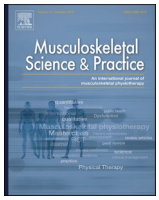
Musculoskeletal Science and Practice 77 (2025) 103317

Contents lists available at ScienceDirect

Musculoskeletal Science and Practice

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



Review article

Diagnostic accuracy of neurodynamic tests in upper-limb entrapment neuropathies: A systematic review and meta-analysis

Daniel Albert-Lucena [a], Marcos José Navarro-Santana [a,b,*], María José Díaz-Arribas [a,b], Gabriel Rabanal-Rodríguez [a], Juan Antonio Valera-Calero [a,b], César Fernández-de-las-Peñas [c,d], Chad Cook [e], Gustavo Plaza-Manzano [a,b]

[a] Department of Radiology, Rehabilitation and Physiotherapy, Faculty of Nursery, Physiotherapy and Podiatry, Complutense University of Madrid, 28040, Madrid, Spain
[b] Grupo InPhysio, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos (IdISSC), 28040, Madrid, Spain
[c] Department of Physical Therapy, Occupational Therapy, Rehabilitation and Physical Medicine, Universidad Rey Juan Carlos, 28922, Alcorcón, Spain
[d] Cátedra Institucional en Docencia, Clínica e Investigación en Fisioterapia: Terapia Manual, Punción Seca y Ejercicio Terapéutico, Universidad Rey Juan Carlos, 28922, Alcorcón, Spain
[e] Department of Orthopaedics, Duke University, Department of Population Health Sciences, Duke Clinical Research Institute, Durham, NC, USA

ARTICLE INFO

Keywords:
Upper limb neurodynamic tests
Neurodynamic
Radiculopathy
Carpal tunnel syndrome
Neuropathic pain
Diagnostic accuracy

ABSTRACT

Background: Upper-limb neurodynamic tests are commonly used to diagnose neuropathies in this area, including cervical radiculopathy and carpal tunnel syndrome, although their diagnostic accuracy remains uncertain across different conditions and criteria.
Objective: To assess the diagnostic accuracy of upper-limb neurodynamic tests and their variations and criteria for upper-limb entrapment neuropathies.
Methods: A systematic review with meta-analysis was conducted in different databases (for their inception in February 2025), including studies evaluating the diagnostic accuracy of these tests. Sensitivity, specificity, likelihood ratios (LR), diagnostic odds ratios, diagnostic accuracy and the area under the curve (AUC) were calculated using a bivariate and univariate meta-analysis. The quality of evidence was evaluated using the GRADE approach, and meta-regression was performed to examine the influence of diagnostic criteria.
Results: Twelve studies were included. Likelihood ratios for neuropathic pain conditions were LR+:1.65 and LR-:0.57, for cervical radiculopathy were LR+:2 and LR-:0.47, and for carpal tunnel syndrome were LR+:1.45 and LR-:0.66. The upper-limb neurodynamic test 2A showed the highest diagnostic accuracy (AUC: 0.76), with LR+:2.59 and LR-:0.42 for cervical radiculopathy, while test 3 had the highest specificity (0.92; LR+:7, LR-:0.48). Diagnostic accuracy for carpal tunnel syndrome was lower (AUC: 0.62). Meta-regression showed significant diagnostic criteria interaction, favoring structural differentiation maneuvers ($p = 0.002$).
Conclusion: Upper-limb neurodynamic tests show moderate sensitivity and low to moderate specificity for diagnosing upper-limb entrapment neuropathies, with diagnostic accuracy varying across conditions. The certainty of evidence ranges from very low to moderate, emphasizing the need for cautious clinical interpretation. Diagnostic reference criteria significantly influence test performance.

1. Introduction

Upper-limb entrapment neuropathies (UEN) arise when nerves are compressed or irritated as they pass through narrow anatomical corridors (Schmid et al., 2020). Examples of UEN are carpal tunnel syndrome (CTS) and cervical radiculopathy (CR). The diagnosis and management of these conditions can be difficult since these are related to a variety of pathomechanisms and clinical presentations (Faktorovich et al., 2021; Schmid et al., 2020). An optimal diagnostic approach likely includes a combination of patient history, clinical tests, and technical investigations such as electrophysiological or imaging studies (Schmid et al., 2013). Clinical examination has gained importance in the social and health care field to shorten diagnostic time and examination-related

* Corresponding author. Pl. de Ramón y Cajal, 3, 28040, Madrid, Spain.
E-mail addresses: daalbert@ucm.es (D. Albert-Lucena), marconav@ucm.es (M.J. Navarro-Santana), mjdiazar@med.ucm.es (M.J. Díaz-Arribas), grabanal@ucm.es (G. Rabanal-Rodríguez), juavaler@ucm.es (J.A. Valera-Calero), cesar.fernandez@urjc.es (C. Fernández-de-las-Peñas), chad.cook@duke.edu (C. Cook), gusplaza@ucm.es (G. Plaza-Manzano).

https://doi.org/10.1016/j.msksp.2025.103317
Received 21 July 2024; Received in revised form 4 March 2025; Accepted 18 March 2025
Available online 22 March 2025
2468-7812/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

List of abbreviations

AUC	Area Under the Curve
CR	Cervical Radiculopathy
CTS	Carpal Tunnel Syndrome
DOR	Diagnostic Odds Ratio
FN	False Negative
FP	False Positive
GRADE	Grading of Recommendations Assessment, Development and Evaluation
LR	Likelihood Ratio
MRI	Magnetic Resonance Imaging
PRISMA-DTA	Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy studies
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
SROC:	Shape of a Receiver Operator Curve
TN	True Negative
TP	True Positive
ULNT	Upper-Limb Neurodynamic Test
UEN	Upper-limb Entrapment Neuropathy

costs (Wainner et al., 2005).

Upper limb neurodynamic tests (ULNTs) were designed to evaluate the mechanosensitivity of the brachial plexus and to test the involvement of neural structures as a possible source of pain (Elvey, 1979). ULNTs consist of a series of movements that elongate the nerve bedding, thereby increasing the pressure in and around peripheral nerves (Butler and Gifford, 1989; Leoni et al., 2016). ULNTs are designed to predominantly stress the nerve associated with each test (median: ULNT1 and ULNT2A; radial: ULNT2B; cubital: ULNT3) at the arm (shoulder, elbow and wrist), signifying that the mechanosensitivity of a specific nerve in proximity to these joints can be best evaluated using its corresponding ULNT (Nee et al., 2012a). It has been described that ULNTs could provoke symptoms that implicate neural tissue sensitivity; however, symptoms are also present in asymptomatic subjects (Covill and Petersen, 2012) and when the patient's pain is not related to central pain mechanisms (Smart et al., 2012; Nee et al., 2012) reported two criteria of a positive ULNT to discriminate patients with UEN from sensory responses on healthy subjects: (1) at least partial reproduction of the patient's symptoms; and (2) a change in these symptoms with structural differentiation, which is defined as the increase or decrease in symptom response with movement of distal segments (e.g. contralateral or ipsilateral sidebending of cervical spine) (Butler and Gifford, 1989; Shacklock, 2005).

Structural differentiation is not routinely performed in many diagnostic utility studies of ULNT (Grondin et al., 2021), despite being important for establishing the focus on neural tissue rather than competing musculoskeletal structures (Shacklock, 2005). Wainner et al. (2005), instead of using structural differentiation maneuvers, defined the following criteria for a positive test (including any one of them): (1) reproduction of the patient's symptoms, (2) side-to-side differences greater than 10° in elbow extension or wrist flexion upon completion of all motion sequences, and (3) on the symptomatic limb side, either contralateral neck side-bending increases symptoms or ipsilateral side-bending decreases symptoms.

ULNTs have been widely incorporated in clinical practice because they are quick and easy to perform, but their usefulness has not been comprehensively studied. Diagnostic utility, which is investigated in a diagnostic accuracy study, determines the evidence of test's abilities to correctly identify or rule out a disease (Mallett et al., 2012). Although there have been several systematic reviews and meta-analyses investigating the diagnostic utility of ULNTs in the last years (Beddaa et al., 2022; De Arenas-Arroyo et al., 2022; Koulidis et al., 2019; Shen et al., 2023), differences in the way the tests are performed and/or interpreted suggest the need for an additional review focusing on these considerations. Elucidation of study findings requires careful examination since the possibility of misclassifying UEN with elevated nerve mechanosensitivity instead of conduction loss is very high (Nee et al., 2012a). Previous systematic reviews have not specifically evaluated the diagnostic accuracy of different neurodynamic tests in UEN and its employed method.

These gaps in the literature call for a more focused meta-analysis to assess the diagnostic acuity of individual neurodynamic tests, including ULNT1, ULNT2a, ULNT2b, and ULNT3. Therefore, the aim of this study is to conduct a systematic review and meta-analysis for analyzing the diagnostic accuracy (sensitivity, specificity, likelihood ratios, diagnostic odds ratio, and diagnostic accuracy) of ULNTs, individually evaluating specific variants for the diagnosis of UEN (CR and CTS).

2. Methods

This systematic review was reported according to Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy studies (PRISMA-DTA)(McInnes et al., 2018). The protocol was based on the Cochrane Handbook for Diagnostic Test Accuracy studies (Leeflang et al., 2013) and registered on Open Science Framework https://doi.org/10.17605/OSF.IO/QF26X.

2.1. Search strategy

PubMed (via MEDLINE), Web of Science, Scopus and the Cochrane Library were systematically searched from inception to March 2023. The following keywords and combinations were used: ("upper limb neurodynamic test" OR "upper-limb neurodynamic test" OR "upper limb tension test" OR "neural tension" OR "neurodynamic" OR "neural provocation") AND ("diagnostic accuracy" OR "validity" OR "diagnostic" OR "predictive value" OR "likelihood ratio") AND ("neuropathic pain" OR "neuropathy" OR "cervical radiculopathy" OR "carpal tunnel syndrome" OR "radial tunnel syndrome" OR "cubital tunnel syndrome" OR "entrapment neuropathy" OR "radicular pain"). The search was augmented by manually revising relevant articles cited in the reference list of included studies (grey literature) and the bibliographic research was extended until February 2025 at the request of the reviewers.

2.2. Eligibility criteria

The recommendations of the Cochrane Handbook for Diagnostic Test Accuracy studies (Leeflang et al., 2013) and the SPIDER search concept (Cooke et al., 2012) were followed for the eligibility criteria, as a previous systematic review addressing a similar topic did (Koulidis et al., 2019).

Inclusion criteria were: sample (S): patients aged>18 years with arm and/or neck symptoms with suspected UEN; phenomenon of interest (PI): accuracy of the upper-limb neurodynamic tests (ULNTs) (detailed explanation of the tests is provided in SUPPLEMENTARY S1); design (D): diagnostic accuracy study; evaluation (E): comparison of the index test (ULNT) to a reference standard (electrophysiologic exam like electromyography or nerve conduction studies or advance imaging via magnetic resonance imaging (MRI)). History and physical examination, in combination with MRI and nerve conduction studies, are the most recommended diagnostic procedures for detecting entrapment disorders such as cervical radiculopathy and carpal tunnel syndrome (Abbed and Coumans, 2007; Padua et al., 2016); research type (R): cross-sectional or cohort studies.

Exclusion criteria included: insufficient data for a 2 × 2 table; case series, case-reports, cadaveric studies, or full text not available. Two independent reviewers (MJNS, DAL) screened titles and abstracts of the identified studies for eligibility using these inclusion/exclusion criteria.

2.3. Reference standards for CR and CTS diagnoses

Reference standards involving nerve conduction velocity or dedicated imaging methods such as MRI were accepted. These have been recommended in previous studies (Alanazy, 2017; Kimura, 1978; Kuijper et al., 2009).

2.4. Data extraction

One reviewer (DAL) extracted data using a pre-determined, self-developed data extraction form. The following data were extracted: author, country and year of publication, study design, eligibility criteria, characteristics of the participants, index tests and reference standard, and accuracy data including sensitivity, specificity and/or likelihood ratios. When values for true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) were not directly reported in the studies, they were calculated using empirical data related to sensitivity and specificity. These data were audited by a second reviewer (MJNS), and, in case of disagreements, a third reviewer (GPM) made the final decision. Collected data were presented in tabular form for additional analysis.

2.5. Methodological quality assessment

Two reviewers (DAL, MJNS) independently evaluated the methodological quality using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool (Reitsma et al., 2011). It assesses four domains: patient selection, index test, reference standard and flow of the participants and timing of the tests. Each domain is marked as "high", "low" or "unclear" risk. A study is rated "at risk" if it has one or more "unclear" and/or "high" rates. The QUADAS-2 also assesses the patient selection, the index test and the reference standard in terms of applicability of the results with "no concerns" or "with concerns". If a study is rated as "with concerns" it has one or more "unclear" and/or "high" ratings. If consensus could not be reached, a third reviewer (GPM) was consulted for a final decision. By using the QUADAS-2, primary diagnostic accuracy studies can be more transparently rated for bias and applicability (Reitsma et al., 2011). Inter-rater agreement between the two investigators for the QUADAS-2 assessment was assessed using Cohen's kappa statistic, yielding a kappa value of 0.85, which indicates a high level of agreement.

2.6. Statistical analysis and data synthesis

The statistical analysis was performed by Meta-DiSc 1.4, Meta-DiSc 2.0, and RevMan 5.4. For analyses involving more than four tests, a binary method was employed, which facilitates the simultaneous examination of multiple diagnostic metrics like sensitivity, specificity, positive and negative likelihood ratios (LR), and diagnostic odds ratios (DORs). Conversely, for analyses involving four or fewer tests, a univariate method was utilized. This simpler approach focuses on analyzing one diagnostic metric at a time (e.g., sensitivity or specificity) and is generally applied when the number of studies is limited, offering a more straightforward statistical framework, less susceptible to the challenges posed by sparse data.

Two-by-two tables were summarized for each index test in each study based on the TP, FP, TN and FN. If the included studies did not provide detailed data, TP, FP, TN, and FN were calculated based on the sensitivity, specificity, and sample size. From the two-by-two tables, we calculated sensitivity, specificity, positive and negative LR, diagnostic odds ratios (DOR), and diagnostic accuracy for individual ULNTs. DORs refer to the ratio between the odds of a positive test result for an individual with the disease and the odds of a positive test result for a person without the disease (Schlattmann, 2023). DORs are used to estimate the discriminative ability of diagnostic procedures and are used to compare the diagnostic accuracies of two or more diagnostic tests within a meta-analysis. It is calculated by the proportion (TP/FN)/(FP/TN) (Šimundić, 2009), where higher values suggest improved discriminatory performance of the test (Schlattmann, 2023; Šimundić, 2009). To estimate the discriminative power of a test (diagnostic accuracy), we used the shape of a receiver operator curve (sROC) and the area under the curve (AUC). A test that achieves perfect discrimination between the two groups will display a curve situated near the upper-left corner of the plot. This implies the absence of FP and FN and is also indicative of a larger area under the receiver operating characteristic curve (AUC). The AUC can range between 0 and 1 and it is a reliable measure of the quality of the test (Schlattmann, 2023; Šimundić, 2009). A test with an AUC value of 0.9–1.0 is considered to have excellent diagnostic accuracy, 0.8–0.89 is considered to have very good diagnostic accuracy, 0.7–0.79 is deemed to offer good diagnostic accuracy, 0.6–0.69 provides sufficient diagnostic accuracy, 0.5–0.59 indicates poor diagnostic accuracy, and any value under 0.5 suggests the test is not useful (Šimundić, 2009).

According to Schneiders et al. (2012) there are no uniform criteria for interpretation of sensitivity and specificity, and the authors arbitrarily classified values used a rating scale ("low": <0.50; "low to moderate": 0.51–0.64; "moderate": 0.65–0.74; "moderate to high": 0.75–0.84; "high": >0.85). There is also an arbitrary classification-boundary from Jaeschke et al. (1994) for the values of the LR about the discriminatory properties of the tests: Conclusive evidence (LR+ >10 and LR- <0.1), strong diagnostic evidence (LR+ 5–10 and LR- 0.1-0.2), weak diagnostic evidence (LR+ 2–5 and LR- 0.2-0.5) and negligible evidence (LR+ 1–2 and LR- 0.5-1).

A challenge when summarizing diagnostic studies is that the findings may differ depending on methodological and/or clinical differences. For clinical comparison, the study design, population, comparable diagnostic data, and reference standard were considered. The evaluation of heterogeneity aimed to determine whether the studies were appropriate for conjoining in an analysis (Lijmer et al., 2002). Heterogeneity was calculated with the I^2 using a bivariate (I^2_{BIV}) or a univariate (I^2_{UNI}) method (Zhou and Dendukuri, 2014). Values of I^2 range between 0 and 1 where higher values indicate higher heterogeneity across studies than within-study sampling error. Higgins et al. (Higgins and Thompson, 2002) proposed a classification of I^2 values with the purpose of helping to interpret its magnitude. Thus, $I^2 < 0.50$ is considered low, I^2 between 0.5 and 0.75 is considered moderate, and $I^2 > 75$ is considered high heterogeneity (Higgins and Thompson, 2002).

2.7. Certainty of the evidence

Certainty of the evidence for diagnostic accuracy of the ULNTs was evaluated using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach (Schünemann et al., 2008). According to the GRADE approach, the level of evidence can be classified as high, moderate, low or very low quality based on the QUADAS-2, indirectness of evidence, inconsistency of results or unexplained heterogeneity, imprecision of results and high probability of publication bias (Austin et al., 2014). The certainty of the evidence is defined as the confidence we have that the estimate of an effect is adequate to make a recommendation, and it is downgraded depending on the items mentioned before (for example: lack of randomization, wide variability of the results, large confidence intervals or small sample sizes) (Aguayo-albasini et al., 2016). This method was independently performed by two authors, with participation of a third when discrepancy occurred. An agreement of 85 % on quality assessment/evidence level between the two authors was observed.

3. Results

3.1. Study identification

During the systematic search, 966 studies were identified. After reviewing their titles and abstracts, only 21 studies were considered for

further evaluation. Out of these, nine studies were excluded for failing to meet the eligibility criteria. Twelve (12) studies (Apelby-Albrecht et al., 2013; Beddaa et al., 2022; Bueno-Gracia et al., 2016, 2024; Ghasemi et al., 2013; Grondin et al., 2021; Sleijser-Koehorst et al., 2021; Trillos et al., 2018; Vanti et al., 2011, 2012; Wainner et al., 2003, 2005) were considered for analysis. Fig. 1 shows the identification process.

3.2. Characteristics of the studies

Details of the included studies are summarized in Table 1. Five of twelve (41.66 %) studies targeted suspected CR (Apelby-Albrecht et al., 2013; Ghasemi et al., 2013; Grondin et al., 2021; Sleijser-Koehorst et al., 2021; Wainner et al., 2003) whereas seven of twelve (58.33 %) were targeted for patients with suspected CTS (Beddaa et al., 2022; Bueno-Gracia et al., 2016, 2024; Trillos et al., 2018; Vanti et al., 2011, 2012; Wainner et al., 2005). All the included studies used ULNT as an index test.

3.3. Cervical radiculopathy

Three studies were labeled as "low" concerns ((Apelby-Albrecht et al., 2013; Grondin et al., 2021; Wainner et al., 2003) and two (Ghasemi et al., 2013; Wainner et al., 2003) had concerns regarding applicability. All the studies (Apelby-Albrecht et al., 2013; Ghasemi et al., 2013; Grondin et al., 2021; Sleijser-Koehorst et al., 2021; Wainner et al., 2003) included the evaluation of the ULNT1 (median nerve) in their assessments. Two (Apelby-Albrecht et al., 2013; Grondin et al., 2021) also evaluated the ULNT2A, three (Apelby-Albrecht et al., 2013; Grondin et al., 2021; Wainner et al., 2003) evaluated the ULNT2B (radial nerve) and two (Apelby-Albrecht et al., 2013; Grondin et al., 2021) evaluated the ULNT3 (cubital nerve). Only one evaluated the combination of all ULNT tests with high sensitivity (0.97, 0.85–1.00) and moderate specificity (0.69, 0.41–0.89)(Apelby-Albrecht et al., 2013). Three studies used a blend of patient history, clinical examination, and MRI results as the reference standard (Apelby-Albrecht et al., 2013; Grondin et al., 2021; Sleijser-Koehorst et al., 2021) whereas two studies used nerve conduction study (NCS) and needle electromyography (Ghasemi et al., 2013; Wainner et al., 2003). (Table 2)

Criteria for a positive ULNT varied across the studies. Three studies used Nee's criteria (as mentioned before)(Apelby-Albrecht et al., 2013; Grondin et al., 2021; Sleijser-Koehorst et al., 2021) whereas Wainner et al. considered the Wainner's criteria (Wainner et al., 2003). Ghasemi et al. considered "reproduction of pain in any step" for its criterion (Ghasemi et al., 2013).

3.4. Carpal tunnel syndrome

Five studies were considered with low risk ((Beddaa et al., 2022; Bueno-Gracia et al., 2024; Trillos et al., 2018; Vanti et al., 2012; Wainner et al., 2005) and only three did not have concerns of applicability (Beddaa et al., 2022; Bueno-Gracia et al., 2016, 2024). Six studies used a combination of patient's history plus clinical examination and NCS as the reference standard (Beddaa et al., 2022; Bueno-Gracia et al., 2024; Trillos et al., 2018; Vanti et al., 2011, 2012; Wainner et al., 2005) while the remain study only used NCS (Bueno-Gracia et al., 2016). Six studies included ULNT1 in their analyses (Bueno-Gracia et al., 2016, 2024; Trillos et al., 2018; Vanti et al., 2011, 2012; Wainner et al., 2005), one included the ULNT2A (Beddaa et al., 2022) and other included the ULNT2B (Wainner et al., 2005). (Table 3)

As we see in CR, criteria for a positive test varied across the studies.

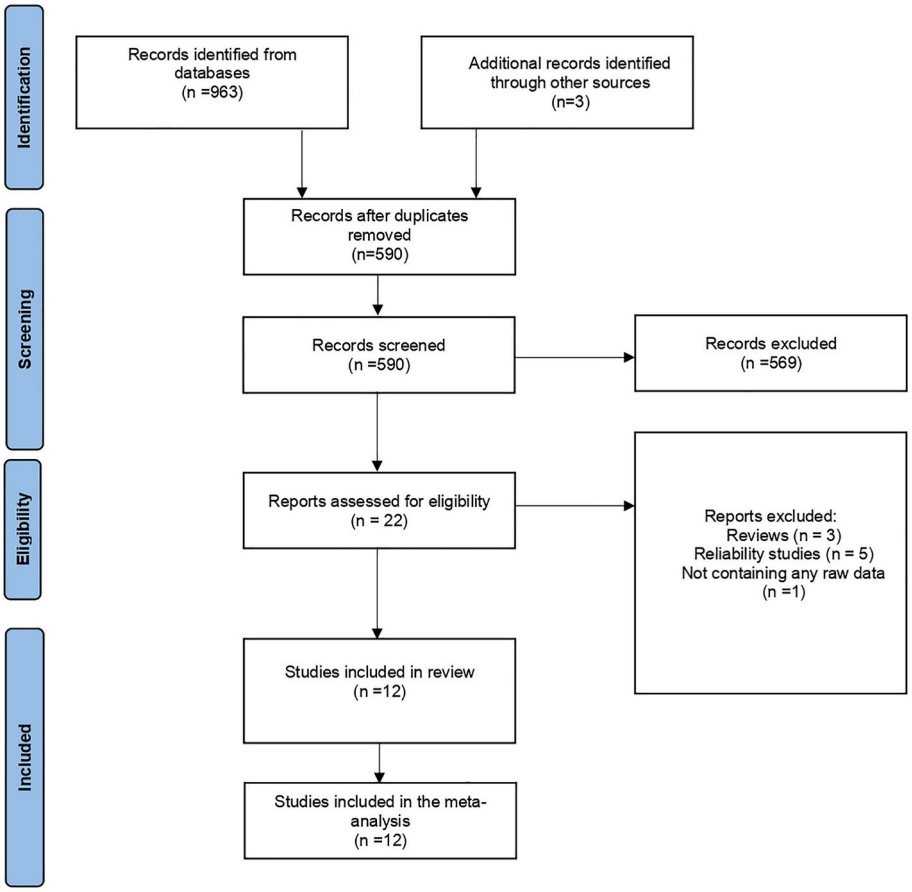


Fig. 1. Flow diagram.

Table 1
Characteristics of the included studies.

Author (year), country	Clinical Setting	Study Design	Target Condition	Participants (number, age, gender)	Index test	Reference Standard	Risk of Bias
Apelby-Albrecht et al. (2013) Sweeden	Center for spinal surgery	Prospective cohort study	CR	N = 58 Woman = 27 Age (mean) = 51 (21–63)	ULNT1 (median) ULNT2A (median) ULNT2B (radial) ULNT3 (cubital) and Combined	Patient History, clinical examination and MRI findings	Low
Grondin et al. (2021) France	Neurosurgery Department	Prospective cohort study	CR	N = 85 Woman = Not reported Age (mean) = 44.61	ULNT1 (median) ULNT2A (median) ULNT2B (radial) ULNT3 (cubital)	Patient history, clinical examination and MRI findings	Low
Wainner et al. (2003) USA	University of Pittsburgh, Wilford Hall USAF Medical Center, Brooke Army Medial Center, and Blanchfield Army Community Hospital	Prospective cohort study	CR	N = 82 Woman = 41 Age (mean) = 48.21 (8)	ULNT 1 (median) ULNT 2B (radial)	Needle EMG and Nerve conduction study	Low
Ghasemi et al. (2013) Iran	Kashani hospital of Isfahan	Cross-sectional study	CR	N = 97 Woman = 72 Age (mean) = 46.14	ULNT 1 (Median)	Nerve conduction study	At risk
Sleijser-Koehorst et al. (2021) Nertherlands	Multidisciplinary clinic	Prospective cohort study	CR	N = 134 Woman = 69 Age (mean) = 49.9 (10.7)	ULNT 1 (Median)	Clinical presentation and MRI findings.	At risk
Beddaa et al. (2022) Morocco	Rabat Specialty Hospital	Prospective cohort study	CTS	N = 94 Woman = 94 Age (mean) = 48.87	ULNT 2A (median)	Nerve conduction study	Low
Wainner et al. (2005) USA	Multicenter medical center and community hospital	Prospective cohort study	CTS	N = 82 Woman = 41 Age (mean) = 45 (13)	ULNT 1 (Median) ULNT 2B (radial)	Nerve conduction study	Low
Vanti et al. (2012) Italy	Department of Internal Medicine, Geriatrics and Nephrology, Alma Mater Studiorum, University of Bologna	prospective cohort study	CTS	N = 47 W oman = 35 Age (mean) = 45.91	ULNT 1 (Median)	Nerve conduction study	Low
Vanti et al. (2011) Italy	Clinic of Occupational Medicine of the University of Bologna (Italy)	Prospective cohort study	CTS	N = 44 Woman = 33 Age (mean) = 46.3 (10.8)	ULNT 1 (Median)	Nerve conduction study	At risk
Bueno-Gracia et al. (2016) Spain	Not reported	Prospective cohort study	CTS	N = 58 Woman = 42 Age (mean) = 54.3 14.5	ULNT 1 (median)	Nerve conduction study	At risk
Trillos et al. (2018) Colombia	Health institution	Prospective cohort study	CTS	N = 118 Woman = 98 Age (mean) = 50.51	ULNT 1 (Median)	Clinical examination and Nerve conduction study	Low
Bueno-Gracia et al. (2024) Spain	Neurophysiology Department	Prospective cohort study	CTS	N = 58 Woman = 42 Age (mean) = 54.6	ULNT1 (Median)	Clinical presentation and nerve conduction study	Low

Abbreviatures: CR: cervical radiculopathy; CTS: carpal tunnel syndrome; ULNT: upper-limb neurodynamic test.

Five studies (Beddaa et al., 2022; Bueno-Gracia et al., 2016, 2024; Vanti et al., 2011, 2012) used Nee's criteria while the remaining two (Trillos et al., 2018; Wainner et al., 2005) used Wainner's criteria.

3.5. Methodological quality of included studies

Two reviewers (DAL, MJNS) achieved complete consensus on the assessment of QUADAS-2 approach through discussion. The results are shown in Table 4. Eight studies (75 %) were assessed as "low risk" of bias (Apelby-Albrecht et al., 2013; Beddaa et al., 2022; Bueno-Gracia et al., 2024; Grondin et al., 2021; Trillos et al., 2018; Vanti et al., 2012; Wainner et al., 2003, 2005). Patient selection procedure (16.66 %) and flow and timing (25 %) were the main reasons for bias in the four studies with "high risk" (Bueno-Gracia et al., 2016; Ghasemi et al., 2013; Sleijser-Koehorst et al., 2021; Vanti et al., 2011). In patient selection domain, two studies (Bueno-Gracia et al., 2016; Ghasemi et al., 2013) did not specify if consecutive or random sampling was used. Also, in flow and timing domain, three studies (Ghasemi et al., 2013; Sleijser--Koehorst et al., 2021; Vanti et al., 2011) did not include all patients in the analysis. The index test domain had high concerns in only one study (Sleijser-Koehorst et al., 2021), as it did not specify the threshold used beforehand. Six studies (50 %) were catalogued as "No concern" regarding the applicability (Apelby-Albrecht et al., 2013; Beddaa et al., 2022; Bueno-Gracia et al., 2016, 2024; Grondin et al., 2021; Sleijser--Koehorst et al., 2021). The rest of the studies (50 %) (Ghasemi et al., 2013; Trillos et al., 2018; Vanti et al., 2011, 2012; Wainner et al., 2003, 2005) had concerns of applicability in the index test domain. This is justified because Nee's criteria for a positive ULNT were considered when interpreting the index test domain: at least partial reproduction of the patient's symptoms and a change in these symptoms with structural differentiation (Nee et al., 2012a). Other criteria (Wainner's criteria and others) were considered as "other criteria" since structural differentiation was not incorporated in the interpretation of the tests. Proportions of QUADAS-2 assessment are shown in Fig. 2.

Table 2
Diagnostic accuracy data for CR.

Author (year)	Test and positive criteria	TP	FN	FP	TN	SEN (95 % CI)	SP (95 % CI)	LR+ (95 % CI)	LR- (95 % CI)	DOR (95 % CI)
Apelby-Albrecht et al. (2013)	Nee's criteria									
	ULNT1	29	6	4	12	0.83 (0.66–0.93)	0.75 (0.48–0.93)	3.32 (1.40–7.85)	0.22 (0.10–0.50)	14.5 (3.46–60.78)
	ULNT2A	23	12	4	12	0.66 (0.48–0.81)	0.75 (0.48–0.93)	2.64 (1.09–6.35)	0.45 (0.28–0.75)	5.75 (1.52–21.73)
	ULNT2B	15	20	4	12	0.43 (0.26–0.61)	0.75 (0.48–0.93)	1.72 (0.68–4.35)	0.76 (0.55–1.06)	2.25 (0.60–8.38)
	ULNT3	25	10	2	14	0.71 (0.54–0.85)	0.87 (0.62–0.98)	5.68 (1.54–21.24)	0.32 (0.19–0.56)	17.5 (3.35–91.41)
	Combined	34	1	5	11	0.97 (0.85–1.00)	0.69 (0.41–0.89)	3.11 (1.50–6.44)	0.04 (0.02–0.04)	74.8 (7.88–711.18)
Grondin et al. (2021)	Nee's criteria									
	ULNT1	50	35	21	64	0.59 (0.39–0.78)	0.75 (0.63–0.86)	2.46 (1.41–4.27)	0.54 (0.33–0.87)	4.35 (2.26–8.39)
	ULNT2A	60	25	23	62	0.70 (0.5–0.86)	0.73 (0.59–0.83)	2.55 (1.57–4.14)	0.41 (0.22–0.75)	6.47 (3.32–12.62)
	ULNT2B	47	38	21	64	0.65 (0.35–0.75)	0.75 (0.63–0.86)	2.30 (1.30–4.06)	0.59 (0.38–0.92)	5.55 (2.78–11.1)
	ULNT3	35	50	6	79	0.41 (0.22–0.61)	0.93 (0.83–0.98)	5.91 (2.07. 16.87)	0.64 (0.46–0.88)	9.21 (3.61–23.5)
Wainner et al. (2003)	Wainner's criteria									
	ULNT1	80	2	64	18	0.97 (0.90–1.0)	0.22 (0.12–0.33)	1.3 (1.1–1.5)	0.12 (0.01–1.9)	4.095 (1.97–8.50)
	ULNT2B	59	23	55	27	0.72 (0.52–0.93)	0.33 (0.21–0.45)	1.1 (0.77–1.5)	0.85 (0.37–1.9)	10.25 (2.52–50.29)
Ghasemi et al. (2013)	Reproduction of pain in any step									
	Acute CR-ULNT1	59	38	58	39	0.61 (0.50–0.71)	0.40 (0.30–0.50)	1.02 (0.81–1.28)	0.98 (0.69–1.38)	1.04 (0.59–1.86)
	Chronic CR-ULNT1	34	63	58	39	0.35 (0.26–0.45)	0.40 (0.30–0.50)	0.59 (0.43–0.80)	1.61 (1.21–2.14)	0.36 (0.20–0.65)
Sleijser-Koehorst et al. (2021)	Nee's criteria									
	ULNT1	43	21	22	44	0.67 (0.54–0.78)	0.67 (0.54–0.78)	1.95 (1.33–2.86)	0.48 (0.33–0.69)	1.26 (0.65–2.45)

CR: cervical radiculopathy; TP: true positive; FN: false negative; FP: false positive; TN: true negative; SEN: sensitivity; SP: specificity; LR: likelihood ratio; DOR: diagnostic odds ratio; CI: confidence intervals; ULNT: Upper limb neurodynamic test.

Table 3
Diagnostic accuracy data for CTS.

Author (year)	Test and positive criteria	TP	FN	FP	TN	SEN (95 % CI)	SP (95 % CI)	LR+ (95 % CI)	LR- (95 % CI)	DOR (95 % CI)
Beddaa et al. (2021)	Nee's criteria									
	ULNT2A	91	33	34	30	0.73 (0.62–0.84)	0.47 (0.17–0.78)	1.38 (1.12–1.65)	0.57 (0.21–1.15)	2.43 (1,29-4,58)
Vanti et al. (2011)	Wainner's criteria									
	ULNT1	22	2	17	3	0.92 (0.74–0.98)	0.15 (0.05–0.36)	1.08 (0.38–3.08)	0.56 (0.19–1.59)	1,94 (0.29–12.95)
	Criterion B: presence of symptoms reproduction in the first, second or third digit of the affected arm.									
	ULNT1	13	11	6	14	0.54 (0.35–0.72)	0.70 (0.48–0.85)	1.81 (1.13–2.88)	0.65 (0.41–1.04)	2,76 (0.79–9.61)
Trillos et al. (2018)	Wainner's criteria									
	ULNT1	186	14	28	2	0.93 (88.21–96.8)	0.068 (0.0–33.59)	1.00 (0.90-1-10)	1.05 (0.25–4.89)	0.95 (0.21–4.4)
Bueno-Gracia et al. (2016)	Nee's criteria									
	ULNT1	33	24	6	32	0.50 (0.45–0.71)	0.84 (0.72–0.96)	3.67 (1.70–7.89)	0.50 (0.36–0.70)	5.18 (1.91–14)
	Criterion B: symptoms appeared at the wrist or the first three digits of the affected hand and changed during SD regardless of the reproduction of patient's clinical symptoms.									
	ULNT1	42	15	19	19	0.74 (0.61–0.83)	0.50 (0.35–0.65)	1.47 (1.03–2.10)	0.53 (0.31–0.90)	2.8 (1.18–6.66)
Wainner et al. (2005)	Wainner's criteria									
	ULNT1	62	20	72	10	0.75 (0.58–0.92)	0.12 (0.04–0.22)	0.86 (0.67–1.1)	1.9 (0.72–5.1)	0.43 (0.19–1)
	ULNT2B	53	29	55	27	0.64 (0.45–0.83)	0.33 (0.17–0.42)	0.91 (0.65–1.3)	1.2 (0.62–2.4)	0.90 (0.47–1.71)
Vanti et al. (2012)	ULNT 1									
	Criterion A: symptoms in fingers I, II, or III.	14	21	10	39	0.40 (0.26–0.56)	0.80 (0.66–0.75)	1.96 (1.28–3.01)	0.75 (0.49–1.16)	2.6 (1–6.86)
	Criterion B: Criterion A + symptoms increased with contralateral cervical side bending.	10	25	8	37	0.29 (0.14–0.45)	0.82 (0.69–0.91)	1.61 (0.93–2.76)	0.87 (0.51–1.49)	1.85 (0.64–5.33)
	Criterion C: Criterion A + symptoms decreased with ipsilateral cervical side bending.	2	33	3	42	0.05 (0.02–0.19)	0.93 (0.82–0.98)	0.86 (0.22–3.3)	1.01 (0.26–3.89)	0.85 (0.13–5.38)
Bueno-Gracia et al. (2024)	Nee's Criteria									
	ULNT1	23	12	1	22	0.66 (0.48–0.81)	0.96 (0.78–0.99)	15.1 (2.2–104.3)	0.36 (0.22–0.57)	42.17 (5.05–352.03)

CTS: carpal tunnel syndrome; TP: True positive; FN: false negative; FP: false positive; TN: true negative; SEN: Sensitivity; SP: specificity; LR: Likelihood ratio; DOR: diagnostic odds ratio; CI: confidence intervals; ULNT: upper limb neurodynamic test; SD: structural differentiation.

Table 4
QUADAS-2 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	
Grondin et al. (2021)	Low	Low	Low	Low	**Low risk**	Low concern	Low concern	Low concern	No concern
Beddaa et al. (2021)	Low	Low	Low	Low	**Low Risk**	Low concern	Low concern	Low concern	No concern
Sleijser-Koehorst et al. (2021)	Low	High	Low	High	**At risk**	Low concern	Low concern	Low concern	No concern
Bueno-Gracia et al. (2016)	High	Low	Low	Low	**At risk**	Low concern	Low concern	Low concern	No concern
Bueno-Gracia et al. (2024)	Low	Low	Low	Low	**Low Risk**	Low concern	Low concern	Low concern	No concern
Apelby-Albrecht et al. (2013)	Low	Low	Low	Low	**Low risk**	Low concern	Low concern	Low concern	No concern
Trillos et al. (2018)	Low	Low	Low	Low	**Low risk**	Low concern	High concern	Low concern	With concern
Ghasemi et al. (2013)	High	Low	Unclear	High	**At risk**	Low concern	High concern	Low concern	With concern
Vanti et al. (2012)	Low	Low	Low	Low	**Low Risk**	Low concern	High concern	Low concern	With concern
Vanti et al. (2011)	Low	Low	Low	High	**At risk**	Low concern	High concern	Low concern	With concern
Wainner et al. (2005)	Low	Low	Low	Low	**Low risk**	Low concern	High concern	Low concern	With concern
Wainner et al. (2003)	Low	Low	Low	Low	**Low risk**	Low concern	High concern	Low concern	With concern

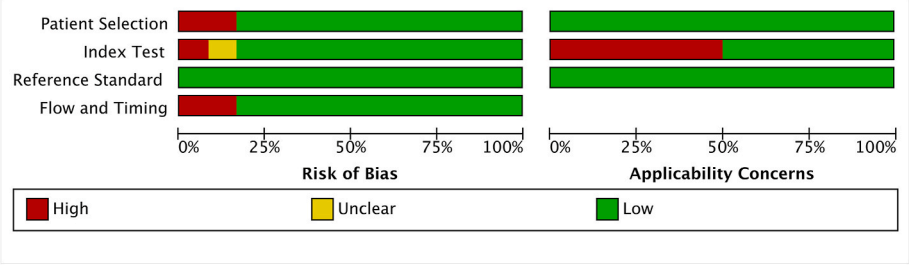


Fig. 2. QUADAS-2 assesment.

3.6. Meta-analysis (Table 5)

3.6.1. Upper-limb entrapment neuropathy conditions

The meta-analysis of different ULNTs for the diagnoses of UEN (CR or CTS) showed moderate sensitivity (0.66 [95 % CI: 0.55–0.75]) and low to moderate specificity (0.62 [95 % CI: 0.48–0.74]) with a high heterogeneity (I^2_{Biv} = 0.87). The discriminatory properties of the test showed negligible evidence (LR+: 1.73 [95 % CI: 1.27–2.20, LR-: 0.55 [95 % CI: 0.42–0.68]) and dOR 3.17 (95 % CI: 2.02–4.58). The AUC showed sufficient diagnostic accuracy (AUC:0.67; SE:0.03) (Fig. 3) (see Table 5).

ULNT1 showed moderate sensitivity (0.65 [95 % CI: 0.49; 0.79]) with negligible evidence (LR+: 2.58 [95 % CI: 1.44–4.60, LR-: 0.60 [95 % CI: 0.43–0.83]) and ULNT3 showed the highest specificity (0.92 [95 % CI: 0.85; 0.96]) with strong diagnostic evidence (LR+: 7 [95 % CI: 4.59–15.19, LR-: 0.48 [95 % CI: 0.29–0.80]) but it was only evaluated in CR. Combined tests were evaluated in one study (Apelby-Albrecht et al., 2013) and showed high sensitivity (0.97 [95 % CI: 0.85; 0.99)]) and moderate specificity (0.69 [95 % CI: 0.41; 0.89)]). ULNT2a demonstrated a higher diagnostic accuracy (AUC: 0.76; SE: 0.03) in comparison with other ULNTs (AUC 0.6 to <0.7). When each ULNT was analyzed in isolation they showed a sufficient diagnostic accuracy (AUC: 0.6 to <0.7). The certainty of evidence using the GRADE approach was rated as very low for the general diagnosis of neuropathy conditions and ULNT1, while for ULNT2A it was rated as moderate, and for ULNT Combined as low. The certainty of evidence was downgraded due to high heterogeneity (I^2 >70 %) across studies (ULNTs, ULNT1, ULNT2a, ULNT2b and ULNT3), concerns related to the QUADAS assessment (ULNTs, ULNT1 and ULNT2b) and imprecision (ULNT2a, ULNT2b, ULNT3 and ULNT combined). The complete GRADE evaluation is presented in Table 6 and a summary of the findings with further explanations and interpretation is provided in SUPPLEMENTARY S.2.

3.6.2. Cervical radiculopathy

The meta-analysis of different ULNTs for the diagnoses of CR showed moderate sensitivity (0.69 [95 % CI: 0.54; 0.81]) and specificity (0.65 [95 % CI: 0.48; 0.79)) with a high heterogeneity (I^2_{Biv} = 0.86 %). The discriminatory properties of the test showed negligible evidence (LR+: 1.99 [95 % CI: 1.42; 2.81) LR-: 0.47 [95 % CI: 0.33; 0.69]) and dOR 4.22 (95 % CI: 2.24; 7.97). The AUC showed good diagnostic accuracy (AUC: 0.73 SE: 0.05). For individual tests, ULNT2a showed weak diagnostic evidence (LR+: 2.59 [95 % CI: 1.83–3.65, LR-: 0.42 [95 % CI: 0.31–0.56]) and dOR 6.15 (95 % CI: 3.42–11.06). As we mentioned before, ULNT3 showed strong diagnostic evidence (LR+: 7 [95 % CI: 4.59–15.19, LR-: 0.48 [95 % CI: 0.29–0.80]). The certainty of evidence using the GRADE approach was rated as very low for the general diagnosis of cervical radiculopathy and ULNT1, while for ULNT2A and ULNT Combined it was rated as low. The certainty of evidence was downgraded due to high heterogeneity (I^2 >70 %) across studies (ULNTs, ULNT1, ULNT2a, ULNT2b and ULNT3), concerns related to the QUADAS

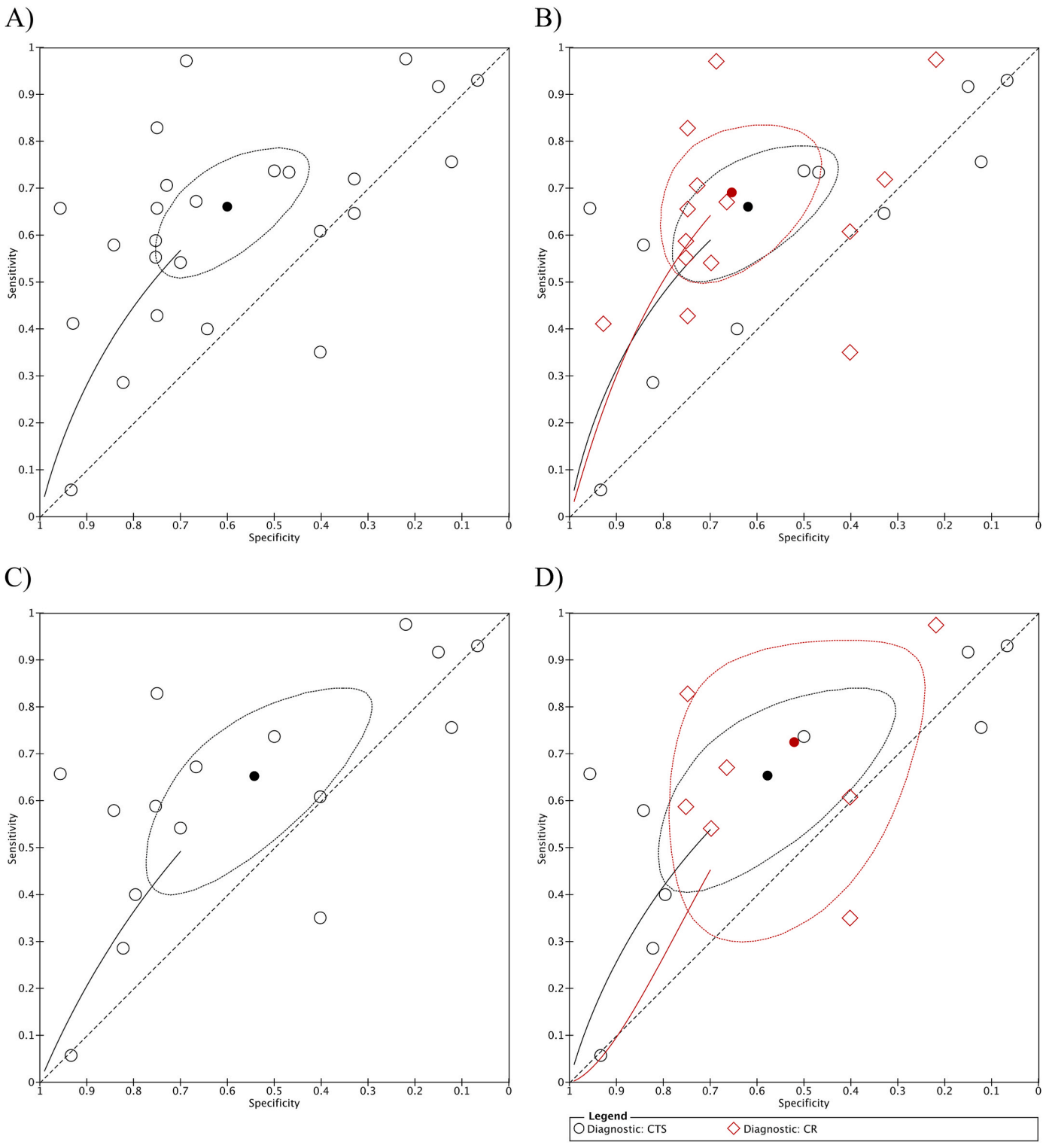


Fig. 3. sROC curve plot. A)ULNTs without subgroup B) ULNTs with subgroup C) ULNT1 without subgroups D) ULNT1 with subgroups.

assessment (ULNTs, ULNT1 and ULNT2b) and imprecision (ULNT2a, ULNT2b, ULNT3 and ULNT combined) (see Table 6 and SUPPLEMENTARY S.2.)

3.6.3. Carpal Tunnel Syndrome

The meta-analysis of different ULNTs for the diagnoses of CTS showed low to moderate sensitivity (0.63 [95 % CI: 0.45; 0.77]) and low to moderate specificity (0.57 [95 % CI: 0.33; 0.78]) with a high heterogeneity ($I^2_{Biv} = 0.82$). The discriminatory properties of the test showed negligible evidence (LR+: 1.44 [95 % CI: 0.99; 2.10]; LR-: 0.66 [95 % CI: 0.51; 0.86]). The AUC showed bad diagnostic accuracy (AUC:0.62 SE:0.04). The dOR value was 2.20 (95 % CI: 1.23–3.91). Also, ULNT1 showed bad diagnostic accuracy (AUC:0.63; SE:0.04). Other AUC for ULNT was not calculated due to lack of studies (n < 3). The interpretation of values between the different ULNT did not differ, all ULNT showed a moderate sensitivity (0.62–0.73) and low to moderate

Table 5
Metanalysis of diagnostic accuracy for ULNT.

Diagnostic	Test	N/T	SEN (95 % CI)	SP (95 % CI)	LR+ (95 % CI)	LR- (95 % CI)	DOR (95 % CI)	I^2	AUC [SE]
UEN	General	12/26	0.66 (0.55; 0.75)	0.62 (0.48; 0.74)	1.73 (1.27; 2.20)	0.55 (0.42; 0.68)	3.17 (2.02; 4.58)	0.87	0.67 [0.03]
	ULNT1	11/16	0.65 (0.49; 0.79)	0.58 (0.38; 0.72)	1.55 (1.11; 2.15)	0.60 (0.43; 0.83)	2.58 (1.44; 4.60)	0.87	0.64 [0.04]
	ULNT2A[a]	3/3	0.71 (0.65; 0.77)	0.64 (0.48; 0.78)	2 (1.29; 3.1)	0.45 (0.33; 0.61)	4.47 (2.16; 9.27)	0 (Sen) 0.82 (Spe)	0.76 [0.03]
	ULNT2B[a]	4/4	0.60 (0.50; 0.70)	0.54 (0.32; 0.74)	1.30 (0.78; 2.20)	0.74 (0.46; 1.21)	1.86 (0.94; 3.66)	0.71 (Sen) 0.93 (Spe)	0.64 [0.06]
	ULNT3[a]	2/2	0.55 (0.34; 0.75)	0.92 (0.85; 0.96)	7 (3.23; 15.19)	0.48 (0.30; 0.80)	14.46 (4.59; 45.58)	0.88 (Sen) 0 (Spe)	NA
	ULNT Combined	1/1	0.97 (0.85; 1)	0.69 (0.41; 0.89)	3.11 (1.5; 6.44)	0.04 (0.01; 0.30)	74.8 (7.87; 711.18)	NA	NA
CR	General	5/14	0.69 (0.54; 0.81)	0.65 (0.48; 0.79)	2 (1.42; 2.81)	0.47 (0.33; 0.69)	4.22 (2.24; 7.97)	0.864	0.727 [0.05]
	ULNT1	5/6	0.73 (0.48; 0.88)	0.52 (0.35; 0.69)	1.51 (1.04; 2.19)	0.53 (0.26; 1.07)	2.87 (1.04; 7.93)	0.894	0.64 [0.1]
	ULNT2A[a]	2/2	0.69 (0.60; 0.77)	0.73 (0.64; 0.81)	2.59 (1.83; 3.65)	0.42 (0.31; 0.56)	6.15 (3.42; 11.06)	0 (sen) 0 (Spe)	NA
	ULNT2B[a]	3/3	0.58 (0.45; 0.71)	0.61 (0.36; 0.82)	1.50 (0.76; 2.95)	0.68 (0.41; 1.14)	2.20 (0.68; 7.12)	0.79 (Sen) 0.94 (Spe)	0.66 [0.08]
	ULNT3[a]	2/2	0.55 (0.34; 0.75)	0.92 (0.85; 0.96)	7 (3.23; 15.19)	0.484 (0.29; 0.80)	14.46 (4.59; 45.58)	0.88 (Sen) 0 (Spe)	NA
	ULNT Combined	1/1	0.97 (0.85; 1)	0.69 (0.41; 0.89)	3.11 (1.5; 6.44)	0.04 (0.01; 0.30)	74.8 (7.87; 711.18)	NA	NA
CTS	General	7/12	0.62 (0.45; 0.77)	0.57 (0.33; 0.78)	1.45 (0.99; 2.11)	0.66 (0.51; 0.86)	2.20 (1.23; 3.90)	0.82	0.62 [0.04]
	ULNT1	6/10	0.61 (0.40; 0.79)	0.61 (0.32; 0.83)	1.55 (0.96; 2.51)	0.64 (0.47; 0.87)	2.43 (1.22; 4.85)	0.81	0.63 [0.04]
	ULNT2A	1/1	0.73 (0.65; 0.81)	0.47 (0.34; 0.60)	1.38 (1.07; 1.78)	0.57 (0.38; 0.84)	2.43 (1.29; 4.58)	NA	NA
	ULNT2B	1/1	0.65 (0.53; 0.75)	0.33 (0.23; 0.44)	0.96 (0.77; 1.20)	1.07 (0.70; 1.64)	0.90 (0.47; 1.71)	NA	NA

[a] Univariate model was used for calculation. ULNT: Upper limb neural test; NA: Not Applicable; CI: Confidence Interval; N: Number of studies; T: Number of test included in the metanalysis; SEN: Sensibility; SP: Specificity; LR: Likelihood Ratio; DOR: Diagnostic Odds Ratio; AUC: Area Under the Curve; UEN: Upper-limb entrapment neuropathies; CR: Cervical Radiculopathy; CTS: Carpal Tunnel Syndrome.

specificity (0.32–0.52). The certainty of evidence using the GRADE approach for CTS diagnosis and ULNT1 was rated as very low, while for ULNT2A and ULNT2B it was rated as low. The certainty of evidence was downgraded due to high heterogeneity (I^2 >70 %) across studies (ULNTs, ULNT1, ULNT2a and ULNT2b), concerns related to the QUADAS assessment (ULNTs, ULNT1 and ULNT2b) and imprecision (ULNT2a and ULNT2b) (see Table 6 and SUPPLEMENTARY S.2.)

3.6.4. *Meta-regression (Table 7)*

Meta-regression did not show a statistically significant relationship by diagnostic (CR or CTS) ($p = 0.35$) or methodological quality ($p = 0.31$) but showed a statistical interaction by diagnostic test criteria ($p < 0.01$). Reproduction of any symptom during test or difference in range of motion without structural differentiation showed a moderate to high sensitivity (0.73 [95 % CI: 0.57–0.85]) and low specificity (0.32 [95 % CI: 0.21–0.46]) and using structural differentiation showed low to moderate sensitivity (0.61 [95 % CI: 0.47–0.73]) and moderate to high specificity (0.78 [95 % CI: 0.68–0.85]) and a weak diagnostic evidence (LR+: 2.71 [95 % CI: 2.05; 3.59] and LR-: 0.50 [95 % CI: 0.38; 0.67]).

4. Discussion

The aim of this research was to analyze the usefulness of ULNTs as a diagnostic tool for upper-limb entrapment neuropathies such as CR and CTS confirmed using a reference standard method of diagnostic imaging or NCS. We gave careful attention to both the quality of the work and how the study performed and interpreted the test and adopted the method advocated by Nee and colleagues as the recommended process. In general, ULNTs showed very low quality of evidence for a moderate sensitivity and low to moderate specificity with negligible diagnostic evidence and sufficient diagnostic accuracy. This is the first meta-analysis estimating diagnostic performance of ULNTs for CR and for CTS.

Furthermore, the present meta-analysis is the first study analyzing individually the diagnostic accuracy of specific neurodynamic tests variants in CR or CTS. As hypothesized, the results showed significant diagnostic accuracy differences among the variants. For instance, while ULNT3 showed a high sensitivity, ULNT2a demonstrated moderate specificity. Similarly, ULNT2b had a low specificity. This variation suggests that no single ULNT is universally superior and the choice of ULNTs should be tailored to the specific diagnostic context.

4.1. *Cervical radiculopathy*

The meta-analysis showed a good diagnostic accuracy of ULNTs for CR, being slightly superior to UEN, with moderate sensitivity and moderate specificity. Consistent with our findings, previous systematic reviews (Koulidis et al., 2019; Shen et al., 2023) found ULNTs provided some diagnostic utility for ruling out CR. However, Koulidis et al. did not assess the diagnostic accuracy based on quantitative analysis, and Shen et al. only reported data regarding ULNT1 (Koulidis et al., 2019; Shen et al., 2023). The diagnostic performance of ULNT for CR is of particular interest because CR frequently exhibits comparable clinical symptoms with other conditions, creating difficulties in diagnosis (Thoomes et al., 2012).

As stated, the differences observed in the diagnostic performance of various ULNTs may be attributable to the unique anatomical and physiological characteristics of the nerves being tested (Kleinrensink et al., 2000). In this context, it is worth noting that the ULNT1 alone, often incorporated into a cluster of tests for detecting CR, showed a very low quality of evidence for negligible diagnostic evidence and a moderate sensitivity and low to moderate specificity in our analysis.

Table 6
Certainty of evidence using the GRADE approach.

Test	Sen. (95 % CI)	Sp. (95 % CI)	Nº. Studies (Nº of participants)	Factors that may decrease certainty of evidence					Certainty
				RoB	Ind.	Inc.	Impr.	Others	
Upper-limb entrapment neuropathies									
General	0.66 (0.55; 0.75)	0.62 (0.48; 0.74)	12 studies TP & FN: 1727 TN & FP: 1380	Serious[a] Serious[a]	Not serious Not serious	Very serious[c] (I^2 = 0.87) Very serious[c] (I^2 = 0.87)	Serious[d] Serious[d]	None None	⨁◯◯◯ Very low ⨁◯◯◯ Very low
ULNT1	0.65 (0.49; 0.79)	0.58 (0.38; 0.72)	11 studies TP & FN: 1044 TN & FP: 833	Serious[a] Serious[a]	Not serious Not serious	Very serious[c] (I^2 = 0.87) Very serious[c] (I^2 = 0.87)	Serious[d] Serious[d]	None None	⨁◯◯◯ Very low ⨁◯◯◯ Very low
ULNT2A	0.71 (0.65; 0.77)	0.64 (0.48; 0.78)	3 studies TP & FN: 244 TN & FP: 165	Not serious Not serious	Not serious Not serious	Not serious (I^2 = 0) Very serious[c] (I^2 = 0.82)	Serious[d] Very serious[e]	None None	⨁⨁⨁◯ Moderate ⨁◯◯◯ Very low
ULNT2B	0.60 (0.50; 0.70)	0.54 (0.32; 0.74)	4 studies TP & FN: 284 TN & FP: 265	Not serious Not serious	Not serious Not serious	Serious[b] (I^2 = 0.71) Very serious[c] (I^2 = 0.93)	Very serious[e] Very serious[e]	None None	⨁◯◯◯ Very low ⨁◯◯◯ Very low
ULNT3	0.55 (0.34; 0.75)	0.92 (0.85; 0.96)	2 studies TP & FN: 120 TN & FP: 101	Not serious Not serious	Not serious Not serious	Very serious[c] (I^2 = 0.88) Not serious (I^2 = 0)	Very serious[e] Very serious[e]	None None	⨁◯◯◯ Very low ⨁⨁◯◯ Low
ULNT Combined	0.97 (0.85; 1)	0.69 (0.41; 0.89)	1 study TP & FN: 35 TN & FP: 16	Not serious Not serious	Not serious Not serious	Not serious Not serious	Very serious[e] Very serious[e]	None None	⨁⨁◯◯ Low ⨁⨁◯◯ Low
Cervical radiculopathy									
General	0.69 (0.54; 0.81)	0.65 (0.48; 0.79)	5 studies TP & FN: 937 TN & FP: 844	Serious[a] Serious[a]	Not serious Not serious	Very serious[c] (I^2 = 0.86) Very serious[c] (I^2 = 0.86)	Serious[d] Serious[d]	None None	⨁◯◯◯ Very low ⨁◯◯◯ Very low
ULNT1	0.73 (0.48; 0.88)	0.52 (0.35; 0.69)	5 studies TP & FN: 460 TN & FP: 443	Serious[a] Serious[a]	Not serious Not serious	Very serious[c] (I^2 = 0.89) Very serious[c] (I^2 = 0.89)	Serious[d] Serious[d]	None None	⨁◯◯◯ Very low ⨁◯◯◯ Very low
ULNT2A	0.69 (0.60; 0.77)	0.73 (0.64; 0.81)	2 studies TP & FN: 120 TN & FP: 101	Not serious Not serious	Not serious Not serious	Not serious (I^2 = 0) Not serious (I^2 = 0)	Very serious[e] Very serious[e]	None None	⨁⨁◯◯ Low ⨁⨁◯◯ Low
ULNT2B	0.58 (0.45; 0.71)	0.61 (0.36; 0.82)	3 studies TP & FN: 202 TN & FP: 183	Not serious Not serious	Not serious Not serious	Very serious[c] (I^2 = 0.79) Very serious[c] (I^2 = 0.94)	Very serious[e] Very serious[e]	None None	⨁◯◯◯ Very low ⨁◯◯◯ Very low
ULNT3	0.55 (0.34; 0.75)	0.92 (0.85; 0.96)	2 studies TP & FN: 120 TN & FP: 101	Not serious Not serious	Not serious Not serious	Very serious[c] (I^2 = 0.88) Not serious (I^2 = 0)	Very serious[e] Very serious[e]	None None	⨁◯◯◯ Very low ⨁⨁◯◯ Low
ULNT Combined	0.97 (0.85; 1)	0.69 (0.41; 0.89)	1 study TP & FN: 35 TN & FP: 16	Not serious Not serious	Not serious Not serious	Not serious Not serious	Very serious[e] Very serious[e]	None None	⨁⨁◯◯ Low ⨁⨁◯◯ Low
Carpal tunnel syndrome									
General	0.62 (0.45; 0.77)	0.57 (0.33; 0.78)	7 studies TP & FN: 790 TN & FP: 536	Serious[a] Serious[a]	Not serious Not serious	Very serious[c] Very serious[c]	Serious[d] Serious[d]	None None	⨁◯◯◯ Very low ⨁◯◯◯ Very low
ULNT1	0.61 (0.40; 0.79)	0.61 (0.32; 0.83)	6 studies TP & FN: 584 TN & FP: 390	Serious[a] Serious[a]	Not serious Not serious	Very serious[c] Very serious[c]	Serious[d] Serious[d]	None None	⨁◯◯◯ Very low ⨁◯◯◯ Very low
ULNT2A	0.55 (0.34; 0.75)	0.92 (0.85; 0.96)	1 study TP & FN: 124 TN & FP: 64	Not serious Not serious	Not serious Not serious	Not serious Not serious	Very serious[e] Very serious[e]	None None	⨁⨁◯◯ Low ⨁⨁◯◯ Low
ULNT2B	0.97 (0.85; 1)	0.69 (0.41; 0.89)	1 study TP & FN: 82 TN & FP: 82	Not serious Not serious	Not serious Not serious	Not serious Not serious	Very serious[e] Very serious[e]	None None	⨁⨁◯◯ Low ⨁⨁◯◯ Low

ULNT: Upper Limb Neural Test; Sen: Sensitivity; Sp: Specificity; CI: Confidence Interval; TN: True Negative; TP: True Positive; FP: False Positive; FN: False Negative; RoB: Risk of Bias.
Reasons.
[a] >30–49 % studies with high risk of bias.
[b] $I^2 > 0.50$–0.74.
[c] $I^2 > 0.75$.

[d] Small sample size (<300 participants) or wide CI.
[e] Small sample size (<300 participants) and wide CI or very small sample size (<150 participants or <3 studies).

Similarly, Shen et al. (2023) found, in accordance with this review, limited evidence supporting a fair diagnostic accuracy for ULNT1. However, the authors did not include any other test variant in their analyses and no further discussion is possible. It is plausible that the other tests in the cluster could positively influence the diagnostic performance of the ULNT1 (Wainner et al., 2003). For instance, the median nerve tested in ULNT2a may be more susceptible to mechanical tension or compression generated over medial or lateral cord, which implies all roots of brachialis plexus, leading to more sensitivity for radiculopathy than other ULNT (Kleinrensink et al., 2000). The differences observed with ULNT1 and ULNT2a could be caused by scapular depression in the different abduction grades and sequences producing a higher tension in brachial plexus than the ULNT1. For example, Nee et al. found that the first stage in a sequence of distal to proximal (similar to ULNT2a) exhibited a greater strain than other movements order in comparison to ULNT1 (Nee et al., 2010).

The finding that the ULNT3 has higher specificity could be related to the ulnar nerve's unique anatomical pathway, which may render it more susceptible to being impinged or irritated (C8-T1 roots or medial cord), thereby leading to a more distinct and consistent symptom profile upon testing (Kleinrensink et al., 2000). However, the nerve roots implicated are lower than the median nerve and, in consequence, less sensitivity is presented. Apelby-Albrecht et al. (2013) suggested initiating the neurodynamic examination with the ULNT3 test, given its high specificity. Based on our findings, when a patient has a positive ULNT3 test, it is highly probable that the patient is suffering from CR because of the strong diagnostic evidence. This suggests that the ULNT3 may provide benefit for diagnosing CR even when symptoms are not representative of an ulnar nerve issue. Moreover, the combined use of both ULNT2a and ULNT3 may provide substantial value in these cases. This notion is supported by Apelby-Albrecht et al. (2013), who reported enhanced sensitivity when combined ULNT1-3 together. By integrating the unique benefits of each test, a more thorough assessment of the patient's condition might be possible. The moderate specificity of ULNT3 could aid in confirming the condition when the test is positive, whereas the high sensitivity of ULNT2a could assist in ruling out the condition when the test is negative.

Based on these findings, combining multiple ULNTs is encouraged as this strategy could increase the overall clinical accuracy because different nerves may be affected to varying degrees in entrapment neuropathy conditions (Grondin et al., 2021). A single test might not be sensitive enough to detect subtle or early-stage changes in nerve function, but a combination of tests could increase the likelihood of detecting these changes. This principle is analogous to the use of test batteries or diagnostic clusters in other areas of medicine, where multiple tests are used together to improve diagnostic sensitivity (Thoomes et al., 2018). It is essential to consider that increasing sensitivity through test combinations may come at the cost of reducing specificity, leading to a higher risk of false-positive results (Parikh et al., 2008). Clinicians should, therefore, use their clinical judgment when interpreting test results and consider other diagnostic information when making clinical decisions. Further research is warranted to establish more definitive guidelines on the optimal use and interpretation of ULNTs in diagnosing CR.

4.2. Carpal tunnel syndrome

The diagnostic accuracy of ULNTs for CTS was poor, with limited clinical utility (moderate sensitivity and low specificity). These findings indicate that ULNTs are less effective in detecting CTS compared to UEN or CR, suggesting that their role in CTS diagnosis should be reconsidered. Similarly, previous systematic reviews (De Arenas-Arroyo et al., 2022; Koulidis et al., 2019) concluded that ULNTs have no diagnostic accuracy when used in isolation to identify patients with CTS. A previous meta-analysis (De Arenas-Arroyo et al., 2022) about accuracy of tests for the diagnostic of CTS showed a 1.78 (95 % CI: 0.61; 5.19) dOR for ULNT1 and our metanalysis showed a similar result. Based on our results for the examination of suspected CTS, the test should be supplemented with other evaluations with higher dORs, such as Durkan test or the hand elevation test (De Arenas-Arroyo et al., 2022). An isolated ULNT test may not provide any significant information for diagnosis, but its combination with other tests could enhance diagnostic validity. Further studies are needed to explore the collective application of tests, along with the ULNTs, to assess whether this enhances diagnostic acuity.

The ULNT2a demonstrated slightly higher sensitivity than ULNT1 however, the diagnostic acuity of the test has only been reported by one study (Beddaa et al., 2022). In terms of specificity, ULNT1 exhibited higher specificity (0.55). In addition, altering the test sequences can affect the location of mechanical tension. This change can result in either increased compression or distension at either proximal or distal points. Consequently, if movements are ordered from a distal to proximal direction, mechanical tension occurs more rapidly during distal excursions. This could result in heightened tension in distal areas such as

Table 7
Meta-regression by diagnostic, structural differentiation or Nee's criteria and methodological quality.

Diagnostic				
Parameter	Estimate	LCI	UCI	p-value
Relative sensitivity CTS vs CR	0.909	0.662	1.248	0.55
Relative specificity level CTS vs CR	0.87	0.571	1.324	0.51
Global test comparison				0.353
Criteria				
Parameter	Estimate	LCI	UCI	p-value
Relative sensitivity level SD vs NSD	0.83	0.623	1.105	0.217
Relative specificity level SD vs NSD	2.432	1.616	3.662	$P < 0.01$
Global test comparison				$P < 0.01$
Methodological Quality				
Parameter	Estimate	LCI	UCI	p-value
Relative sensitivity level Low vs High	1.04	0.713	1.52	0.832
Relative specificity level Low vs High	1.33	0.737	2.407	0.293
Global test comparison				0.311

LCI: Lower confidence interval; UCI: Upper Confidence Interval; CTS: Carpal Tunnel Syndrome; CR: Cervical Radiculopathy; NSD: Non- Structural Differentiation; SD: Structural Differentiation.

the hand or carpal tunnel (Nee et al., 2010). Based on these findings, performing a sequence that generates more distal tension first and then proximal could be recommended as seems to be more effective for the diagnosis of patients with CTS and explain the differences in sensitivity.

4.3. Structural differentiation

One of the main findings is that we found a significant interaction on criteria for a positive test. Nee's criteria or structural differentiation showed better values than the other criteria or without structural differentiation. Nee's criteria included structural differentiation as a main criteria while the other criteria did not (Nee et al., 2012a). This is relevant as structural differentiation may show an increase in nerve mechanosensitivity that could explain patient's symptoms by adding or removing sensitizing maneuvers (Butler and Gifford, 1989). ULNTs with structural differentiation are more accurate than those without, particularly due to their higher specificity. Based on these findings, we recommend the use of structural differentiation for ULNTs. Further diagnostic accuracy studies including Nee's criteria or structural differentiation are needed.

4.4. Study limitations

One of the key limitations of our systematic review and meta-analysis on the diagnostic accuracy of ULNTs relates to the amalgamation of different tests from the same studies. This approach may have introduced a degree of heterogeneity that could affect the overall results and interpretations. Different tests may have varying levels of sensitivity, specificity and likelihood ratios, and combining them could potentially dilute the accuracy of individual tests. Also, certain meta-analyses were conducted with a limited number of studies, often comprising only 2 or 3 studies.

Furthermore, the diagnostic categorization of upper-limb entrapment neuropathies was broad, encompassing both radiculopathy and carpal tunnel syndrome. These conditions, while both involving nerve impairment, have distinct pathophysiological mechanisms and clinical presentations. The lumping together of these diagnoses may have led to an overgeneralization of the results, potentially obscuring the specific diagnostic accuracy of the neurodynamic tests for each condition. However, meta-regression did not identify any interaction between diagnosis, and we reported the results by diagnosis (CR or CTS).

In addition, many of the studies included exhibited biases, which could have influenced the results of our meta-analysis. These biases, which could range from selection bias to information bias, may have skewed the findings and limited the reliability of our conclusions. Therefore, the quality of the evidence in most cases was rated as very low due to study biases, imprecision and heterogeneity. The imprecision and heterogeneity of the studies could be attributed to the variability in study design, population characteristics, and neurodynamic test protocols among the included studies. For example, meta-regression showed differences between using or not a structural differentiation.

In light of these limitations, caution should be exercised when interpreting the findings of this review and meta-analysis. Future research should aim to minimize these limitations by ensuring rigorous study design, precise diagnostic categorization, and the use of standardized neurodynamic test protocols.

5. Conclusion

In conclusion, there is very low evidence supporting the diagnostic accuracy of UEN and CR, with moderate sensitivity and specificity, and sufficient and good diagnostic utility, respectively. The diagnostic accuracy of ULNTs showed moderate sensitivity and low to moderate specificity, but insufficient overall diagnostic accuracy. The likelihood ratios varied depending on the diagnostic criteria used (with or without structural differentiation). The diagnostic utility values of ULNTs varied significantly, with no single test proving universally superior. Therefore, the choice of ULNT should be tailored to the specific diagnostic context. Further research is needed to enhance the quality of evidence and refine the diagnostic utility of ULNTs for peripheral entrapment neuropathy conditions.

CRediT authorship contribution statement

Daniel Albert-Lucena: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Marcos José Navarro-Santana:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **María José Díaz-Arribas:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Conceptualization. **Gabriel Rabanal-Rodríguez:** Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Juan Antonio Valera-Calero:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Chad Cook:** Writing – review & editing, Writing – original draft, Validation, Investigation, Formal analysis, Conceptualization. **Gustavo Plaza-Manzano:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Formatting of funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Gabriel Rabanal-Rodríguez is recipient of a researcher training scholarship from the Universidad Complutense de Madrid and Banco Santander (CT25/24).

Conflict of interest

There is no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.msksp.2025.103317.

References

Abbed, K.M., Coumans, J.V.C.E., 2007. Cervical radiculopathy: pathophysiology, presentation, and clinical evaluation. Neurosurgery 60. https://doi.org/10.1227/01.NEU.0000249223.51871.C2.

Aguayo-albasini, L., Flores-Pastor, B., Soria-Aledo, V., 2016. Sistema GRADE : clasificación de la fuerza de la evidencia y graduación recomendación. Cirugía Española 92, 82–88.

Alanazy, M.H., 2017. Clinical and electrophysiological evaluation of carpal tunnel syndrome: approach and pitfalls. Neurosciences 22, 169–180. https://doi.org/10.17712/NSJ.2017.3.20160638.

Apelby-Albrecht, M., Andersson, L., Kleiva, I.W., Kvåle, K., Skillgate, E., Josephson, A., 2013. Concordance of upper limb neurodynamic tests with medical examination and magnetic resonance imaging in patients with cervical radiculopathy: a diagnostic cohort study. J. Manipulative Physiol. Therapeut. 36, 626–632. https://doi.org/10.1016/j.jmpt.2013.07.007.

Austin, T.M., Richter, R.R., Sebelski, C.A., 2014. Introduction to the GRADE approach for guideline development: considerations for physical therapist practice. Phys. Ther. 94, 1652–1659. https://doi.org/10.2522/PTJ.20130627.

Beddaa, H., Kably, B., Mouhi, I., Marzouk, B., Marfak, A., Nafai, S., Ouazzani, R., Birouk, N., 2022. The validity of the upper limb neurodynamic test 2A in women with a clinical diagnosis of carpal tunnel syndrome: a prospective diagnostic accuracy study. Pan African Med. J. 42, 61. https://doi.org/10.11604/pamj.2022.42.61.30119.

Bueno-Gracia, E., Fanlo-Mazas, P., Malo-Urriés, M., Rodriguez-Mena, D., Montaner-Cuello, A., Ciuffreda, G., Shacklock, M., Estébanez-de-Miguel, E., 2024. Diagnostic accuracy of the upper limb neurodynamic test 1 using neurodynamic sequencing in diagnosis of carpal tunnel syndrome. Musculoskeletal Sci. Pract. 69, 102897. https://doi.org/10.1016/J.MSKSP.2023.102897.

Bueno-Gracia, E., Tricás-Moreno, J.M., Fanlo-Mazas, P., Malo-Urriés, M., Haddad-Garay, M., Estébanez-de-Miguel, E., Hidalgo-García, C., Krauss, J.R., 2016. Validity of the upper limb neurodynamic test 1 for the diagnosis of carpal tunnel syndrome. The role of structural differentiation. Man. Ther. 22, 190–195. https://doi.org/10.1016/j.math.2015.12.007.

Butler, D., Gifford, L., 1989. The concept of adverse mechanical tension in the nervous system Part 1: testing for "dural tension.". Physiotherapy 75, 622–629. https://doi.org/10.1016/S0031-9406(10)62374-7.

Cooke, A., Smith, D., Booth, A., 2012. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. Qual. Health Res. 22, 1435–1443. https://doi.org/10.1177/1049732312452938.

Covill, L.G., Petersen, S.M., 2012. Upper extremity neurodynamic tests: range of motion asymmetry may not indicate impairment. Physiother. Theory Pract. 28, 535–541. https://doi.org/10.3109/09593985.2011.641198.

De Arenas-Arroyo, S.N., Cavero-Redondo, I., Torres-Costoso, A., Reina-Gutiérrez, S., José Guzmán-Pavón, M., Martínez-Vizcaíno, V., 2022. Accuracy of the most common provocation tests for diagnosing carpal tunnel syndrome: a systematic review with meta-analysis. J. Orthop. Sports Phys. Ther. 52, 522–531. https://doi.org/10.2519/jospt.2022.10828.

Elvey, R., 1979. Brachial plexus tension tests and the pathoanatomical origin of arm pain. Aspects Manip. Ther. 105–110.

Faktorovich, S., Filatov, A., Rizvi, Z., 2021. Common compression neuropathies. Clin. Geriatr. Med. 37, 241–252. https://doi.org/10.1016/J.CGER.2021.01.011.

Ghasemi, M., Golabchi, K., Mousavi, S.A., Asadi, B., Rezvani, M., Shaygannejad, V., Salari, M., 2013. The value of provocative tests in diagnosis of cervical radiculopathy. J. Res. Med. Sci. 18, 35–38.

Grondin, F., Cook, C., Hall, T., Maillard, O., Perdrix, Y., Freppel, S., 2021. Diagnostic accuracy of upper limb neurodynamic tests in the diagnosis of cervical radiculopathy. Musculoskeletal Sci. Pract. 55. https://doi.org/10.1016/j.msksp.2021.102427.

Higgins, J.P.T., Thompson, S.G., 2002. Quantifying heterogeneity in a meta-analysis. Stat. Med. 21, 1539–1558. https://doi.org/10.1002/sim.1186.

Jaeschke, R., Guyatt, G.H., Sackett, D.L., Bass, E., Edwards, P.B., Browman, G., Cook, D., Farkouh, M., Gerstein, H., Haynes, B., Hayward, R., Holbrook, A., Juniper, E., Lee, H., Levine, M., Moyer, V., Nishikawa, J., Oxman, A., Patel, A., Philbrick, J., Richardson, W.S., Sauve, S., Sinclair, J., Trout, K.S., Tugwell, P., Tunis, S., Walter, S., Wilson, M., 1994. 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, 703–707. https://doi.org/10.1001/JAMA.1994.03510330081039.

Kimura, J., 1978. A method for determining median nerve conduction velocity across the carpal tunnel. J. Neurol. Sci. 38, 1–10. https://doi.org/10.1016/0022-510X(78)90240-X.

Kleinrensink, G.J., Stoeckart, R., Mulder, P.G.H., Hoek, G.V.D., Broek, T., Vleeming, A., Snijders, C.J., 2000. Upper limb tension tests as tools in the diagnosis of nerve and plexus lesions - anatomical and biomechanical aspects. Clin. BioMech. 15, 9–14. https://doi.org/10.1016/S0268-0033(99)00042-X.

Koulidis, K., Veremis, Y., Anderson, C., Heneghan, N.R., 2019. Diagnostic accuracy of upper limb neurodynamic tests for the assessment of peripheral neuropathic pain: a systematic review. Musculoskelet Sci Pract 40, 21–33. https://doi.org/10.1016/j.msksp.2019.01.001.

Kuijper, B., Tans, J.T.J., Schimsheimer, R.J., Van Der Kallen, B.F.W., Beelen, A., Nollet, F., De Visser, M., 2009. Degenerative cervical radiculopathy: diagnosis and conservative treatment. A review. Eur. J. Neurol. 16, 15–20. https://doi.org/10.1111/J.1468-1331.2008.02365.X.

Leeflang, M.M.G., Deeks, J.J., Takwoingi, Y., Macaskill, P., 2013. Cochrane diagnostic test accuracy reviews. Syst. Rev. 2, 82. https://doi.org/10.1186/2046-4053-2-82.

Leoni, D., Storer, D., Gatti, R., Egloff, M., Barbero, M., 2016. Upper limb neurodynamic test 1 on healthy individuals: intra-and intersession reliability of the angle between pain onset and submaximal pain. Pain Res. Manag. https://doi.org/10.1155/2016/9607262, 2016.

Lijmer, J.G., Bossuyt, P.M.M., Heisterkamp, S.H., 2002. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. Stat. Med. 21, 1525–1537. https://doi.org/10.1002/sim.1185.

Mallett, S., Halligan, S., Matthew Thompson, G.P., Collins, G.S., Altman, D.G., 2012. Interpreting diagnostic accuracy studies for patient care. BMJ 345. https://doi.org/10.1136/BMJ.E3999.

McInnes, M.D.F., Moher, D., Thombs, B.D., McGrath, T.A., Bossuyt, P.M., Clifford, T., Cohen, J.F., Deeks, J.J., Gatsonis, C., Hooft, L., Hunt, H.A., Hyde, C.J., Korevaar, D. A., Leeflang, M.M.G., Macaskill, P., Reitsma, J.B., Rodin, R., Rutjes, A.W.S., Salameh, J.P., Stevens, A., Takwoingi, Y., Tonelli, M., Weeks, L., Whiting, P., Willis, B.H., 2018. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. JAMA 319, 388–396. https://doi.org/10.1001/JAMA.2017.19163.

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, 413–424. https://doi.org/10.2519/jospt.2012.3988.

Nee, R.J., Yang, C.H., Liang, C.C., Tseng, G.F., Coppieters, M.W., 2010. Impact of order of movement on nerve strain and longitudinal excursion:A biomechanical study with implications for neurodynamic test sequencing. Man. Ther. 15, 376–381. https://doi.org/10.1016/j.math.2010.03.001.

Padua, L., Coraci, D., Erra, C., Pazzaglia, C., Paolasso, I., Loreti, C., Caliandro, P., Hobson-Webb, L.D., 2016. Carpal tunnel syndrome: clinical features, diagnosis, and management. Lancet Neurol. 15, 1273–1284. https://doi.org/10.1016/S1474-4422(16)30231-9.

Parikh, R., Mathai, A., Parikh, S., Sekhar, G.C., Thomas, R., 2008. Understanding and using sensitivity, specificity and predictive values. Indian J. Ophthalmol. 56, 45. https://doi.org/10.4103/0301-4738.37595.

Reitsma, J.B., Leeflang, M.M.G., Sterne, J.A.C., Bossuyt, P.M.M., Whiting, P.F., Rutjes, A. W.S.S., Westwood, M.E., Mallet, S., Deeks, J.J., Reitsma, J.B., Leeflang, M.M.G., Sterne, J.A.C., Bossuyt, P.M.M., 2011. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann. Intern. Med. 155, 529–536.

Schlattmann, P., 2023. Tutorial: statistical methods for the meta-analysis of diagnostic test accuracy studies. Clin. Chem. Lab. Med. 1–18. https://doi.org/10.1515/cclm-2022-1256.

Schmid, A.B., Fundaun, J., Tampin, B., 2020. Entrapment neuropathies: a contemporary approach to pathophysiology, clinical assessment, and management. Pain Reports 5. https://doi.org/10.1097/PR9.0000000000000829.

Schmid, A.B., Nee, R.J., Coppieters, M.W., 2013. Reappraising entrapment neuropathies-Mechanisms, diagnosis and management. Man. Ther. 18, 449–457. https://doi.org/10.1016/j.math.2013.07.006.

Schneiders, A.G., Sullivan, S.J., Hendrick, P.A., Hones, B.D.G.M., McMaster, A.R., Sugden, B.A., Tomlinson, C., 2012. The ability of clinical tests to diagnose stress fractures: a systematic review and meta-analysis. J. Orthop. Sports Phys. Ther. 42, 760–771. https://doi.org/10.2519/jospt.2012.4000.

Schünemann, H.J., Oxman, A.D., Brozek, J., Glasziou, P., Jaeschke, R., Vist, G.E., Williams, J.W., Kunz, R., Craig, J., Montori, V.M., Bossuyt, P., Guyatt, G.H., 2008. Rating Quality of Evidence and Strength of Recommendations: GRADE: Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ Br. Med. J. (Clin. Res. Ed.) 336, 1106. https://doi.org/10.1136/BMJ.39500.677199.AE.

Shacklock, M., 2005. Improving application of neurodynamic (neural tension) testing and treatments: a message to researchers and clinicians. Man. Ther. 10, 175–179. https://doi.org/10.1016/j.math.2005.03.001.

Shen, P., Chi-Chung Tsang, R., Liang, Y., Chen, X., 2023. Diagnostic accuracy of the upper limb neurodynamic test with median bias (ULNT1) for cervical radiculopathy: a systematic review and meta-analysis. Physiotherapy 120, 17–25. https://doi.org/10.1016/j.physio.2023.06.001.

Šimundić, A.-M., 2009. Measures of diagnostic accuracy: basic definitions. EJIFCC 19, 203.

Sleijser-Koehorst, M.L.S., Coppieters, M.W., Epping, R., Rooker, S., Verhagen, A.P., Scholten-Peeters, G.G.M., 2021. Diagnostic accuracy of patient interview items and clinical tests for cervical radiculopathy. Physiotherapy 111, 74–82. https://doi.org/10.1016/J.PHYSIO.2020.07.007.

Smart, K.M., Blake, C., Staines, A., Thacker, M., Doody, C., 2012. Mechanisms-based classifications of musculoskeletal pain: part 1 of 3: symptoms and signs of central sensitisation in patients with low back (± leg) pain. Man. Ther. 17, 336–344. https://doi.org/10.1016/J.MATH.2012.03.013.

Thoomes, E.J., Scholten-Peeters, G.G.M., De Boer, A.J., Olsthoorn, R.A., Verkerk, K., Lin, C., Verhagen, A.P., 2012. Lack of uniform diagnostic criteria for cervical radiculopathy in conservative intervention studies: a systematic review. Eur. Spine J. : Offic. Publ. Euro. Spine Soc. Euro. Spinal Deformity Soc. Euro. Section Cervical Spine Res. Soc. 21, 1459–1470. https://doi.org/10.1007/S00586-012-2297-9.

Thoomes, E.J., van Geest, S., van der Windt, D.A., Falla, D., Verhagen, A.P., Koes, B.W., Thoomes-de Graaf, M., Kuijper, B., Scholten-Peeters, W.G.M., Vleggeert-Lankamp, C. L., 2018. Value of physical tests in diagnosing cervical radiculopathy: a systematic review. Spine J. 18, 179–189. https://doi.org/10.1016/j.spinee.2017.08.241.

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, 333–338. https://doi.org/10.1016/j.jht.2017.05.004.

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. Man. Ther. 16, 258–263. https://doi.org/10.1016/j.math.2010.11.003.

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. Manipulative Physiol. Therapeut. 35, 54–63. https://doi.org/10.1016/j.jmpt.2011.09.008.

Wainner, R.S., Fritz, J.M., Irrgang, J.J., Boninger, M.L., Delitto, A., Allison, S., 2003. Reliability and diagnostic accuracy of the clinical examination and patient self-report measures for cervical radiculopathy. Spine 28, 52–62. https://doi.org/10.1097/00007632-200301010-00014.

Wainner, R.S., Fritz, J.M., Irrgang, J.J., Delitto, A., Allison, S., Boninger, M.L., 2005. Development of a clinical prediction rule for the diagnosis of carpal tunnel syndrome. Arch. Phys. Med. Rehabil. 86, 609–618. https://doi.org/10.1016/j.apmr.2004.11.008.

Zhou, Y., Dendukuri, N., 2014. Statistics for quantifying heterogeneity in univariate and bivariate meta-analyses of binary data: the case of meta-analyses of diagnostic accuracy. Stat. Med. 33, 2701–2717. https://doi.org/10.1002/SIM.6115.