Musculoskeletal Science and Practice 40 (2019) 21–33

Contents lists available at ScienceDirect

Musculoskeletal Science and Practice

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

Systematic Review

Diagnostic accuracy of upper limb neurodynamic tests for the assessment of peripheral neuropathic pain: A systematic review

Konstantinos Koulidis[a], Yannis Veremis[a], Christina Anderson[a], Nicola R. Heneghan[b,*]

[a] School of Sport, Exercise & Rehabilitation Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK
[b] Centre of Precision Rehabilitation for Spinal Pain, School of Sport, Exercise & Rehabilitation Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

ARTICLE INFO

Keywords:
Entrapment neuropathies
Carpal tunnel syndrome
Cervical radiculopathy
Upper limb neurodynamics
Validity

ABSTRACT

Background: Upper limb neurodynamic tests (ULNTs) are used to identify a neuropathic pain component in patients' presenting with arm and/or neck pain. Clinical tests with established diagnostic accuracy are required to not only to inform clinical management but also minimise costs associated with expensive medical investigations.
Objective: To evaluate the role of ULNTs in assessment of peripheral neuropathic pain and to inform their value in clinical practice when assessing patients with arm and/or neck symptoms.
Design: Systematic review was undertaken according to published guidelines, and reported in line with PRISMA-DTA.
Method: Key databases were searched up to 21/11/2017. Inclusion criteria: Patient population experiencing arm and/or neck symptoms with suspected peripheral neuropathic involvement, studies that compared ULNT to a reference standard, any study design using primary diagnostic accuracy data. Two reviewers independently assessed risk of bias (ROB) using QUADAS-2. The overall quality of evidence was evaluated using GRADE.
Results: Of eight included studies (n = 579), four were assessed as low ROB, although all had concerns regarding applicability. For carpal tunnel syndrome, ULNT1 sensitivity values ranged 0.4–0.93, specificity 0.13–0.93, positive likelihood ratio 0.86–3.67 and negative likelihood ratio 0.5–1.9. For cervical radiculopathy ULNT1 and the combined use of four ULNTs had sensitivity of 0.97 (95%CI 0.85–1.00) whereas the ULNT3 was the most specific (0.87, 95%CI 0.62–0.98). Positive likelihood ratio ranged 0.58–5.68 and negative likelihood ratio 0.12–1.62.
Conclusion: Based on the available evidence ULNTs cannot be utilised as a stand-alone test for the diagnosis of CTS. Limited evidence suggests that ULNTs may be clinically relevant for the diagnosis of CR, but only as a "ruling out" strategy. However, the overall quality of the body of evidence after applying the GRADE approach was low to very low across studies. Further higher quality research is needed to establish firm conclusions.

1. Introduction

Peripheral neuropathic pain (PNP) is a term used to describe pain that results from a lesion or disease affecting the somatosensory nervous system (Finnerup et al., 2016). PNP can arise when a peripheral nerve trunk or a nerve root has been subject to injury, compression, inflammation or ischemia resulting in reduced physical capabilities of the nervous system (Nee and Butler, 2006). Symptoms and signs in neuropathies can be classified as positive (gain of function) or negative (loss of function). Positive symptoms include pain, paresthesia, dysesthesia, hyperalgesia and allodynia and indicate abnormal excitability in the nervous system, whereas negative symptoms, such as hypoesthesia or anesthesia and weakness reflect reduced impulse conduction (Woolf, 2004).

The most common conditions affecting the peripheral nervous system are entrapment neuropathies (EN), with carpal tunnel syndrome (CTS), cubital tunnel syndrome and cervical radiculopathy (CR) being examples which contribute considerably to the socioeconomic burden of occupational related musculoskeletal complaints and the associated costs. Individually EN have been associated with severe pain, depression and functional limitations (Fernández-de-las-Peñas et al., 2015). CTS is often observed in activities involving repetitive manual tasks, forceful wrist movements or with direct pressure on the wrist, estimated to affect 2–15% of workers (Atroshi et al., 1999) and costing more than

* Corresponding author.
E-mail addresses: kostaskoul23@gmail.com (K. Koulidis), veremis.y@gmail.com (Y. Veremis), c.b.anderson@bham.ac.uk (C. Anderson), n.heneghan@bham.ac.uk (N.R. Heneghan).

https://doi.org/10.1016/j.msksp.2019.01.001
Received 16 May 2018; Received in revised form 18 December 2018; Accepted 2 January 2019

2 billion dollars each year in the USA (work absenteeism, medical evaluation, treatment) (Saint-Lary et al., 2015). In the case of CR, the data regarding the prevalence and the epidemiology of the condition are very limited. The reported annual incident of CR is 83.2 per 100.000 persons (107.3 for men and 63.5 for women) with a peak incidence in the fifth and sixth decade for both genders (Radhakrishnan et al., 1994).

The diagnosis of EN is based on information received during the subjective (history taking) and physical examination, which is then confirmed via diagnostic imaging or electrophysiological studies. Clinical examination of EN encompasses a variety of tests (sensation, muscle strength and reflexes) assessing the integrity and ability of the nervous system to conduct afferent or efferent impulses (loss of function) (Baselgia et al., 2017). In addition, a thorough examination includes evaluation of increased mechanical sensitivity of the nervous system, since PNP can be present without or with minimal loss of nerve conduction (Schmid et al., 2009). Diagnostic imaging and electrophysiological studies are most commonly used to establish a diagnosis of EN (Wainner et al., 2003). For most clinicians, these methods are accessible but given the waiting time for patients and the high cost for the society it would be useful to establish accurate clinical examination tests for the diagnosis of EN.

Neurodynamic tests are used by musculoskeletal physiotherapists in order to identify changes of mechanosensitivity in the nervous system, thus assessing gain of function (Baselgia et al., 2017). Due to the interdependence of the mechanical, electrical and chemical properties of the nervous system, changes in one of these features may affect the others (Butler, 2008). Impairments in the surrounding musculoskeletal structures could apply mechanical or chemical stimuli to a nerve, resulting in venous congestion, impaired axoplasmic flow, inflammation and development of mechanosensitive abnormal impulse generating sites (Nee and Butler, 2006).

For disorders affecting the upper limbs four different neurodynamic tests have been proposed to assess mechanosensitivity of the brachial plexus, medial, radial and ulnar nerve (Elvey, 1980) (Table 1). Where symptoms are not related to central pain mechanisms (broader distribution of symptoms due to central sensitization e.g. in case of persistent pain) a positive test response could be associated with neural or non-neural tissue sensitivity. A neurodynamic test is considered positive if it can reproduce the patient's own symptoms and if those symptoms can be altered through structural differentiation (Butler, 2000). Schmid et al. (2009) assessed the reliability of ULNTs and found that those tests have moderate reliability. Wainner et al. (2003, 2005) reported substantial to almost perfect reliability for the interpretation of theULNT1 (median) and ULNT2b (radial).

Although used by clinicians the diagnostic accuracy of upper limb neurodynamic tests (ULNTs) has not yet been fully established and is important to optimise patient care. A recent systematic review has summarised the evidence on diagnostic performance of tests (including ULNTs) which are utilised for the identification of CR and concluded that when consistent with patient history, a combined result of four negative ULNTs (high sensitivity) and a negative Arm Squeeze test could be used to rule out the disorder (Thoomes et al., 2017). Likewise an earlier systematic review, concluded that a positive Spurling's, traction/neck distraction, and Valsalva's test might be indicative of CR, while a negative ULNT1 might be used to rule it out (high sensitivity) (Rubinstein et al., 2007). Of the eight included studies in this systematic review only two had assessed the diagnostic accuracy of ULNTs. Finally in a previous clinical commentary the authors attempted to summarise the available evidence in regard to the diagnostic usefulness of neurodynamic tests (Nee et al., 2012). The authors, based on biomechanical and experimental studies, concluded that ULNTs can potentially distinguish pain related to neural mechanosensitivity from pain arising from other tissues, and therefore could detect PNP. In the view of the growing body of evidence, a systematic review is required to evaluate the quality and synthesis the available current evidence of the diagnostic accuracy of ULNTs and to inform clinical practice. The aim therefore of this study was to examine the intended role of ULNTs in assessment of PNP, by answering the following research question: What is the diagnostic accuracy of ULNTs when compared to diagnostic imaging or electrophysiologic studies, and how results from ULNTs can be interpreted when assessing patients with arm and/or neck symptoms?

2. Design and methods

This systematic review was conducted according to a pre-defined protocol based on the Cochrane Handbook for Diagnostic Test Accuracy studies (Deeks et al., 2013) and the Center for Reviews and Dissemination (CRD, 2009). In addition, the study is reported according to Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA) (McInnes et al., 2018). (Appendix 1)

2.1. Search strategy

Informed by subject (NH, KK, YV) and methodological experts (NH, CA) key bibliographic databases were searched independently by two reviewers (KK, YV). The search employed sensitive topic-based strategies designed for each database from inception to 21st November 2017. Databases of interest were: PEDro, MEDLINE (through PubMed), AMED, CINAHL, Cochrane Library, and EMBASE. The search strategy, informed by scoping search included MeSH terms and text words, as well as a combination of both for a comprehensive search. The following keywords and combination of them were used: upper limb neurodynamic test, neural provocation test, upper limb tension test, diagnosis, peripheral neuropathic pain, peripheral entrapment neuropathy, radicular pain, cervical radiculopathy, brachial plexus, carpal tunnel syndrome, cubital tunnel syndrome, accuracy, specificity, sensitivity, validity.

The search was augmented using reference lists of included studies, as well as searching the grey literature. Box 1 details the MEDLINE search strategy.

Table 1
ULNT procedure.

Order of movements	ULNT1 (median)	ULNT2a (median)	ULNT2b (radial)	ULNT3 (ulnar)
1	Shoulder depression	Shoulder depression	Shoulder depression	Shoulder depression
2	Shoulder abduction 110°	Elbow extension	Elbow extension	Shoulder abduction 100°
3	Wrist and fingers extension	Lateral rotation of the arm	Medial rotation arm	Lateral rotation arm
4	Forearm supination	Wrist and finger extension	Wrist and fingers flexion	Forearm pronation
5	Shoulder lateral rotation	Shoulder abduction 10°	Shoulder abduction	Elbow flexion
6	Elbow extension	Contralateral lateral flexion of the cervical spine	Contralateral lateral flexion of the cervical spine	Wrist and fingers extension
7	Contralateral lateral flexion of the cervical spine			Contralateral lateral flexion of the cervical spine

Box 1
MEDLINE search strategy

1. peripheral neuropathic pain.mp or exp Neuralgia/
2. radicular pain.mp or exp Hereditary Sensory and Autonomic Neuropathies/
3. peripheral entrapment neuropathy.mp
4. cervical radiculopathy.mp or exp Radiculopathy/
5. carpal tunnel syndrome.mp or exp Carpal tunnel syndrome/
6. cubital tunnel syndrome.mp or exp Cubital tunnel syndrome/
7. brachial plexus neuropathies.mp or exp Brachial plexus neuropathies/
8. exp Nerve compression syndromes/
9. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8
10. upper limb neurodynamic test.mp
11. upper limb tension test.mp
12. neural provocation test.mp
13. exp Diagnosis/
14. exp Pain measurements/
15. exp Neurologic examination/
16. exp Physical examination/
17. 10 or 11 or 12 or 13 or 14 or 15 or 16
18. diagnostic accuracy.mp
19. sensitivity and specificity.mp or exp Sensitivity and specificity/
20. validity.mp
21. exp Reproducibility of results/
22. exp Predictive value of tests/
23. 18 or 19 or 20 or 21 or 22
24. 9 and 17 and 23

2.2. Eligibility criteria

Eligibility criteria were established following the recommendations of The Cochrane Handbook for Diagnostic Test Accuracy studies (Leeflang et al., 2008) and informed using the SPIDER search concept (Cooke et al., 2012). Titles and abstract of the identified studies were screened by two independent reviewers (KK, YV) for eligibility using pre-specified inclusion criteria.

Inclusion criteria (based on SPIDER) included that the sample (S) comprised populations aged > 18 years with arm and/or neck symptoms with suspected peripheral neuropathic involvement (signs and symptoms suggesting excitability in the nervous system such as pain, paresthesia, dysesthesia, spasm or reduced impulse conduction such as hypoesthesia or anesthesia and weakness) (Nee and Butler, 2006); the phenomenon of interest (PI) was the diagnostic accuracy of ULNTs; investigated using a diagnostic accuracy study design (D); with comparison of the index test (ULNTs) to a reference standard, such as, electrophysiologic examination (electromyography and nerve conduction studies) or advanced imaging (e.g. Magnetic Resonance Imaging (MRI), CT, myelography) (E). Although not perfect, these tests are considered to be the most accurate diagnostic tests available (Wainner et al., 2003; Jablecki et al., 1993, 2002; Kuijper et al., 2009).

Exclusion criteria: case series, case reports, surgical or cadaveric studies; publications for which full text not available.

2.3. Quality assessment

Two reviewers (KK, YV) independently conducted the risk of bias (ROB) assessment using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) - tool, a development of the original tool (Whiting et al., 2011). It consists of four key domains: patient selection, index test, reference standard and, flow and timing. All key areas are assessed for ROB, whereas the first three are also assessed in terms of applicability to the review question. Each domain is judged as “high risk”, “low risk” or “unclear risk” based on signaling questions aiming to assist judgment (Whiting et al., 2011). Overall, a study can be judged as having “low risk of bias” if every domain has been ranked as “low risk”. Assessment of applicability is based on the first three domains and whether they are in line with the review question. The study is judged as having “no concerns” regarding applicability if these domains are in line with the review question and “with concerns” if deviates from the review objective. The QUADAS-2 has been used in recent systematic reviews (Grodahl et al., 2016; Hegedus et al., 2012) and is recommended by the Cochrane Collaboration and the U.K National Institute for Health and Clinical Excellence (Reitsma et al., 2009).

2.4. Data extraction

Diagnostic accuracy data and study characteristics were extracted by one reviewer (KK) using a pre-designed data extraction sheet which covered five areas. The data were audited by a second reviewer (YV) for accuracy. The following data were extracted: authors and publication details, studies' methods (aim of study, study design, method of recruitment, eligibility criteria, and ethical approval), participant details, diagnostic test data (sensitivity, specificity, predictive values, likelihood ratios and other). Finally, the fifth section was 2×2 contingency tables for the diagnostic tests.

2.5. Summary measures

Sensitivity, specificity, likelihood ratios (LR) and predictive values (PV) were the outcomes for which data were sought. True positive, false positive, true negative and false negative values were summarised. In cases where only incomplete or raw data were presented, a 2×2 contingency table was used to re-estimate these values. Sensitivity and specificity were graded as low (< 0.50), low/moderate (0.51–0.64), moderate (0.65–0.74), moderate/high (0.75–0.84) and high (> 0.85) in line with previous systematic reviews of diagnostic accuracy studies (Grodahl et al., 2016; Schneiders et al., 2012). Clinical interpretation of likelihood ratios was based on Jaeschke et al. (1994a,b) as follows: conclusive evidence (LR+ > 10 and LR- < 0.1), strong diagnostic evidence (LR+ 5 to 10 and LR- 0.1 to 0.2), weak diagnostic evidence (LR

+2 to 5 and LR- 0.2 to 0.5) and negligible evidence (LR+ 1 to 2 and LR- 0.5 to 1).

2.6. Data analysis

Homogeneity among studies was explored to evaluate if the studies were suitable for combining in a meta-analysis. Areas of exploration were: study designs, patient population, comparable reference tests and diagnostic data, no differences in diagnostic thresholds (Burgess et al., 2011). In addition, quality assessment of the included studies was conducted, since studies with high ROB often over-estimate the performance of a test (Lijmer et al., 2002). Given the heterogeneity of the included studies a narrative synthesis was undertaken.

2.7. Quality of evidence across studies

Quality of evidence, including risk of bias across studies was evaluated using GRADE (Schünemann et al., 2008) for individual tests. Quality of overall body of evidence is influenced by amongst other factors, study design, patient populations, precision, consistency, directness and as such each outcome was evaluated by both reviewers independently (Schünemann et al., 2008).

3. Results

3.1. Study identification

The searches identified 1802 studies with screening of title and abstract resulting in 15 studies that were retrieved for full-text evaluation and 8 studies (n = 579) meeting the eligibility requirements for inclusion. (Fig. 1). There was 100% of agreement between the reviewers on selecting studies.

3.2. Study description

Table 2 summarises the specific characteristics of all eight studies. Three studies investigated the diagnostic accuracy of ULNTs in individuals with suspected CR (Wainner et al., 2003; Apelby-Albrecht et al., 2013; Ghasemi et al., 2013). Two of the studies used electrophysiologic procedures as the reference standard (Wainner et al., 2003; Ghasemi et al., 2013). One study used MRI, clinical examination and history as a reference standard (Apelby-Albrecht et al., 2013). Five studies investigated the diagnostic accuracy of ULNTs in individuals with suspected CTS with nerve conduction studies as the reference standard (Wainner et al., 2005; Vanti et al., 2011; 2012; Bueno-Gracia et al., 2016; Trillos et al., 2018).

3.3. Risk of bias assessment

Agreement of risk of bias following discussion was excellent (100%). Four studies were assessed as “low risk of bias” (ROB) (Wainner et al., 2003, 2005; Vanti et al., 2012; Trillos et al., 2018), but all of them had concerns with regards to applicability (Table 3). Patient selection procedures and poor reporting of flow and timing were the main areas of ROB. Only two studies were assessed as no concerns for applicability (Fig. 2) (Apelby-Albrecht et al., 2013; Bueno-Gracia et al., 2016). Interpretation of the index test was the main reason for concern regarding applicability since it was not in agreement with our review question. In our study an ULNT is considered positive only when it reproduces the patient's clinical symptoms and those symptoms are modified with structural differentiation (Nee et al., 2012; Butler, 2000; Coppieters et al., 2002).

3.4. Synthesis of results

The main limitations for performing a meta-analysis were the heterogeneity in terms of the reference standard utilised, as well as in the interpretation of the index test and the methodological quality of the included studies. Since a meta-analysis was not possible, diagnostic accuracy data (sensitivity, specificity, predictive values and likelihood ratios) are presented using a narrative approach. The overall body of the evidence in terms of ROB, inconsistency, indirectness, imprecision, and the presence of potential reported bias after applying the GRADE approach was low to very low across studies and across outcomes. Diagnostic accuracy for all clinical indicators is summarised in Tables 4 and 5 and outcome of GRADE evaluation in Tables 6 and 7.

3.5. Diagnostic accuracy of upper limb neurodynamic tests

3.5.1. Carpal tunnel syndrome

Five studies examined the diagnostic accuracy of ULNTs in patients with suspected CTS (Wainner et al., 2005; Vanti et al., 2011; 2012; Bueno-Gracia et al., 2016; Trillos et al., 2018). From these studies two were at ROB (Vanti et al., 2011; Bueno-Gracia et al., 2016) and four had concerns regarding applicability (Wainner et al., 2005; Vanti et al., 2011; 2012; Trillos et al., 2018). Those at ROB had limitations related to patient selection and flow and timing. The study of Vanti et al. (2011) was at ROB because the number of patients enrolled in the study was different from the number of patients that were included in the analysis (Whiting et al., 2011), whereas in the study by Bueno-Gracia et al. (2016) the authors provided limited information in regards to the methods used for the enrollment of the sample (consecutive or random sample). The studies that had concerns regarding applicability used a definition for a positive ULNT that differs from that being used in this review.

Three studies assessed the validity of ULNT1 (median) considering the test positive in the presence of only one of the following criteria: 1) reproduction of patient's symptoms; 2) side to side differences (> 10°) in elbow extension; 3) contralateral neck side-flexion increased symptoms or ipsilateral side-flexion decreased symptoms (Wainner et al., 2005; Vanti et al., 2011; Trillos et al., 2018). Sensitivity was moderate/high 0.75 (95%CI 0.58–0.92) (Wainner et al., 2005) to high 0.91 (95%CI 0.74–0.98) (Vanti et al., 2011) and 0.93 (95%CI 0.88–0.96) (Trillos et al., 2011). Specificity was low in all 3 studies: 0.13 (95%CI 0.04–0.22) (Wainner et al., 2005), 0.15 (95%CI 0.05–0.36) (Vanti et al., 2011) and 0.06 (95%CI 0.0–0.33) (Trillos et al., 2018). In the study by Vanti et al. (2011) the authors conducted a second analysis in which “reproduction of patient's symptoms” changed to “reproduction of symptoms in the first, second or third digit”, but again only one of the three criteria was required for a positive ULNT1. The second analysis revealed low to moderate sensitivity (0.54, 95%CI 0.35–0.72) and moderate specificity (0.70, 95%CI 0.48–0.85). Overall, none of the interpretations of ULNT1 was capable of ruling in or ruling out a diagnosis of CTS because LRs were between 0.5 and 2.0.

Two studies examined the diagnostic accuracy of ULNT1 using a different interpretation for a positive test. In these studies the test was considered positive if it was able to reproduce patient's symptoms and these symptoms were altered with structural differentiation (Vanti et al., 2012; Bueno-Gracia et al., 2016). Sensitivity ranged from low 0.05 (95%CI 0.02–0.19) (Vanti et al., 2012) to low/moderate 0.58 (95%CI 0.45–0.71) (Bueno-Gracia et al., 2016). Specificity ranged from moderate/high 0.84 (95%CI 0.72–0.96) (Bueno-Gracia et al., 2016) to high 0.93 (95%CI 0.82–0.98) (Vanti et al., 2012). Bueno-Gracia and colleagues (2016) suggested that the ULNT1 may be clinically useful to determine patients with CTS due to high +LR (3.67). However the high number of false negatives results challenges this notion (Table 4).

3.5.2. Cervical radiculopathy

Three studies investigated the concordance of ULNT1 with a reference standard in patients with suspected CR (Wainner et al., 2003; Apelby-Albrecht et al., 2013; Ghasemi et al., 2013). The reference standard in two of these studies was NCS and needle electromyography

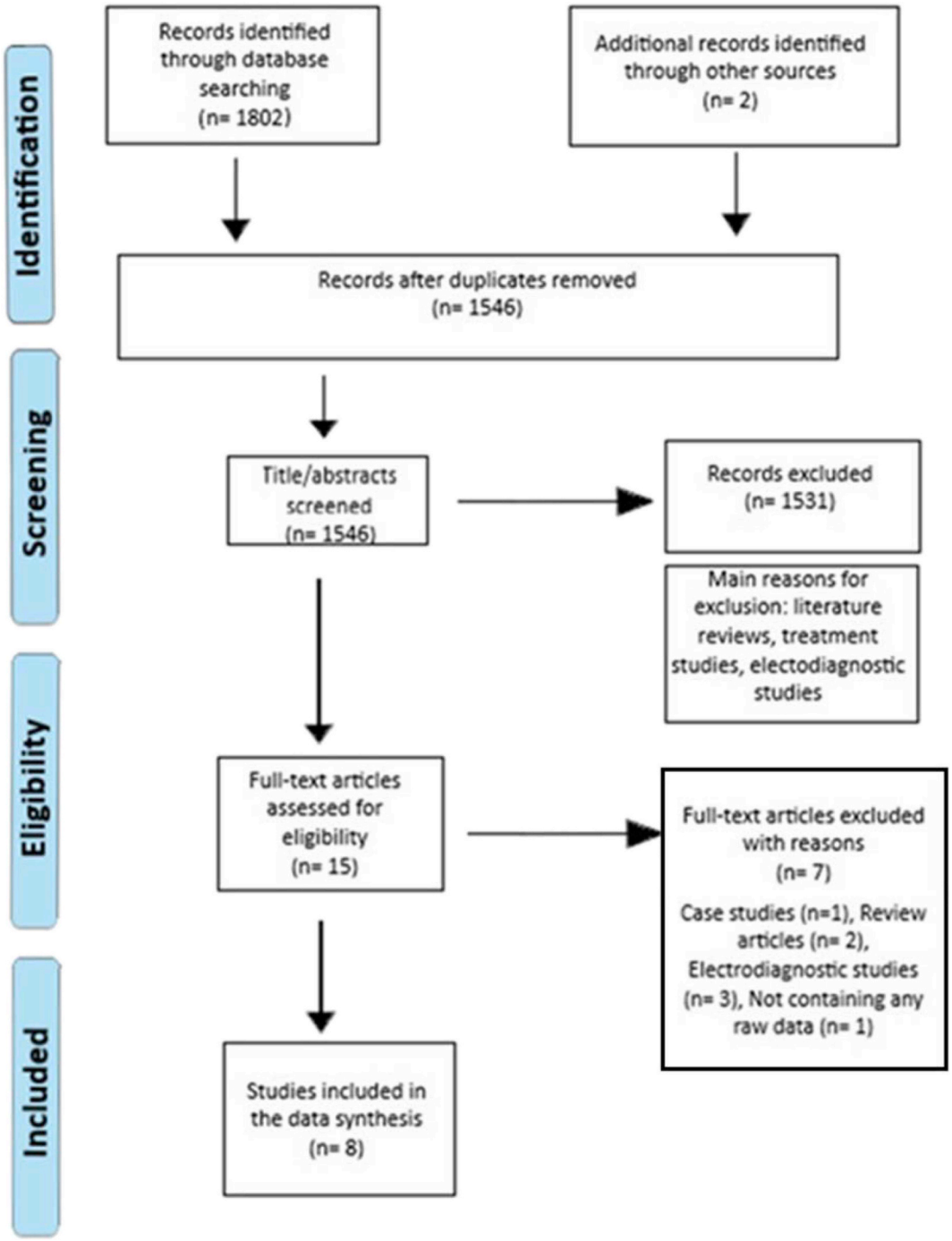

Fig. 1. PRISMA flow diagram for systematic reviews.

(Wainner et al., 2003; Ghasemi et al., 2013), whereas in the third study the authors used the combination of patient history, clinical examination and MRI findings as the reference standard (Apelby-Albrecht et al., 2013). In two of these studies ULNT1 showed moderate to high (0.83, 95%CI 0.66–0.93) and high sensitivity (0.97, 95%CI 0.90–1.0) (Apelby-Albrecht et al., 2013; Wainner et al., 2003) whereas in the third study the sensitivity was low 0.35 for chronic CR and low/moderate 0.6 for acute CR (Ghasemi et al., 2013). Specificity ranged from low 0.22 (95%CI 0.12–0.33) (Wainner et al., 2003) and 0.4 (Ghasemi et al., 2013) to moderate/high 0.75 (95%CI 0.48–0.93) (Apelby-Albrecht et al., 2013). Moreover, in the study of Wainner et al. (2003) the ULNT1 demonstrated negative likelihood ratio (LR) of 0.12, meaning that a negative ULNT1 could rule out CR. This study had low ROB, but had concerns regarding applicability related to the different interpretation of the index test from the authors compared with the review question (Whiting et al., 2011). In addition, due to wide 95% CI the results of this study should be interpreted cautiously. Wide CIs reduce the strength of evidence by influencing the precision of the pooled estimates.

The validity of ULNT2b (radial) was assessed by two studies (Wainner et al., 2003; Apelby-Albrecht et al., 2013). Sensitivity was moderate in both studies: 0.66 (95%CI 0.48–0.81) (Apelby-Albrecht et al., 2013) and 0.72 (95%CI 0.52–0.93) (Wainner et al., 2003). Specificity ranged from low 0.33 (95%CI 0.21–0.45) (Wainner et al., 2003) to moderate/high 0.75 (95%CI 0.48–0.93) (Apelby-Albrecht et al., 2013).

Table 2
Characteristics of included studies.

Author. (year), country	Type of study	Pathology	Setting	Inclusion/Exclusion criteria	Population (Number, gender, age)	Outcome measures	Reference Standard	ROB
Apelby-Albrecht et al. (2013) Sweden	Prospective cohort	Cervical radiculopathy	Center for spinal surgery	Inclusion: neck/arm pain Exclusion: History of multitrauma, malignant, system disease with possible neuropathy, or patients whose general condition (physically or/ and psychologically) could influence the results.	N = 51 Women n = 27 Men n = 24 Mean age: 51 (25–67) years	ULNT (1, 2a, 2b, 3) Combined and individually	MRI, Clinical examination, Patient history	At risk
Ghasemi et al. (2013) Iran	Cross-sectional	Cervical radiculopathy	Electrodiagnostic center (hospital)	Inclusion: Aged > 20 years, symptoms of neck/ radicular pain > 3 weeks Exclusion: History of neck trauma, prior surgery, tumors or congenital abnormality of cervical spine, any systemic situation known to cause peripheral neuropathies and known cases of rheumatoid arthritis	N = 97 Women n = 72 Men n = 25 Mean age: Women 46.14 ± 11.45 Men 46.32 ± 13.97 years	ULNT 1 (median)	NCS	At risk
Wainner et al. (2003)	Prospective cohort	Cervical radiculopathy	University of Pittsburgh, Wilford Hall USAF Medical Center, Brooke Army Medical Center, and Blanchfield Army Community Hospital	Inclusion: signs and symptoms compatible with CR or CTS Exclusion: systemic disease, primary report of bilateral radiating arm pain, history of conditions involving the affected upper extremity or surgical procedures for pathologies giving rise to neck pain or CTS, discontinuation of work > 6 months, previous EMG and NCS testing the symptomatic limb for CR, CTS, or both	N = 82 Women n = 41 Men n = 41 Mean age: 45 ± 12 years	ULNT 1 (median), ULNT 2b (radial)	Needle EMG and NCS	Low risk
Bueno-Gracia et al. (2016) Spain	Prospective cohort	Carpal tunnel syndrome	Not reported	Inclusion: patients with hand, wrist or forearm symptoms Exclusion: any ROM limitations of the upper limb, inability to lie supine, any physical contraindications for physical therapy, presence of any cognitive or communicative deficits	N = 58 Women n = 42 Men n = 16 Mean age: 54.3 ± 14.5 years	ULNT 1 (median)	NCS and clinical presentation	At risk
Trillos et al. (2018) Colombia	Prospective cohort	Carpal tunnel syndrome	Health service institution	Inclusion: age 18–86, referred with a clinical diagnosis of CTS Exclusion: upper limb joint and cervical spine pathologies, patients with history of rheumatoid arthritis, anterior shoulder dislocation, CRPS, Raynaud's syndrome, breast cancer, RC injuries, patients with cervical spinal stenosis, or cognitive deficits	N = 118 Women n = 98 Men n = 20 Mean age: 50.51 ± 11.1 years	ULNT 1 (median)	NCS	Low risk
Vanti et al. (2011) Italy	Prospective cohort	Carpal tunnel syndrome	Clinic of Occupational Medicine of the University of Bologna (Italy)	Inclusion: individuals with suspected CTS Exclusion: upper limb joint pathologies inflammatory, infective or systemic pathologies, history of surgical procedure for CTS, CR, cognitive deficits	N = 44 Women n = 33 Men n = 11 Mean age: 46.3 ± 10.8 years	ULNT 1 (median)	NCS	At risk
Vanti et al. (2012) Italy	Prospective cohort	Carpal tunnel syndrome	Occupational Medicine of the Department of Internal Medicine, Geriatrics and Nephrology, Alma Mater Studiorum, University of Bologna (Italy)	Inclusion: individuals with suspected CTS Exclusion: upper limb joint pathologies that could significantly limit the ROM of the upper limbs; inflammatory, systemic, or infectious diseases; history of surgical intervention for CTS; CR; and cognitive deficits	N = 47 Women n = 35 Men n = 12 Mean age: 45.9 ± 10.6 years	ULNT 1 (median)	NCS	Low risk

(*continued on next page*)

Table 2 (*continued*)

Author. (year), country	Type of study	Pathology	Setting	Inclusion/Exclusion criteria	Population (Number, gender, age)	Outcome measures	Reference Standard	ROB
Wainner et al. (2005)	Prospective cohort	Carpal tunnel syndrome	Multicenter medical center and community hospital	Inclusion: signs and symptoms compatible with CR or CTS Exclusion: systemic disease, primary report of bilateral radiating arm pain, history of conditions involving the affected upper extremity or surgical procedures for pathologies giving rise to neck pain or CTS, discontinuation of work > 6 months, previous EMG and NCS testing the symptomatic limb for CR, CTS, or both	N = 82 Women n = 41 Men n = 41 Mean age: 45 ± 12 years	ULNT 1 (median) ULNT 2b (radial)	NCS and clinical presentation	

ROM: Range of motion, ULNT: Upper limb neurodynamic test, NCS: Nerve conduction studies, CTS:Carpal tunnel syndrome, CRPS: Complex regional pain syndrome, RC: Rotator cuff, CR:Cervical radiculopathy, EMG: Electromyography.

Apelby-Albrecht et al. (2013) also examined the diagnostic accuracy of ULNT2a (median), ULNT3 (ulnar) and ULNTs combined as a single test. This study was assessed as at ROB due to the time lapse between the MRI and the neurodynamic testing (up to six months) (Whiting et al., 2011); however, no concerns regarding applicability were identified. Combined ULNTs showed high sensitivity (0.97, 95%CI 0.85–1.00) and moderate specificity (0.69, 95%CI 0.41–0.89) whereas the ULNT3 (ulnar) was the most specific (0.87, 95%CI 0.62–0.98) (Apelby-Albrecht et al., 2013) (Table 5).

4. Discussion

The purpose of this study was to evaluate the role of ULNTs in the assessment of PNP and to reflect on their value in clinical practice in the assessment and diagnosis of patients with arm and/or neck symptoms. Current research suggests that ULNTs cannot be used in isolation for the diagnosis of PNP. Specifically, ULNTs cannot be utilised as a stand-alone test in the clinical setting for the diagnosis of CTS. Limited evidence suggests that ULNTs demonstrate better diagnostic accuracy and may be clinically relevant for the diagnosis of CR, but only as a “ruling out” strategy. However, the overall body of the evidence after applying the GRADE approach was low to very low for all outcomes, therefore any interpretation of these findings should be made cautiously.

4.1. Carpal tunnel syndrome

Overall, the five studies that examined the validity of ULNT1 are characterised by diversity in the interpretation of the index test. From these studies only the interpretation by Bueno-Gracia et al. (2016) is in agreement with the review question, that is, the ULNT1 is considered positive only when it reproduces the patient's clinical symptoms and those symptoms are modified with structural differentiation. This criterion is supported by several authors, who suggest that structural differentiation is necessary in order to distinguish between neuropathic pain and pain that arises from other somatic sources (Nee et al., 2012; Butler, 2000, Coppieters et al., 2002). Using the above definition for a positive test Bueno-Gracia et al. (2016) found that the ULNT1 may has strong ability to identify patients who do not have CTS (high specificity).

Using a different definition of a positive test Wainner et al. (2005), Vanti et al. (2011) and Trillos et al. (2018) found that the ULNT1 had moderate/high to high sensitivity. However, the low specificities and LRs that have been obtained in these studies decrease the diagnostic accuracy of ULNT1 and suggest that they cannot be considered adequate for the diagnosis of CTS.

4.2. Cervical radiculopathy

The diagnostic accuracy of ULNTs seems more promising for the diagnosis of CR. Apelby-Albrecht et al. (2013) investigated the validity of ULNTs combined and individually, using the same definition for a positive test as this review. Individually, ULNT1 and ULNT3 were the most valid tests for detecting CR. Combining the tests increased the diagnostic accuracy of ULNTs further, giving an accurate diagnosis in 88.2% of patients.

Whilst findings by Wainner et al. (2003) are in agreement with the study by Apelby-Albrecht et al. (2013) the authors used a more liberal definition of a positive test. In their study, the ULNT1 was highly sensitive and had LR-of 0.12 meaning that when the test is negative, CR can be ruled out. In these studies the vast majority of patients with CR presented with nerve root compression at C6-C7 level, therefore the diagnostic properties of ULNTs may be different when the C5 or C8 root level is involved.

Overall, following analysis of the available evidence, ULNTs seem to have no diagnostic accuracy to inform clinical practice in patients with suspected CTS. In contrast, ULNTs may be more useful for the diagnosis

Table 3
Risk of bias assessment of included studies.

Study	RISK OF BIAS				Summary	APPLICABILITY CONCERNS			Summary
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING		PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	
Apelby-Albrecht et al. (2013)	☺	☺	☺	☹	At risk	☺	☺	☺	No concern
Bueno-Gracia et al. (2016)	☹	☺	☺	☺	At risk	☺	☺	☺	No concern
Ghasemi et al. (2013)	☹	?	?	?	At risk	☺	☹	☺	With concern
Trillos et al. (2018)	☺	☺	☺	☺	Low risk	☺	☹	☺	With concern
Vanti et al. (2011)	☺	☺	☺	☹	At risk	☺	☹	☺	With concern
Vanti et al. (2012)	☺	☺	☺	☺	Low risk	☺	☹	☺	With concern
Wainner et al. (2003)	☺	☺	☺	☺	Low risk	☺	☹	☺	With concern
Wainner et al. (2005)	☺	☺	☺	☺	Low risk	☺	☹	☺	With concern

of CR, but only as a "ruling out" strategy. Nonetheless, these findings should be interpreted cautiously due to the small number of studies investigating the diagnostic accuracy of ULNTs and the differences between them in regards to the interpretation of a positive test.

There are a number of concerns that may explain some of the results obtained in these studies. Firstly, electrodiagnostic testing provides information in regards to conduction loss in large myelinated motor neurons and Aβ fibres (Schmid et al., 2013). Increased mechanosensitivity, however, is related to increased excitability of small-diameter afferents and sensitization of nociceptors in the nervi nervorum and sinuvertebral nerves (Baron et al., 2010). Moreover, recent evidence suggests that damage of small axons is more common in entrapment neuropathies than previously believed (Chien et al., 2008; Schmid et al., 2012) and may occur even before any dysfunction of large axons (Tamburin et al., 2011). Thus, it becomes apparent that the inability of the criterion standard to identify neuropathies related to small axons damage may have led to false-negative results in cases where NCS classified a patient as not having the condition whereas the ULNTs were positive.

Secondly, in a recent study Baselgia et al. (2017) found that > 54% of patients with CTS had negative ULNT1 despite a clear dysfunction in the median nerve, as proven with NCS. The authors advocated that the non-reproduction of symptoms during neurodynamic testing can be a sign of a more severe neural dysfunction of the unmyelinated fibres (Baselgia et al., 2017). These findings, could explain some of the false-negative results that have been obtained in the included studies in cases where the NCS confirmed a diagnosis but the neurodynamic testing was negative.

4.3. Future direction

A reference standard should be comprehensive enough to accurately inform clinicians in regards to the diagnostic accuracy of an index test. Given the insufficiency of electrodiagnostic tests to provide information about the integrity of small-diameter nerve fibres (Schmid et al., 2013), it becomes apparent that diagnostic accuracy studies need a supplementary test that will increase the criterion validity of the reference standard. Quantitative Sensory Testing (QST) provides information for both loss and gain of function, in large myelinated (Aβ) and thinly myelinated (Aδ) or unmyelinated fibres (C-fibres) (Rolke et al., 2006). QST protocols include tests that investigate thermal, mechanical and pain thresholds, and based on the results clinicians could be informed in regards to which type of nerve fibres might be involved. Incorporating QST in protocols, may enhance their ability to correctly classify patients with PNP. Additionally future diagnostic accuracy studies aiming to investigate the validity of ULNTs in patients with CTS could adopt the principle of "neurodynamic sequencing" and alter the order of joint movement. Various studies have shown that the range of motion and the symptoms can be modified by altering the testing sequence during straight leg raise (Boland and Adams, 2000), slump test (Johnson and Chiarello, 1997) and ULNT1 (Coppieters et al., 2001). Moving the wrist to extension first during ULNT1 testing may increase the likelihood of a positive neurodynamic test (Baselgia et al., 2017). Moreover, consensus as to what defines a positive test would be useful. Standardisation of the performance and the interpretation of ULNTs are essential to draw safe inferences for the true diagnostic accuracy of the tests (Nee et al., 2012). Finally, future diagnostic accuracy studies should evaluate the

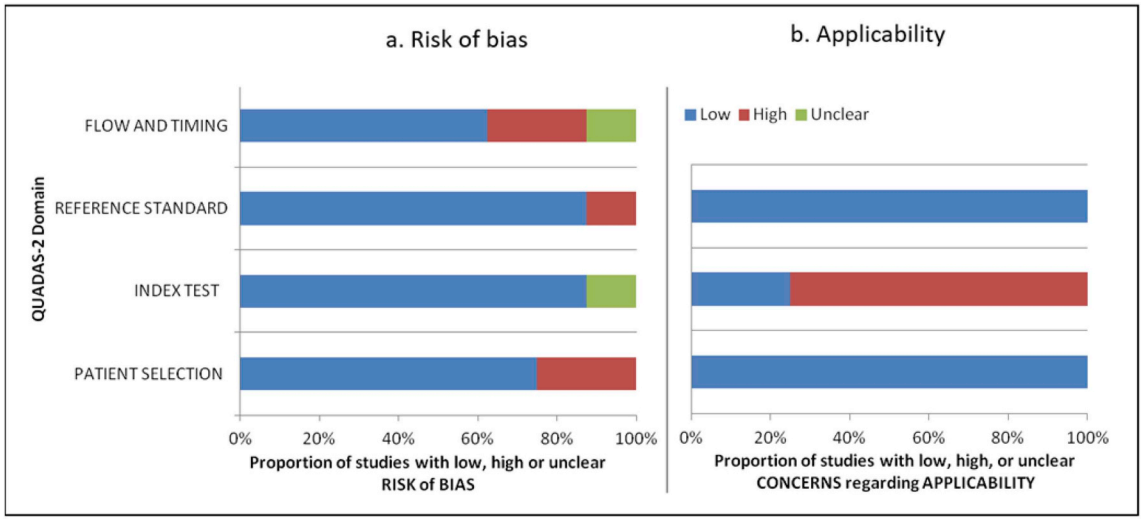

Fig. 2. Proportion of studies assessed as low, high or unclear ROB and/or applicability.

Table 4
Diagnostic ULNTs accuracy data for CTS.

Author (Year)	Test (Positive test criteria)	SN (95% CI)	SP (95% CI)	+LR (95% CI)	-LR (95% CI)	PPV (95% CI)	NPV (95% CI)
Bueno-Gracia et al. (2016)	ULNT1	0.58 (0.45–0.71)	0.84 (0.72–0.96)	3.67 (1.70–7.89)	0.50 (0.36–0.70)	0.85 (0.71–92)	0.43 (36–51)
	Criterion A -Patient's symptoms reproduced and changed with SD	0.74 (0.61–0.83)	0.50 (0.35–0.65)	1.47 (1.03–2.10)	0.53 (0.31–0.90)	0.69 (61–75)	0.44 (32–45)
	Criterion B -Reproduction of symptoms in the wrist and first three digits that changed with SD, regardless of the reproduction of patient's clinical symptoms						
Trillos et al. (2018)	ULNT1 -Any one of the following: (1) patient's symptoms reproduced; (2) side to side differences (> 10°) in elbow extension on completion of all motion sequences; (3) change of symptoms with SD	0.93 (0.88–0.96)	0.06 (0.0–0.33)	1.00	1.05	0.87 (?)	0.12 (?)
Vanti et al. (2011)	ULNT1	0.91 (0.74–0.98)	0.15 (0.05–0.36)	1.07 (0.38–3.08)	0.55 (0.19–1.59)	0.56 (?)	0.40 (?)
	Criterion A -Any one of the following: (1) reproduction of patient's symptoms; (2) side to side differences (> 10°) in elbow extension on completion of all motion sequences; (3) change of symptoms with SD	0.54 (0.35–0.72)	0.70 (0.48–0.85)	1.8 (1.13–2.88)	0.65 (0.41–1.04)	0.68 (?)	0.44 (?)
	Criterion B -Side to side differences (> 10°) in elbow extension on completion of all motion sequences, but (1) and (3) positive only in presence of symptoms reproduction in the 1st, 2nd and 3rd digit of the affected arm						
Vanti et al. (2012)	ULNT1	0.4 (0.26–0.56)	0.79 (0.66–0.88)	1.96 (1.27–3.01)	0.75 (0.49–1.16)	0.58 (0.39–0.75)	0.65 (0.52–0.76)
	Criterion A	0.28 (0.16–0.45)	0.82 (0.69–0.91)	1.6 (0.93–2.76)	0.86 (0.50–1.49)	0.55 (0.34–0.75)	0.59 (0.47–0.70)
	-symptoms in fingers I, II or III	0.05 (0.02–0.19)	0.93 (0.82–0.98)	0.85 (0.22–3.30)	1.01 (0.26–3.89)	0.4 (0.12–0.77)	0.56 (0.45–0.67)
	Criterion B -A + symptoms increased with controlateral cervical side bending						
	Criterion C -A + symptoms decreased with ipsilateral cervical side bending						
Wainner et al. (2005)	ULNT1	0.75 (0.58–0.92)	0.13 (0.04–0.22)	0.86 (0.67–1.0)	1.9 (0.72–5.1)	(?)	(?)
	-Any one of the following: (1) patient's symptoms reproduced; (2) side to side differences (> 10°) in elbow extension on completion of all motion sequences; (3) change of symptoms with SD	0.64 (0.45–0.83)	0.30 (0.17–0.42)	0.91 (0.65–1.3)	1.2 (0.62–2.4)	(?)	(?)
	ULNT2b -Any one of the following: (1) patient's symptoms reproduced; (2) side to side differences (> 10°) in elbow extension on completion of all motion sequences; (3) change of symptoms with SD						

SD: structural differentiation, SN: sensitivity, SP: specificity, +LR: positive likelihood ratio, -LR: negative likelihood ratio, PPV: positive predictive value, NPV: negative predictive value, CI: confidence intervals,?: Data not available, authors have been contacted but did not respond.

Table 5
Diagnostic ULNTs accuracy data for CR.

Author (Year)	Test (Positive test criteria)	SN (95% CI)	SP (95% CI)	+LR (95% CI)	-LR (95% CI)	PPV (95% CI)	NPV (95% CI)
Apelby-Albrecht et al. (2013)	ULNT1	0.83 (0.66–0.93)	0.75 (0.48–0.93)	3.32	0.22	0.88 (0.72–0.97)	0.67 (0.41–0.87)
	-(1) Reproducible neurogenic pain in neck and arm, (2) increased/decreased symptoms with SD, (3) difference in painful radiation between right and left sides	0.66 (0.48–0.81)	0.75 (0.48–0.93)	2.64	0.45	0.85 (0.66–0.96)	0.50 (0.29–0.71)
		0.43 (0.26–0.61)	0.75 (0.48–0.93)	1.72	0.76	0.79 (0.54–0.94)	0.37 (0.21–0.56)
	ULNT2a	0.71 (0.54–0.85)	0.87 (0.62–0.98)	5.68	0.32	0.93 (0.76–0.99)	0.58 (0.37–0.78)
	-(1) Reproducible neurogenic pain in neck and arm, (2) increased/decreased symptoms with SD, (3) difference in painful radiation between right and left sides	0.97 (0.85–1.00)	0.69 (0.41–0.89)	3.11	0.04	0.87 (0.73–0.96)	0.92 (0.62–1.00)
	ULNT2b						
	-(1) Reproducible neurogenic pain in neck and arm, (2) increased/decreased symptoms with SD, (3) difference in painful radiation between right and left sides						
	ULNT3						
	-(1) Reproducible neurogenic pain in neck and arm, (2) increased/decreased symptoms with SD, (3) difference in painful radiation between right and left sides						
	ULNTcomb.						
	-(1) Reproducible neurogenic pain in neck and arm, (2) increased/decreased symptoms with SD, (3) difference in painful radiation between right and left sides						
Ghasemi et al. (2013)	ULNT1	0.6	0.4	1.0	1.0	0.68 (?)	0.32 (?)
	-Reproduction of pain in any step	0.35	0.4	0.58	1.62	0.50 (?)	0.27 (?)
	Acute CR						
	Chronic CR						
Wainner et al. (2003)	ULNT1	0.97 (0.90–1.0)	0.22 (0.12–0.33)	1.3 (1.1–1.5)	0.12 (0.01–1.9)	(?)	(?)
	-Any one of the following: (1) patient's symptoms reproduced; (2) side to side differences (> 10°) in elbow extension on completion of all motion sequences; (3) change of symptoms with SD	0.72 (0.52–0.93)	0.33 (0.21–0.45)	1.1 (0.77–01.5)	0.85 (0.37–1.9)	(?)	(?)
	ULNT2b						
	-Any one of the following: (1) patient's symptoms reproduced; (2) side to side differences (> 10°) in elbow extension on completion of all motion sequences; (3) change of symptoms with SD						

SD: structural differentiation, SN: sensitivity, SP: specificity, +LR: positive likelihood ratio, -LR: negative likelihood ratio, PPV: positive predictive value, NPV: negative predictive value, CI: confidence intervals,?: Data not available, authors have been contacted but did not respond.

Table 6
GRADE assessment of evidence (CR).

	No of studies (No of patients)	Accuracy measures	RoB	Indirectness	Inconsistency	Imprecision	Publication bias	Quality of evince
ULNT1	3 studies (n = 230)	Sensitivity	Serious	Serious	Serious	Serious	Undetected	Very low
	(Apelby-Albrecht et al., 2013; Ghasemi et al., 2013; Wainner et al., 2003)	Specificity	Serious	Serious	Very serious	Very serious	Undetected	Very low
ULNT2a	1 study (n = 51)	Sensitivity	Serious	No	No	Very serious	Undetected	Very low
	(Apelby-Albrecht et al., 2013)	Specificity	Serious	No	No	Very serious	Undetected	Very low
ULNT2b	2 studies (n = 133)	Sensitivity	Serious	Serious	Very serious	Very serious	Undetected	Very low
	(Apelby-Albrecht et al., 2013; Wainner et al., 2003)	Specificity	Serious	Serious	Very serious	Serious	Undetected	Very low
ULNT3	1 study (n = 51)	Sensitivity	Serious	No	No	Very serious	Undetected	Very low
	(Apelby-Albrecht et al., 2013)	Specificity	Serious	No	No	Very serious	Undetected	Very low
ULNT (combined)	1 study (n = 51)	Sensitivity	Serious	No	No	Serious	Undetected	Low
	(Apelby-Albrecht et al., 2013)	Specificity	Serious	No	No	Very serious	Undetected	Very low

ULNT: upper limb neurodynamic test, RoB: risk of bias, CR: cervical radiculopathy.

Table 7
GRADE assessment of evidence (CTS).

	No of studies (No of patients)	Accuracy measures	RoB	Indirectness	Inconsistency	Imprecision	Publication bias	Quality of evince
ULNT1	5 studies (n = 349)	Sensitivity	Serious	Serious	Serious	Very serious	Undetected	Very low
	Wainner et al. (2005); Vanti et al. (2011, 2012); Bueno-Gracia et al. (2016); Trillos et al. (2018)	Specificity	Serious	Serious	Very serious	Very serious	Undetected	Very low
ULNT2b	1 study (n = 82)	Sensitivity	No	Serious	Serious	Very serious	Undetected	Very low
	(Wainner et al., 2005)	Specificity	No	Serious	Serious	Very serious	Undetected	Very low

ULNT: upper limb neurodynamic test, RoB: risk of bias, CTS: carpal tunnel syndrome.

diagnostic utility of ULNTs for ulnar nerve EN, since currently there are limited evidence regarding to the validity of ULNTs in pathologies such cubital syndrome.

4.4. Strengths and limitations

The strengths of this review are that provides clear recommendations for future studies and emphasises the importance of precisely reported methodologically robust studies. Among the limitations of this systematic review is that it includes studies only written in English which may have introduced bias (Song et al., 2002). Whilst we have adopted the grading of sensitivity and specificity using parameters based on existing reviews we acknowledge interpretation is context specific; further research is required to validate these categories.

5. Conclusion

Based on the available evidence, ULNTs have no diagnostic accuracy to identify patients with CTS when used in isolation. Limited evidence suggests that ULNTs demonstrate better diagnostic accuracy and may be clinically relevant for the diagnosis of CR, but only in a "ruling out" strategy. However, the overall quality of the body of evidence after applying the GRADE approach was low to very low across studies. Further higher quality research is needed to establish firm conclusions regarding to the value of ULNTs in the assessment and diagnosis of patients with arm and/or neck symptoms.

Conflict of interest

None declared.

Ethical approval

None required.

Funding

None.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

None required.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.msksp.2019.01.001.

Appendix 1

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
TITLE/ABSTRACT			
Title	1	Identify the report as a systematic review (+/− meta-analysis) of diagnostic test accuracy (DTA) studies.	1
Abstract	2	Abstract: See PRISMA-DTA for abstracts.	1

INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	2–4
Clinical role of index test	D1	State the scientific and clinical background, including the intended use and clinical role of the index test, and if applicable, the rationale for minimally acceptable test accuracy (or minimum difference in accuracy for comparative design).	2–4
Objectives	4	Provide an explicit statement of question(s) being addressed in terms of participants, index test(s), and target condition(s).	4
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	N/A
Eligibility criteria	6	Specify study characteristics (participants, setting, index test(s), reference standard(s), target condition(s), and study design) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5–6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	4–5
Search	8	Present full search strategies for all electronic databases and other sources searched, including any limits used, such that they could be repeated.	5
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	7–8
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	6–7
Definitions for data extraction	11	Provide definitions used in data extraction and classifications of target condition(s), index test(s), reference standard(s) and other characteristics (e.g. study design, clinical setting).	6–7
Risk of bias and applicability	12	Describe methods used for assessing risk of bias in individual studies and concerns regarding the applicability to the review question.	6
Diagnostic accuracy measures	13	State the principal diagnostic accuracy measure(s) reported (e.g. sensitivity, specificity) and state the unit of assessment (e.g. per-patient, per-lesion).	7
Synthesis of results	14	Describe methods of handling data, combining results of studies and describing variability between studies. This could include, but is not limited to: a) handling of multiple definitions of target condition. b) handling of multiple thresholds of test positivity, c) handling multiple index test readers, d) handling of indeterminate test results, e) grouping and comparing tests, f) handling of different reference standards	14
Meta-analysis	D2	Report the statistical methods used for meta-analyses, if performed.	N/A
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			
Study selection	17	Provide numbers of studies screened, assessed for eligibility, included in the review (and included in meta-analysis, if applicable) with reasons for exclusions at each stage, ideally with a flow diagram.	7–8
Study characteristics	18	For each included study provide citations and present key characteristics including: a) participant characteristics (presentation, prior testing), b) clinical setting, c) study design, d) target condition definition, e) index test, f) reference standard, g) sample size, h) funding sources	8–12
Risk of bias and applicability	19	Present evaluation of risk of bias and concerns regarding applicability for each study.	13–14
Results of individual studies	20	For each analysis in each study (e.g. unique combination of index test, reference standard, and positivity threshold) report 2 × 2 data (TP, FP, FN, TN) with estimates of diagnostic accuracy and confidence intervals, ideally with a forest or receiver operator characteristic (ROC) plot.	?
Synthesis of results	21	Describe test accuracy, including variability; if meta-analysis was done, include results and confidence intervals.	14–18
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression; analysis of index test: failure rates, proportion of inconclusive results, adverse events).	N/A
DISCUSSION			
Summary of evidence	24	Summarise the main findings including the strength of evidence.	21–22
Limitations	25	Discuss limitations from included studies (e.g. risk of bias and concerns regarding applicability) and from the review process (e.g. incomplete retrieval of identified research).	23
Conclusions	26	Provide a general interpretation of the results in the context of other evidence. Discuss implications for future research and clinical practice (e.g. the intended use and clinical role of the index test).	24
FUNDING			
Funding	27	For the systematic review, describe the sources of funding and other support and the role of the funders.	24

Adapted From: McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, The PRISMA-DTA Group (2018). Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA. 2018 Jan 23; 319(4):388–396. https://doi.org/10.1001/jama.2017.19163.

For more information, visit: www.prisma-statement.org.

References

Apelby-Albrecht, M., Andersson, L., Kleiva, I.W., et al., 2013. Concordance of upper limb neurodynamic tests with medical examination and magnetic resonance imaging in patients with cervical radiculopathy: a diagnostic cohort study. J. Manip. Physiol. Therapeut. 36 (9), 626–632.

Atroshi, I., Gummesson, C., Johnsson, R., Ornstein, E., Ranstam, J., Rosén, I., 1999. Prevalence of carpal tunnel syndrome in a general population. Jama 282 (2), 153–158.

Baron, R., Binder, A., Wasner, G., 2010. Neuropathic pain: diagnosis, pathophysiological mechanisms, and treatment. Lancet Neurol. 9 (8), 807–819.

Baselgia, L.T., Bennett, D.L., Silbiger, R.M., et al., 2017. Negative neurodynamic tests do not exclude neural dysfunction in patients with entrapment neuropathies. Arch. Phys. Med. Rehabil. 98 (3), 480–486.

Boland, R.A., Adams, R.D., 2000. Effects of ankle dorsiflexion on range and reliability of straight leg raising. Aust. J. Physiother. 46 (3), 191–200.

Bueno-Gracia, E., Tricás-Moreno, J.M., Fanlo-Mazas, P., Malo-Urriés, M., Haddad-Garay, M., Estébanez-de-Miguel, E., ... Krauss, J.R., 2016. Validity of the Upper Limb Neurodynamic Test 1 for the diagnosis of Carpal Tunnel Syndrome. The role of structural differentiation. Manual Ther. 22, 190–195.

Burgess, R.M., Rushton, A., Wright, C., Daborn, C., 2011. The validity and accuracy of clinical diagnostic tests used to detect labral pathology of the hip: a systematic review. Man. Ther. 16 (4), 318–326.

Butler, D., 2008. The Neurodynamic Techniques. A Definitive Guide from the Noigroup Team. Noigroup Publication, Adelaide.

Butler, D.S., 2000. The Sensitive Nervous System. Noigroup publications.

Chien, A., Eliav, E., Sterling, M., 2008. Whiplash (grade II) and cervical radiculopathy share a similar sensory presentation: an investigation using quantitative sensory testing. Clin. J. Pain 24 (7), 595–603.

Cooke, A., Smith, D., Booth, A., 2012. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. Qual. Health Res. 22 (10), 1435–1443.

Coppieters, M., Stappaerts, K., Janssens, K., et al., 2002. Reliability of detecting 'onset of pain' and 'submaximal pain' during neural provocation testing of the upper quadrant. Physiother. Res. Int. 7 (3), 146–156.

Coppieters, M.W., Stappaerts, K.H., Everaert, D.G., et al., 2001. Addition of test components during neurodynamic testing: effect on range of motion and sensory responses. J. Orthop. Sports Phys. Ther. 31 (5), 226–237.

CRD, 2009. Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care. Centre for Reviews and Dissemination (Chapter 2).

Deeks, J.J., Wisniewski, S., Davenport, C., 2013. Chapter 4: guide to the contents of a Cochrane diagnostic test accuracy protocol. In: Deeks, J.J., Bossuyt, P.M., Gatsonis, C.

(Eds.), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration Available from: http://srdta.cochrane.org/.
Elvey, R.L., 1980. Brachial plexus tension tests and the pathoanatomical origin of arm pain. In: Idczak, R.M. (Ed.), Aspects of Manipulative Therapy. Lincoln Institute of Health Sciences, Melbourne, pp. 105–110.
Fernández-de-las-Peñas, C., Fernández-Muñoz, J.J., Palacios-Ceña, M., Navarro-Pardo, E., Ambite-Quesada, S., Salom-Moreno, J., 2015. Direct and Indirect effects of function in associated variables such as depression and severity on pain intensity in women with carpal tunnel syndrome. Pain Med. 16 (12), 2405–2411.
Finnerup, N.B., Haroutounian, S., Kamerman, P., et al., 2016. Neuropathic pain: an updated grading system for research and clinical practice. Pain 157 (8), 1599–1606.
Ghasemi, M., Golabchi, K., Mousavi, S.A., et al., 2013. The value of provocative tests in diagnosis of cervical radiculopathy. J. Res. Med. Sci. Off. J. Isfahan Univ. Med. Sci. 18 (Suppl. 1), S35–S38.
Grodahl, L.H., Fawcett, L., Nazareth, M., et al., 2016. Diagnostic utility of patient history and physical examination data to detect spondylolysis and spondylolisthesis in athletes with low back pain: a systematic review. Man. Ther. 24, 7–17.
Hegedus, E.J., Goode, A.P., Cook, C.E., et al., 2012. Which physical examination tests provide clinicians with the most value when examining the shoulder? Update of a systematic review with meta-analysis of individual tests. Br. J. Sports Med. 46 (14), 964–978.
Jablecki, C.K., Andary, C.M.T., So, Y.T., Wilkins, D.E., Williams, F.H., 1993. Literature review of the usefulness of nerve conduction studies and electromyography for the evaluation of patients with carpal tunnel syndrome. Muscle Nerve Off. J. Am. Assoc. Electrodiagn. Med. 16 (12), 1392–1414.
Jablecki, C.K., Andary, M.T., Floeter, M.K., et al., 2002. Practice parameter: electrodiagnostic studies in carpal tunnel syndrome. Report of the American association of electrodiagnostic medicine, American academy of neurology, and the American academy of physical medicine and rehabilitation. Neurology 58 (11), 1589–1592.
Jaeschke, R., Guyatt, G.H., Sackett, D.L., Guyatt, G., Bass, E., Brill-Edwards, P., Haynes, B., 1994a. Users' guides to the medical literature: III. How to use an article about a diagnostic test B. What are the results and will they help me in caring for my patients? Jama 271 (9), 703–707.
Jaeschke, R., Guyatt, G.H., Sackett, D.L., 1994b. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. Jama 271 (9), 703–707.
Johnson, E.K., Chiarello, C.M., 1997. The slump test: the effects of head and lower extremity position on knee extension. J. Orthop. Sports Phys. Ther. 26 (6), 310–317.
Kuijper, B., Tans, J.T., Schimsheimer, R.J., et al., 2009. Degenerative cervical radiculopathy: diagnosis and conservative treatment. A review. Eur. J. Neurol. 16 (1), 15–20.
Leeflang, M.M., Deeks, J.J., Gatsonis, C., Bossuyt, P.M., 2008. Systematic reviews of diagnostic test accuracy. Ann. Internal Med. 149 (12), 889–897.
Lijmer, J.G., Bossuyt, P.M., Heisterkamp, S.H., 2002. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. Stat. Med. 21 (11), 1525–1537.
McInnes, M.D., Moher, D., Thombs, B.D., McGrath, T.A., Bossuyt, P.M., Clifford, T., et al., 2018. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. Jama 319 (4), 388–396.
Nee, R.J., Butler, D., 2006. Management of peripheral neuropathic pain: Integrating neurobiology, neurodynamics, and clinical evidence. Phys. Ther. Sport 7 (1), 36–49.
Nee, R.J., Jull, G.A., Vicenzino, B., et al., 2012. The validity of upper-limb neurodynamic tests for detecting peripheral neuropathic pain. J. Orthop. Sports Phys. Ther. 42 (5), 413–424.
Radhakrishnan, K., Litchy, W.J., O'Fallon, W.M., et al., 1994. Epidemiology of cervical radiculopathy. A population-based study from Rochester, Minnesota, 1976 through 1990. Brain J. Neurol. 117 (Pt 2), 325–335 (Pt 2).
Reitsma, J.B., Rutjes, A.W.S., Whiting, P., Vlassov, V.V., Leeflang, M.M.G., Deeks, J.J., 2009. Chapter 9: assessing methodological quality. In: Deeks, J.J., Bossuyt, P.M., Gatsonis, C. (Eds.), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration Available from: http://srdta.cochrane.org/.
Rolke, R., Baron, R., Maier, C., et al., 2006. Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): standardized protocol and reference values. Pain 123 (3), 231–243.
Rubinstein, S.M., Pool, J.J., van Tulder, M.W., Riphagen, I.I., de Vet, H.C., 2007. A systematic review of the diagnostic accuracy of provocative tests of the neck for diagnosing cervical radiculopathy. Eur. Spine J. 16 (3), 307–319.
Saint-Lary, O., Rébois, A., Mediouni, Z., Descatha, A., 2015. Carpal tunnel syndrome: primary care and occupational factors. Front. Med. 2, 28.
Schmid, A.B., Brunner, F., Luomajoki, H., et al., 2009. Reliability of clinical tests to evaluate nerve function and mechanosensitivity of the upper limb peripheral nervous system. BMC Muscoskelet. Disord. 10 11-2474-10-11.
Schmid, A.B., Nee, R.J., Coppieters, M.W., 2013. Reappraising entrapment neuropathies–mechanisms, diagnosis and management. Man. Ther. 18 (6), 449–457.
Schmid, A.B., Soon, B.T., Wasner, G., et al., 2012. Can widespread hypersensitivity in carpal tunnel syndrome be substantiated if neck and arm pain are absent? Eur. J. Pain 16 (2), 217–228.
Schneiders, A.G., Sullivan, S.J., Hendrick, P.A., et al., 2012. The ability of clinical tests to diagnose stress fractures: a systematic review and meta-analysis. J. Orthop. Sports Phys. Ther. 42 (9), 760–771.
Schünemann, H.J., Oxman, A.D., Brozek, J., Glasziou, P., Jaeschke, R., Vist, G.E., Williams Jr., J.W., Kunz, R., Craig, J., Montori, V.M., Bossuyt, P., Guyatt, G.H., GRADE Working Group, 2008. Rating quality of evidence and strength of recommendations: GRADE: Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ British Med. J. 336 (7653), 1106.
Song, F., Khan, K., Dinnes, J., Sutton, A., 2002. Asymmetrical funnel pots and publication bias in meta-analyses of diagnostic accuracy. Int. J. Epidemiol. (31), 88e95.
Tamburin, S., Cacciatori, C., Praitano, M.L., et al., 2011. Median nerve small- and large-fiber damage in carpal tunnel syndrome: a quantitative sensory testing study. J. Pain 12 (2), 205–212.
Thoomes, E.J., van Geest, S., van der Windt, D.A., Falla, D., Verhagen, A.P., Koes, B.W., Vleggeert-Lankamp, C.L., 2017. Value of physical tests in diagnosing cervical radiculopathy: a systematic review. Spine J.
Trillos, M.C., Soto, F., Briceno-Ayala, L., 2018. Upper limb neurodynamic test 1 in patients with clinical diagnosis of carpal tunnel syndrome: A diagnostic accuracy study. J. Hand Ther. 31 (3), 333–338.
Vanti, C., Bonfiglioli, R., Calabrese, M., Marinelli, F., Guccione, A., Violante, F.S., Pillastrini, P., 2011. Upper limb neurodynamic test 1 and symptoms reproduction in carpal tunnel syndrome. A validity study. Manual Ther. 16 (3), 258–263.
Vanti, C., Bonfiglioli, R., Calabrese, M., Marinelli, F., Violante, F.S., Pillastrini, P., 2012. Relationship between interpretation and accuracy of the upper limb neurodynamic test 1 in carpal tunnel syndrome. J. Manipulat. Physiol. Ther. 35 (1), 54–63.
Wainner, R.S., Fritz, J.M., Irrgang, J.J., et al., 2003. Reliability and diagnostic accuracy of the clinical examination and patient self-report measures for cervical radiculopathy. Spine 28 (1), 52–62.
Wainner, R.S., Fritz, J.M., Irrgang, J.J., et al., 2005. Development of a clinical prediction rule for the diagnosis of carpal tunnel syndrome. Arch. Phys. Med. Rehabil. 86 (4), 609–618.
Whiting, P.F., Rutjes, A.W., Westwood, M.E., et al., 2011. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann. Intern. Med. 155 (8), 529–536.
Woolf, C.J., 2004. Dissecting out mechanisms responsible for peripheral neuropathic pain: implications for diagnosis and therapy. Life Sci. 74 (21), 2605–2610.