercise and Rehabilitation Sciences, University of Birmingham, Birmingham, United Kingdom
[c] Physiotherapy Department, Faculty of Nursing, Physiotherapy and Podiatry, University of Seville, Seville, Spain

ARTICLE INFO

Keywords:
Whiplash-associated disorders
Physical testing
Test-retest reliability
Neuromusculoskeletal disorders

ABSTRACT

Background: People with whiplash-associated disorders (WAD) commonly present with a variety of physical impairments. However, the reliability of physical tests has not been established for patients with acute WAD.
Objective: To assess test-retest reliability of different physical tests in acute WAD.
Design: Intra-rater test-retest reliability.
Methods: Patients with acute WAD were recruited. Physical tests were used to evaluate articular, muscular and neural systems in two blocks of measurements separated by 10 min. Bland-Altman plots were performed to assess intrarater agreement, which included calculation of the mean difference (*d*) between rates, the 95% CI for *d*, the standard deviation of the differences and the 95% limits of agreement. Reliability was calculated via the standard error of measurement, the minimal detectable change, percent of agreement, the intraclass-correlation coefficient, and kappa coefficient.
Results: 47 patients participated. Test-retest reliability was excellent or good for almost all measures, except for extension ROM, ULTT for the radial nerve, and active cervical extension and upper cervical rotation performed in 4-point kneeling, which presented moderate reliability. Systematic bias was found in cervical ROM in flexion, left and right lateral-flexion, left and right rotation; left ULTT for radial nerve; right trapezius, suboccipitalis and temporalis muscles, left temporalis; C3, both sides of C1–C2, left C3–C4.
Conclusion: The majority of physical tests achieved good or excellent test-retest intra-rater reliability when tested in patients with acute WAD. Findings must be considered with caution for those tests which demonstrated systematic bias. Additional research is warranted to evaluate inter-rater reliability.

1. Introduction

Whiplash-associated disorders (WAD) are a disabling and costly condition (Spitzer et al., 1995); approximately 50% of individuals suffering from this condition will continue to report symptoms after the initial injury (Sterling, 2014). The Quebec Task Force (QTF) classification of whiplash injuries (Spitzer et al., 1995) is the classification method most commonly adopted. However, diagnosis of peripheral pathology in people with WAD is challenging since specific tissue damage or peripheral lesions are often not visible (Carroll et al., 2008). Nevertheless, there is evidence describing the presence of peripheral lesions in some individuals after a whiplash injury (Sterling et al., 2011). Among these peripheral injuries which could lead to nociception, different structures can be affected including the zygapophysial joints and capsules, ligaments, discs, muscles, or nerves, (Curatolo et al., 2011; Ettlin et al., 2008; Fundaun et al., 2021).

Given the challenges in identify a pathoanatomical source of pain in people with WAD, much attention has focused on characterising their physical and psychological impairments (Carroll et al., 2008). Several disturbances in physical function have been identified in people with acute and chronic WAD including decreased maximum angular velocity, range and smoothness of active neck movement (Alalawi et al., 2022; Baydal-Bertomeu et al., 2011) as well as reduced strength and endurance of neck muscles (Pearson et al., 2009; Jull, 2011). Additionally, an impairment in local mechanical hyperalgesia at the cervical spine and at remote sites, increased sensitivity on upper limb neurodynamic testing

* Corresponding author. Department of Physiotherapy, Faculty of Nursing, Physiotherapy and Podiatry, University of Seville, Spain.
E-mail address: cleofas@us.es (C. Rodriguez-Blanco).

https://doi.org/10.1016/j.msksp.2023.102738
Received 29 September 2022; Received in revised form 18 January 2023; Accepted 27 February 2023
Available online 28 February 2023

(ULNT) and a greater prevalence of trigger points in cervical muscles has been observed (Fernández-Pérez et al., 2012). Therefore, a thorough physical examination is essential to identify targets for rehabilitation (Huhn et al., 2019). Recently, recommendations for a core outcome domain set for WAD were published, which included neck posture, range of motion, muscle endurance and pain thresholds as recommended measures in clinical studies (Chen et al., 2019). An essential requirement for all outcome measures is that they are valid and reproducible or reliable (de Vet et al., 2006). Although many studies include physical testing of people with neck pain (Jull et al., 2011) and physical tests are commonly applied in clinical practice, there is a lack of studies examining the reliability of physical testing specifically in people with WAD. In particular, due to the heterogeneity in acute WAD clinical presentation, the use of reliable tests to help in the identification of patients who could benefit from tailored physical therapy interventions in their early management is relevant (Jonsson and Rasmussen-Barr, 2018). Therefore, the objective of the present study is to assess the intra-rater within-session test-retest reliability of a battery of tests performed by a physical therapist, evaluating articular, muscular and neural structures and function, in patients with acute WAD.

2. Methods

2.1. Study design

An intra-rater within-session test-retest reliability study was carried out involving patients with acute pain attributed to a whiplash injury due to a traffic accident who were attending a Traumatology Clinic in Madrid, Spain, from September 2020 to February 2021. Ethical approval was granted by the institutional human research ethics committee from University Rey Juan Carlos, Madrid, Spain (Ref: 1003202108121). All participants gave their written informed consent to participate in this study. The study was conducted according to the Declaration of Helsinki and is reported in accordance with GRRAS Guidelines (Kottner et al., 2011).

2.2. Participants

Consecutive patients with a diagnosis of acute WAD were recruited from the Traumatology Department of the Clinic. After being diagnosed by a physician, who then informed patients about the study, those that agreed to participate provided written informed consent and were referred to the Physiotherapy Department.

Inclusion criteria consisted of Grade II WAD, as defined by The Quebec Task Force on Whiplash-Associated Disorders (Spitzer et al., 1995) between 7 and 30 days after the accident and aged between 18 and 65 years old, to avoid the inclusion of older people who may have declined physical function. Individuals were excluded if they were diagnosed with fibromyalgia or had a history of generalized pain, had experienced a previous whiplash injury, had been diagnosed with osteoporosis, cervical myelopathy, had a diagnosed temporomandibular disorder (TMD), vertebral fractures and/or, inflammatory or rheumatic diseases, had a known psychological disorder or congenital disturbances, had undergone previous surgery in the cervical region, had received physical therapy treatment after the accident but before participation in the study, or were not able to complete patient-reported outcome measures. In addition, with the aim of excluding people suffering from concussion, we followed the criteria of the International Headache Society (Headache Classification Committee of the, 2018) and excluded people that had experienced one or more of the following signs and/or symptoms: confusion, disorientation or impaired consciousness; loss of memory for events immediately before or after the accident; and one or more of the following: nausea, vomiting, visual disturbances, dizziness and/or vertigo, gait and/or postural imbalance, and impaired memory and/or concentration.

The sample size estimation was performed using the Grammo calculator v.7.12. Using the method developed by Shoukri et al. (2004), a sample size of 37 participants was required to detect an ICC between 0.60 and 0.80 estimating an alpha risk of 0.05 and a beta risk of 20% (0.20).

2.3. Procedures

All measurements were collected in a single session conducted at a research center within the Physiotherapy Department by the same rater, who was a physical therapist with four years of experience and with a Master's Degree in Orthopaedic Manual Therapy.

All measures were evaluated twice, and the mean of both was used for the analysis. When tests were performed bilaterally then the outcomes were considered separately for each side. Where test performance was defined as positive/negative or correct/incorrect, we considered that the person was able to do the test when one of the two repetitions was positive/correct.

2.4. Outcome measures

Age, sex, height, and weight were recorded for all participants.

2.4.1. Articular system

Cervical Range of Motion (CROM,°). Flexion, extension, lateral-flexion and rotation were assessed in a relaxed sitting position using a smartphone Xiaomi® MiA2. Participants were asked to sit comfortably on a chair with back support with both feet flat on the floor, and their hips and knees at 90° (Fernández-de-las-Peñas et al., 2006). Smartphone apps (Android Clinometer Application for the frontal plane and Smartphone Compass Application for the horizontal plane) were used for this purpose, as previously described (Tousignant-Laflamme et al., 2013; Ghorbani et al., 2020).

Passive Accessory Intervertebral Movements (PAIVMs). Central and bilateral posterior-anterior intervertebral movements were applied as a grade III over C1–C3 (central, spinous processes) and C0–C1/C3–C4 (bilateral, zygapophyseal joints). The pain intensity provoked through the movement was recorded on via a Numeric Rating Scale (NRS), ranging from 0 (no pain) to 10 (worst pain imaginable) (Hengeveld and Banks, 2013; Luedtke et al., 2018).

Flexion-Rotation Test (FRT). The participant lay in supine on the plinth. They were asked to relax while their neck was moved to end range cervical flexion by the examiner. In this flexed position, the head and neck were passively rotated as far as possible within comfortable limits, and the number of degrees (°) of rotation was recorded (Hall and Robinson, 2004) with the Smartphone Compass Application. The test was performed bilaterally.

Forward Head Posture (FHP). FHP was assessed in a relaxed standing and sitting position via a lateral photograph taken from a distance of 1.5 m. Reference markers were placed on the spinous process of C7 and on the tragus of the ear, which were identified through palpation (Shaghayegh Fard et al., 2016). The smartphone image was introduced in FHPapp to obtain the calculation of the Cranio-Vertebral Angle (CVA) (Gallego-Izquierdo et al., 2020). FHP was defined as a CVA smaller than 48°, as described previously (Shaghayegh Fard et al., 2016).

2.4.2. Muscular system

Muscle palpation. Palpation was performed at predetermined points over different muscles. A single location in the middle of muscle belly was palpated and the participants were asked to rate their pain intensity upon palpation over the upper trapezius, suboccipitalis, masseter, temporalis and sternocleidomastoid (SCM) bilaterally. Pincer palpation was performed for upper trapezius and SCM whereas pressure palpation was performed for the remaining muscles. Pain intensity was recorded via a NRS, ranging from 0 (no pain) to 10 (the worst pain imaginable). All points were assessed with the subject laying supine with the neck in a neutral position.

Cranio-Cervical Flexion Test (CCFT). The participant lay supine with the neck in a neutral position, supported by towels as needed. An uninflated pressure cuff (Chattanooga Stabilizer Group Inc., Hixson, TN, USA) was placed behind the neck so that it abutted the occiput and was then inflated to a stable baseline pressure of 20 mmHg, filling the space between the testing surface and the neck without pushing the neck into a lordosis (Jull et al., 2008). The highest level of the five stages of the cranio-cervical flexion test (22–30 mmHg) that was held for 10 s without substitution using excessive superficial neck muscle activity was recorded as described previously (Luedtke et al., 2018). The highest level they achieved over the two repetitions of the test was used for analysis.

Neck flexor endurance. The test was performed with the participant positioned in supine on the plinth. The participant's head was positioned in slight upper cervical flexion by the examiner who placed his left hand on the table just below the participant's occiput. The participant was then asked to gently flex his/her upper neck and lift his/her head off the examiner's hand while retaining upper cervical flexion. Verbal feedback ("tuck your chin in" or "hold your head up") was given to the participant when their head touched the examiner's hand during the test. The test was terminated if the participant was unable to maintain the position of their head off the examiner's hand despite verbal encouragement or if they reached the maximum holding time of 30 s (Edmondston et al., 2008).

Neck extensor endurance. This test measured the time, in seconds, to keep the head steady, while lying in a prone position with the head over the edge of the plinth in a neutral position (Ris et al., 2017). This test was terminated if the participant lost the position despite verbal encouragement or reached a maximum holding time of 30 s.

Active cervical extension in 4-point kneeling (4K Extension). Positioned in 4-point kneeling on a plinth, the participant was asked to look between their hands, then look down to flex the head and the neck together as far as they could go and then curl their neck and head back up into extension as far as they could whilst keeping their gaze fixed between their hands. The test aimed to assess the quality of cervical extension while keeping the cranio-cervical region in a neutral position. Poor performance was considered if the patient was unable to dissociate mid-lower from upper cervical extension, as described previously (Segarra et al., 2015). The test was considered successful if the participant was able to perform the test in at least one of the two repetitions.

Active upper cervical rotation in 4-point kneeling (4K Upper Rotation). Positioned in 4-point kneeling on a plinth, the participant was asked to perform small ranges of cranio-cervical rotation to both sides (no greater than 40°), while maintaining their neck in a neutral position. Poor performance was considered when the patient was unable to dissociate upper cervical rotation movement from movement at the typical cervical region i.e., excessive motion of the typical region occurs (Segarra et al., 2015). The test was considered successful if the participant was able to perform the test in at least one of the two repetitions.

2.4.3. Neural system

Mechanosensitivity of the median, radial and ulnar nerves. Upper limb tension tests (ULTT) for the median (ULTT1), radial (ULTT2) and ulnar (ULTT3) nerves were assessed as described previously (Ris et al., 2017). The elbow was the last joint moved (extension for median and radial, flexion for ulnar nerve) during each test and the range of elbow extension was recorded in degrees (°) with a standard goniometer at the point where the patient reported discomfort (Nee et al., 2012).

Mechanosensitivity during Upper Limb Tension Testing (ULTT) combined with Cranio-Cervical Flexion (CCF). The patient was asked to perform active CCF and then the ULTT1 was performed as described previously (Zito et al., 2006). The elbow was the last joint moved and the range of elbow extension was recorded in degrees (°) with a standard goniometer at the point where the patient reported discomfort.

Pain Pressure Thresholds over the median, radial, ulnar, supra-orbital and greater occipital nerve. Pressure pain thresholds were measured bilaterally using a digital algometer (Force Ten™-Model FDX, Wagner, Greenwich, USA) with a surface area of round tip of 1 cm^2 and were recorded in N/cm^2. The supra-orbital nerve was tested over the supra-orbital notch (at the junction between the medial third and the two lateral thirds of the upper part of the margin of the orbit); the median nerve was located over the cubital fossa medial to and adjacent to the biceps tendon; the radial nerve was marked where it passes through the lateral intermuscular septum between the medial and lateral heads of the triceps brachii to enter the mid to lower third of the humerus; the ulnar nerve was located in the groove between the medial epicondyle and the olecranon; the greater occipital nerve was assessed approximately 2 cm medial to the greater occipital protuberance (Szikszay et al., 2018; Fernández-de-Las-Peñas et al., 2011).

Ten minutes of rest was provided between repeated testing for the assessment of intra-rater reliability. During this time, participants sat on a chair resting. The order of testing was comparable between sets and endurance/motor control tests were always performed at the end of each set to avoid the possible influence of hypoalgesic effects of exercise. The order of testing was chosen to minimise change in the patients position and adhered to the following sequence: CROM, FHP, PAIVMs, PPT over the greater occipital nerve, FRT, muscle palpation, PPTs over the other nerves, ULTTs and ULTT + CCF, CCFT, neck flexor endurance, neck extensor endurance, 4K Extension and 4K Upper Rotation.

2.5. Statistical analysis

SPSS software was used for all statistical analyses (IBM SPSS 25 for Mac, Armonk, NY, USA). Firstly, Shapiro Wilk's test was used to assess normality of the data. Student's t-test was applied for test-retest when the data had a parametric distribution. Logarithmic10 transformation was applied to data in the presence of non-normality and normality was retested to ensure that previous assumptions were met.

Bland Altman plots were performed to detect systematic biases and 95% limits of agreement (LOA) for each measurement using a scatter graph of the differences and the means of assessments (Bland and Altman, 1999). The difference in means (d) between test and retest and the standard deviation (SD) for this difference (SDd) was quantified to obtain 95% LOA. Next, d±1.96SDd was calculated, indicating the total error (bias and random error together, corresponding to 95% LOA). The presence of bias is estimated by calculating the 95%CI for d. If zero lies outside the 95% CI of d, there are systematic biases between the measurements (Smidt et al., 2002; Bland and Altman, 1986).

2.5.1. Numerical data - absolute reliability

Absolute reliability was assessed using the standard error of measurement (SEM). SEM represents the within-subject variation and is defined as "the standard deviation of errors of measurement that is associated with the test scores for a specific group of test takers" (Atkinson and Nevill, 1988; Harvill, 1991). SEM was calculated as:

$$SD \times \sqrt{1 - ICC}$$

Responsiveness was assessed using the minimal detectable change (MDC). MDC_{90} expresses the minimal change required to be 90% confident that the observed change between the two measures reflects the real change and not measurement error (Lexell and Downham, 2005); it is calculated as:

$$SEM \times \sqrt{2} \times 1.96$$

The measurement of SEM and MDC were also presented as percentage of pooled means (average of test and retest measurements) designated as %SEM and %MDC, which allow comparisons between studies and facilitate interpretation.

In categorical data, absolute reliability was calculated as the percentage (%) of agreement between measurements.

2.5.2. Numerical data - relative reliability

The relative reliability was calculated using the intraclass correlation coefficient (ICC). ICC's model 2,1 was used since each subject was assessed by the same rater and the rater represents the population of possible raters (Jonsson and Rasmussen-Barr, 2018). ICC 95% confidence interval was calculated to represent ICC variability. An ICC >0.80 was considered "excellent"; between 0.61 and 0.80, good; between 0.41 and 0.60, "moderate"; between 0.21 and 0.40, "acceptable"; between 0 and 0.20, "poor" (Brennan and Silman, 1992).

2.5.3. Categorical data - relative reliability

The kappa value was calculated via Cohen's kappa (k) coefficient. K-values were categorized as low (≤0,40), moderate (0,41–0,60), good (0,61–0,80); and excellent reliability (0,81–1,00). The percent of agreement was also calculated (Brennan and Silman, 1992).

2.5.4. Categorical data - absolute reliability

The percent of agreement was calculated according to the following formula: A/Nx100, where A reflects the number of tests where agreement was found, and N is sample size (Brennan and Silman, 1992).

3. Results

Forty-nine people were recruited and after the exclusion of two because of a history of previous neck surgery, 47 patients remained and participated in the study. Descriptive data is presented in Table 1.

Group means and standard deviation (SD) for each test measures are presented in Table 2. Twenty of the variables (posture in sitting, right lateral-flexion, left and right FRT, PPT over right median nerve and right greater occipital nerve, upper limb tension test for right median and ulnar nerves, CCFT, and palpation over left and right SCM and trapezius, right masseter, spinous process of C1, and zygapophyseal joints of right C0–C1, left and right C1–C2, and left and right C3–C4) required logarithmic transformation for the application of a parametric test for the assessment of significance between means. Significant differences were found for ROM when left lateral flexion (<0.001) and left rotation (0.049) were performed and also when left CCF + BPPT (0.029) was assessed.

The results for the Bland Altman plots can be found in Table 3 and Appendix 1. Systematic bias was observed for cervical ROM in flexion (95%LOA = −14.54, 6.56), left lateral-flexion (95%LOA = −15.60, 12.50), right lateral-flexion (95%LOA = - 13.21, 4.79), left rotation (95%LOA = −16.13, 6.44) and right rotation (95%LOA = −13.42, 7.63); left ULTT for radial nerve (95%LOA = −10.82, 7.76) and FCC + ULLT1 (95%LOA = - 16.08, 5.96); right trapezius (95%LOA = −1.38, 0.86), right suboccipitalis muscle (95%LOA = −0.50, −0.09), left temporalis (95%LOA = −1.29, 0.87), right temporalis (95%LOA = −1.07, 0.69); C3 (95%LOA = −1.29, 0.78), right C0–C1 (95%LOA = −0.53, −0.18), left C1–C2 (95%LOA = −1.25, 0.79), right C1–C2 (95%LOA = −0.85, 1.19), left C3–C4 (95%LOA = −1.01, 0.67).

The data for numerical variables are presented in Table 4. All variables showed excellent reliability (ICC>0.81) except for: cervical ROM in extension (ICC [95%CI] = 0.489 [0.242–0.677], left lateral-flexion (ICC [95%CI] = 0.621 [0.093–0.831] and right lateral-flexion (ICC [95%CI] = 0.717 [0.479–0.846]; left ULNT for the median nerve (ICC [95%CI] = 0.787 [0.649–0.875], ULNT for left (ICC [95CI] = 0.310 [0.042–0.540] and right (ICC [95%CI] = 0.374 [0.099–0.596] radial

Table 1
Descriptive statistics. Data expressed on mean (SD) (n = 47).

Age (years)	39.93 (10.99)
Sex (Male/Female)	27/20
Height (cm)	175.53 (9.22)
Weight (kg)	73.1 (10.52)
Days from the accident	12.7 (4.1)

Table 2
Mean, SD and test-retest comparison of measures for every test at each testing set.

Outcome	Site	Test		Retest		Test vs Retest
		Mean	SD	Mean	SD	p-value
Posture, °	Standing	51.03	5.72	50.95	4.84	0.473
	Sitting	47.69	4.20	47.78	4.50	0.466[b]
ROM, °	Flexion	52.89	13.33	57.04	13.57	0.069
	Extension	31.64	7.13	33.19	7.14	0.147
	Left LF	29.51	6.19	33.85	7.00	<0.001[a]
	Right LF	32.51	6.77	35.23	6.91	0.044[b]
	Left Rot	52.87	13.65	57.7	14.46	0.049[a]
	Right Rot	54.68	14.65	57.57	14.83	0.172
	Left FRT	31.85	6.76	32.17	7.26	0.481[b]
	Right FRT	32.91	7.32	32.11	7.23	0.329[b]
PPT, N/cm^2	Left Median	17.93	4.58	18.24	4.65	0.370
	Right Median	18.53	6.12	19.47	5.95	0.194[b]
	Left Radial	22.28	6.93	22.31	6.80	0.493
	Right Radial	22.38	8.16	22.56	7.87	0.457
	Left Ulnar	18.08	5.82	18.40	5.69	0.382[b]
	Right Ulnar	18.07	5.97	18.33	6.37	0.422
	Left Sup-Orb	9.07	2.92	9.22	2.80	0.398
	Right Sup-Orb	9.49	3.03	9.43	3.28	0.461
	Right GON	10.16	3.63	10.26	3.41	0.443
	Right GON	9.91	3.83	10.11	3.65	0.500[b]
ULTT, °	Left ULTT1	142.66	13 52	144.45	12.76	0.256
	Right ULTT1	18.66	6.42	19.09	6.31	0.347[b]
	Left ULTT2	29.32	3.71	30.85	4.43	0.036
	Right ULTT2	30.66	3.33	31.04	4.21	0.313
	Left ULTT3	103.77	20.31	106.26	20.28	0.277
	Right ULTT3	18.08	5.97	18.33	6.37	0.446[b]
	Left CCF + ULTT1	137.21	13.78	142.27	11.86	0.029[a]
	Right CCF + ULTT1	141.23	11.86	143.34	11.24	0.190
Endurance, s	Flexors	11.12	4.95	11.74	6.02	0.287
	Extensors	14.21	7.21	15.02	7.51	0.500
	CCFT (stage)	24.72	1.69	24.89	1.81	0.327[b]
NRS (0–10)	Left SCM	4.45	1.8	4.51	1.89	0.464[b]
	Right SCM	5.09	1.69	5.11	1.70	0.474[b]
	Left Trap	5.30	1.78	5.36	1.79	0.443[b]
	Right Trap	5.32	1.89	5.57	1.94	0.290[b]
	Left SO	4.96	1.99	5.09	2.11	0.382
	Right SO	4.87	2.08	5.17	2.09	0.245
	Left MAS	4.51	1.46	4.60	1.48	0.390
	Right MAS	4.68	1.66	4.79	1.67	0.392[b]
	Left TEMP	4.57	1.77	4.79	1.84	0.299
	Right TEMP	4.40	1.60	4.60	1.69	0.324
	C1	5.26	1.75	5.38	1.77	0.363
	C2	5.47	1.51	5.66	1.52	0.262[b]
	C3	5.28	1.64	5.53	1.61	0.224
	Left C0–C1	4.68	2.00	4.85	2.06	0.343
	Right C0–C1	4.74	1.76	5.1	1.9	0.263[b]
	Left C1–C2	5.17	1.64	5.4	1.72	0.283[b]
	Right C1–C2	5.19	1.76	5.36	1.71	0.296[b]
	Left C2–C3	5.43	1.51	5.68	1.45	0.203
	Right C2–C3	4.97	1.67	5.2	1.69	0.251
	Left C3–C4	5.32	1.41	5.49	1.48	0.350[b]
	Right C3–C4	5.21	1.37	5.34	1.26	0.320[b]

LF (Lateral-Flexion), Rot (Rotation), FRT (Flexion-Rotation Test), Sup-Orb (Supraorbitaire), GON (Greater Occipital Nerve), PPT (Pressure Pain-Threshold), ULTT1 (Upper Limb Tension Test for Median), ULTT2 (Upper Limb Tension Test for Radial), ULTT3 (Upper Limb Tension Test for Ulnar), CCF (Cranio-Cervical Flexion), CCFT (Cranio-Cervical Flexion Test, SCM (Sterno-cleidomastoid), Trap (Trapezius), SO (Suboccipitalis), MAS (Masseter) TEMP (Temporalis).

[a] p-value<0,05 obtained after the application of t-Student.

[b] Significance was calculated with based on logarithmic transformation data.

Table 3
Results from the Bland-Altman Plots: mean difference between test-retest, 95% CI of the mean, upper and lower limits of agreement and systematic bias.

Test	Site	d (SDd)	95%CI of the mean	95% LoA	Systematic bias
ROM, °	Flexion	−4.15 (5.47)	−5.75, −2.54	−14.56, 6.56	Yes
	Extension	−1.55 (7.17)	−3.66, 0.55	−15.60, 12.50	No
	Left LF	−4.34 (4.66)	−5.71, −2.97	−13.47, 4.79	Yes
	Right LF	−2.72 (5.35)	−4.29, −1.15	−13.21, 7.77	Yes
	Left Rot	−4.85 (5.76)	−6.54, −3.16	−16.13, 6.44	Yes
	Right Rot	−2.89 (5.37)	−4.47, −1.31	−13.42, 7.63	Yes
	Left FRT	−0.32 (3.2)	−1.26, 0.62	−6.59, 5.95	No
	Right FRT	0.8 (2.51)	−0.11, 1.42	−4.2, 5.72	No
Posture	Sitting	−0.1 (2.26)	−0.76, 0.57	−4.53, 4.33	No
	Standing	0.07 (1.9)	−0.48, 0.63	−3.65, 3.79	No
PPT, N/kg^2	Left	−0,32	−0.64,	−2.51,	No
ULTT, °	Median	(1.12)	0.01	1.88	
Endurance, s	Right Median	−0.43 (1.81	−0.96, 0.10	−3.98, 3.12	No
	Left Radial	−0.03 (2.25)	−0.69, 0.63	4.44, 4.38	No
	Right Radial	−0.18 (1.83)	−0.72, 0.36	−3.77, 3.41	No
	Left Ulnar	−0.32 (1.93)	−0.89, 0.24	−3.46, 4.1	No
	Right Ulnar	−0.25 (1.74)	−0.76, 0.26	−3.66, 3.16	No
	Left SupOrb	−0.15 (0.85)	−0.40, 0.09	−1.82, 1.52	No
	Right SupOrb	0.06 (1.48)	−0.37, 0.49	−2.84, 2.96	No
	Left GON	−0.10 (0.96)	−0.39, 0.18	−1.98, 1.78	No
	Right GON	−0.20 (0.79)	−0.43, 0.03	−1.75, 1.35	No
	Left ULTT1	−1.79 (8.50)	−4.28, 0.71	−18.45, 14.87	No
	Right ULTT1	−2.13 (8.00)	−4.48, 0.22	−17.81, 13.55	No
	Left ULTT2	−1.53 (4.74)	−2.92, −0.14	−10.82, 7.76	Yes
	Right ULTT2	−0.38 (4.26)	−1.63, 0.87	−8.73, 7.97	No
	Left ULTT3	−2.49 (6.86)	−3.90, 0.28	−15.94, 10.96	No
	Right ULTT3	−1.26 (7.57)	−3.48, 0.97	−16.10, 13.58	No
	Left CCF + ULTT1	−5.06 (5.62)	−6.71, −3.41	−16.08, 5.96	Yes
	Right CCF + ULTT1	−2.11 (8.43)	−4.58, 0.37	−18.63, 14.41	No
	Flexors	−0.62 (1.86)	−1.36, 0.17	−4.27, 3.02	No
	Extensors	−0.01 (2.06)	−0.61, 0.59	−4.05, 4.03	No
	CCFT	−0.17 (0.70)	−0.38, 0.03	−1.54, 1.20	No
NRS	Left SCM	−0.64 (0.44)	−0.19, 0.06	−1.5, 0.22	No
	Right SCM	−0.02 (0.32)	−0.12, 0.08	−0.64, 0.61	No
	Left Trapezius	−0.64 (0.25)	−0.14, 0.01	−1.13, −0.15	No
	Right Trapezius	−0.26 (0.57)	−0.42, −0.09	−1.38, 0.86	Yes
	Left SO	−0.13 (0.58)	−0.30, 0.04	−1.27, 1.01	No
	Right SO	−0.30 (0.69)	−0.50, −0,09	−1.65, 1.05	Yes
	Left MAS	−0.86 (0.46)	−0.22, 0.04	−1.76, 0.04	No
	Right MAS	−0.11 (0.52)	−0.26, 0.04	−1.13, 0.91	No
	Left TEMP	−0.21 (0.55)	−0.37, −0.05	−1.29, 0.87	Yes
	Right TEMP	−0.19 (0.45)	−0.32, −0.06	−1.07, 0.69	Yes
	C1	−0.13 (0.61)	−0.31, 0.05	−1.33, 1.06	No
	C2	−0.19 (0.50)	−0.41, 0.03	−1.17, 0.79	No
	C3	−0.26 (0.53)	−0.41, −0.09	−1.29, 0.78	Yes
	Left C0–C1	−0.17 (0.56)	−0.34, 0.11	−1.27, 0.93	No
	Right C0–C1	−0.36 (0.61)	−0.53, −0.18	−1.56, 0.84	Yes
	Left C1–C2	−0.23 (0.52)	−0.39, −0.09	−1.25, 0.79	Yes
	Right C1–C2	0.17 (0.52)	−0.31, −0.03	−0.85, 1.19	Yes
	LeftC2-C3	−0.26 (0.67)	−0.51, 0.01	−1.53, 1.05	No
	Right C2–C3	0.23 (0.63)	−0.51, 0.04	−1.00, 1.47	No
	Left C3–C4	−0.17 (0.43)	−0.30, −0.04	−1.01, 0.67	Yes
	Right C3–C4	−0.13 (0.53)	−0.29, 0.03	−1.17, 0.91	No

d (mean differences), SDd (Standard Deviation of d), LoA (Limits of Agreement) (LF (Lateral-Flexion), Rot (Rotation), FRT (Flexion-Rotation Test), Sup-Orb (Supraorbitaire), GON (Greater Occipital Nerve), PPT (Pressure Pain-Threshold), ULTT1 (Upper Limb Tension Test for Median), ULTT2 (Upper Limb Tension Test for Radial), ULTT3 (Upper Limb Tension Test for Ulnar), CCF (Cranio-Cervical Flexion), CCFT (Cranio-Cervical Flexion Test, SCM (Sterno-cleiomastoid), SO (Suboccipitalis), MAS (Masseter) TEMP (Temporalis).

nerves, and right ULNT + CCF ICC [95%CI] = 0.734 [0.568–0.843]. Results for the SEMs and MDC_{90} can be found in Table 4.

For the data which did not show a normal distribution, we present reliability data both before and after logarithmic transformation since logarithmic data cannot be extrapolated to clinical practice.

Intra-rater reliability for categorical data can be found in Table 5. Active cervical extension in 4-point kneeling (k = 0.443) and active cervical rotation in 4-point kneeling (k = 0.575) showed moderate reliability; forward head posture in sitting showed good reliability (k = 0.612), while forward head posture in standing showed excellent reliability (k = 0.844).

4. Discussion

This study is the first to assess the test-retest reliability of a battery of tests evaluating the neuromusculoskeletal system in people with acute WAD. This is of significance since clinical tests that show acceptable reliability facilitate valid clinical decision making (Brennan and Silman, 1992). The results of this study show that the vast majority of the tests that were evaluated are reliable when test-retest measurements are taken on people with acute WAD.

We identified only two other studies examining test-retest reliability of measures in people with acute WAD: the first evaluated self-perceived change and self-perceived recovery (Ngo et al., 2010) and the other examined the morphology and quality of the deep extensor muscles via

Table 4
Reliability of numerical data.

Test	Site	ICC (95%CI)	SEM		MDC90	
			Measure	%	Measure	%
Posture, °	Standing	0.937(0.889–0.964)	1.33	2.83	3.1	6.6
	Sitting[a]	0.859(0.761–0.919)	1.58	3.38	3.7	7.88
	SittingLog10	0.867(0.774–0.924)	1.03	2.20	1.08	2.30
ROM, °	Flexion	0.877(0.602–0.949)	4.717	10.03	11.01	23.41
	Extension	0.489(0.242–0.677)	5.1	10.84	11.89	25.30
	Left Lateral Flexion	0.621(0.093–0.831)	4.07	8.65	9.49	20.20
	Right Lateral Flexion[a]	0.647(0.386–0.801)	4.06	8.64	9.48	20.17
	RightLFLog10	0.717(0.479–0.846)	0.06	0.13	0.14	0.29
	Left Rotation	0.866(0.509–0.947)	5.15	10.95	12.01	25.55
	Right Rotation	0.917(0.809–0.960)	4.25	9.04	9.91	21.1
	Left FRT[a]	0.896(0.822–0.902)	2.26	4.8	5.27	11.22
	Left FRTLog10	0.918(0.857–0.953)	0.06	0.12	0.14	0.29
	Right FRT[a]	0.936(0.883–0.964)	1.83	3.91	4.29	9.13
	Right FRTLog10	0.728(0.561–0.839)	0.06	0.14	0.15	0.32
PPT, N/kg^2 ULTT Endurance, s	Left Median	0.969(0.944–0.983)	1.16	2.48	2.71	5.79
	Right Median[a]	0.929(0.856–0.963)	1.61	3.42	3.75	7.98
	Right MedianLog10	0.913(0.820–0.955)	0.04	0.09	0.103	0.22
	Left Radial	0.947(0.907–0.970)	1.58	3.36	3.69	7.84
	Right Radial	0.974(0.954–9.986)	1.29	2.75	3.01	6.41
	Left Ulnar[a]	0.944(0.901–0.968)	1.36	2.88	3.16	6.73
	Left UlnarLog10	0.941(0.896–0.966)	0.03	0.08	0.09	0.18
	Right Ulnar	0.960(0.930–0.978)	1.324	2.82	3.09	6.57
	Left SupOrb	0.956(0.922–0.975)	0.60	1.27	1.40	2.98
	Right SupOrb	0.893(0.815–0.939)	1.03	2.20	2.41	5.13
	Left GON	0.963(0.935–0.979)	0.68	1.44	1.58	3.36
	Right GON[a]	0.977(0.958–0.987)	0.57	1.20	1.32	2.81
	Right GON Log10	0.973(0.950–0.985)	0.03	0.06	0.07	0.15
	Left ULTT1	0.787(0.649–0.875)	6.06	12.90	14.15	30.11
	Right ULTT1[a]	0.958(0.926–0.976)	1.30	3.04	3.04	6.47
	Right ULTT1Log10	0.946(0.904–0.970)	0.03	0.07	0.08	0.17
	Left ULTT2	0.310(0.042–0.540)	3.38	7.19	7.89	16.78
	Right ULTT2	0.374(0.099–0.596)	2.98	6.35	6.96	14.81
	Left ULTTU3	0.937(0.882–0.966)	5.10	10.84	11.89	25.30
	Right ULTT3[a]	0.960(0.930–0.978)	1.23	2.62	2.88	6.13
	Right ULTT3Log10	0.977(0.958–0.987)	0.02	0.05	0.06	0.12
	Left FCC + BPPT	0.904(0.834–0.945)	3.96	8.43	9.25	19.69
	Right FCC + BPPT	0.734(0.568–0.843)	5.93	12.63	13.85	29.47
	Flexors	0.933(0.878–0.963)	1.42	3.02	3.31	7.05
	Extensors	0.962(0.933–0.979)	1.43	3.05	3.34	7.12
	CCFT[a]	0.920(0.860–0.954)	0.49	1.05	1.15	2.46
	CCFTLog10	0.929(0.875–0.960)	0.01	0.02	0.02	0.04
NRS	Left SCM[a]	0.972(0.950–0.984)	0.3	0.67	0.74	1.57
	Left SCMLog10	0.984(0.971–0.991)	0.03	0.05	0.06	0.12
	Right SCM[a]	0.981(0.967–0.990)	0.23	0.50	0.55	1.16
	Right SCMLog10	0.985(0.974–0.992)	0.02	0.04	0.04	0.09
	Left Trap[a]	0.990(0.982–0.994)	0.179	0.417	0.38	0.89
	Left TrapLog10	0.993(0.987–0.996)	0.015	0.03	0.035	0.074
	Right Trap[a]	0.948(0.893–0.973)	0.44	0.93	1.02	2.17
	Right TrapLog10	0.953(0.908–0.975)	0.04	0.09	0.10	0.21
	Left SO	0.960(0.929–0.977)	0.41	0.88	0.96	2.04
	Right SO	0.937(0.873–0.967)	0.52	1.11	1.22	2.60
	Left MAS	0.951(0.914–0.972)	0.33	0.69	0.76	1.62
	Right MAS[a]	0.950(0.912–0.972)	0.37	0.79	0.87	1.84
	Right MASLOG10	0.956(0.922–0.975)	0.04	0.07	0.08	0.18
	Left TEMP[a]	0.948(0.900–0.972)	0.40	0.86	0.94	2.00
	Left TEMPLOG10	0.952(0.906–977)	0.04	0.08	0.09	0.20
	Right TEMP[a]	0.957(0.913–0.978)	0.34	0.72	0.80	1.70
	Right TEMPLOG10	0.972(0.943–0.986)	0.03	0.07	0.07	0.16
	C1	0.938(0.892–0.965)	0.44	0.93	1.02	2.18
	C2[a]	0.929(0.867–0.961)	0.40	0.86	0.94	1.99
	C2LOG10	0.940(0.886–0.968)	0.03	0.07	0.08	0.17
	C3	0.936(0.863–0.968)	0.41	0.88	0.96	2.04
	Left C0–C1	0.959(0.925–0.977)	0.41	0.87	0.96	2.04
	Right C0–C1[a]	0.929(0.814–0.967)	0.48	1.03	1.14	2.42
	Right C0–C1LOG10	0.965(0.911–0.984)	0.04	0.09	0.10	0.21
	Left C1–C2[a]	0.944(0.884–0.971)	0.40	0.85	0.93	1.97
	Left C1–C2LOG10	0.945(0.895–0.970)	0.04	0.08	0.09	0.19
	Right C1–C2[a]	0.945(0.897–0.970)	0.41	0.87	0.95	2.02
	Right C1–C2LOG10	0.958(0.921–0.977)	0.04	0.08	0.09	0.18
	Left C2–C3	0.885(0.789–0.937)	0.50	1.06	1.17	2.49
	Right C2–C3	0.922(0.854–0.957)	0.47	1.00	1.09	2.33
	Left C3–C4[a]	0.950(0.903–0.973)	0.33	0.69	0.76	1.62
	Left C3–C4LOG10	0.977(0.954–0.988)	0.02	0.05	0.06	0.12

(*continued on next page*)

Table 4 (*continued*)

Test	Site	ICC (95%CI)	SEM		MDC90	
			Measure	%	Measure	%
	Right C3–C4[a]	0.914(0.850–0.951)	0.38	0.82	0.90	1.90
	Right C3–C4LOG10	0.922(0.863–0.956)	0.04	0.09	0.10	0.22

SEM: Standard Error of Measurement; MDC90: Minimal Detectable Change.
LF (Lateral-Flexion), Rot (Rotation), FRT (Flexion-Rotation Test), Sup-Orb (Supraorbitaire), GON (Greater Occipital Nerve), PPT (Pressure Pain-Threshold), ULTT1 (Upper Limb Tension Test for Median), ULTT2 (Upper Limb Tension Test for Radial), ULTT3 (Upper Limb Tension Test for Ulnar), CCF (Cranio-Cervical Flexion), BPPT (Braquial PLes Provocation Test), CCFT (Cranio-Cervical Flexion Test, SCM (Sternocleiomastoid), Trap (Trapezius), SO (Suboccipitalis), MAS (Masseter) TEMP (Temporalis).
In those data who did not follow normal distribution, we included both data, non-transformed and logarithmically transformed, to improve data presentation.
[a] Reliability outcomes are presented for non-transformed data distributions in order to facilitate interpretation, but consideration of these data must be done cautiously because the assumption of the normality of the distribution was violated.

Table 5
Reliability of categorical data.

TEST	POSITIVE TEST	Cohen's Kappa	% AGREEMENT
4K EXTENSION	26/47	0.443	72.34%
4K UPPER ROTATION	25/47	0.575	78.72%
FHP Sitting	25/47	0.612	90.85%
FHP Standing	35/47	0.844	93.61%

Positive test: proportion of patients with a positive test.
4K Extension: Active cervical extension in 4-point kneeling; 4K Upper Rotation: Active upper cervical rotation in 4-point kneeling. Both were considered positive if patients were not able to perform the test.
FHP: Forward Head Posture. Forward Head Posture was considered to be present when Cranio-Vertebral Angle was smaller than $<48°$.

ultrasound (Valera-Calero et al., 2022). In contrast, intra-rater reliability of physical testing has been investigated in people with chronic WAD. For instance, a clinical test to assess isometric cervical strength showed excellent intra-rater reliability (ICC = 0.91) when tested in people with chronic WAD (Habberfield et al., 2022). In addition, although not performed in people with WAD, neck flexor and extensor endurance tests showed excellent intra-rater reliability, as also shown in the current study, in patients with mechanical neck pain (Edmondston et al., 2008), and inter-rater reliability was established for the same neck flexor endurance test in people with chronic WAD (Kumbhare et al., 2005). In the current study, we chose to limit our endurance assessment tests to 30 s. Previous research has shown that patients with chronic neck pain are likely to be able to hold the position during each test for more than 30 s (Edmondston et al., 2008; Ris et al., 2017). However, no previous studies were performed in people with acute WAD. Therefore, due to the number of tests to be performed, and with the objective of not aggravating their pain, we chose to limit our test to 30 s. Our decision was supported by the fact that, for flexion endurance, only one participant achieved 30 s (retest), and for extension, only 3 participants (1 the test and 2 the retest) achieved 30 s.

Good and excellent reliability has also been demonstrated for the assessment of cervical ROM in patients with neck pain, including those with neck pain attributed to a traumatic event. However, only inter-rater reliability was evaluated (Cleland et al., 2006). Nonetheless, another study assessed intra- and inter-rater reliability for a battery of tests in subjects with chronic neck pain, with good and excellent reliability established for the CCFT and ROM (Jørgensen et al., 2014). Previous studies have assessed the use of smartphones to assess cervical ROM, as has been done in the current study. A recent systematic review concluded that smartphone apps are a reliable and valid method for measuring neck ROM in people with and without neck pain (Elgueta-Cancino et al., 2002). Nonetheless, two studies (Ghorbani et al., 2020; Schmid et al., 2009) found good or excellent reliability for all movements except for rotation to both sides, hypothesizing that the poorer reliability for rotation could be due to the interaction of magnetic fields with the magnetometer of the Compass Application. This should be considered when interpreting the results of the current study. However, it is worth adding that this interaction of magnetic fields may have been avoided in our study since we performed the ROM assessment and then followed with the remaining tests, with a 10 min rest between sets, and re-calibration of the Compass Application which may have avoided this increased measurement error.

Previous studies have assessed the reliability of ULNT for the median, radial and ulnar nerve based on the reproducibility of negative or positive findings during the assessment (Riley et al., 2020). In the current study we quantified the range of elbow movement during each test and show that ULNT for the median and ulnar nerves can be assessed in a reliable way however, assessment of radial nerve should be interpreted with caution. Although not in patients with WAD, most of the previous studies examining the reliability of ULNT have assessed the median nerve and have shown good/excellent intertester reliability (Vanti et al., 2010; Jull et al., 1997). In addition, and in line with other research in different patient populations (Szikszay et al., 2018; Fernández-de-Las-Peñas et al., 2011), PPT over nerves is also reliable when assessed in patients with acute WAD.

Intra-rater reliability was excellent when using the FHPapp to assess head posture, and these results are in line with those obtained previously in patients with and without neck pain (Gallego-Izquierdo et al., 2020). In the current study we assessed FHP in both sitting and standing unlike the previous study which was limited to assessment in standing only. The level of agreement in relation to the identification of the presence of FHP, was 90.85% and 93.61% for sitting and standing positions, respectively.

Test-retest reliability was lower for the tests of active cervical extension and cervical rotation tests performed in a 4-point kneeling position (k = 0.443 and k = 0.575, respectively). In contrast with our results, a previous study evaluating these tests found, found excellent intratester reliability (k = 0.86 and k = 0.80) in patients with and without neck pain (Segarra et al., 2015). Possible explanations for this include the discomfort in performing these tests given that the patients enrolled in the current study had acute symptoms.

When assessing the reliability of manual palpation, some studies have focused on the ability of the physical therapist to detect a hypomobile segment or the affected structure when evaluating the articular system (Luedtke et al., 2018; Rathbone et al., 2017), or to detect the presence of trigger points when evaluating the muscle system, but high levels of reliability have not been established (Jull and Hall, 2018). Previous work has shown that manual spinal examination (articular tests incorporating an assessment of pain provocation combined with the quality of movement) is reliable and valid in the diagnosis of cervical facet joint pain (Schneider et al., 2013; Schneider et al., 2014). We chose to examine PAIVMS in quantitative way by assessing pain intensity with numerical data. It should be noted however, that the results may have differed if we had of considered categorical versus numerical data. Overall, our results for test-retest reliability were excellent, with ICC ranging from 0.937 to 0.993 for muscle palpation and from 0.885 to 0.977 for PAIVMs. Although some limitations may arise from evaluating

pain perception during palpation, evaluation of pain sensitivity via palpation may prove more reliable in patients with acute symptoms (Rathbone et al., 2017).

4.1. Limitations

This study was limited to the evaluation of intra-rater reliability. Due to the nature of the study (performed on participants attending physical therapy treatment for their post-whiplash symptoms), and the number of tests performed, inter-rater evaluation could not be assessed. The physical therapist who performed testing had expertise in Orthopaedic Manual Therapy and therefore the results may not transfer to novice physical therapists. A further limitation was method of muscle palpation since the extent of pressure was not objectively measured. Moreover, we did not evaluate the presence or absence of trigger points, but rather, the pain intensity produced by palpation over the muscle belly. Nevertheless, our results show that the test-retest reliability of muscle palpation is high. In addition, during the assessment of ULNT, we did not assess structural differentiation and thus we cannot determine whether the onset of symptoms which was assessed was attributed to neural or the musculoskeletal factors.

5. Conclusion

The majority of physical tests achieved good or excellent test-retest intra-rater reliability when tested in patients with acute WAD. However, some of them were found to present systematic bias, such as cervical ROM in flexion, left and right lateral-flexion, left and right rotation; left ULTT for radial nerve; right trapezius, right suboccipitalis muscle, left temporalis, right temporalis; C3, both sides of C1–C2, left C3–C4. Therefore, clinicians should take caution when interpreting these last tests. Further testing is warranted to evaluate inter-rater reliability of these tests in patients with acute WAD.

Funding

This research received no external funding.

Author's contributions

EAL performed the data acquisition and data management. CBU and CRB were involved in methodological issues and statistical analysis. DF, CRB and EAL contributed to the development of the manuscript. CBU and DF contributed to supervision and conceptualization. All authors have read and agreed the final version of the manuscript.

Declaration of competing interest

Authors declare no conflict of interest.

Acknowledgements

We would like to thank all patients involved in this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.msksp.2023.102738.

References

Alalawi, A., Luque-Suarez, A., Fernandez-Sanchez, M., Tejada-Villalba, R., Navarro-Martin, R., Devecchi, V., Gallina, A., Falla, D., 2022. Perceived pain and disability but not fear of movement are associated with altered cervical kinematics in people with acute neck pain following a whiplash injury. Musculoskelet. Sci. Pract. 62, 102633. Aug 4.

Atkinson, G., Nevill, A.M., 1988. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med. 26 (4), 217–238. Oct.

Baydal-Bertomeu, J.M., Page, A.F., Belda-Lois, J.M., Garrido-Jaén, D., Prat, J.M., 2011. Neck motion patterns in whiplash-associated disorders: quantifying variability and spontaneity of movement. Clin. Biomech. 26 (1), 29–34. Ja.

Bland, J.M., Altman, D.G., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1 (8476), 307–310. Feb 8.

Bland, J.M., Altman, D.G., 1999. Measuring agreement in method comparison studies. Stat. Methods Med. Res. 8 (2), 135–160. Jun.

Brennan, P., Silman, A., 1992. Statistical methods for assessing observer variability in clinical measures. BMJ 304 (6840), 1491–1494. Jun 6.

Carroll, L.J., Holm, L.W., Hogg-Johnson, S., Côté, P., Cassidy, J.D., Haldeman, S., Nordin, M., Hurwitz, E.L., Carragee, E.J., van der Velde, G., Peloso, P.M., Guzman, J., 2008. Bone and joint decade 2000-2010 Task Force on neck pain and its associated disorders. Course and prognostic factors for neck pain in whiplash-associated disorders (WAD): results of the bone and joint decade 2000-2010 Task Force on neck pain and its associated disorders. Spine (Phila Pa 1976) 33 (4 Suppl. l), S83–S92. Feb 15.

Chen, K., Andersen, T., Carroll, L., Connelly, L., Côté, P., Curatolo, M., Elliott, J., Grant, G., Jull, G., Kasch, H., MacDermid, J., Malmström, E.M., Maujean, A., McLean, S.A., Nielsen, M., Rebbeck, T., Söderlund, A., Sterling, J., Treleaven, J., Walton, D.M., Westergren, H., Sterling, M., 2019. Recommendations for core outcome domain set for whiplash-associated disorders (CATWAD). Clin. J. Pain 35 (9), 727–736. Sep.

Cleland, J.A., Childs, J.D., Fritz, J.M., Whitman, J.M., 2006. Interrater reliability of the history and physical examination in patients with mechanical neck pain. Arch. Phys. Med. Rehabil. 87 (10), 1388–1395. Oct.

Curatolo, M., Bogduk, N., Ivancic, P.C., McLean, S.A., Siegmund, G.P., Winkelstein, B.A., 2011. The role of tissue damage in whiplash-associated disorders: discussion paper 1. Spine (Phila Pa 1976) 36 (25 Suppl. l), S309–S315. Dec 1.

de Vet, H.C., Terwee, C.B., Knol, D.L., Bouter, L.M., 2006. When to use agreement versus reliability measures. J. Clin. Epidemiol. 59 (10), 1033–1039. Oct.

Edmondston, S.J., Wallumrød, M.E., Macléid, F., Kvamme, L.S., Joebges, S., Brabham, G. C., 2008. Reliability of isometric muscle endurance tests in subjects with postural neck pain. J. Manip. Physiol. Ther. 31 (5), 348–354. Jun.

Elgueta-Cancino, E., Rice, K., Abichandani, D., Falla, D., 2002. Measurement properties of smartphone applications for the measurement of neck range of motion: a systematic review and meta-analyses. BMC Muscoskel. Disord. 23 (1), 138. Feb 10.

Ettlin, T., Schuster, C., Stoffel, R., Brüderlin, A., Kischka, U., 2008. A distinct pattern of myofascial findings in patients after whiplash injury. Arch. Phys. Med. Rehabil. 89 (7), 1290–1293. Jul.

Fernández-de-las-Peñas, C., Alonso-Blanco, C., Cuadrado, M.L., Pareja, J.A., 2006. Forward head posture and neck mobility in chronic tension-type headache: a blinded, controlled study. Cephalalgia 26 (3), 314–319. Mar.

Fernández-de-Las-Peñas, C., Ortega-Santiago, R., Cuadrado, M.L., López-de-Silanes, C., Pareja, J.A., 2011. Bilateral widespread mechanical pain hypersensitivity as sign of central sensitization in patients with cluster headache. Headache 51 (3), 384–391. Mar.

Fernández-Pérez, A.M., Villaverde-Gutiérrez, C., Mora-Sánchez, A., Alonso-Blanco, C., Sterling, M., Fernández-de-Las-Peñas, C., 2012. Muscle trigger points, pressure pain threshold, and cervical range of motion in patients with high level of disability related to acute whiplash injury. J. Orthop. Sports Phys. Ther. 42 (7), 634–641. Jul.

Fundaun, J., Kolski, M., Baskozos, G., Dilley, A., Sterling, M., Schmid, A., 2021. Nerve pathology and neuropathic pain after whiplash injury. Pain. https://doi.org/10.1097/j.pain.0000000000002509. October 12.

Gallego-Izquierdo, T., Arroba-Díaz, E., García-Ascoz, G., Val-Cano, M.D.A., Pecos-Martin, D., Cano-de-la-Cuerda, R., 2020. Psychometric proprieties of a mobile application to measure the craniovertebral Angle a validation and reliability study. Int. J. Environ. Res. Publ. Health 17 (18), 6521. Sep 8.

Ghorbani, F., Kamyab, M., Azadinia, F., 2020. 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. 19 (1), 38–48. Mar.

Habberfield, J., Schneider, G., Schneider, K., et al., 2022. A clinical test to assess isometric cervical strength in chronic whiplash associated disorder (WAD): a reliability study. BMC Muscoskel. Disord. 23, 736.

Hall, T., Robinson, K., 2004. The flexion-rotation test and active cervical mobility–a comparative measurement study in cervicogenic headache. Man. Ther. 9 (4), 197–202. Nov.

Harvill, L.M., 1991. An NCME instructional module on. Educ. Meas. 10, 33–41.

Headache classification committee of the international headache society (IHS) the international classification of headache disorders. In: third ed. Cephalalgia, 2018, pp. 1–211 38, 1.

Hengeveld, E., Banks, K. (Eds.), 2013, Maitland's Vertebral Manipulation: Management of Neuromusculoskeletal Disorders, eighth ed., vol. 1. Elsevier, Amsterdam.

Huhn, K., Gilliland, S.J., Black, L.L., Wainwright, S.F., Christensen, N., 2019. Clinical reasoning in physical therapy: a concept analysis. Phys. Ther. 99 (4), 440–456. Apr 1.

Jonsson, A., Rasmussen-Barr, E., 2018. Intra- and inter-rater reliability of movement and palpation tests in patients with neck pain: a systematic review. Physiother. Theory Pract. 34 (3), 165–180. Mar.

Jørgensen, R., Ris, I., Falla, D., Juul-Kristensen, B., 2014. Reliability, construct and discriminative validity of clinical testing in subjects with and without chronic neck pain. BMC Muscoskel. Disord. 15, 408. Dec 4.

Jull, G.A., 2011. Considerations in the physical rehabilitation of patients with whiplash-associated disorders. Spine (Phila Pa 1976) 36 (25 Suppl. l), S286–S291. Dec 1.

Jull, G., Hall, T., 2018. Cervical musculoskeletal dysfunction in headache: how should it be defined? Musculoskelet. Sci. Pract. 38, 148–150. Dec.
Jull, G., Zito, G., Trott, P., Potter, H., Shirley, D., 1997. Inter-examiner reliability to detect painful upper cervical joint dysfunction. Aust. J. Physiother. 43 (2), 125–129.
Jull, G.A., O'Leary, S.P., Falla, D.L., 2008. Clinical assessment of the deep cervical flexor muscles: the craniocervical flexion test. J. Manip. Physiol. Ther. 31 (7), 525–533. Sep.
Jull, G.A., Söderlund, A., Stemper, B.D., Kenardy, J., Gross, A.R., Côté, P., Treleaven, J., Bogduk, N., Sterling, M., Curatolo, M., 2011. Toward optimal early management after whiplash injury to lessen the rate of transition to chronicity: discussion paper 5. Spine (Phila Pa 1976) 36 (25 Suppl. l), S335–S342. Dec 1.
Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B.J., Hróbjartsson, A., Roberts, C., Shoukri, M., Streiner, D.L., 2011. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. J. Clin. Epidemiol. 64 (1), 96–106. Jan.
Kumbhare, D.A., Balsor, B., Parkinson, W.L., Harding Bsckin, P., Bedard, M., Papaioannou, A., Adachi, J.D., 2005. Measurement of cervical flexor endurance following whiplash. Disabil. Rehabil. 27 (14), 801–807. Jul 22.
Lexell, J.E., Downham, D.Y., 2005. How to assess the reliability of measurements in rehabilitation. Am. J. Phys. Med. Rehabil. 84 (9), 719–723. Sep.
Luedtke, K., Starke, W., May, A., 2018. Musculoskeletal dysfunction in migraine patients. Cephalalgia 38 (5), 865–875.
Nee, R.J., Jull, G.A., Vicenzino, B., Coppieters, M.W., 2012. The validity of upper-limb neurodynamic tests for detecting peripheral neuropathic pain. J. Orthop. Sports Phys. Ther. 42 (5), 413–424. May.
Ngo, T., Stupar, M., Côté, P., Boyle, E., Shearer, H., 2010. A study of the test-retest reliability of the self-perceived general recovery and self-perceived change in neck pain questions in patients with recent whiplash-associated disorders. Eur. Spine J. 19 (6), 957–962. Jun.
Pearson, I., Reichert, A., De Serres, S.J., Dumas, J.P., Côté, J.N., 2009. 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. 39 (3), 179–187. Mar.
Rathbone, A.T.L., Grosman-Rimon, L., Kumbhare, D.A., 2017. Interrater agreement of manual palpation for identification of myofascial trigger points: a systematic review and meta-analysis. Clin. J. Pain 33 (8), 715–729. Aug.
Riley, S.P., Grimes, J.K., Calandra, K., Foster, K., Peet, M., Walsh, M.T., 2020. Agreement and reliability of median neurodynamic test 1 and resting scapular position. J. Chiropr. Med. 19 (4), 203–212. Dec.
Ris, I., Juul-Kristensen, B., Boyle, E., Kongsted, A., Manniche, C., Søgaard, K., 2017. Chronic neck pain patients with traumatic or non-traumatic onset: differences in characteristics. A cross-sectional study. Scand. J. Pain 14, 1–8. Jan.
Schmid, A.B., Brunner, F., Luomajoki, H., Held, U., Bachmann, L.M., Künzer, S., Coppieters, M.W., 2009. Reliability of clinical tests to evaluate nerve function and mechanosensitivity of the upper limb peripheral nervous system. BMC Muscoskel. Disord. 10, 11. Jan 21.
Schneider, G.M., Jull, G., Thomas, K., Smith, A., Emery, C., Faris, P., Schneider, K., Salo, P., 2013. Intrarater and interrater reliability of select clinical tests in patients referred for diagnostic facet joint blocks in the cervical spine. Arch. Phys. Med. Rehabil. 94 (8), 1628–1634. Aug.
Schneider, G.M., Jull, G., Thomas, K., Smith, A., Emery, C., Faris, P., Cook, C., Frizzell, B., Salo, P., 2014. Derivation of a clinical decision guide in the diagnosis of a cervical facet joint pain. Arch. Phys. Med. Rehabil. 95 (9), 1695–1701. Sep.
Segarra, V., Dueñas, L., Torres, R., Falla, D., Jull, G., Lluch, E., 2015. Inter-and intra-tester reliability of a battery of cervical movement control dysfunction tests. Man. Ther. 20 (4), 570–579. Aug.
Shaghayegh Fard, B., Ahmadi, A., Maroufi, N., Sarrafzadeh, J., 2016. Evaluation of forward head posture in sitting and standing positions. Eur. Spine J. 25 (11), 3577–3582. Nov.
Shoukri, M.M., Asyali, M.H., Donner, A., 2004. Sample size requirements for the design of reliability study: review and new results. Stat. Methods Med. Res. 13, 251–271.
Smidt, N., van der Windt, D.A., Assendelft, W.J., Mourits, A.J., Devillé, W.L., de Winter, A.F., Bouter, L.M., 2002. Interobserver reproducibility of the assessment of severity of complaints, grip strength, and pressure pain threshold in patients with lateral epicondylitis. Arch. Phys. Med. Rehabil. 83 (8), 1145–1150. Aug.
Spitzer, W.O., Skovron, M.L., Salmi, L.R., Cassidy, J.D., Duranceau, J., Suissa, S., Zeiss, E., 1995. Scientific monograph of the Quebec Task Force on whiplash-associated disorders: redefining "whiplash" and its management. Apr 15 Spine (Phila Pa 1976) 20 (8 Suppl. l), 1S–73S. Erratum in: Spine 1995 Nov 1;20(21):2372.
Sterling, M., 2014. Physiotherapy management of whiplash-associated disorders (WAD). J. Physiother. 60 (1), 5–12. Mar.
Sterling, M., McLean, S.A., Sullivan, M.J., Elliott, J.M., Buitenhuis, J., Kamper, S.J., 2011. Potential processes involved in the initiation and maintenance of whiplash-associated disorders: discussion paper 3. Spine (Phila Pa 1976) 36 (25 Suppl. l), S322–S329. Dec 1.
Szikszay, T.M., Luedtke, K., Harry von, P., 2018. Increased mechanosensivity of the greater occipital nerve in subjects with side-dominant head and neck pain - a diagnostic case-control study. J. Man. Manip. Ther. 26 (4), 237–248. Aug.
Tousignant-Laflamme, Y., Boutin, N., Dion, A.M., Vallée, C.A., 2013. Reliability and criterion validity of two applications of the iPhone™ to measure cervical range of motion in healthy participants. J. NeuroEng. Rehabil. 10 (1), 69. Jul 5.
Valera-Calero, J.A., Fernández-de-Las-Peñas, C., Cleland, J.A., Varol, U., Ortega-Santiago, R., Arias-Buría, J.L., 2022. Ultrasound assessment of deep cervical extensors morphology and quality in populations with whiplash associated disorders: an intra- and inter-examiner reliability study. Musculoskelet. Sci. Pract. 59, 102538. Jun.
Vanti, C., Conteddu, L., Guccione, A., Morsillo, F., Parazza, S., Viti, C., Pillastrini, P., 2010. The Upper Limb Neurodynamic Test 1: intra- and intertester reliability and the effect of several repetitions on pain and resistance. J. Manip. Physiol. Ther. 33 (4), 292–299. May.
Zito, G., Jull, G., Story, I., 2006. Clinical tests of musculoskeletal dysfunction in the diagnosis of cervicogenic headache. Man. Ther. 11 (2), 118–129. May.