Value of physical tests in diagnosing cervical radiculopathy:

Thoomes, Erik J; van Geest, Sarita; van der Windt, Danielle A; Falla, Deborah; Verhagen, Arianne P; Koes, Bart W; Thoomes-de Graaf, Marloes; Kuijper, Barbara; Scholten-Peeters, Wendy Gm; Vleggeert-Lankamp, Carmen L

DOI:
10.1016/j.spinee.2017.08.241

Document Version
Peer reviewed version

Citation for published version (Harvard):
Thoomes, EJ, van Geest, S, van der Windt, DA, Falla, D, Verhagen, AP, Koes, BW, Thoomes-de Graaf, M, Kuijper, B, Scholten-Peeters, WG & Vleggeert-Lankamp, CL 2017, 'Value of physical tests in diagnosing cervical radiculopathy: a systematic review', *The Spine Journal*. https://doi.org/10.1016/j.spinee.2017.08.241

Accepted Manuscript

Title: Value of physical tests in diagnosing cervical radiculopathy: a systematic review

Author: Erik J Thoomes, Sarita van Geest, Danielle A van der Windt, Deborah Falla, Arianne P Verhagen, Bart W Koes, Marloes Thoomes-de Graaf, Barbara Kuijper, Wendy GM Scholten-Peeters, Carmen L Vleggeert-Lankamp

PII: S1529-9430(17)30918-X
DOI: http://dx.doi.org/doi: 10.1016/j.spinee.2017.08.241
Reference: SPINEE 57470

To appear in: *The Spine Journal*

Received date: 24-4-2017
Revised date: 7-8-2017
Accepted date: 15-8-2017

Please cite this article as: Erik J Thoomes, Sarita van Geest, Danielle A van der Windt, Deborah Falla, Arianne P Verhagen, Bart W Koes, Marloes Thoomes-de Graaf, Barbara Kuijper, Wendy GM Scholten-Peeters, Carmen L Vleggeert-Lankamp, Value of physical tests in diagnosing cervical radiculopathy: a systematic review, *The Spine Journal* (2017), http://dx.doi.org/doi: 10.1016/j.spinee.2017.08.241.

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[d], PhD; Arianne P Verhagen[e], PhD; Bart W Koes[e], PhD; Marloes Thoomes-de Graaf[a], MSc; Barbara Kuijper[f], MD PhD; Wendy GM Scholten-Peeters[g], PhD; Carmen L Vleggeert-Lankamp[b], MD PhD.

a) Fysio-Experts, Hazerswoude, The Netherlands
b) Leiden University Medical Centre, Dept. Neurosurgery, Leiden, The Netherlands
c) Arthritis Research UK Primary Care Centre, Institute for Primary Care and Health Sciences, Keele Staffordshire, United Kingdom
d) Centre of Precision Rehabilitation for Spinal Pain (CPR Spine), School of Sport, Exercise and Rehabilitation Sciences, University of Birmingham, Birmingham, UK
e) Department of General Practice, Erasmus Medical Centre University, Rotterdam, The Netherlands
f) Maasstad Hospital, Rotterdam, The Netherlands
g) Department of Human Movement Sciences, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam Movement Sciences, The Netherlands.

Word count:

- Abstract: 260
- Manuscript: 4000 (including in-text references)

Tables: 4
Figures: 3

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

Acknowledgements

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

ABSTRACT

Background context

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

Purpose

To summarize and update the evidence on diagnostic performance of tests carried out during a physical examination for the diagnosis of cervical radiculopathy**.**

Study design

Review of the accuracy of diagnostic tests.

Study Sample

Diagnostic studies comparing results of tests performed during a 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. 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 0.89-1.00 (95%CI: 0.59-1.00); sensitivity varied from 0.38-0.97 (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 the 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 negative combined neurodynamic testing and an Arm Squeeze test could be used to rule out the disorder.

Keywords: cervical radiculopathy; diagnostic accuracy; Spurling; shoulder physical examination; Arm Squeeze test; neurodynamic testing

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 for 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 due to inflammation of the nerve (Bogduk, 2009).

A 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, 2007a; Wainner, 2000) include tendon reflexes, manual muscle testing of key muscles for weakness or atrophy and testing for sensory deficits, the assessment of range of motion (ROM) and provocative test like the foraminal compression test or Spurling's test (Spurling RG, 1944), shoulder abduction (relief) test (Davidson, 1981), Upper Limb Tension Test (ULTT) or 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 the physical examination for the identification of cervical radiculopathy (Bono, 2011; Ellenberg, 1994; Nordin, 2008; Rubinstein, 2007a; Wainner, 2000). Two reviews included an assessment of the methodological quality of the primary studies

(Rubinstein, 2007a) 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, 2007a). Clusters of tests were generally considered to be more accurate (Bono., 2011).

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 (Gumina, 2013) or combinations of tests (Apelby-Albrecht, 2013) have been described and a commonly used test (i.e. Spurling's test) has been further assessed (Shabat, 2012). Therefore, this 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. 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 than18 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 (i.e. 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 the 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 the physical examination were compared to 1) diagnostic imaging: MRI or 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 *Appendix A*. No language restrictions were applied. Reference lists of relevant publications were checked for grey literature and a forward citation was performed searching relevant articles using the PubMed related articles feature.

Assessment of methodological quality

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

With respect to the QUADAS-2 risk of bias domain related the reference standard, a tiered scoring system was devised. A combination of history taking, physical examination including neurological assessment and MRI or CT-myelography imaging (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 “unclear” due to the inappropriate high number of false positives (Ernst, 2005; Kuijper, 2011; Siivola, 2002). 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 (Nunally, 1994). 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, false positive, true negative, and false negative 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 (e.g. 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 synthesis. 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 (Portney, 2009). 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 based on the extracted number of true positives [TP], false negatives [FP], true negatives [TN] and false positives [FP]. Results in terms of sensitivity and specificity and 95% CI for each test were presented in a forest-plot. Results were entered into Review Manager 5.3®. 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 where no pooled estimate could be calculated.

Investigations of heterogeneity

Heterogeneity was examined by considering study characteristics, visual inspection of (the confidence intervals 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 5 were retrieved from searching through grey 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 regards to the reasons for exclusion out of these 87 manuscripts. Disagreements were resolved through minor discussion and arbitration through the 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

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 the results were not reported by one author (Viikari-Juntura, 1989) due to poor inter-examiner reliability. The other index tests (arm squeeze test, shoulder abduction (relief) test and traction test) were all assessed in single studies only.

Please insert Table 1

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).
The initial agreement between both raters on the score per domain 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 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 due to an unclear process for excluding patients or what tests had been conducted prior to inclusion in the study as exclusions seemed likely to have taken place after history taking and the 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 4 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, while another study (Apelby-Albrecht, 2013) reported an inappropriate time between reference and index test. Other studies did not report on time between the reference and index test.

Please insert Table 2:

Please insert *Figure 2*

Results

Positivity thresholds for index tests 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, are presented in Table 3. Descriptions of the execution of the tests are described in Table 4.

Please insert Table 3:

Please insert Table 4:

Provocative tests:

Spurling's test

Three studies (n=350) evaluated the diagnostic accuracy of the Spurling's test, but all performed slightly different movements before adding downward axial compression to the cervical spine (Shabat, 2012; Shah, 2004; Viikari-Juntura, 1989). Shah et al (Shah, 2004) reported using cervical extension and ipsilateral lateral flexion. 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 (Viikari-Juntura, 1989) combined ipsilateral lateral flexion and rotation but did not specify adding cervical extension, although they did depict it as such in their manuscript. 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 (Shabat, 2012) used cervical extension combined with ipsilateral rotation and used two different positive test results. Evaluation showed both 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 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, as several studies mention pain in the peri-scapular region as one of the more patient-specific findings during history taking (Tanaka, 2006; Wainner, 2003a; Yoss, 1957). Only the radicular symptoms test results are presented in pooling of results (see Figure 3).

Please insert Figure 3

Upper Limb Neural Tension test

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

- 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;
- differences 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 a specificity of 0.69 (95%CI: 0.41-0.88). Individually, the ULNT 3 (ulnar) had the highest specificity of 0.88 (95%CI: 0.60-0.98) while the ULNT 1 (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 not 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 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) of this test (Viikari-Juntura, 1989).

Traction test

One study evaluated the 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 10-15kg was applied. A sensitivity of 0.33 (95%CI: 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 more painful than other areas of the shoulder and upper arm (Gumina, 2013). The authors defined a positive test when the pain score (on a 0-10 visual analogue scale or VAS) 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 due to 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) were 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 ULNT 3 (ulnar) individually showing high specificity. The included recent study (Apelby-Albrecht, 2013) showed higher specificity than previously reported (Rubinstein, 2007b).

No studies were found that assessed the diagnostic accuracy of widely used neurological tests such as key muscle strength, tendon reflexes and sensory impairments. But eight studies were identified that retrospectively evaluated neurological symptoms prior to surgical management (Chen, 2000; Conradie, 2006;

Henderson, 1983; Kuijper, 2011; Post, 2006; Rainville, 2016; Rainville, 2007; Yoss, 1957).

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 has 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 false positives (Ernst, 2005; Matsumoto, 1998; Okada, 2011; Siivola, 2002). 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 false positives due to reproducing somatic referred pain from compression of degenerative zygapophyseal joints of a population generally in their 5^{th} or 6^{th} 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 and/or physical examination in the identification of cervical radiculopathy, with those of a reference standard of imaging or surgical findings. But since relying only on imaging in a diagnostic process has a risk of an inappropriate high number of false positives (Ernst, 2005; Kuijper, 2011; Siivola, 2002), it can only assist the clinician in his/ her clinical reasoning process. We consider 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 North American Spine Society (NASS) guideline for the diagnosis and treatment of cervical radiculopathy from degenerative disorders suggests that MRI, CT or CT myelography are suitable for identifying the affected level in patients with cervical radiculopathy, prior to 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, 2003a), 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 the 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, 2013; Kwast Rabben, 2011; 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 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 those 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 review retrieved, multivariate regression is not yet an option. A test item cluster has been proposed for indicating the presence of cervical radiculopathy (Wainner, 2003b). From the results of our review, it is proposed that, when consistent with history and other physical findings, a combination of a positive Spurling's test, axial traction test and Arm Squeeze test may be used to increase the likelihood of a cervical radiculopathy while a negative outcome of combined ULNTs and Arm Squeeze test may be used to decrease the likelihood. More high-quality research however is needed to further develop a test item cluster and to improve point estimate precision.

REFERENCES

Apelby-Albrecht, M., Andersson, L., Kleiva, I. W., Kvale, 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 Manip Physiol Ther, 36*(9), 626-632. doi: 10.1016/j.jmpt.2013.07.007

Ashkan, K., & Johnston, P. (2002). A comparison of magnetic resonance imaging and neurophysiological studies in the assessment of cervical radiculopathy. *British journal of ….*

Bogduk, N. (2009). On the definitions and physiology of back pain, referred pain, and radicular pain. *Pain, 147*(1-3), 17-19.

Bono, C. M., Ghiselli, G., Gilbert, T. J., Kreiner, D. S., Reitman, C., Summers, J. T., . . . Toton, J. F. (2011). An evidence-based clinical guideline for the diagnosis and treatment of cervical radiculopathy from degenerative disorders. *Spine J, 11*(1), 64-72.

Bono., C. M., Ghiselli., G., Gilbert., T. J., & Kreiner., D. S. (2011). An evidence-based clinical guideline for the diagnosis and treatment of cervical radiculopathy from degenerative disorders. *The Spine Journal*.

Bussieres, A. E., Taylor, J. A., & Peterson, C. (2008). Diagnostic imaging practice guidelines for musculoskeletal complaints in adults-an evidence-based approach-part 3: spinal disorders. *J Manipulative Physiol Ther, 31*(1), 33-88.

Chen, T. Y. (2000). The clinical presentation of uppermost cervical disc protrusion. *Spine, 25*(4), 439-442. doi: 10.1097/00007632-200002150-00008

Conradie, M., Bester, M. M., Crous, L. C., & Kidd, M. (2006). Do clinical features and MRI suggest the same nerve root in acute cervical radiculopathy? *South African Journal of Physiotherapy, 62*(2), 12-17.

Davidson, R. I., Dunn, E. J., & Metzmaker, J. N. (1981). The shoulder abduction test in the diagnosis of radicular pain in cervical extradural compressive monoradiculopathies. *Spine (Phila Pa 1976), 6*(5), 441-446.

Ellenberg, M. R., Honet, J. C., & Treanor, W. J. (1994). Cervical radiculopathy. *ARCH PHYS MED REHABIL, 75*(3), 342-352. doi: 10.1016/0003-9993(94)90040-x

Elvey, R. L. (1997). Physical evaluation of the peripheral nervous system in disorders of pain and dysfunction. *J Hand Ther, 10*(2), 122-129.

Ernst, C. W., Stadnik, T. W., Peeters, E., Breucq, C., & Osteaux, M. J. (2005). Prevalence of annular tears and disc herniations on MR images of the cervical spine in symptom free volunteers. *Eur J Radiol, 55*(3), 409-414. doi: 10.1016/j.ejrad.2004.11.003

Govindarajan, R., Kolb, C., & Salgado, E. (2013). Sensitivity and specificity of MRI and EMG in diagnosing clinically evident cervical radiculopathy: A retrospective study. *Neurology, 80*(1).

Gumina, S., Carbone, S., Albino, P., Gurzi, M., & Postacchini, F. (2013). Arm Squeeze Test: a new clinical test to distinguish neck from shoulder pain. *EUR SPINE J, 22*(7), 1558-1563.

Guttmann, A., Li, X., Feschet, F., Gaudart, J., Demongeot, J., Boire, J.-Y., & Ouchchane, L. (2015). Cluster Detection Tests in Spatial Epidemiology: A Global Indicator for Performance Assessment. *PLoS ONE, 10*(6), e0130594. doi: 10.1371/journal.pone.0130594

Henderson, C. M., Hennessy, R. G., Shuey, H. M., Jr., & Shackelford, E. G. (1983). Posterior-lateral foraminotomy as an exclusive operative technique for cervical

radiculopathy: a review of 846 consecutively operated cases. *Neurosurgery, 13*(5), 504-512.

Jull, G. (2015). *Grieve's modern musculoskeletal physiotherapy*. Edinburgh ; New York: Elsevier.

Kim, H. J., Nemani, V. M., Piyaskulkaew, C., Vargas, S. R., & Riew, K. D. (2016). Cervical Radiculopathy: Incidence and Treatment of 1,420 Consecutive Cases. *Asian Spine J, 10*(2), 231-237. doi: 10.4184/asj.2016.10.2.231

Kuijper, B., Tans, J. T. J., Beelen, A., Nollet, F., & de Visser, M. (2009). Cervical collar or physiotherapy versus wait and see policy for recent onset cervical radiculopathy: randomised trial. *British Medical Journal, 339*.

Kuijper, B., Tans, J. T. J., van der Kallen, B. F., Nollet, F., Nijeholt, G., & de Visser, M. (2011). Root compression on MRI compared with clinical findings in patients with recent onset cervical radiculopathy. *Journal of Neurology Neurosurgery and Psychiatry, 82*(5), 561-563.

Kwast-Rabben, O., Heikkila, H., & Fagerlund, M. (2013). Electromyography (EMG) and magnetic resonance imaging (MRI) in evaluation of root injury in symptomatic cervical spine disorders (CSD). Specificity and sensitivity test. *J Neurol, 260*, S158. doi: 10.1007/s00415-013-6924-0

Kwast Rabben, O., Heikkila, H., & Fagerlund, M. (2011). 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, 122*, S78. doi: 10.1016/s1388-2457(11)60267-8

Lin, C. H., Tsai, Y. H., Chang, C. H., Chen, C. M., Hsu, H. C., Wu, C. Y., & Hong, C. Z. (2013). The comparison of multiple F-wave variable studies and magnetic resonance imaging examinations in the assessment of cervical radiculopathy. *Am J Phys Med Rehabil, 92*(9), 737-745.

Malanga, G. A. (1997). The diagnosis and treatment of cervical radiculopathy. *Med Sci Sports Exerc, 29*(7 Suppl), S236-245.

Matsumoto, M., Fujimura, Y., Suzuki, N., Nishi, Y., Nakamura, M., Yabe, Y., & Shiga, H. (1998). MRI of cervical intervertebral discs in asymptomatic subjects. *J Bone Joint Surg Br, 80*(1), 19-24.

Merskey H, B. N. (1994). *Classification of chronic pain. Descriptions of chronic pain syndromes and definitions of pain terms* (2nd ed.). Seattle: IASP Press.

Nicotra, A., Khalil, N. M., & O'Neill, K. (2011). Cervical radiculopathy: discrepancy or concordance between electromyography and magnetic resonance imaging? *Br J Neurosurg, 25*(6), 789-790. doi: 10.3109/02688697.2011.594189

Nordin, M., Carragee, E. J., Hogg-Johnson, S., Weiner, S. S., Hurwitz, E. L., Peloso, P. M., . . . Haldeman, S. (2008). 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 (Phila Pa 1976), 33*(4 Suppl), S101-122.

Nunally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (Vol. 3). New York: McGraw-Hill.

Okada, E., Matsumoto, M., Fujiwara, H., & Toyama, Y. (2011). Disc degeneration of cervical spine on MRI in patients with lumbar disc herniation: comparison study with asymptomatic volunteers. *Eur Spine J, 20*(4), 585-591. doi: 10.1007/s00586-010-1644-y

Portney, L. G., & Watkins, M. P. . (2009). *Foundations of clinical research: Applications to practice* (3rd ed.). Upper Saddle River, N.J: Pearson/Prentice Hall.

Post, N. H., Cooper, P. R., Frempong-Boadu, A. K., & Costa, M. E. (2006). Unique features of herniated discs at the cervicothoracic junction: clinical presentation, imaging, operative management, and outcome after anterior decompressive operation in 10 patients. *Neurosurgery, 58*(3), 497-501; discussion 497-501. doi: 10.1227/01.NEU.0000197118.86658.A6

Radhakrishnan, K., Litchy, W. J., O'Fallon, W. M., & Kurland, L. T. (1994). Epidemiology of cervical radiculopathy. A population-based study from Rochester, Minnesota, 1976 through 1990. *Brain, 117 (Pt 2)*, 325-335.

Rainville, J., Laxer, E., Keel, J., Pena, E., Kim, D., Milam, R. A., & Carkner, E. (2016). Exploration of sensory impairments associated with C6 and C7 radiculopathies. *Spine Journal, 16*(1), 49-54.

Rainville, J., Noto, D. J., Jouve, C., & Jenis, L. (2007). Assessment of forearm pronation strength in C6 and C7 radiculopathies. *Spine, 32*(1), 72-75. doi: 10.1097/01.brs.0000251002.39417.2c

Reza Soltani, Z., Sajadi, S., & Tavana, B. (2014). A comparison of magnetic resonance imaging with electrodiagnostic findings in the evaluation of clinical radiculopathy: a cross-sectional study. *Eur Spine J, 23*(4), 916-921. doi: 10.1007/s00586-013-3164-z

Rubinstein, S. M., Pool, J. J., van Tulder, M. W., Riphagen, II, & de Vet, H. C. (2007a). A systematic review of the diagnostic accuracy of provocative tests of the neck for diagnosing cervical radiculopathy. *Eur Spine J, 16*(3), 307-319.

Rubinstein, S. M., Pool, J. J. M., Van Tulder, M. W., Riphagen, I. I., & De Vet, H. C. W. (2007b). A systematic review of the diagnostic accuracy of provocative tests of the neck for diagnosing cervical radiculopathy. *Eur Spine J, 16*(3), 307-319. doi: 10.1007/s00586-006-0225-6

Shabat, S., Leitner, Y., David, R., & Folman, Y. (2012). The Correlation between Spurling Test and Imaging Studies in Detecting Cervical Radiculopathy. *J Neuroimaging, 22*(4), 375-378. doi: 10.1111/j.1552-6569.2011.00644.x

Shah, K. C., & Rajshekhar, V. (2004). Reliability of diagnosis of soft cervical disc prolapse using Spurling's test. *Br J Neurosurg, 18*(5), 480-483. doi: 10.1080/02688690400012350

Siivola, S. M., Levoska, S., Tervonen, O., Ilkko, E., Vanharanta, H., & Keinanen-Kiukaanniemi, S. (2002). MRI changes of cervical spine in asymptomatic and symptomatic young adults. *Eur Spine J, 11*(4), 358-363. doi: 10.1007/s00586-001-0370-x

Spurling RG, S. W. (1944). Lateral rupture of the cervical intervertebral disks: a common cause of shoulder and arm pain. *Surg Gynecol Obstet, 78*, 350-358.

Tanaka, Y., Kokubun, S., Sato, T., & Ozawa, H. (2006). Cervical roots as origin of pain in the neck or scapular regions. *Spine (Phila Pa 1976), 31*(17), E568-573. doi: 10.1097/01.brs.0000229261.02816.48

Thoomes, E. J., Scholten-Peeters, G. G., 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*.

Viikari-Juntura, E. (1987). Interexaminer reliability of observations in physical examinations of the neck. *PHYS THER, 67*(10), 1526-1532.

Viikari-Juntura, E., Porras, M., & Laasonen, E. M. (1989). Validity of clinical tests in the diagnosis of root compression in cervical disc disease. *SPINE, 14*(3), 253-257.

Wainner, R. S., Fritz, J. M., Irrgang, J. J., Boninger, M. L., Delitto, A., & Allison, S. (2003a). Reliability and diagnostic accuracy of the clinical examination and patient self-report measures for cervical radiculopathy. *Spine, 28*(1), 52-62. doi: 10.1097/00007632-200301010-00014

Wainner, R. S., Fritz, J. M., Irrgang, J. J., Boninger, M. L., Delitto, A., & Allison, S. (2003b). Reliability and diagnostic accuracy of the clinical examination and patient self-report measures for cervical radiculopathy. *Spine (Phila Pa 1976), 28*(1), 52-62. doi: 10.1097/01.BRS.0000038873.01855.50

Wainner, R. S., & Gill, H. (2000). Diagnosis and nonoperative management of cervical radiculopathy. *J Orthop Sports Phys Ther, 30*(12), 728-744.

Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., . . . Bossuyt, P. M. (2011). QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med, 155*(8), 529-536.

Yoss, R. E., Corbin, K. B., Maccarty, C. S., & Love, J. G. (1957). Significance of symptoms and signs in localization of involved root in cervical disk protrusion. *Neurology, 7*(10), 673-683.

Figure 1. PRISMA Flow Diagram of included studies

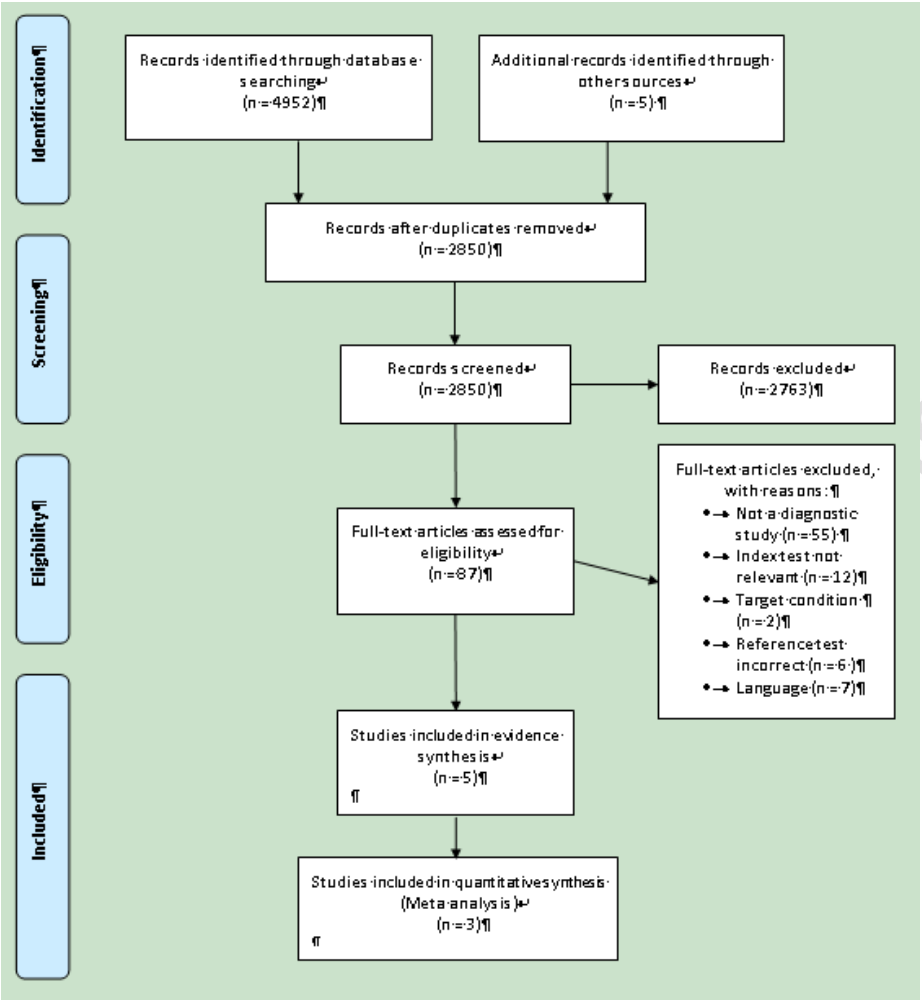


Figure 2. QUADAS-2. Proportion of studies with low, high or unclear risk of bias

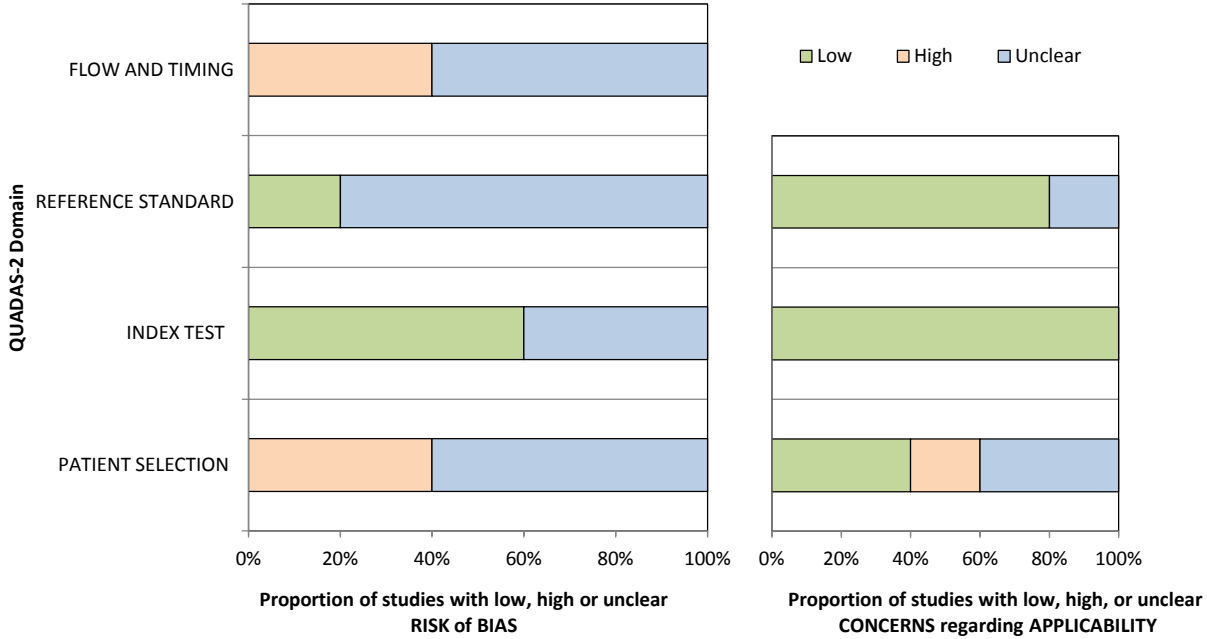
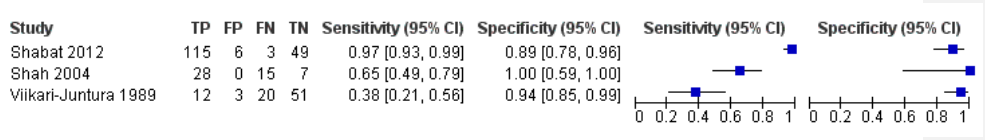


Figure 3 Forest plot – Spurling's test



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

1 Table 1: Characteristics of included studies

Author /year	Clinical Feature and setting	Participants	Study design	Target condition and Reference standard(s)	Index and comparator tests	Notes
Apelby-Albrecht, 2013	Center for spinal surgery, Sweden	51 consecutive patients referred for clinical investigation of cervical and/or arm pain	Diagnostic cohort study	Cervical radiculopathy; MRI, medical history, and clinical examination (dermatomes, reflex testing and Spurlings' test), in patients with cervical radiculopathy.	4 Upper Limb Neurodynamic Tests: ULNT1 (median), ULNT2a (median), ULNT2b (radial) and ULNT3 (ulnar)	
Gumina, 2013	Shoulder Clinical Office and Orthopedic Spine Ambulatory. Italy	1,567 patients with pain localized at the shoulder girdle including patients with neck and arm pain	Cohort study	Cervical radiculopathy; 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	Arm Squeeze test	
Shabat, 2012	Spinal Surgery Unit, Israel	257 patients with symptoms of unilateral cervical radiculopathy lasting for at least 4 weeks.	Cohort study	Unilateral cervical radiculopathy; Complete physical examination for range of motion, motor and sensory examination, and reflex examination.	Spurling (extension+ rotation + axial compression) and physical examination for range of motion, motor and sensory examination, and reflex examination	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)

Comment [A1]: AUTHOR: Two different versions of Table 1 caption were provided and the one in the manuscript has been used. Please check and confirm that it is correct.

						eliciting nonspecific radicular pain radiating to scapular or occipital region.
Shah, 2004	Neurosurgical unit, India	50 patients with neck and arm pain suggestive of radicular pain	Prospective cohort study	Cervical radiculopathy; 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.	Spurling: extension + lateral flexion towards involved side + axial pressure	
Viikari-Juntura, 1989	Neurosurgery department Finland	69 patients sent for cervical myelography	Prospective cohort study	Cervical disc disease (spondylosis and/or disc herniation); Cervical myelography combined with conventional neurological examination (sensory, motor and reflex testing)	Spurling (lateral flexion,+ rotation+ axial compression); cervical distraction and shoulder abduction relief (Davidson's test)	Brachial plexus tension test discarded due to poor inter-examiner reliability, although only one rater examined.

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, 2013	?	+	+	-	+	+	+
Gumina, 2013	-	+	?	?	?	+	+
Shabat, 2012	?	?	?	?	?	+	?
Shah, 2004	?	?	?	-	+	+	+
Viikari-Juntura, 1989	-	+	?	-	-	+	+

+Low Risk -High Risk ? Unclear Risk

1 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)

1 Table 4: Execution of index tests

Index test (Author, Year)	**Description of execution**
Spurling's test	
Shabat, 2012	Patient sitting. The examiner performed cervical extension and ipsilateral rotation and then added axial compression. An increase in symptoms was considered a positive outcome
Shah, 2004	Patient sitting. The examiner performed cervical extension and ipsilateral lateral flexion and then added axial pressure. An increase in symptoms was considered a positive outcome
Viikari-Juntura, 1989	Patient sitting. The examiner performed cervical ipsilateral lateral flexion and ipsilateral rotation and then added axial compression. An increase in symptoms was considered a positive outcome
Upper Limb Neurodynamic Test	
Apelby-Albrecht, 2013	Passive movements in the following order of movements, specific for each of the 4 Upper Limb Neurodynamic Tests, were performed to provide a progressive tension of the nerve. An increase or decrease in symptoms with structural differentiation was considered a positive outcome. ULNT1 (median nerve bias) shoulder depression, shoulder abduction 110°, wrist & finger extension, shoulder lateral rotation, elbow extension, contralateral lateral flexion of the cervical spine. ULNT2a (median nerve bias) Shoulder depression, elbow extension, lateral rotation of the arm, wrist & finger extension, shoulder abduction 10°, contralateral lateral flexion of the cervical spine. ULNT2b (radial nerve bias) Shoulder depression, elbow extension, medial rotation of the arm, wrist & finger flexion, shoulder abduction 10°, contralateral lateral flexion of the cervical spine. ULNT3 (ulnar nerve bias) shoulder depression, shoulder abduction 110°, lateral rotation of the arm, forearm pronation, elbow flexion, wrist & finger extension, contralateral lateral flexion of the cervical spine.
Arm Squeeze test	
Gumina, 2013	The examiner squeezed the patient's middle third of the upper arm with his own hand [with simultaneous thumb and fingers compression]; the thumb from posterior on the triceps muscle and the fingers from anterior on the biceps muscle. The test was considered as positive when the score was 3 points or higher on pressure on the middle third of the upper arm compared with to the other two areas (difference between results in middle third of the upper arm area and in the AC joint and subacromial area).
Shoulder abduction (relief) test	
Viikari-Juntura, 1989	In a sitting position, the patient positions his/her afflicted hand above their head. A decrease in symptoms was considered a positive outcome.
Traction-Distraction test	
Viikari-Juntura, 1989	In a supine position, the examiner applied an axial traction force corresponding to 10-15 kgs. to the patient's neck. A decrease in symptoms with traction and an increase or return of symptoms with the release of traction (distraction) was considered an positive outcome.

2