Elsevier required licence: © <2018>. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

VALUE OF PHYSICAL TESTS IN DIAGNOSING CERVICAL RADICULOPATHY; A SYSTEMATIC REVIEW

Erik J Thoomes[a], MSc; Sarita van Geest[b], MD; Danielle A van der Windt[c], PhD; Deborah Falla[f], PhD; Arianne P Verhagen[e], PhD; Bart W Koes[3], PhD; Marloes Thoomes-de Graaf[a], MSc; Barbara Kuijper[d], MD PhD; Wendy G.M. Scholten-Peeters, PhD; Carmen L Vleggeert-Lankamp[b], MD PhD.

a) OMPT Clinic Thoomes, Hazerswoude, The Netherlands
b) Leiden University Medical Centre, Dept. Neurosurgery, Leiden, The Netherlands
c) Arthritis Research Campaign National Primary Care Centre, Keele University, Keele, Staffordshire, United Kingdom
d) Maasstad Hospital, Rotterdam, The Netherlands
e) Department of General Practice, Erasmus Medical Centre University, Rotterdam, The Netherland
f) School of Sport, Exercise and Rehabilitation Sciences, University of Birmingham, Birmingham, UK

Word count:

- Abstract: 241
- Manuscript: 4900 (including in-text references)

Tables: 3
Figures: 4

Corresponding author:
Erik J Thoomes, Fysio-Experts, Rijndijk 137, 2394 AG Hazerswoude, The Netherlands
Tel: +31 6 2919 3359
E-mail: erikthoomes@gmail.com

ABSTRACT

Background context

In clinical practice, the diagnosis of cervical radiculopathy is based on information from patient history, physical examination and diagnostic imaging. Various physical tests are performed, but their diagnostic accuracy is unknown.

Purpose

This study aimed to summarize and update evidence on diagnostic performance of tests carried out during physical examination for the diagnosis of cervical radiculopathy**.**

Study design

A review of the accuracy of diagnostic tests was carried out.

Study Sample

The study sample comprised diagnostic studies comparing results of tests performed during physical examination in diagnosing cervical radiculopathy with a reference standard of imaging or surgical findings.

Outcome measures

Sensitivity, specificity, likelihood ratios are presented, together with pooled results for sensitivity and specificity.

Methods

A literature search up to March 2016 was performed in CENTRAL, PubMed (MEDLINE), EMBASE, CINAHL, Web of Science and Google Scholar. The methodological quality of studies was assessed using the QUADAS-2.

Results

Five diagnostic accuracy studies were identified. Only Spurling's test was evaluated in more than one study, showing high specificity ranging from 0.89 to 1.00 (95% confidence interval [CI]: 0.59-1.00); sensitivity varied from 0.38 to 0.97 (95% CI: 0.21-0.99). No studies were found that assessed the diagnostic accuracy of widely used neurological tests such as key muscle strength, tendon reflexes and sensory impairments.

Conclusions

There is limited evidence for accuracy of physical examination tests for the diagnosis of cervical radiculopathy. When consistent with patient history, clinicians may use a combination of Spurling's, axial traction, and an Arm Squeeze test to increase the

likelihood of a cervical radiculopathy, whereas a combined result of four negative neurodynamics tests and an Arm Squeeze test could be used to rule the disorder out.

Key Words: cervical radiculopathy, radiating pain, arm pain, neck pain diagnostic accuracy, physical examination, Spurling, shoulder abduction relief test, cervical axial traction test, Arm Squeeze test, ULNT.

BACKGROUND

Cervical radiculopathy is a term used to describe pain radiating into the arm corresponding to the dermatome of the involved cervical nerve root (Kuijper, 2009; Thoomes, 2012).

The incidence and prevalence of cervical radiculopathy is unclear and epidemiological data are sparse. In the only large retrospective population-based study, the annual age-adjusted incidence rate was 83.2 per 100,000 persons (107.3 for men and 63.5 for women) with a peak incidence in the 5th and 6th decade in both genders (Radhakrishnan, 1994). The most commonly affected levels are C6 (66%) and C7 (62%) (Kim, 2016).

Radiculopathy is differentiated from radicular pain, where radiculopathy is a neurological state in which conduction is limited or blocked along a spinal nerve or its roots. Radiculopathy and radicular pain commonly occur together (Bogduk, 2009; Merskey H, 1994). Radicular pain is usually caused by compression of the nerve root due to cervical disc herniation or degenerative spondylotic changes, but radicular symptoms can also occur without evident compression, for instance, because of inflammation of the nerve (Bogduk, 2009).

A recent systematic review concluded that criteria used to select patients with cervical radiculopathy varied widely. There was consensus only on the presence of pain, but not on the exact location of pain (Thoomes, 2012).

The diagnosis of radiculopathy is based on information received during the subjective (history taking) and physical examination, which is then confirmed via diagnostic imaging or supported by surgical findings (Bussieres, 2008). The most commonly used physical tests (Bono, 2011; Rubinstein, 2007; Wainner, 2000) include tendon reflexes, manual muscle testing of key muscles for muscle weakness or atrophy, and testing for sensory deficits, the assessment of range of motion (ROM), and provocative tests like the foraminal compression test or Spurling's test (Spurling RG, 1944), shoulder abduction (relief) test (Davidson, 1981), Upper Limb Tension Test (ULTT), Upper Limb Neural Tension test (ULNT) (Elvey, 1997), neck traction/distraction test, and Valsalva maneuver (Jull, 2015).

Some previous reviews have summarized the results of studies on the diagnostic accuracy of physical examination for the identification of cervical radiculopathy (Bono, 2011; Rubinstein, 2007; Wainner, 2000; Ellenberg, 1994; Nordin, 2008). Two reviews included an assessment of the methodological quality of the primary studies

(Rubinstein, 2007) and one review offered a qualitative summary of the findings (Bono, 2011). These reviews noted that some provocative tests (e.g. Spurling's test, traction/distraction, Valsalva maneuver) may have low to moderate sensitivity and high specificity, but the diagnostic accuracy of individual tests varied considerably between individual studies. Only one test (ULNT) showed high sensitivity and low specificity (Bono, 2011; Rubinstein, 2007). Clusters of tests were generally considered to be more accurate.

However, these reviews are limited either because they did not apply contemporary methods for quality appraisal and data synthesis (Wainner, 2000), were narrative reviews (Ellenberg, 1994; Malanga, 1997), or did not specifically address cervical radiculopathy (Nordin, 2008).

The most recent systematic review was aimed at producing a North American Spine Society (NASS) clinical guideline (Bono, 2011). Since then, new tests [18] or a combination of tests [19] have been described and a commonly used test (ie., Spurling's test) has been further assessed [20].

Therefore, the present study aims to summarize and update the evidence on the diagnostic performance of specific tests carried out during the physical examination for the diagnosis of cervical radiculopathy. Moreover, a quality assessment was performed to assess the influence of potential sources of bias.

METHODS

Inclusion criteria

Studies were included that involved patients who were greater than 18 years of age and were suspected of having a cervical radiculopathy from nerve root compression due to cervical disc herniation or degenerative spondylotic changes. The diagnostic accuracy of physical examination tests had to be assessed in the study (ie., how well a test, or a series of tests, was able to correctly identify patients with cervical radiculopathy). Studies carried out in primary as well as secondary care were eligible. Only results from full reports were included.

Index tests

Studies on all items that have been proposed as a diagnostic test during physical examination for identifying cervical radiculopathy were eligible for inclusion. Primary

diagnostic studies were considered only if they compared the results of tests performed during physical examination for the identification of cervical radiculopathy with those of imaging or surgical findings. Studies were included in which the diagnostic performances of individual aspects of the physical examination were evaluated separately, or in combination. In case of a combination, the study should have clearly described which tests were included in the combination and how it was performed.

Reference standards

Studies were included when the results of physical examination were compared with (1) diagnostic imaging: magnetic resonance imaging (MRI) or computed tomography (CT) myelography or (2) findings during surgery.

Search methods

Electronic searches

A search strategy was developed in collaboration with a librarian, according to guidelines set by the Cochrane Diagnostic Test Accuracy group. A search was performed through CENTRAL (The Cochrane Library), PubMed (including MEDLINE), EMBASE, CINAHL, Web of Science and Google Scholar for eligible diagnostic studies from their inception to March 2016. The search strategy for EMBASE is presented in Supplementary *Appendix S1*. No language restrictions were applied. Reference lists of relevant publications were checked for gray literature and a forward citation was performed searching of relevant articles using the PubMed related articles feature.

Assessment of methodological quality

Three sets of review authors (ET, SG and either AV, BK or DvdW) assessed the methodological quality in each study, using the Quality Assessment of Diagnostic Accuracy Studies list (QUADAS-2) (Whiting, 2011). Specifically to this review, tailored guidelines for the assessment of of the four bias domains were made available to the review authors (Supplementary *Appendix S2*).

With respect to the QUADAS-2 criterion of risk of bias domain related reference standard, a tiered scoring system was devised. A combination of history taking, physical examination including neurological assessment and MRI or CT-myelography

(or surgical findings) was considered to be a true diagnostic gold standard, resulting in a “yes”, whereas a reference standard of only assessing MRI of CT-myelography imaging should result in an “unclear” because of the inappropriate high number of false positives (FPs) (Kuijper, 2011; Siivola, 2002; Ernst, 2005). Potential incorporation bias was avoided by the index test never being part of the reference test set. An intraclass coefficient (ICC) was calculated to assess the initial agreement between both raters on the overall score per domain; an ICC higher than 0.70 was considered good [25]. Disagreements were resolved by consensus and, if necessary, through arbitration by a third review author (CV-L). Both a tabular (Table 2) as well as a graphical (Figure 2) display was used to summarize the QUADAS-2 assessments.

Data collection and analysis

Selection of studies

Two review authors (ET, SG) independently screened titles, abstracts and the full text of potentially relevant articles. Disagreements on inclusion were initially resolved by discussion or, if necessary, through arbitration by a third review author (CV-L).

Data extraction and management

Characteristics of participants, the index tests and reference standard, and aspects of study methods for each included study were extracted using a standardized form.

- *Characteristics of participants:* setting (primary /secondary care); numbers enrolled in the study, receiving index test and reference standard, for whom results were reported in the two-by-two table and reasons for withdrawal; duration of radicular symptoms and neurological signs.
- *Test characteristics*: the type of test, role of the test in the diagnostic pathway, method of execution, experience and expertise of the assessors, type of reference standard, and cut-off points for diagnosing cervical radiculopathy due to cervical disc herniation or to degenerative spondylotic changes, definitions of positive outcomes for the reference tests.
- *Aspects of study methods*: the design of the study, time and treatment between index test and reference standard, and risks of bias (see section on assessment of methodological quality).

Two review authors (ET, SG) independently extracted data and diagnostic two-by-two tables (true positive [TP], false positive [FP], true negative [TN], and false

negative [FN] index test results, likelihood ratios and predictive values) for each study. Two-by-two tables were reconstructed if they were not available, using information on relevant parameters (eg, sensitivity and specificity). Both a narrative synthesis as well as a quantitative analysis was performed. Eligible studies were not included in the quantitative analyses when the diagnostic two-by-two table could not be reconstructed, but their results were included in the narrative sysnthesis. A three-point rating scale (“low”: 0.0-0.33; “moderate”: 0.34-0.66 and “high”: 0.67-1.0) was used to classify sensitivity/specificity [28]. Prior probability (prevalence) of nerve root compression was calculated as the proportion of patients in the cohort diagnosed with nerve root compression according to the reference standard. Disagreements were resolved by consensus or arbitration of a third reviewer (CV-L).

Statistical analysis and data synthesis

Two-by-two tables were constructed for each index test evaluated in each study from the reported number of TPs, FPs, TNs and FPs. Results in terms of sensitivity and specificity and 95% confidence interval (CI) for each test were presented in a forest-plot. Results were entered into Review Manager 5.3 (The Nordic Cochrane Centre, Copenhagen, Denmark). Pooled estimates of sensitivity and specificity were only presented if studies showed clinical homogeneity (similar reference standard and index test, similar definition of nerve root compression and the same cut-off points used). The range of sensitivity and specificity for each index test are presented in cases were no pooled estimate could be calculated.

Investigations of heterogeneity

Heterogeneity was examined by considering study characteristics, visual inspection of (the Cis of) forest plots of sensitivities and specificities. The findings of the review are summarized in table 3, including a summary estimate of sensitivity, specificity, and likelihood ratios for relevant tests and subgroups of studies (e.g. studies on patients in primary or secondary care, and studies using different reference standards). The prevalence of the target condition (cervical nerve root compression) in the study populations is presented along with measures of diagnostic performance.

RESULTS

Search results

The search identified 2845 unique citations. Another five were retrieved from searching through gray literature. After screening titles and abstracts, 87 manuscripts were retrieved for a full text assessment. Initial agreement between authors was almost perfect (IRR=95%) with regard to the reasons for exclusion out of these 87 manuscripts. Disagreements were resolved through minor discussion and arbitration through a third author was not necessary. Five of the 87 manuscripts (Apelby-Albrecht, 2013; Gumina, 2013; Shabat, 2012; Shah, 2004; Viikari-Juntura, 1989) met all eligibility criteria and were included in the quantitative synthesis (Figure 1).

Please insert figure 1: PRISMA flowchart of included studies

Description of the studies

Details on the design, setting, population, reference standard and definition of the target condition are provided in Table 1. All studies were conducted in a hospital setting. Only two studies (Apelby-Albrecht, 2013; Gumina, 2013) used a combination of history taking, clinical examination and imaging as a reference standard. Spurling's test was an index test in three studies (Shabat, 2012; Shah, 2004; Viikari-Juntura, 1989) and neurodynamic tests in two studies (Apelby-Albrecht, 2013; Viikari-Juntura, 1989) but results were not reported by one author (Viikari-Juntura, 1989) due to poor inter-examiner reliability. Other index tests (Arm Squeeze test, shoulder abduction (relief) test, and traction test) were all assessed in single studies only.

Please insert table 1: *Characteristics of included studies* near here

Methodological quality of included studies

Overall, the quality of the studies was poor to moderate (see Table 2), as all studies had a 'high' or 'unclear' risk of bias in at least one category (see Figure 2).
Initial agreement between both raters on the score was good (ICC two way random agreement=0.92 [95% CI: 0.78-0.98]); arbitration through the third author was not necessary.
For the patient selection domain, two studies had a high risk of bias: one study (Gumina, 2013) strongly resembled a case control study type and the other study (Viikari-Juntura, 1989) had inappropriate exclusion criteria. Regarding the

applicability to the review question, one study (Viikari-Juntura, 1989) raised serious concerns caused by an unclear process for excluding patients or what tests had been conducted before inclusion in the study, as exclusions seemed likely to have taken place after history taking and physical examination. This does not reflect the intended use of the index test. Two studies (Gumina, 2013; Shabat, 2012) were unclear in this domain.

For the index test domain, no studies had a high risk of bias and four studies (Apelby-Albrecht, 2013; Gumina, 2013; Shabat, 2012; Viikari-Juntura, 1989) specified a positivity threshold (interpretation of “positive” results). There were no concerns regarding the applicability for any of the studies.

With respect to the reference standard, only one study (Apelby-Albrecht, 2013) was considered to have an appropriate reference test (low risk of bias) and only one study assessed the root canal diameter on MRI for all patients, and for a portion of patients, the results at surgery (Shah, 2004). The remaining studies did not include information on the type of physical examination with the information in their (MRI or CT-myelography) reference standard conclusion, or were unclear with respect to blinding of assessors, resulting in an unclear score.

The most common methodological concerns were with respect to the patient flow and timing. Two studies used different reference tests for some patients (Shabat, 2012; Shah, 2004). One study (Viikari-Juntura, 1989) had too many missing patients and not all included patients received the same reference standard or index test, whereas another study (Apelby-Albrecht, 2013) reported an inappropriate time between reference and index test. Other studies did not report on time between reference and index test.

Please insert Table 2: *tabular presentation for QUADAS-2 results* near here

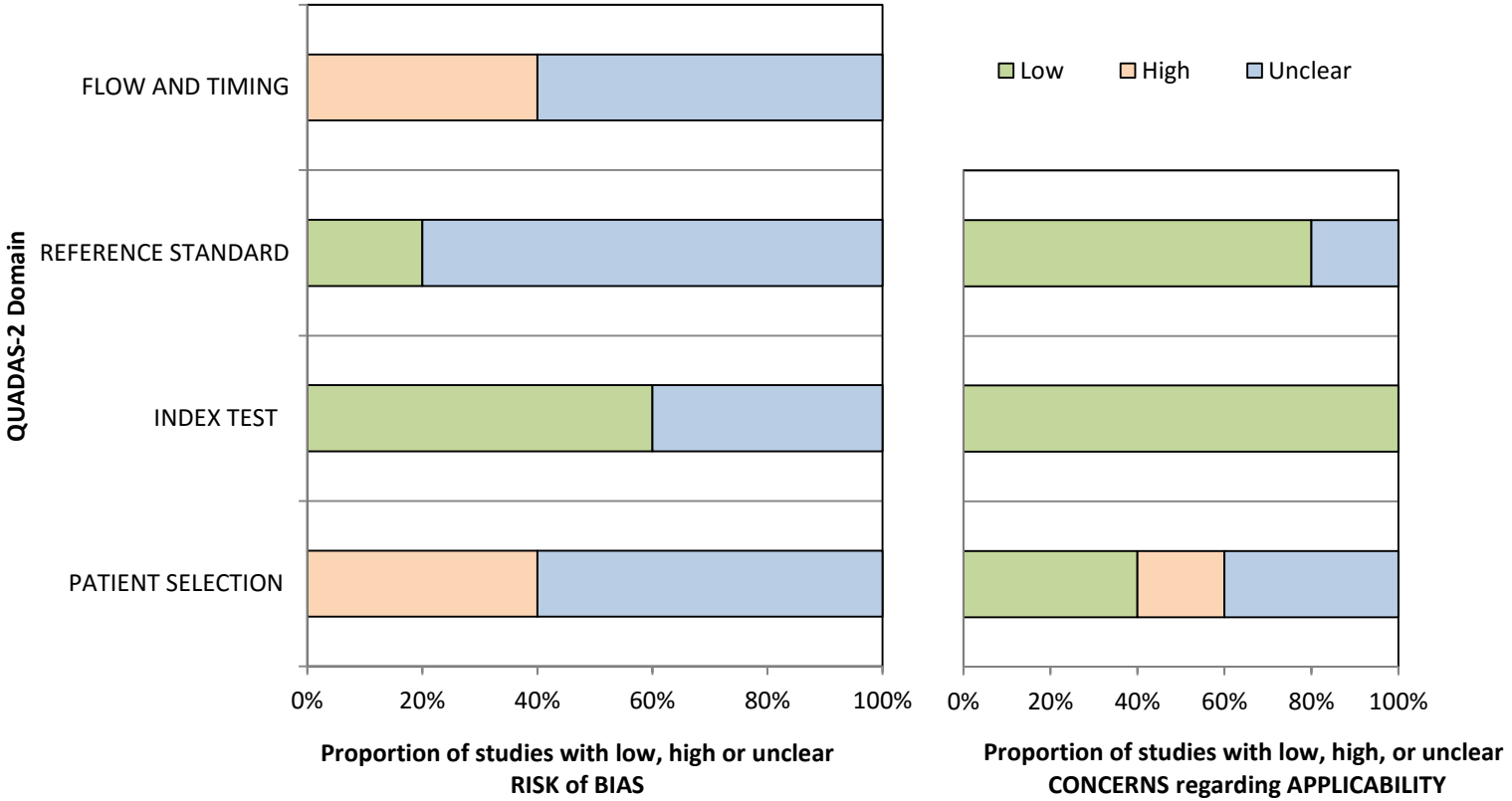
Table 2. Tabular presentation for QUADAS-2 results

Study	RISK OF BIAS				APPLICABILITY CONCERNS		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Apelby-Albrecht	?	☺	☹☺	☹	☺	☺	☺
Gumina	☹	☺	?	?	?	☺	☺
Shabat	?	?	?	?	?	☺	?
Shah	?	?	?	☹	☺	☺	☺
Viikari-Juntura	☹	☺	?	☹	☹	☺	☺


Low Risk
High Risk
? Unclear Risk

Please insert *Figure 2: graphical presentations summarizing QUADAS-2 assessments* near here

Figure 2 QUADAS-2. Proportion of studies with low, high or unclear risk of bias



Findings

Positivity thresholds varied across studies, and some studies presented diagnostic performance of an index test at several different cut-off points. Data were extracted regarding cut-off points most commonly used by studies in the review. There were no disagreements on the extracted data. Results regarding diagnostic accuracy (TP, FP, FN, TN, sensitivity and specificity) from five studies (Apelby-Albrecht, 2013; Gumina, 2013; Shabat, 2012; Shah, 2004; Viikari-Juntura, 1989), all assessing provocative tests is presented in table 3. Descriptions of the execution of the tests are described in table 4.

Please insert Table 3: *Diagnostic accuracy of included studies* near here

Provocative tests:

Spurling's test

Three studies (n=350) evaluated the diagnostic accuracy of the ***Spurling's test***, but all performed a slightly different movements before adding downward axial compression to the cervical spine (Shabat, 2012; Shah, 2004; Viikari-Juntura, 1989). Shah and Rajshekhar reported using cervical extension and ipsilateral lateral flexion (Shah, 2004). Analyses showed a moderate sensitivity and high specificity (Se 0.65, 95% CI: 0.49-0.79; Sp 1.00, 95% CI: 0.56-1.00). Viikari-Juntura et al combined ipsilateral lateral flexion and rotation but did not specify adding cervical extension, although they did depict it as such in their manuscript (Viikari-Juntura, 1989). A moderate sensitivity and high specificity was found (Se 0.38, 95% CI: 0.22-0.56; Sp 0.94, 95% CI: 0.83-0.99).

Shabat et al used cervical extension combined with ipsilateral rotation and used two different positive test results (Shabat, 2012). Evaluation showed a high sensitivity and specificity. The proposed test could either provoke "true radicular symptoms": radiating into the upper extremity along the distribution of a specific dermatome (Se 0.98, 95% CI: 0.92-0.99; Sp 0.89, 95% CI: 0.77-0.96) or a nonspecific radicular pain that radiated to the scapula or occiput region (Se 0.99, 95% CI: 0.95-1.00; Sp 0.85, 95% CI: 0.72-0.92). Both outcomes are presented in table 3. Only the radicular symptoms test results are presented in pooling of results (see Figure 3).

Please insert figure 3 *Forest Plot* near here

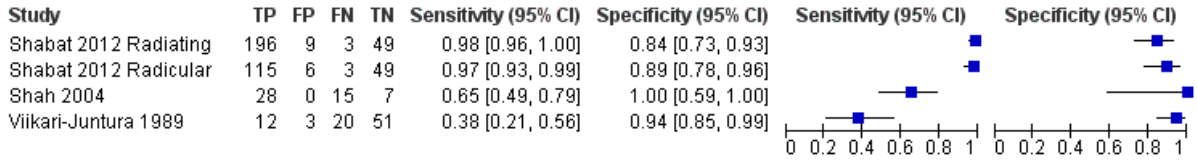


Figure 3 Forest plot – Spurling's test

TP=true positive; FP=false positive; FN=false negative; TN=true negative

Upper Limb Neural Tension test

One study evaluated the concordance of four separate ULNTs (with a bias for the median [ULNT1], radial [ULNT2a and 2b], and ulnar nerve [ULNT3], respectively) as well as the combined results (Apelby-Albrecht, 2013). In this study, a positive test was defined as follows:

- Reproduction of neurogenic pain (defined as: 'burning' or 'lightning like' pain, tingling sensation, according to dermatome pattern in nerve root pathology) in neck and arm and;
- Increased/decreased symptoms with structural differentiation and;
- Difference in painful radiation between right and left sides.

The combined use of four ULNTs had a sensitivity of 0.97 (95% CI: 0.83-1.00) and specificity of 0.69 (95% CI: 0.41-0.88). Individually, the ULNT3 (ulnar) had the highest specificity 0.88 (95% CI: 0.60-0.98) whereas the ULNT1 (median) showed the highest sensitivity of 0.83 (95% CI: 0.66-0.93). One other study set out to evaluate the brachial plexus test but decided to analyze the results due to poor inter-examiner reliability (Viikari-Juntura, 1989).

Shoulder abduction (relief) test

One study evaluated the diagnostic accuracy in 13 patients (Viikari-Juntura, 1989). The authors defined a positive test as when radicular symptoms decreased or disappeared when the patient lifted the affected hand above the head. The study showed a moderate sensitivity of 0.47 (95% CI: 0.22-0.73) and high specificity of 0.85 (95% CI: 0.54-0.97) (Viikari-Juntura, 1989).

Traction test

One study evaluated its diagnostic accuracy of traction in 24 patients (Viikari-Juntura, 1989). The authors defined a positive test as when radicular symptoms decreased or disappeared when an axial traction force of 10-15 kg was applied. A sensitivity of 0.33 (95%C I: 0.13-0.61) and specificity of 0.97 (95% CI: 0.83-0.99) was computed for this test.

Arm Squeeze test

The “arm squeeze test” is a newly devised test working on the proposition that, in the presence of a pathologic compression of a cervical nerve root, one or more nerves of the arm are painful and a moderate compression of the brachial biceps and triceps area should be therefore more painful than other areas of the shoulder and upper arm (Gumina, 2013). The authors defined a positive test as when the pain score (on a 0-10 visual analogue scale) was 3 points or higher during pressure on the middle third of the upper arm, compared with two other (acromioclavicular and anterolateral-subacromial) areas. In trying to differentiate between patients with pain caused by either shoulder pathology or cervical nerve root compression and pain free controls a high sensitivity of 0.97 (95% CI: 0.93-0.98) and specificity of 0.97 (95% CI: 0.95-0.98) was reported (Gumina, 2013).

DISCUSSION

This study summarizes the evidence on the value of specific tests carried out during the physical examination for the diagnosis of cervical radiculopathy confirmed by diagnostic imaging or surgery.

No prospective studies comparing an index test to the findings at surgery were found, although one study (Shah, 2004) did so with a portion of patients and several studies retrospectively reported their clinical findings (Post, 2006; Yoss, 1957). The Spurling’s test was the only test which had the diagnostic accuracy evaluated previously in more than a single study. This seriously limits the level of evidence and also limited the possibility to study the influence of sources of heterogeneity. The sensitivity of Spurling’s test varied from moderate to high while its specificity was high. The recently described Arm Squeeze test showed both high specificity and

sensitivity in the one study in which it is first presented and proposed. The axial traction test and the shoulder abduction test both showed high specificity but moderate sensitivity. The combined ULNTs showed high sensitivity and moderate specificity, with the ULNT3 (ulnar) individually showing high specificity. The included recent study (Apelby-Albrecht, 2013) showed higher specificity than previously reported (Rubinstein, 2007).

No studies were found that assessed the diagnostic accuracy of widely used neurological tests such as key muscle strength, tendon reflexes, and sensory impairments. However eight studies were identified that retrospectively evaluated neurological symptoms before surgical management (23,31,32,34-38).

Factors affecting interpretation

The diagnostic value of physical examination in the diagnosis of cervical radiculopathy can be influenced by many factors, which include the setting in which the examination is performed (primary or secondary care), the characteristics of the study population, the reproducibility (inter-observer variation of the tests), and the reference standard against which the tests are compared (neurophysiological testing, diagnostic imaging or surgical findings).

Population and setting

As all evaluated studies were carried out in a secondary care setting, findings could be an overestimation of diagnostic performance as these studies are more susceptible to selection and verification bias. The large differences in prevalence between studies also have an impact on the accuracy.

Reference tests

Several studies have shown that a substantial proportion of asymptomatic people have disc herniations or degenerative changes on MRI or CT imaging, leading to FPs (Siivola, 2002; Ernst, 2005; Matsumoto, 1998; Okada, 2011). The studies in this review included only symptomatic patients, but none used a meaningful predefined definition of a positive result indicating the relevant presence of a herniated disc or foraminal encroachment with clear nerve root impingement.

Index tests

The large variability in sensitivity of Spurling's test (from 0.38 to 0.98) in three studies (Shabat, 2012; Shah, 2004; Viikari-Juntura, 1989) might be a result of the different ways of executing the procedure, combined with the potential of FPs due to reproducing somatic referred pain from compression of degenerative zygapophyseal joints of a population generally in their fifth or sixth decade of life.

Reliability

Adequate inter- and intra-observer reliability is a prerequisite for good performance of diagnostic tests, but a synthesis of evidence on reliability was not included in the scope of the present review. Our study did show that the procedures for provocative tests were often poorly described and it was not always clear if and what thresholds were used to define positive test results. Only three studies defined a positive test result (Apelby-Albrecht, 2013; Shabat, 2012; Shah, 2004), two studies provided some information on training (Apelby-Albrecht, 2013; Gumina, 2013) and only one, in a related study, on the reliability of examiners (Viikari-Juntura, 1987).

Strengths and Limitations

Studies were only included in this review if they compared the results of tests performed during history taking or physical examination in the identification of cervical radiculopathy, with those of a reference standard of imaging or surgical findings. But because relying only on imaging in a diagnostic process has a risk of an inappropriate high number of FPs (Kuijper, 2011; Siivola, 2002; Ernst, 2005), it can only assist the clinician in his or her clinical reasoning process. We considered a composite reference standard (a combination of history taking, physical examination including neurological assessment and MRI or CT myelography imaging) to be the best available diagnostic gold standard and therefore used this in a tiered scoring of the QUADAS-2. The NASS guideline for the diagnosis and treatment of cervical radiculopathy from degenerative disorders suggests MRI, CT, or CT myelography are suitable for identifying the affected level in patients with cervical radiculopathy, before surgical decompression (Bono, 2011).

Studies using neurophysiological testing (i.e. electromyography [EMG]) as a reference standard, such as the widely referred study of Wainner et al (Wainner, 2003) were excluded. Neurophysiological testing studies the physiological effects of nerve root compression and will thus only be positive if active changes are occurring;

the timing of testing will greatly alter the test's usefulness (Ashkan, 2002). Neurophysiological changes of denervation develop within the first to third week after compression; re-innervation changes may be seen at around 3–6 months. Neurophysiological testing may therefore be negative if performed before denervation has occurred or when re-innervation is complete (Ashkan, 2002). When there is discordance between EMG and MRI findings, EMG might help in the guidance of patient selection for surgical intervention because it provides information of nerve root lesion (Nicotra, 2011). However, a retrospective study reviewing patients operated on for cervical radiculopathy during a 10-year period, concluded neurophysiological testing had limited additional diagnostic value (Ashkan, 2002). A recent study on the diagnostic utility of multiple F-wave variables in the prediction of cervical radiculopathy concluded there was a low correlation between F-wave studies and MRI examinations and could therefore not support its use as such (Lin, 2013). The NASS proposes there is insufficient evidence to make a recommendation for or against the use of EMG for patients in whom the diagnosis of cervical radiculopathy is unclear after clinical examination and MRI (Bono, 2011). So for now, the usefulness of electrodiagnosis is still under debate (Govindarajan, 2013; Kwast Rabben, 2011; Kwast-Rabben, 2013; Reza Soltani, 2014).

Applicability of findings to clinical practice

Although eight studies evaluated neurological symptoms (motor, reflex, and/or sensory changes) as a result of diminished nerve conduction, it is of interest to note that no studies were found that assessed the diagnostic accuracy of these widely used neurological assessment tests.

As there is a paucity of evidence on the diagnostic accuracy of the individual tests, perhaps clustering of the ones that have been studied is a best evidence option for clinicians. Clustering of provocative tests has been proposed to increase diagnostic accuracy (Guttmann, 2015). It also more closely reflects how many clinicians make decisions because it takes into account a number of findings from the clinical assessment. The goal when clustering tests is to determine the best combination estimates that produce the strongest likelihood ratios and to do so, multivariate modeling is required. Due to the limited number of studies this study retrieved, multivariate regression is not yet an option. A test item cluster has been proposed for indicating a cervical radiculopathy (Wainner, 2003). From the results of our review,

this study proposes that, when consistent with history and other physical findings, a combination of Spurling's test, axial traction test and Arm Squeeze test can be used to increase the likelihood of a cervical radiculopathy while a negative outcome of combined ULNTs and Arm Squeeze test can be used to rule the disorder out. More high-quality research is needed to further develop a test item cluster and to improve point estimate precision.

Acknowledgements

We would like to acknowledge the invaluable assistance of Mr. Wichor Bramer, biomedical information specialist of the Erasmus MC Medical Library, Rotterdam, the Netherlands.

REFERENCES

1 Kuijper B, Tans JTJ, Beelen A, et al. Cervical collar or physiotherapy versus wait and see policy for recent onset cervical radiculopathy: randomised trial. Br Med J. 2009 Oct;339. PubMed PMID: WOS:000270719200001.

2 Thoomes EJ, Scholten-Peeters GG, de Boer AJ, et al. Lack of uniform diagnostic criteria for cervical radiculopathy in conservative intervention studies: a systematic review. Eur Spine J. 2012 Apr 25. PubMed PMID: 22531897. Eng.

3 Bogduk N. On the definitions and physiology of back pain, referred pain, and radicular pain. Pain. 2009 Dec 15;147(1-3):17-9. PubMed PMID: 19762151. eng.

4 Merskey H BN. Classification of chronic pain. Descriptions of chronic pain syndromes and definitions of pain terms. 2nd ed. Merskey H BN, editor. Seattle: IASP Press; 1994.

5 Radhakrishnan K, Litchy WJ, O'Fallon WM, Kurland LT. Epidemiology of cervical radiculopathy. A population-based study from Rochester, Minnesota, 1976 through 1990. Brain. 1994 Apr;117 (Pt 2):325-35. PubMed PMID: 8186959. eng.

6 Kim HJ, Nemani VM, Piyaskulkaew C, et al. Cervical Radiculopathy: Incidence and Treatment of 1,420 Consecutive Cases. Asian spine journal. 2016 Apr;10(2):231-7. PubMed PMID: 27114762. Pubmed Central PMCID: PMC4843058. Epub 2016/04/27. Eng.

7 Bussieres AE, Taylor JA, Peterson C. Diagnostic imaging practice guidelines for musculoskeletal complaints in adults-an evidence-based approach-part 3: spinal disorders. Journal of manipulative and physiological therapeutics. 2008 Jan;31(1):33-88. PubMed PMID: 18308153. eng.

8 Bono CM, Ghiselli G, Gilbert TJ, et al. An evidence-based clinical guideline for the diagnosis and treatment of cervical radiculopathy from degenerative disorders. Spine J. 2011 Jan;11(1):64-72. PubMed PMID: 21168100. eng.

9 Rubinstein SM, Pool JJ, van Tulder MW, et al. A systematic review of the diagnostic accuracy of provocative tests of the neck for diagnosing cervical radiculopathy. Eur Spine J. 2007 Mar;16(3):307-19. PubMed PMID: 17013656. eng.

10 Wainner RS, Gill H. Diagnosis and nonoperative management of cervical radiculopathy. The Journal of orthopaedic and sports physical therapy. 2000 Dec;30(12):728-44. PubMed PMID: 11153552. eng.

11 Spurling RG SW. Lateral rupture of the cervical intervertebral disks: a common cause of shoulder and arm pain. Surg Gynecol Obstet. 1944;78:350-8.

12 Davidson RI, Dunn EJ, Metzmaker JN. The shoulder abduction test in the diagnosis of radicular pain in cervical extradural compressive monoradiculopathies. Spine. 1981 Sep-Oct;6(5):441-6. PubMed PMID: 7302677. eng.

13 Elvey RL. Physical evaluation of the peripheral nervous system in disorders of pain and dysfunction. J Hand Ther. 1997 Apr-Jun;10(2):122-9. PubMed PMID: 9188031. eng.

14 Jull G. Grieve's modern musculoskeletal physiotherapy. Jull GA, editor. Edinburgh ; New York: Elsevier; 2015.

15 Elvey R. Brachial plexus tension tests for the pathoanatomical origin of arm pain in "Aspects of Manipulative Therapy" Melbourne, Australia: Lincoln Institute of Health Sciences; 1979.

16 Butler D. The upper limb tension test revisited. In: Clinics in physical therapy. Physical therapy of the cervical thoracic spine. 2nd ed. Edinburgh: Churchill Livingstone; 1994.

17 Ellenberg MR, Honet JC, Treanor WJ. Cervical radiculopathy. Archives of physical medicine and rehabilitation. 1994 Mar;75(3):342-52. PubMed PMID: 8129590. eng.

18 Nordin M, Carragee EJ, Hogg-Johnson S, et al. Assessment of neck pain and its associated disorders: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. Spine. 2008 Feb 15;33(4 Suppl):S101-22. PubMed PMID: 18204385. eng.

19 Malanga GA. The diagnosis and treatment of cervical radiculopathy. Medicine and science in sports and exercise. 1997 Jul;29(7 Suppl):S236-45. PubMed PMID: 9247921. eng.

20 Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of internal medicine. 2011 Oct 18;155(8):529-36. PubMed PMID: 22007046. eng.
21 Kuijper B, Tans JTJ, van der Kallen BF, et al. Root compression on MRI compared with clinical findings in patients with recent onset cervical radiculopathy. J Neurol Neurosurg Psychiatry. 2011 May;82(5):561-3. PubMed PMID: WOS:000289234500020.
22 Siivola SM, Levoska S, Tervonen O, et al. MRI changes of cervical spine in asymptomatic and symptomatic young adults. Eur Spine J. 2002 Aug;11(4):358-63. PubMed PMID: 12193998. eng.
23 Ernst CW, Stadnik TW, Peeters E, et al. Prevalence of annular tears and disc herniations on MR images of the cervical spine in symptom free volunteers. European journal of radiology. 2005 Sep;55(3):409-14. PubMed PMID: 16129249. eng.
24 Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic reviews of diagnostic test accuracy. Annals of internal medicine. 2008 Dec 16;149(12):889-97. PubMed PMID: 19075208. eng.
25 Apelby-Albrecht M, Andersson L, Kleiva IW, et al. 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 Ther. 2013;36(9):626-32.
26 Gumina S, Carbone S, Albino P, et al. Arm Squeeze Test: a new clinical test to distinguish neck from shoulder pain. EUR SPINE J. 2013 Jul;22(7):1558-63. PubMed PMID: 23604976.
27 Shabat S, Leitner Y, David R, Folman Y. The Correlation between Spurling Test and Imaging Studies in Detecting Cervical Radiculopathy. J Neuroimaging. 2012;22(4):375-8.
28 Shah KC, Rajshekhar V. Reliability of diagnosis of soft cervical disc prolapse using Spurling's test. Br J Neurosurg. 2004;18(5):480-3.
29 Viikari-Juntura E, Porras M, Laasonen EM. Validity of clinical tests in the diagnosis of root compression in cervical disc disease. SPINE. 1989;14(3):253-7.
30 Chen TY. The clinical presentation of uppermost cervical disc protrusion. Spine. 2000;25(4):439-42.
31 Conradie M, Bester MM, Crous LC, Kidd M. Do clinical features and MRI suggest the same nerve root in acute cervical radiculopathy? South African Journal of Physiotherapy. 2006;62(2):12-7. PubMed PMID: 2009967697. Language: English. Entry Date: 20080822. Revision Date: 20091218. Publication Type: journal article.
32 Post NH, Cooper PR, Frempong-Boadu AK, Costa ME. Unique features of herniated discs at the cervicothoracic junction: clinical presentation, imaging, operative management, and outcome after anterior decompressive operation in 10 patients. Neurosurgery. 2006 Mar;58(3):497-501; discussion 497-501. PubMed PMID: 16528189.
33 Rainville J, Noto DJ, Jouve C, Jenis L. Assessment of forearm pronation strength in C6 and C7 radiculopathies. Spine. 2007;32(1):72-5.
34 Yoss RE, Corbin KB, Maccarty CS, Love JG. Significance of symptoms and signs in localization of involved root in cervical disk protrusion. Neurology. 1957 Oct;7(10):673-83. PubMed PMID: 13477342.
35 Rainville J, Laxer E, Keel J, et al. Exploration of sensory impairments associated with C6 and C7 radiculopathies. Spine J. 2016 Jan;16(1):49-54. PubMed PMID: WOS:000366656100031. English.
36 Henderson CM, Hennessy RG, Shuey HM, Jr., Shackelford EG. Posterior-lateral foraminotomy as an exclusive operative technique for cervical radiculopathy: a review of 846 consecutively operated cases. Neurosurgery. 1983 Nov;13(5):504-12. PubMed PMID: 6316196. eng.
37 Dinnes J, Deeks J, Kirby J, Roderick P. A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. Health technology assessment (Winchester, England). 2005 Mar;9(12):1-113, iii. PubMed PMID: 15774235. eng.

38 Tanaka Y, Kokubun S, Sato T, Ozawa H. Cervical roots as origin of pain in the neck or scapular regions. Spine (Phila Pa 1976). 2006 Aug 1;31(17):E568-73. PubMed PMID: 16924193.
39 Wainner RS, Fritz JM, Irrgang JJ, et al. Reliability and diagnostic accuracy of the clinical examination and patient self-report measures for cervical radiculopathy. Spine. 2003;28(1):52-62.
40 Rubinstein SM, Pool JJM, Van Tulder MW, et al. A systematic review of the diagnostic accuracy of provocative tests of the neck for diagnosing cervical radiculopathy. Eur Spine J. 2007;16(3):307-19.
41 Matsumoto M, Fujimura Y, Suzuki N, et al. MRI of cervical intervertebral discs in asymptomatic subjects. J Bone Joint Surg Br. 1998 Jan;80(1):19-24. PubMed PMID: 9460946. eng.
42 Okada E, Matsumoto M, Fujiwara H, Toyama Y. Disc degeneration of cervical spine on MRI in patients with lumbar disc herniation: comparison study with asymptomatic volunteers. Eur Spine J. 2011 Apr;20(4):585-91. PubMed PMID: 21127918. Pubmed Central PMCID: PMC3065617. Epub 2010/12/04. eng.
43 Viikari-Juntura E. Interexaminer reliability of observations in physical examinations of the neck. PHYS THER. 1987;67(10):1526-32.
44 Bono CM, Ghiselli G, Gilbert TJ, Kreiner DS. An evidence-based clinical guideline for the diagnosis and treatment of cervical radiculopathy from degenerative disorders. The Spine Journal. 2011.
45 Ashkan K, Johnston P. A comparison of magnetic resonance imaging and neurophysiological studies in the assessment of cervical radiculopathy. British journal of …. 2002.
46 Nicotra A, Khalil NM, O'Neill K. Cervical radiculopathy: discrepancy or concordance between electromyography and magnetic resonance imaging? British journal of neurosurgery. 2011 Dec;25(6):789-90. PubMed PMID: 21899381. eng.
47 Lin CH, Tsai YH, Chang CH, et al. The comparison of multiple F-wave variable studies and magnetic resonance imaging examinations in the assessment of cervical radiculopathy. Am J Phys Med Rehabil. 2013 Sep;92(9):737-45. PubMed PMID: 23291601.
48 Govindarajan R, Kolb C, Salgado E. Sensitivity and specificity of MRI and EMG in diagnosing clinically evident cervical radiculopathy: A retrospective study. Neurology. 2013;80(1).
49 Kwast Rabben O, Heikkila H, Fagerlund M. Specificity and sensitivity of electromyography in evaluation of C6 and C7 root involvement in patients with symptomatic cervical spine disorders. Correlation analysis between electromyography and magnetic resonance imaging. Clin Neurophysiol. 2011;122:S78.
50 Kwast-Rabben O, Heikkila H, Fagerlund M. Electromyography (EMG) and magnetic resonance imaging (MRI) in evaluation of root injury in symptomatic cervical spine disorders (CSD). Specificity and sensitivity test. J Neurol. 2013;260:S158.
51 Reza Soltani Z, Sajadi S, Tavana B. A comparison of magnetic resonance imaging with electrodiagnostic findings in the evaluation of clinical radiculopathy: a cross-sectional study. Eur Spine J. 2014 Apr;23(4):916-21. PubMed PMID: 24413775. Pubmed Central PMCID: PMC3960424. Epub 2014/01/15. eng.
52 Guttmann A, Li X, Feschet F, et al. Cluster Detection Tests in Spatial Epidemiology: A Global Indicator for Performance Assessment. PloS one. 2015 06/18 10/06/received
05/22/accepted;10(6):e0130594. PubMed PMID: PMC4472237.
53 Wainner RS, Fritz JM, Irrgang JJ, et al. Reliability and diagnostic accuracy of the clinical examination and patient self-report measures for cervical radiculopathy. Spine. 2003 Jan 1;28(1):52-62. PubMed PMID: 12544957.

Figure 1. PRISMA Flow Diagram of included studies

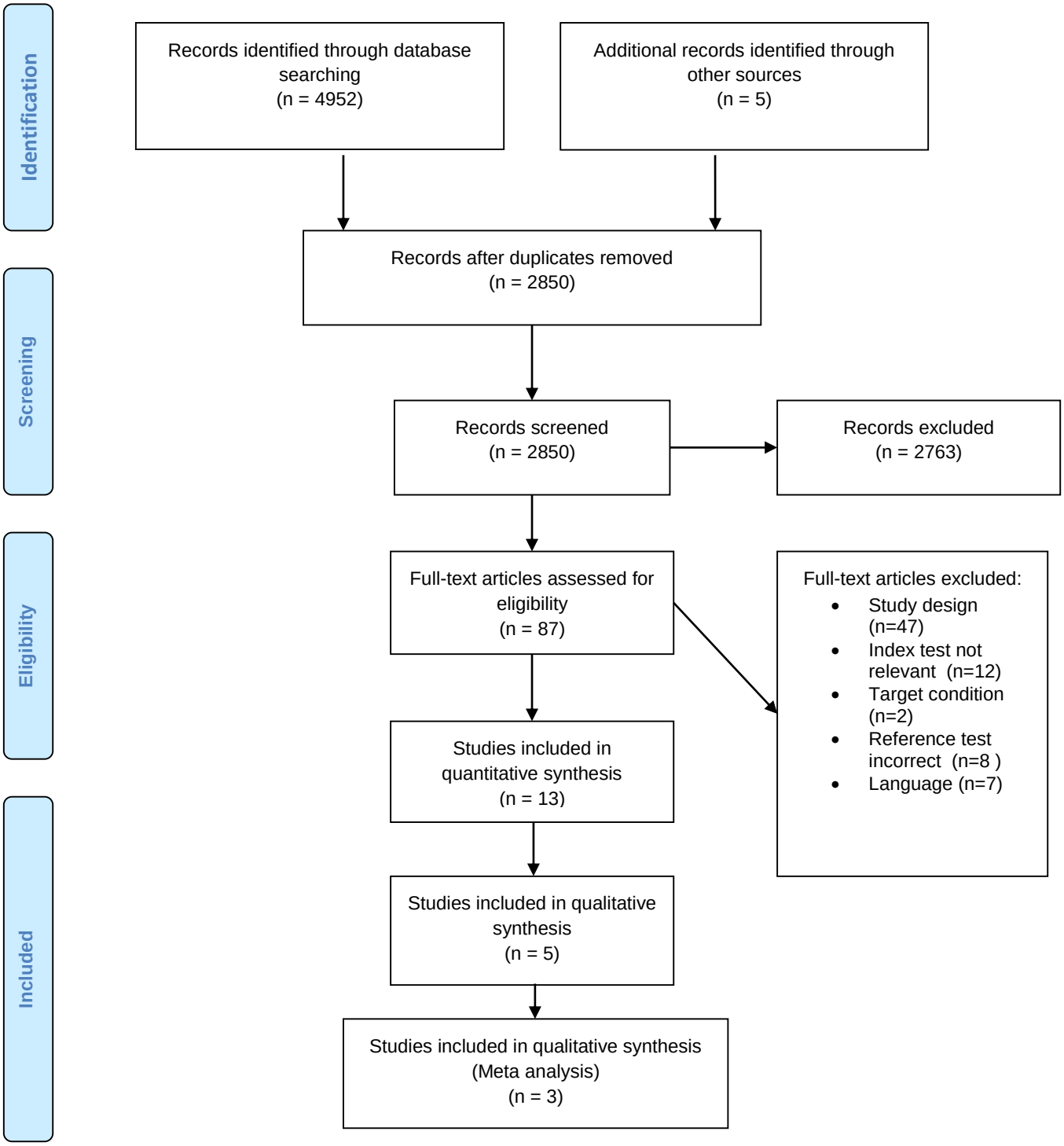


Table 1. Characteristics of included diagnostic accuracy studies

Author /year	**Apelby-Albrecht, 2013**
Clinical Feature and setting	Center for spinal surgery, Sweden
Participants	51 consecutive patients referred for clinical investigation of cervical and/or arm pain
Study design	Diagnostic cohort study
Target condition and Reference standard(s)	Patients with cervical and/or arm pain; MRI, medical history, and clinical examination (dermatomes, reflex testing and Spurling's test), in patients with cervical radiculopathy.
Index and comparator tests	4 Upper Limb Neurodynamic Tests: ULNT1 (median), ULNT2a (median), ULNT2b (radial) and ULNT3 (ulnar)
Notes	

Author /year	**Gumina, 2013**
Clinical Feature and setting	Shoulder Clinical Office and Orthopedic Spine Ambulatory. Italy
Participants	1,567 patients with pain localized at the shoulder girdle
Study design	Cohort study
Target condition and Reference standard(s)	Patients with neck and arm pain; Clinical examination of the cervical spine, of the shoulder and of the upper limb; electromyography (for C5 to T1 roots); X-rays (AP and lateral view); MRI of the cervical spine
Index and comparator tests	Arm Squeeze test
Notes	

Author /year	**Shabat, 2012**
Clinical Feature and setting	Spinal Surgery Unit, Israel
Participants	257 patients with symptoms of unilateral cervical radiculopathy lasting for at least 4 weeks.
Study design	Cohort study
Target condition and Reference standard(s)	Unilateral cervical radiculopathy lasting for at least 4 weeks; Complete physical examination for range of motion, motor and sensory examination, and reflex examination.
Index and comparator tests	Spurling's test (extension+ rotation + axial compression) and physical examination for range of motion, motor and sensory examination, and reflex examination
Notes	Patients were divided into 3 groups: 1) true positive test (radicular pain radiating into the upper extremity, along the distribution of a specific dermatome; 2) negative test; 3) elicting nonspecific radicular pain radiating to scapular or occipital region.

Author /year	Shah, 2004
Clinical Feature and setting	Neurosurgical unit, India
Participants	50 patients with neck and arm pain suggestive of radicular pain
Study design	Prospective cohort study
Target condition and Reference standard(s)	Patients with neck & arm pain suggestive of radicular pain; MRI, the effective root canal diameter was measured at the entry point of root in the canal on T2W axial MR image at the level of the disc prolapse and compared with that of the unaffected side.
Index and comparator tests	Spurling's test: extension + lateral flexion towards involved side + axial pressure
Notes	

Author /year	Viikari-Juntura, 1989
Clinical Feature and setting	Neurosurgery department Finland
Participants	69 patients sent for cervical myelography
Study design	Prospective cohort study
Target condition and Reference standard(s)	Cervical disc disease (spondylosis and/or disc herniation); Cervical myelography combined with conventional neurological examination (sensory, motor and reflex testing)
Index and comparator tests	Spurling's test: (lateral flexion,+ rotation+ axial compression); cervical distraction and shoulder abduction relief (Davidson's test)
Notes	Brachial plexus tension test discarded due to poor inter-examiner reliability, although only one rater examined.

Table 3 Diagnostic accuracy of included studies

Author, year, N	Reference test(s)	Index Test(s)	TP	FP	FN	TN	Sens (95%CI)	Spec (95%CI)	LR+ (95%CI)	LR- (95%CI)	PPV	NPV	Prevalence
Apelby-Albrecht, 2013, n=51	MRI	Upper Limb Neural Tension tests:											0.69 (0.54-0.81)
		ULNT1 median	29	4	6	12	0.83 (0.66-0.93)	0.75 (0.48-0.93)	3.31 (1.40-7.85)	0.23 (0.10-0.50)	0.88 (0.71-0.96)	0.67 (0.41-0.86)	
		ULNT2a median	23	4	12	12	0.66 (0.48-0.80)	0.75 (0.47-0.92)	2.63 (1.09-6.35)	0.46 (0.28-0.75)	0.85 (0.65-0.95)	0.50 (0.29-0.71)	
		ULNT2b radial	15	4	20	12	0.43 (0.27-0.60)	0.75 (0.47-0.92)	1.71 (0.68-4.35)	0.76 (0.55-1.06)	0.79 (0.54-0.93)	0.38 (0.22-0.56)	
		ULNT3 ulnar	25	2	10	14	0.71 (0.54-0.85)	0.88 (0.60-0.98)	5.71 (1.54-21.24)	0.33 (0.19-0.56)	0.93 (0.74-0.99)	0.58 (0.37-0.77)	
		Combined 4 ULNTs	34	5	1	11	**0.97 (0.83-1.00)**	**0.69 (0.41-0.88)**	**3.10 (1.50-6.44)**	**0.04 (0.01-0.30)**	**0.87 (0.72-0.95)**	**0.92 (0.59-1.00)**	
Gumina, 2013, n=1567	MRI	Arm Squeeze test	295	43	10	1219	0.97 (0.93-0.98)	0.97 (0.95-0.98)	28.39 (21.15-38-11)	0.03 (0.02-0.06)	0.87 (0.83-0.91)	0.99 (0.98-0.99)	0.20 (0.18-0.22)
Shabat, 2012, n=257	MRI/ CT	Spurling's test (Ext+Rot): radicular pain	115	6	3	49	0.98 (0.92-0.99)	0.89 (0.77-0.96)	8.93 (4.20-19.02)	0.03 (0.01-0.09)	0.95 (0.89-0.98)	0.94 (0.83-0.99)	0.68 (0.61-0.75)
		Spurling's test: radiating pain	196	9	3	49	0.99 (0.95-1.00)	0.85 (0.72-0.92)	6.35 (3.48-11.57)	0.02 (0.01-0.06)	0.96 (0.92-0.98)	0.94 (0.83-0.99)	0.77 (0.72-0.82)
Shah, 2004, n=50	MRI/ operation	Spurling's test (Ext+LF)	28	0	15	7	0.65 (0.49-0.79)	1.00 (0.56-1.00)	n/a	0.35 (0.23-0.52)	1.00 (0.85-1.00)	0.32 (0.15-0.55)	0.86 (0.73-0.94)
Viikari-Juntura, 1989, n=43	Myelogram	Spurling's test (LF+Rot), n=43:	12	3	20	51	0.38 (0.22-0.56)	0.94 (0.83-0.99)	6.75 (2.06-22.13)	0.67 (0.50-0.87)	0.86 (0.56-0.98)	0.80 (0.51-0.95)	0.37 (0.27-0.48)
		Traction, n=24:	5	1	10	32	0.33 (0.13-0.61)	0.97 (0.83-0.99)	11.00 (1.40-86.17)	0.69 (0.48-0.98)	0.83 (0.37-0.99)	0.76 (0.60-0.87)	0.31 (0.19-0.46)
		Shoulder ABd test, n=13:	7	2	8	11	0.47 (0.22-0.73)	0.85 (0.54-0.97)	3.03 (0.76-12.12)	0.63 (0.38-1.04)	0.78 (0.40-0.96)	0.58 (0.34-0.79)	0.54 (0.34-0.72)

Table 4. Characteristics of included retrospective studies

Author /year	**Conradie, 2006**
Clinical Feature and setting	Department of Physiotherapy, Faculty of Health Sciences, Stellenbosch University, South Africa.
Participants	A convenience sample of 21 consecutive patients referred from private medical practices to a neurosurgeon.
Study design	Prospective study.
Target condition and reference standard(s)	Acute cervical radiculopathy confirmed by MRI.
Index and comparator tests	Distribution patterns of clinical features: motor weakness, pain and paresthesia.
Notes	All included had been diagnosed with cervical radiculopathy before index tests were applied. Overall weakness of key muscles showed a sensitivity of 0.71 (95%CI: 0.51-0.86) and specificity of 0.94 (95%CI: 0.80-0.99); paresthesia a sensitivity of 0.61 (95%CI: 0.41-0.78) and specificity of 0.89 (95%CI: 0.72-0.96) and pain a sensitivity of 0.71 (95%CI: 0.51-0.86) and specificity of 0.80 (95%CI: 0.63-0.91).

Author /year	**Chen, 2000**
Clinical Feature and setting	Department of Neurosurgery, Chang Gung University and Chang Gung Memorial Hospital, Taoyuan, Taiwan.
Participants	8 patients with C2-C3 disc herniation, in 7 as a result of a cranio-vertebral injury.
Study design	Retrospective study
Target condition and reference standard(s)	C2–C3 cervical disc herniation
Index and comparator tests	Motor muscle action & deep tendon reflex decrease and sensory distribution
Notes	All patients had developed a lack of fine motor control of the hands and complained more commonly of sensory changes than of motor and reflex changes. Hypesthesia was more common (87.5%) than allodynia /hyperesthesia (14%) or propioception loss (14%). Three patients (38%) complained of difficulty walking and loss of balance. Decreased muscle power varied from upper limbs to lower limbs. Presence of Hoffman's sign was identified in three patients (38%), positive Lhermitte's sign and Spurling's test were found in five patients (63%).

Author /year	**Henderson, 1983**
Clinical Feature and setting	Dept. of Neurosurgery, University of Maryland Hospital, Baltimore, USA.
Participants	736 cervical radiculopathy patients surgically managed with posterior-lateral foraminotomies.
Study design	Retrospective study
Target condition and reference standard(s)	Simple cervical radiculopathy confirmed by pantopaque myelography
Index and comparator tests	Distribution of pain, paresthesia and muscle weakness
Notes	In 465 cases (53.9%) a dermatomal pattern of pain and/or paresthesia was noted, in 385 (45.5%) a diffuse non-dermatomal pattern and in 5 cases (0.6%) no arm pain at all. In 567 (68%)

	cases a specific motor weakness and in 270 (32%) no deficits were noted. In 602 patients (71.2%) a specific decreased tendon reflex was recorded.

Author /year	**Kuijper, 2011**
Clinical Feature and setting	Department of Neurology, Medical Centre Haaglanden, The Hague, The Netherlands.
Participants	82 patients in whom efficacy of either a cervical collar or physiotherapy was compared with a wait-and-see policy.
Study design	Prospective cohort study.
Target condition and reference standard(s)	Recent onset unilateral cervical radiculopathy.
Index and comparator tests	MRI
Notes	29.5% of cases showed signs of muscle weakness, 48,7% showed diminished reflexes and 89,7% showed sensory abnormalities. Two patients (2,6%) only reported having pain.

Author /year	**Post, 2006**
Clinical Feature and setting	Department of Neurosurgery, New York University School of Medicine, New York, USA.
Participants	10 cases of C7-T1 radiculopathy from a cohort of 268 surgically managed patients.
Study design	Retrospective study.
Target condition and reference standard(s)	C7-T1 radiculopathy.
Index and comparator tests	Motor function of hand intrinsic muscles, finger flexors and finger extensor muscles.
Notes	Nine out of ten patients had hand weakness, generally consistent with C8 nerve root dysfunction as well as shoulder pain radiating into the lateral aspect of the hand. No patient complained of neck pain or paresthesias in the hands.

Author /year	**Rainville, 2007**
Clinical Feature and setting	New England Baptist Hospital, Boston, USA.
Participants	55 consecutive patients with clinical radiculopathies.
Study design	Consecutive case series
Target condition and reference standard(s)	C6 or C7 radiculopathies confirmed by either MRI or CT.
Index and comparator tests	Pronation strength.
Notes	Forearm pronation weakness was present in 72%, but in only 23% of subjects with C7 radiculopathy.

Author /year	**Rainville, 2016**
Clinical Feature and setting	New England Baptist Hospital, Boston, USA.
Participants	55 consecutive patients with clinical radiculopathies.
Study design	Consecutive case series
Target condition and reference standard(s)	C6 or C7 radiculopathies confirmed by either MRI or CT.
Index and comparator tests	Sensory deficits.
Notes	The location of sensory impairments associated with symptomatic C6 and C7 nerve root compression overlap to the extent that caution should be exercised when predicting compression of either

	the C6 or C7 nerve roots based on locations of impaired sensation.

Author /year	**Yoss,1957**
Clinical Feature and setting	Neurology department Mayo clinic, Rochester, USA.
Participants	79 patients complaining of scapular and/or interscapular pain.
Study design	Retrospective study.
Target condition and reference standard(s)	Surgically managed C5, C6, C7or C8 nerve root compression.
Index and comparator tests	Distribution of pain, paresthesia and muscle weakness.
Notes	39% of patients with C7 or C8 nerve root involvement complained of scapular and/or interscapular pain, 100% of patients with C5 and 80% of patients with C6 radiculopathy complained of pain the lateral aspect of the forearm. Diminished reflexes of biceps and brachioradialis combined were noted in 50% of cases with C5 and 32% of cases with C6 radiculopathy, diminished reflex of triceps in 65% of C7 and 60% of C8 radiculopathy respectively.

Appendix A
Embase search strategy

('cervicobrachial neuralgia'/de OR 'brachial plexus neuropathy'/de OR myeloradiculopathy/de OR 'cervical spondylosis'/de OR 'cervical myelopathy'/exp OR (('spinal cord compression'/de OR 'intervertebral disk hernia'/de OR 'vertebral canal stenosis'/de OR 'intervertebral disk degeneration'/de OR stenosis/de OR 'vertebral canal stenosis'/de OR spondylosis/de OR radiculopathy/de OR 'nerve root compression'/de) AND (neck/exp OR 'neck pain'/exp OR 'neck injury'/de OR 'cervical spine'/exp OR 'Cervical Plexus'/de OR 'cervical spine injury'/de OR 'cervical spinal cord'/exp OR 'cervical spinal cord injury'/exp OR 'cervical vertebral canal'/de)) OR (((cervic* OR brachial*) NEAR/3 (neuralg* OR compress* OR radiculop* OR avulsion* OR radiculitis* OR radiculitides* OR syndrome* OR myelopath* OR spondylos* OR osteophytos* OR stenosis* OR degenerat* OR neuritis*)) OR cervicobrachial* OR 'arm neck shoulder' OR 'shoulder arm neck' OR 'neck shoulder arm' OR myeloradicul* OR radiculomyel*):ab,ti) AND ('diagnostic accuracy'/exp OR 'predictive value'/exp OR differentiation/de OR 'differential diagnosis'/exp OR 'diagnostic error'/exp OR recognition/de OR 'sensitivity and specificity'/exp OR 'delayed diagnosis'/de OR 'cognitive bias'/exp OR 'statistical bias'/exp OR reliability/exp OR validity/exp OR 'validation study'/exp OR reproducibility/de OR (((diagnos* OR detect* OR test*) NEAR/6 (accur* OR inaccura* OR possibil* OR error* OR fail* OR advantag* OR better* OR best OR worse* OR worst OR unsuspect* OR qualit* OR poor OR identif* OR utilit* OR adequa* OR inadequa* OR delay* OR appropriat* OR inappropriat* OR pitfall* OR challenge* OR difficul* OR confus* OR effectiv* OR prefer* OR superior* OR inferior* OR missed OR bias*)) OR (predict* NEAR/3 value*) OR differentia* OR misdiagnos* OR undiagnos* OR underdiagnos* OR recogni* OR unrecogni* OR underrecogni* OR ((under OR un OR mis*) NEXT/1 (diagnos* OR recogni*)) OR reliab* OR valid* OR reproducib* OR sensitivit* OR specificit* OR insensitiv* OR unspecific* OR asensitiv* OR aspecific* OR ((positive* OR negative*) NEAR/3 (false* OR true*)) OR ((observer* OR interobserver* OR intraobserver* OR intrarater* OR interrater* OR rater*) NEAR/3 (varia* OR agree* OR bias*))):ab,ti) AND ('physical examination'/de OR 'physical medicine'/exp OR physiotherapist/de OR 'medical examination'/exp OR provocation/de OR 'provocation test'/de OR 'movement (physiology)'/exp OR reflex/de OR 'tendon reflex'/de OR 'manipulative medicine'/exp OR 'sensory dysfunction'/de OR 'abnormal sensation'/de OR 'Valsalva maneuver'/de OR ('foramen magnum'/de AND compression/de) OR 'traction therapy'/exp OR (physical* OR (medical* NEAR/3 examin*) OR provocat* OR movement* OR abduction* OR motion* OR (tendon* NEAR/3 reflex*) OR manipulat* OR manual* OR ((sensor* OR sensat*) NEAR/3 (dysfunction* OR abnormal*)) OR (Valsalva* NEXT/1 maneuv*) OR (foram* NEAR/3 compress*) OR ((spurling* OR relief* OR Davidson* OR traction* OR distraction*) NEAR/3 test*) OR ('Upper Limb' NEXT/1 (Tension* OR nerve*)) OR physiotherap*):ab,ti)

Appendix B
QUADAS-2 Signaling questions

Phase 1: please state the review question:

Patients (setting, intended use of index test, presentation, prior testing): Patients with radicular arm and neck pain in primary or secondary care
Index test(s): specific tests carried out during the physical examination for the diagnosis of cervical radiculopathy: i.e.: Spurling's, Valsalva, ULNT, Shoulder abduction relief, traction, reflex, key muscles
Reference standard and target condition: (Physical examination combined with) MRI / CT and or surgery

Phase 2: Draw a flow diagram for the primary study

Phase 3: Risk of bias and applicability judgments

QUADAS-2 is structured so that 4 key domains are each rated in terms of the risk of bias and the concern regarding applicability to the research question (as defined above). Each key domain has a set of signalling questions to help reach the judgments regarding bias and applicability.

DOMAIN 1: PATIENT SELECTION

A. Risk of Bias

Describe methods of patient selection:
Please describe the method as you understand it from the description in the manuscript.

From the answers to the signaling questions below, please derive a final score. The lowest score should determine the final score

- Was a consecutive or random sample of patients enrolled? Yes/No/Unclear
- Was a case-control design avoided? Yes/No/Unclear
- Did the study avoid inappropriate exclusions? Yes/No/Unclear

Could the selection of patients have introduced bias? RISK: LOW / HIGH / UNCLEAR

B. Concerns regarding applicability

Describe included patients (prior testing, presentation, intended use of index test and setting)**:**
Please describe the included patients as you understand them from the description in the manuscript.

Is there concern that the included patients do not match the review question?

Do you feel the included patients might have disorders not related to the review question? Eg. if the objective is to differentiate between NonSpecificArmPain and cervical radiculopathy that is okay. But not so if the included patients might have completely unrelated disorders or have a spectrum of the

disorder too different from the review question

CONCERN: LOW / HIGH / UNCLEAR

DOMAIN 2: INDEX TEST(S)

If more than one index test was used, please complete for each test

A. Risk of Bias

Describe the index test and how it was conducted and interpreted:

Please describe the index test(s) and the manner af applying them as you understand it from the description in the manuscript

From the answers to the signaling questions below, please derive a final score. The lowest score should determine the final score

- Were the index test results interpreted without knowledge of the results of the reference standard? Yes /No /Unclear
- If a threshold was used, was it pre-specified? Yes /No /Unclear

Could the conduct or interpretation of the index test have introduced bias?

RISK: LOW / HIGH / UNCLEAR

B. Concerns regarding applicability

Is there concern that the index test, its conduct or interpretation differ from the review question?

Do you feel the index test or its manner of applying or interpreting the outcome (pos/ neg scoring) is too different so the review question cannot be answered from the result?

CONCERN: LOW /HIGH/UNCLEAR

DOMAIN 3: REFERENCE STANDARD

A. Risk of Bias

Describe the reference standard and how it was conducted and interpreted:

Please describe the reference standard(s) and the manner of applying and interpreting the outcome (pos/neg) as you understand it from the description in the manuscript.
In the absence of a true gold standard we state that the combination of a neurological examination (consisting of testing of tendon reflexes, manual muscle testing of key muscles for muscle weakness or atrophy and testing for sensory deficits) and results from MRI/ CT imagingand/or the postoperative results is to be seen as correctly classifying the target condition.
A sole assessment of an MRI / CT (eg. by a radiologist) potentially has too many false positives and is therefore usually to be scored as "Unclear"

From the answers to the signaling questions below, please derive a final score. The lowest score should determine the final score

- Is the reference standard likely to correctly classify the target condition? Yes /No /Unclear
- Where the reference standard results interpreted without knowledge of the results of the index test? Yes /No /Unclear

Could the reference standard, its conduct, or its interpretation have introduced bias?

RISK: LOW / HIGH / UNCLEAR

B. Concerns regarding applicability

Is there concern that the target condition as defined by the reference standard does not

match the review question?

Do you feel the reference test itself or its manner of applying or interpreting the outcome (pos/ neg scoring) is too different so the review question cannot be answered from the result?

CONCERN: LOW / HIGH / UNCLEAR

DOMAIN 4: FLOW AND TIMING

A. Risk of Bias

Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram):

Describe the time interval and any interventions between index test(s) and reference standard:

From the answers to the signaling questions below, please derive a final score. The lowest score should determine the final score

- Was there an appropriate interval **(< 1week)** between index test(s) and reference standard? Yes /No /Unclear
- Did all patients receive a reference standard? Yes /No /Unclear
- Did patients receive the same reference standard? Yes /No /Unclear
- Were all patients included in the analysis? Yes /No /Unclear

Could the patient flow have introduced bias? RISK: LOW /HIGH/UNCLEAR