Postprint Version	1.0
Journal website	http://mrw.interscience.wiley.com/cochrane/clsysrev/articles/CD007431/frame.html
Pubmed link	http://www.ncbi.nlm.nih.gov/pubmed/20166095
DOI	10.1002/14651858.CD007431.pub2.

Physical examination for lumbar radiculopathy due to disc herniation in patients with low-back pain

DANIËLLE AWM VAN DER WINDT[1], EMMANUEL SIMONS[2], INGRID I RIPHAGEN[3], CARLO AMMENDOLIA[4], ARIANNE P VERHAGEN[5], MARK LASLETT[6], WALTER DEVILLÉ[7], RICK A DEYO[8], LEX M BOUTER[9], HENRICA CW DE VET[10], BERT AERTGEERTS[2]

[1]Department of Primary Care & Health Sciences, Keele University, Keele, UK. [2]Belgian Branch of the Dutch Cochrane Centre, CEBAM, Leuven, Belgium. [3]Unit for AppliedClinical Research, Faculty ofMedicine,NorwegianUniversity of Science and Technology, Trondheim, Norway. [4]Institute for Work & Health, Toronto, Canada. [5]Department of General Practice, Erasmus Medical Centre University, Rotterdam, Netherlands. [6]School of Rehabilitation and Occupational Studies, Health and Rehabilitation Research Centre, Auckland University of Technology, Auckland, New Zealand. [7]International and Migrant Health, NIVEL, Utrecht, Netherlands.
[8]Evidence-Based Family Medicine, Oregon Health and Science University, Portland, OR, USA. [9]Executive Board of VU University Amsterdam,Amsterdam,Netherlands. [10]Department of Epidemiology andBiostatistics, EMGOInstitute forHealth andCareResearch, Amsterdam, Netherlands

ABSTRACT

Background

Low-back pain with leg pain (sciatica) may be caused by a herniated intervertebral disc exerting pressure on the nerve root.Most patients will respond to conservative treatment, but in carefully selected patients, surgical discectomy may provide faster relief of symptoms.

Primary care clinicians use patient history and physical examination to evaluate the likelihood of disc herniation and select patients for further imaging and possible surgery.

Objectives

(1) To assess the performance of tests performed during physical examination (alone or in combination) to identify radiculopathy due to lower lumbar disc herniation in patients with low-back pain and sciatica; (2) To assess the influence of sources of heterogeneity on diagnostic performance.

Search strategy

We searched electronic databases for primary studies: PubMed (includes MEDLINE), EMBASE, and CINAHL, and (systematic) reviews: PubMed and Medion (all from earliest until 30 April 2008), and checked references of retrieved articles.

Selection criteria

We considered studies if they compared the results of tests performed during physical examination on patients with back pain with those of diagnostic imaging (MRI, CT, myelography) or findings at surgery.

Data collection and analysis

Two review authors assessed the quality of each publication with the QUADAS tool, and extracted details on patient and study design characteristics, index tests and reference standard, and the diagnostic two-by-two table. We presented information on sensitivities and specificities with 95% confidence intervals (95% CI) for all aspects of physical examination. Pooled estimates of sensitivity and specificity were computed for subsets of studies showing sufficient clinical and statistical homogeneity.

Main results

We included 16 cohort studies (median N = 126, range 71 to 2504) and three case control studies (38 to100 cases). Only one study was carried out in a primary care population. When used in isolation, diagnostic performance of most physical tests (scoliosis, paresis or muscle weakness, muscle wasting, impaired reflexes, sensory deficits) was poor. Some tests (forward flexion, hyper-extension test, and slump test) performed slightly better, but the number of studies was small. In the one primary care study, most tests showed higher specificity and lower sensitivity compared to other settings.

Most studies assessed the Straight Leg Raising (SLR) test. In surgical populations, characterized by a high prevalence of disc herniation (58% to 98%), the SLR showed high sensitivity (pooled estimate 0.92, 95% CI: 0.87 to 0.95) with widely varying specificity (0.10 to 1.00, pooled estimate 0.28, 95% CI: 0.18 to 0.40). Results of studies using imaging showed more heterogeneity and poorer sensitivity.

The crossed SLR showed high specificity (pooled estimate 0.90, 95%CI: 0.85 to 0.94) with consistently low sensitivity (pooled estimate 0.28, 95% CI: 0.22 to 0.35).

Combining positive test results increased the specificity of physical tests, but few studies presented data on test combinations.

Authors' conclusions

When used in isolation, current evidence indicates poor diagnostic performance of most physical tests used to identify lumbar disc herniation. However, most findings arise from surgical populations and may not apply to primary care or non-selected populations.

Better performance may be obtained when tests are combined.

BACKGROUND

Target condition being diagnosed

Low-back pain (LBP) is a common cause of disability in Western industrialised countries. Althoughmany people experience at least one episode of low-back pain in their life, in up to 85% of the patients, no specific pathology is identified (Deyo 1992). In patients who report symptoms radiating into the leg (sciatica), clinicians evaluate the possible causes of radiculopathy (compression of the nerve root) through history and physical examination. One of the causes may be a herniated (protruded, extruded or sequestrated) intervertebral disc exerting pressure on the nerve root. Herniated discs often occur without symptoms, as revealed by magnetic resonance imaging studies in asymptomatic people. They are only clinically relevant when they impinge on a nerve root, causing radiculopathy (sciatica, if the lower lumbar roots are involved).

The large majority of these patients (about 90% to 95%) will respond to conservative treatment (Deyo 1990), but in carefully selected patients, surgical discectomy may provide faster relief from the acute attack than conservative management (Gibson 2007).

However, positive or negative effects on the long-term natural history of the underlying disc disease remain unclear (Gibson 2007; Peul 2007; Weinstein 2006), and decisions for surgery generally involve patients' preferences in addition to clinical judgment.

This systematic review concerns the diagnosis of radiculopathy as a result of lumbar disc herniation. A diagnosis of lumbar disc herniation is often based on the results of diagnostic imaging including Magnetic Resonance Imaging (MRI), Computed Tomography (CT), or myelography. The quality of diagnostic imaging as a reference test has been debated as herniated discs can be found on diagnostic imaging in 20% to 30% of symptom-free persons (Boden 1990). Therefore, findings at surgery are often used to confirm the diagnosis. The disadvantage of using surgical findings as a reference standard is that these studies include a highly selective sample of patients who have all been treated by surgery, and are much more likely to show positive signs during physical examination. This results in a high risk of verification bias, which occurs when patients with negative test results are not evaluated with the reference standard.

Index tests

In patients with LBP, physicians or therapists use the information gained during history and physical examination to decide on a management plan. The most commonly used physical tests include the straight leg raising test, Lasègue's test crossed straight leg raising test, tendon reflexes, and signs of weakness, atrophy or sensory deficits (Deyo 1992; Rebain 2002; Rebain 2003; van den Hoogen 1995). Part of this management plan includes making decisions about referral for diagnostic imaging, or the potential value of surgical intervention. Therefore, investigation of diagnostic performance is especially important in primary care settings.

An accurate diagnostic strategy is important because not all radiculopathies are caused by disc herniation, and consequently, not all patients who undergo surgery for suspected disc herniation may experience pain relief after surgery. Some studies suggest that certain findings of physical examination (for example, a positive straight leg raising test) may predict better outcomes of surgery and chemonucleolysis (Kim 2002; Kohlboeck 2004). Therefore, patients who, based on the results of physical examination, have a high likelihood of radiculopathy due to disc herniation may be better candidates for surgery. If not improved after a few weeks of non-surgical care, these patients are usually referred for diagnostic imaging to obtain more information on the location and severity of the lesion.

Rationale

There are many circumstances that can influence the diagnostic performance of physical examination in the diagnosis of radiculopathy, which include the setting in which physical examination is performed (primary or secondary care), characteristics of the study population, the reproducibility (inter-observer variation of the tests), and the reference standard against which the tests are compared (diagnostic imaging or surgical findings).

Several systematic reviews have summarized the results of available studies on the diagnostic performance of the physical examination for the identification of lumbar radiculopathy in these patients (Devillé 2000;Deyo 1992; van denHoogen 1995;Vroomen 1999).

Three of these reviews included an assessment of the risk of bias in primary diagnostic studies (Devillé 2000; Deyo 1992; Vroomen 1999) and two offered a quantitative summary of the findings (Devillé 2000; Vroomen 1999). These systematic reviews show that most physical tests may have adequate sensitivity, but poor specificity in the identification of disc herniation, while some tests have high specificity and low sensitivity. The diagnostic accuracy varied considerably across studies included in these reviews. Given the publication dates of these reviews (between 1992 and 2000), they have not used current methods for quality appraisal and data synthesis, and the results are out of date. Our current systematic review provides updated evidence on the diagnostic performance of several tests carried out during physical examination, includes a quality assessment , and assesses the influence of potential sources of heterogeneity.

O B J E C T I V E S

To determine the diagnostic accuracy of tests performed during physical examination (individual or in combination) for diagnosing radiculopathy due to lumbar disc herniation as established during imaging or surgery in patients with low-back pain and sciatica.

Investigation of sources of heterogeneity

An additional objective was to assess the influence of sources of heterogeneity on the diagnostic accuracy of tests performed during physical examination, in particular the type of reference standard, the health care setting, the spectrum of disease, and the study design.

M E T H O D S

Criteria for considering studies for this review

Types of studies

We considered primary diagnostic studies if they compared the results of tests performed during physical examination in the identification of radiculopathy due to lumbar disc herniation with those of a reference standard. Both cohort studies and case-control studies were found and included in the review. We only included results from full reports.

Participants

We included studies that assessed diagnostic accuracy of physical examination in patients with low-back pain with pain radiating into the leg (sciatica), who were suspected of having radiculopathy due to disc herniation.We included studies carried out in primary as well as secondary care, and examined the potential influence of the setting on diagnostic performance. Results of studies carried out in primary care will be clearly indicated in text and tables.

Index tests

Studies on all relevant physical examination tests were eligible for inclusion, including the straight leg raising test (and the test of Lasègue), crossed straight leg raising test, paresis or muscle weakness, sensory deficits, impaired reflexes, and other aspects of physical examination that have been proposed as a diagnostic test for identifying radiculopathy due to disc herniation. We included studies in which the diagnostic performances of individual aspects of the physical examination were evaluated separately, or in combination.

In the case of a combination, the study should clearly describe which tests are included in the combination, and how.We excluded studies inwhich only a clinical diagnosis (some unknown combination of history and physical examination) was compared with the results of a reference standard.

Target conditions

We selected diagnostic studies if the aim of the diagnostic test was to investigate causes of radiculopathy, and results included cases of lower lumbar disc herniation. We excluded studies that only reported on other causes of low-back pain or radiculopathy (for example, infection, tumour, severe osteoarthritis, or fractures), and diagnostic testing was aimed at identifying these conditions.

Reference standards

We included studies if the results of a physical examination was compared to 1) diagnostic imaging: Magnetic Resonance Imaging (MRI), Computed Tomography (CT), myelography; or 2) findings at surgery. Separate (stratified) analyses were carried out for these two different reference standards, or data were clearly presented separately. The exact definition of a positive outcome of the reference standard varied among studies, including a bulging, protruding, or extruding disc.

Search methods for identification of studies

Electronic searches

A search strategy was developed in collaboration with a medical information specialist (IR). We searched relevant computerised databases for eligible diagnostic studies from their earliest date to 30 April 2008: PubMed (includesMEDLINE), EMBASE (through EMBASE.com), and CINAHL (through EBSCO host).

The search strategy forMEDLINE is presented inAppendix 1, and was adapted for EMBASE (Appendix 2) and CINAHL (Appendix 3). A previous systematic review on the diagnostic performance of the straight leg raising test was used as a point of reference (Devillé 2000). All publications included in that review are indexed in MEDLINE. The search was refined until all publications in the review were identified by our search. The strategy used several combinations of searches related to the patient population, aspects of physical examination, and the target condition. A methodological filter for the identification of primary diagnostic studies (search 4c) was added to some elements of the searches in PubMed and EMBASE to increase the specificity of the search, and to limit the harvest to less than 2000 hits. This filter is highly sensitive and partly based on those proposed by Devillé et al (Devillé 2000a), and Bachman et al (Bachmann 2002; Bachmann 2003). However, because several authors have recommended against amethodological filter to retrieve diagnostic accuracy studies (Doust 2005; Leeflang 2006), we conducted a sensitivity analyses, analysing all additional citations from PubMed and EMBASE based on the same search, but without the use of a methodological filter. All references were managed by ReferenceManager software, and any duplicates removed.

Searching other resources

We checked the reference lists of all retrieved relevant publications (primary diagnostic studies). If studies were reported in abstracts or conference proceedings we searched for full publications.

An additional electronic search was composed to identify relevant (systematic) reviews in MEDLINE and Medion (www.mediondatabase.nl), and their references were checked. In addition, we contacted researchers in the field of low-back pain research to identify additional diagnostic studies. No language restrictions were applied.

DATA COLLECTION AND ANALYSIS

Selection of studies

Two review authors (BA and ES) independently applied the selection criteria to all citations (titles and abstracts) identified by the search strategy described above. Consensus meetings were organised to discuss any disagreement regarding selection. Final selection was based on a review of full publications, which were retrieved for all studies that either met the selection criteria, or for which there was uncertainty regarding selection. A third review author (DvdW) was consulted in cases of persisting disagreement.

Data extraction and management

For each included study, we used a standardised form to extract characteristics of participants, the index tests and reference standard, and aspects of study methods.

- Characteristics of participants included setting (primary / secondary care); inclusion and exclusion criteria; enrolment (consecutive or non-consecutive); number of subjects (including number eligible for the study, number enrolled in the study, number receiving index test and reference standard, number for whom results are reported in the two-by-two table, reasons for withdrawal); duration and history of low-back pain, and presence of sciatica.
- Test characteristics included the type of test, methods of execution, experience and expertise of the assessors, type of reference standard, and cut-off points for diagnosing radiculopathy due to lumbar disc herniation. Positivity thresholds (interpretations of “positive” results) may vary across studies, and some studies may present diagnostic performance of an index test at several different cut-off points. We extracted data regarding cut-off points most commonly used by studies in the review.
- Aspects of study methods included the basic design of the study (case-control, prospective cohort, or historical cohort with data collection based on medical records), time and treatment between index test and reference standard, and quality assessment (see section on assessment of methodological quality and Appendix 4).

We extracted the diagnostic two-by-two table (true positive, false positive, true negative, and false negative index test results) from the publications, or if not available, reconstructed the two-by-two table using information on relevant parameters (sensitivity, specificity or predictive values). Eligible studies for

which the diagnostic two-by-two table could not be reconstructed were presented in the review, but were not included in the quantitative analyses.

Two review authors independently extracted the data (ML and DvdW) to ensure adequate reliability of collected data. Disagreements were resolved by consensus. For each study, we presented aspects of study design, characteristics of the population, index test, reference standard, and diagnostic parameters (sensitivity and specificity) in tables. For cohort studies only, we calculated prior probability (prevalence) of disc herniation as the proportion of patients in the cohort diagnosed with lumbar disc herniation according to the reference standard. We used two diagnostic accuracy studies not included in the review (on the diagnostic accuracy of physical examination in patients with shoulder pain) to pilot the data extraction form. Disagreements were resolved by consensus.

Assessment of methodological quality

Three review authors (AV, CA, DvdW) assessed the methodological quality in each study, using the Quality Assessment of Diagnostic Accuracy Studies list (QUADAS) (Whiting 2004). The Cochrane Diagnostic Test AccuracyWorking Group recommends this checklist (Handbook 2005). TheQUADAS checklist consists of 11 items that refer to internal validity (e.g. blind assessment of index and reference test, or avoidance of verification bias). Three additional items described in the Cochrane Diagnostic Reviewers' Handbook (Handbook 2005) are of relevance to this review and were also scored. These additional items refer to the definition of the positivity threshold of the index test, treatment given between index test and reference standard, and observer variation.

The authors scored each item as 'yes' (positive assessment, high quality), 'no' (negative assessment, low quality), or 'unclear' (insufficient information). Guidelines for the assessment of each criterion were made available to the review authors (Appendix 4).

Again, quality assessment was pre-tested using two studies not included in the review. We quantified inter-observer agreement by computing the percentage agreement for each item of the checklist.

Disagreements were resolved by consensus and, if necessary, by third party (HdV) adjudication.

We did not applyweights to the different items of the checklist, and did not use a summary score to incorporate studies with certain levels of quality in the analysis. We explored the influence of negative scores on important items using subgroup analyses or metaregression analyses (see below). The following items, which have been shown to affect diagnostic performance in previous research (Lijmer 1999; Rutjes 2006) were considered for these analyses: item 1 (spectrum variation / selective sample), item 4 (verification bias), items 7 and 8 (blinded interpretation of reference standard and index test) and item 11 (explanation of withdrawals).

Statistical analysis and data synthesis

The two key and commonly reported parameters of diagnostic test accuracy are sensitivity and specificity. Because a trade-off may exist between these two parameters, they should be analysed jointly.

Sensitivities and specificities for each index test with 95% confidence intervals are presented in forest plots. In addition, a scatterplot of study-specific estimates of sensitivity and 1-specificity was used to display data in Receiver Operating Characteristic (ROC) space.

For pooling of results of sensitivity and specificitywe used bivariate analysis (Reitsma 2005), which accounts for both within-study variation, between-study variation, and any negative correlation that may exist between sensitivity and specificity.

The bivariate model preserves the two-dimensional nature of diagnostic data by directly analysing the logit transformed sensitivity and specificity of each study in a single model. The model produces the following results: a random-effects estimate of the mean sensitivity and specificity with corresponding 95% CIs, the amount of between-study variation for sensitivity and specificity separately, and the strength and shape of the correlation between sensitivity and specificity. Using these results, we calculated a 95% confidence ellipse (i.e. bivariate confidence interval) around the summary estimate of sensitivity and specificity. All the results were transformed back to the original scale, and plotted in ROC space (Reitsma 2005).We presented pooled estimates of sensitivity and specificity if studies showed clinical homogeneity (same reference standard, similar definition of disc herniation) and results of sensitivity and specificity

showed sufficient statistical homogeneity (visual inspection of point estimates and confidence intervals). Bivariate analyses were carried out using STATA software. We used pooled estimates of sensitivity and specificity to calculate the likelihood ratio (LR) of a positive test result as sensitivity/(1-specificity), and the LR of a negative test result as (1-sensitivitity)/specificity.

Investigations of heterogeneity

Several factors (next to variability in the positivity threshold) may contribute to heterogeneity in diagnostic performance across studies.

We used subgroup analyses to investigate the potential influence of differences in the type of reference standard (surgery versus imaging); study population (primary versus secondary care, previous lumbar disc surgery), and study design (prospective cohort or other designs, scores on items 1, 4, 7, 8, and 11 of the QUADAS checklist). Given the small number of studies per test, we studied the influence of no more than one study level covariate at a time.

Finally, we summarized the findings of the review in a summary table (Handbook 2005), which includes 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). If no pooled estimate could be calculated, we presented the range of sensitivity and specificity for each index test. The prevalence of the target condition (lumbar disc herniation) in the study populations is presented along with measures of diagnostic performance.

RESULTS

Results of the search

We identified 1529 citations from the electronic searches in PubMed, 793 additional citations from EMBASE, and 321 from CINAHL. A search aimed at identifying reviews identified 20 citations from the Medion database, and 226 from PubMed. The references of relevant reviews and primary diagnostic studies were checked. After initial evaluation, 134 full papers were retrieved, 20 of which were finally considered eligible for the review. Two papers seemed to report on the same cohort (Vucetic 1996; Vucetic 1999) and information from both publications was used during quality assessment.However, as the number of patients was slightly different between the publications, they have been separately presented in the characteristics of studies table. So, a total of 19 studies were included in the review. A summary of the search results, including the main reason for excluding papers is presented in Figure 1.

Three of the four studies that were not identified by the electronic searches were published in 1961 or earlier. Exclusions mainly concerned the design of the study (not a diagnostic accuracy study), or the relevance of the index test (studies investigating imaging techniques or electromyography). Fifteen studies were excluded as they examined the value of physical examination in determining the level of disc herniation (L5-S1 versus L4-L5) among patients with herniated discs, rather than the presence of a disc herniation.

[FIGURE 1]

Additional search results without use of a methodological filter

When themethodological filterwas removed fromthe search strategy, the search resulted in 329 additional citations from PubMed and 125 from EMBASE. Most of these papers did not concern studies of diagnostic accuracy, did not include patients with back pain, or reported on irrelevant index tests or target conditions.

Twenty full papers were retrieved, but none met the selection criteria for this review: seven were editorials or narrative reviews; seven did not concern a relevant index test; five were not studies of diagnostic accuracy, and one did not address disc herniation as a target condition.

Description of the studies

Details on the design, setting, population, reference standard and definition of the target condition are provided in the Characteristics

of included studies table. The 19 selected studies included three case-control studies (Demircan 2002; Kerr 1988; Majlesi 2008), and 16 cohort studies, five of which used a retrospective design, collecting information from medical records (Charnley 1951; Gurdjian 1961;Hakelius 1972; Knutsson 1961; Spangfort 1972).

Only one studywas clearly conducted in a primary care population (Vroomen 2002 (prim care)). The prevalence (prior probability of lumbar disc herniation) varied widely between 26% (Haldeman 1988) and 98% (Gurdjian 1961). The definition of lumbar disc herniation varied across studies: in most studies the diagnostic criteria included a bulging or protruded disc, whereas in other studies only disc extrusion or sequestration was defined as disc herniation (Albeck 1996). Not all studies clearly explained if the definition involved nerve root compression or impingement. If possible, we presented the results of diagnostic performance of physical examination separately for different definitions of disc herniation.

Surgical findings were used as the reference standard in nine studies, imaging techniques (CT or MRI) in six studies. Two of the three case-control designs used surgery to confirm disc herniation in cases, and imaging techniques (MRI or myelography) to exclude nerve root compression in controls (Demircan 2002; Kerr 1988). The third case control study used MRI in all patients, but used different sets of selection criteria to identify patients with bulging or herniated discs and controls with normalMRI findings (Majlesi 2008).One additional cohort study (Hudgins 1979) used both surgery and clinical follow-up depending on the severity of symptoms; only patients not responding to conservative treatment received surgery.

A wide variety of tests were examined. Diagnostic accuracy of the straight leg raising test or Lasègue's test was most frequently evaluated (15 studies), followed by impaired reflexes (nine studies); paresis or muscle weakness (seven studies); sensory deficits (six studies), and the crossed straight leg raising test or crossed Lasègue's test (six studies). Scoliosis, muscle wasting, forward flexion, slump test, hyperextension test, segmental spasm, and the Bell test were evaluated by four studies or less. Six studies reported findings on diagnostic accuracy of combinations of test results.

Methodological quality of included studies

The results of the quality assessment are presented in Figure 2 (results for individual studies). Many studies, mainly those published before 1990, provided little detail on research methods, resulting in many items being scored as unclear. The majority of publications poorly described the following aspects: time period and treatment between index test and reference standard (items 3 and 13), review bias (items 7 and 8), and information on observer variation (item 14). Most studies scored well on the use of an appropriate reference standard (item 2), avoidance of partial verification bias (item 4), availability of clinical information (item 9), uninterpretable test results (item 10), and reasons for withdrawal (item 11). Only 40% of all studies gave an adequate description of the methods used to conduct and interpret the index test. Five studies performed well, and received a positive assessment of at least 10 out of 14 QUADAS items, including the one primary care study (Poiraudeau 2001; Stankovic 1999; Vroomen 1998; Vroomen 2002 (prim care); Vucetic 1996). The assessment was not easy for some items or publications, resulting in disagreement between reviewers. On average, the reviewers disagreed in four out of 14 items (range 1 to 7 across publications). Disagreements mainly concerned review bias, availability of clinical information, uninterpretable test results, and definition of a positive index test result. All disagreements were resolved during consensusmeetings.

[FIGURE 2]

Findings

The extracted data (two-by-two tables) and sensitivity and specificity for all index tests for each study, are presented in Appendix 4 to Appendix 15, including all cut-off points used for test positivity, and definitions of lumbar disc herniation. The findings of the single primary care study (Vroomen 2002 (prim care)) are presented first, followed by secondary care studies. The primary care study is clearly marked in all forest plots. For presentation of the results in forest plots or ROC plots and for pooled analyses, we tried to generate clinically homogeneous subgroups with respect to cutoff points and definition of disc herniation, and results used for further analyses are shaded grey in Appendix 4 to Appendix 15.

Study findings for a specific index test were entered only once in a (pooled) analysis. Most studies used a broad definition of disc herniation, including sequestration, extrusion and protrusion or bulging of the disc. Therefore, wherever studies reported findings at multiple cut-offs, we used the broader definition for pooling of results and for presenting results in forest or ROC plots. Furthermore, whenever studies reported findings separately for subgroups of patients treated with surgery for the first time, these were selected for pooled analyses or presentation in plots.

Straight Leg Raising test (SLR) or Lasègue's test

The SLRor Lasègue's testwas evaluated in 15 studies (Appendix 5).

Most studies performed the SLR by passive elevation of the leg on the symptomatic sidewith the patient in supine position. Lasègue's test is an extension of the SLRinwhich the leg is lowered five to ten degrees and the foot is dorsiflexed. If pain occurs, then Lasègue's test is considered positive. However, there was great inconsistency regarding the use of these terms. Several of the studies referred to the SLR as Lasègue's sign (see Characteristics of included studies).

One study described the procedure of the Lasègue's test, but used the termSLR(Haldeman 1988); one study used a slightly different manoeuvre (passive flexion of the hip with the knee extended followed by passive flexion of the hip with the knee flexed) and named this Lasègue's sign (Vucetic 1996). One study (Knutsson 1961) indicated they used Lasègue's test, but gave no description of test performance. Given the fact that nearly all studies used passive elevation of the leg (knee extended), we have used the term SLR in this review.

The test is considered positive when pain below the knee (sciatica) occurs upon elevation, but the cut-point used - that is the angle at which pain occurs - varied considerably across studies. Most studies used ipsi-lateral leg pain occurring at any angle to define a positive test result. We extracted these data for further analysis of diagnostic performance of the SLR in case studies that presented findings formultiple cut-points. The results showed heterogeneity with sensitivities ranging between 0.35 and 0.97 and specificities between 0.10 and 1.00 (Appendix 5).

Figure 3 (forest plot) presents the results of the SLR for the subgroup of five studies using imaging as the reference standard (Haldeman 1988; Majlesi 2008; Meylemans 1988; Poiraudeau 2001; Vroomen 2002 (prim care)). The prevalence of disc herniation (prior probability) in these studies was lower compared to the 10 surgical studies (range 26% to 55% versus 58% to 98%).

The plot shows that results of imaging studies, including the one primary care study (Vroomen 2002 (prim care) are close to the 45° line, indicating poor diagnostic performance of the SLR test.

Specificity of the SLR in the study byMeylemans 1988 was 100%, but this was the only study in which it was not fully clear if the target condition was indeed radiculopathy due to disc herniation (or radiculopathy as amore general condition). The one case-control study (Majlesi 2008) showed slightly more favourable results than the four cohort studies. Pooled estimates were not computed because of large statistical heterogeneity.

[FIGURE 3]

In the subgroup of surgical studies (Figure 4 - Forest plot; and Figure 5 -ROCplot), sensitivity was generally high,whereas specificity still showed large statistical heterogeneity. Assessment indicated review bias for one of the two case control studies (Demircan 2002). Specificity reported by this study was higher compared to all other surgical studies (0.82). Excluding this study from the analysis resulted in a pooled estimate (bivariate analysis) of 0.92 (95% CI: 0.87 to 0.95) for sensitivity and of 0.28 (95% CI: 0.18 to 0.40) for specificity (Figure 5). It must be noted that there is still considerable heterogeneity of the results for specificity in this pooled estimate. There was no single explanation for this heterogeneity, but there was variation in the cut points used to define a positive SLR(see Appendix 5), and the criteria to define disc herniation (disc protrusion, extrusion, prolapse, sequestration). The influence of other potential sources of heterogeneity was studied using subgroup analyses. In the subset of studies using surgical findings as the reference standard, these exploratory analyses showed that the specificity of the SLR may be poorer in studies in which patients with previous lumbar disc surgery have been clearly excluded (pooled estimate for specificity 0.24, 95% CI 0.14 to 0.37 versus specificity 0.45, 95% CI 0.22 to 0.70 for other studies), with similar estimates for sensitivity (0.91, 95% CI 0.84 to 0.96 versus 0.94, 95% CI 0.86 to 0.98). Other pre-

defined potential sources of heterogeneity (selection bias, verification bias, test bias, withdrawals) did not influence diagnostic performance of the SLR or the number of studies in subgroups was too small to carry out a pooled analysis.

[FIGURE 4]

[FIGURE 5]

Crossed Straight Leg Raising test (XSLR) or crossed Lasègue's test

The XSLR (sciatica reproduced upon passive extension of the contra-lateral leg) was assessed by five studies (Appendix 6 and Figure 6), only one of which used imaging as a reference standard (Poiraudeau 2001). Figure 7 shows that most study results are clustered in the lower left corner of the ROC space, indicating poor sensitivity (ranging between 0.23 and 0.43) coupled with high specificity (0.83 to 1.00). Kerr 1988 reported higher sensitivity compared to other studies, but it must be noted that verification bias may have influenced results in this case-control study: controls consisted of patients with back pain and sciatica, but had normal myelograms. As the results of the imaging study (Poiraudeau 2001) were very consistent with those of the surgical studies, these results were combined in the meta-analysis (Figure 7). Pooled estimates of sensitivity and specificity were 0.28 (95% CI: 0.22 to 0.35) and 0.90(95% CI: 0.85 to 0.94), respectively.

[FIGURE 6]

[FIGURE 7]

Scoliosis

Four studies investigated scoliosis (using visual inspection) as a potential indicator of lumbar disc herniation. The results of three studies indicate poor diagnostic performance of this index test, with low rates of sensitivity and specificity (see Appendix 7 and Figure 8) (Albeck 1996, Kerr 1988; Kosteljanetz 1988, Vucetic 1996). Again, the results byKerr 1988 seemto bemore favourable, but there was a high risk of verification bias in this case-control study. Because of large heterogeneity pooled estimates are not presented for scoliosis.

[FIGURE 8]

Paresis or muscle weakness

Muscle weakness or paresis was evaluated in seven studies, usually by testing muscle strength during ankle dorsiflexion (L4 radiculopathy) or extension of the big toe (L5 radiculopathy) either against resistance (Kerr 1988) or without resistance (Vucetic 1996).With the lattermethod, tests were considered to be positive if the patientwas unable to extend the ankle or great toe in the same range as on the non-symptomatic side. All but the primary care study (Vroomen 2002 (prim care)) compared the results with surgical findings as the reference standard. Knutsson 1961 separately presented the results for patients receiving surgery for the first time, and for a small subgroup of patients who had been operated on before (Appendix 8). For further analysis we used the results of the subgroup with first-time lumbar surgery. Figure 9 demonstrates poor performance of paresis or muscle weakness in identifying lumbar disc herniation. There was large heterogeneity which precluded pooling of sensitivities and specificities, but the results of most studies are close to the diagonal in ROC space. Again, the case-control study by Kerr 1988 reported more favourable findings compared to the cohort studies. In a primary care population Vroomen 2002 (prim care) reported a higher specificity (0.93, 95% CI: 0.88 to 0.97) coupled with low sensitivity (0.27, 95% CI: 0.20 0.37) of paresis compared to the surgical studies. This study was the only one using imaging to identify patients with lumbar disc herniation.

[FIGURE 9]

Muscle wasting

Three studies assessed muscle wasting with results similar to those for muscle weakness (Appendix 9 and Figure 10). Only Kerr 1988 explained that muscle wasting was assessed by measuring calf circumference, and provided a cut-off for a positive test result (1 cm difference with non-symptomatic side). Sensitivity ranged between 0.15 (Albeck 1996) and 0.38 (Kerr 1988), specificity between 0.50 (Knutsson 1961) and 0.94 (Kerr 1988). The very high specificity was reported by the one case-control study (Kerr 1988).

Because of the small number of studies and large heterogeneity, we decided to refrain from statistical pooling of results.

[FIGURE 10]

Impaired reflexes

Absence or weakness of the tendon reflexes was examined by seven studies. Most of them included an evaluation of the Achilles tendon reflex (S1 radiculopathy). Tests were not described in most studies. In the primary care study by Vroomen 2000, the ankle tendon reflex was tested by the observer with his back to the patient holding the patient's leg (hip and knee flexed 90°) in the observer's axilla. Knutsson 1961 and Gurdjian 1961 also assessed the patellar tendon reflex (L4 radiculopathy), the results of which are only presented in Appendix 10. Results for impaired reflexes were distributed close to the 45° line and mostly in the lower left corner of the ROC space (except for Kerr 1988), indicating poor diagnostic performance, especially in terms of sensitivity (Figure 11). Again, the primary care study Vroomen 2002 (prim care) showed higher specificity (0.93, 95% CI: 0.88 to 0.97) and lower sensitivity (0.15, 95% CI: 0.09 to 0.21) compared to the surgical studies.We did not pool results of impaired reflexes because of the large heterogeneity of results.

[FIGURE 11]

Sensory deficits

The diagnostic performance of sensory deficits, including hypoaesthesia, hypoalgesia, tingling or numbness, has been studied in six studies (Albeck 1996; Kerr 1988; Knutsson 1961; Kosteljanetz 1984; Kosteljanetz 1988; Vroomen 2002 (prim care); Vucetic 1996). Few studies described the methods of tests or positivity criteria. Albeck 1996 mentioned that hypoaesthesia had to have a dermatomal distribution to be considered positive. Vroomen et al. provided a more extensive description of sensory tests in an additional paper (Vroomen 2000). Dermatomes were tested bilaterally and simultaneously by softly striking the skin. The patient (with eyes closed) was asked if the feeling clearly differed between left and right sides. Sensory loss and pain (disturbed, not disturbed) were tested by asking the patient (eyes closed) whether gentle pressure exerted by the observer through a plastic stick was sharp or blunt. The order of sharp and blunt pressure application was random (Vroomen 2000). The forest plot (Figure 12) shows poor diagnostic performance of both sensitivity and specificity.

Statistical pooling was not undertaken, and Appendix 11 presents results for sensitivity and specificity in each individual study. In the primary care study (Vroomen 2002 (prim care)) sensitivity of sensory deficits was low compared to most (but not all) other studies (0.28, 95% CI: 0.21 to 0.36).

[FIGURE 12]

Forward flexion and extension test

Forward flexion (bending forward in standing position) was evaluated in three studies, all using different methods to define a positive test result. None of the studies indicated whether a positive test resultwas defined by limitation of forward flexion due to back/ leg pain or due to restricted mobility. Two studies in surgical populations (Albeck 1996; Charnley 1951) showed high sensitivity (0.90 and 0.85, respectively) coupled with poor specificity (0.16 and 0.29 respectively Appendix 12 and Figure 13. In contrast, Vroomen 2002 (prim care), who enrolled a primary care population and used MRI as the reference standard, reported

lower sensitivity (0.45, 95% CI: 0.37 to 0.53) and higher specificity 0.74 (95% CI: 0.65 to 0.81). Pooling of estimates was not carried out because of large heterogeneity and the small number of studies.

[FIGURE 13]

The (hyper)extension test was only investigated by two studies (Poiraudeau 2001; Stankovic 1999; Table 9). Poiraudeau 2001 performed the test with the patient standing, mobilising the trunk passively over the full range of extension, with the knees extended.

They considered the test positive if sciatica was reproduced or worsened during extension. Stankovic 1999 present the results of this test at three different cut-off points (major, moderate or any loss of extension, measured in prone position by % extension of elbows while keeping pelvis, hips and legs relaxed on the table), and used two definitions of lumbar disc herniation. As expected, specificity increased with the use of a stricter cut-off point, but with considerable loss of sensitivity. Including bulging disc in the definition of disc herniation did not strongly affect diagnostic performance of the extension test (Appendix 13).

Other tests

Two studies (Majlesi 2008; Stankovic 1999) reported results on the slump test (Table 10). During the slump test, the patient sits with head bent forward and leg outstretched, toes pointing upwards.

The examiner gently eases the patient forward to increase stretch on the sciatic nerve. In the slump test (Maitland 1985) the neural structures within the vertebral canal and foramen are put on maximum stretch. Stankovic 1999 present the results of the slump test at different cut-off values (angles at which pain occurred), showing that sensitivity of the slump test was poor (0.44, 95% CI: 0.34 to 0.55), and specificity slightly better (0.58, 95% CI: 0.28 to 0.85) when using a strict cut-off (pain radiating below the knee). Sensitivity increased (and specificity decreased) when using a milder cut-off (pain anywhere). Majlesi 2008 reported similar sensitivity (0.84), but higher specificity (0.83), using an unknown cut-off for a positive test result. The higher specificity might partly be the result of the case control design of this study: patients with back pain were selected as controls if MRI findings were completely normal.

Poor diagnostic performance (Appendix 14) was also reported for the Bell test. The Bell test is positive when the examiner can reproduce or exacerbate the usual leg pain by pressure applied with the thumb between the spinous processes L4/L5 or L5/S1 or in the near paraspinal area (Poiraudeau 2001).

Combination of tests

Four studies investigated the performance of using various combinations of physical examination tests (Appendix 15). Poiraudeau 2001 reported the results of several combinations of tests showing high specificity (range 0.74 to 0.94) coupled with low sensitivity (range 0.16 to 0.28) when combining a positive XSLR with positive results on the Bell test, hyper extension test or SLR. Two other studies reported similar results when combining a positive SLR with positive neurological signs (Majlesi 2008) or with a positive XSLR (Hudgins 1979). Charnley 1951 examined diagnostic performance when combining a positive SLR with limited mobility of the lumbar spine. He also reported higher specificity and lower sensitivity compared to using the SLR in isolation, but in this surgical study sensitivity was relatively high (0.73 to 0.77) and specificity low 0.57 to 0.64).

Two additional studies (Vroomen 2002 (primcare);Vucetic 1999) derived amultivariable model to identify the combination of diagnostic tests which best predicts the presence of lumbar disc herniation.

Both models included aspects of patient history and physical examination. In the primary care study by Vroomen 2002 (prim care) the following combination showed the strongest association with nerve root compression on MRI: age; duration of disease 15 to 30 days; paroxysmal pain; pain worse in leg than in back; typical dermatomal distribution of pain; pain worse on coughing/ sneezing/straining; finger-floor distance; and paresis (area under the ROC curve 0.80 for history alone, and 0.83 for history and physical examination) (Vroomen 2002 (prim care)). Maximum diagnostic performance of themodel occurred at a predicted probability of 62.5% (prior probability of lumbar disc herniation in this cohort was 55.4%), with a sensitivity of 0.72 and specificity of 0.80 (Vroomen 2002 (prim care)). The model developed by Vucetic 1999 included the following factors: high education; no comorbidity; no previous surgery; incapacitating pain; restricted lumbar range of motion; positive crossed SLR; and dislocated dura or root on myelography (only explained variance presented: 0.495). Both studies presented the multivariable models, but did not

propose a decision rule indicating which or how many tests would need to be positive to identify patients with a high likelihood of radiculopathy due to lumbar disc herniation.

Summary of results

[TABLE1]

D I S C U S S I O N

Summary of main results

This reviewaimed to summarize evidence for the accuracy of physical examination in identifying radiculopathy due to lumbar disc herniation. An important finding is that only one of the studies was carried out in a primary care setting, and that most studies were carried out in populations with very high prevalence (prior probabilities) of disc herniation. The results show that diagnostic performance of most physical tests (in particular scoliosis, paresis ormuscleweakness,musclewasting, impaired reflexes, and sensory deficits)was poor, especiallywhen used in isolation. For a fewother tests (forward flexion, hyperextension test, slump test) the results seemed slightly more favourable with either better sensitivity or specificity, but the number of studies assessing diagnostic performance of these tests was small (three or less). The performance of the SLR was evaluated inmost studies. In surgical populations, the SLR showed high sensitivity, with widely varying results for specificity.

The XSLR is usually only positive in patients with major nerve root impingement, and showed consistently high specificity in surgical studies (coupled with low sensitivity). However, it is important to note that these results were obtained in populations characterised by a very high prevalence of disc herniation (>75% in nearly all studies) and a severe spectrum of disease, and cannot be generalised to populations with a lower prevalence of the target condition. This means that there is still insufficient evidence for the clinical usefulness of the SLR and XSLR in the diagnosis of disc herniation in primary care populations and other populations of patients not (yet) referred for surgery.

There were only limited possibilities to study the influences of sources of heterogeneity in this review. The number of studies per index test was small, and studies did not always provide sufficient information about important study characteristics. Our analyses therefore focused on differences in diagnostic performance between studies carried out in different settings (primary versus secondary care) and studies using different reference standards (imaging versus surgery).

Factors affecting interpretation

Population and setting

Most studies were carried out in a secondary care setting, often using a historical design in which medical records were analysed to investigate the association between diagnostic test results and disc herniation.Our findings seemed to indicate an overestimation of diagnostic performance in historical and case control designs.

These studies are more susceptible to selection and verification bias, meaning that not all patients receiving the index test were selected for the study or went on to receive the reference standard.

In most surgical populations, patients had received some form of imaging prior to surgery, although generally only patients with positive findings on imaging are referred for surgery. In particular, patients with negative results on physical examination will often not be subjected to surgery, and may have been excluded from diagnostic studies. The risk of bias is particularly high in studies with a retrospective or case control design (Leeflang 2006; Lijmer 1999; Rutjes 2006), which also appeared to be the case in our review.

Most surgical studies excluded patients with previous lumbar disc surgery. Studies including patients with previous surgery showed better results for specificity for the SLR, which can be explained by the fact that it is very likely that only patients with positive results of physical examination will have been referred for surgery, and that these patients may have more severe disc disease. Where possible, we only analysed the results of patients with first-time surgery.

Reference standard

Several important characteristics of the studies clearly clustered; studies using surgery were carried out in secondary care populations and often showed a high prevalence (prior probability) of lumbar disc herniation with a severe spectrum of disease, while both factors probably affected results of specificity and sensitivity.

A higher prior probability of disc herniation was found in populations of patients treated with surgery, and test results often showed higher sensitivity (e.g. for the SLR). These studies are likely to include a strongly selected population, representing extreme cases in terms of severity and duration of symptoms, although they have the virtue of including only patients with clinically important disc herniations, and not those with irrelevant imaging findings.

Imaging studies were carried out in populations with a generally lower prior probability of disc herniation, and may better reflect the diagnostic value of physical examination in primary care settings, or in patients with less severe symptoms. However, imaging is likely to include more false positive findings. We know that imaging by either MRI or CT shows herniated discs in a substantial fraction of asymptomatic people and may not always be relevant in predicting back problems (Boden 1990; Boden 1996; Borenstein 2001; Jarvik 2005). Depending on age, 20% to 35% of asymptomatic people have protruding or extruded discs, 25% under age 40 and nearly everyone over age 60 have bulging discs.

The studies in this review included only symptomatic patients, but false positive results of imaging may explain part of the reduction in sensitivity in the imaging studies. A meaningful definition of a positive result of imaging may be the presence of a herniated disc with clear nerve root impingement. None of the studies in this review specifically used this definition, although a few studies provided some information on to what extent nerve root compression was considered to be likely, based on the results of imaging (Poiraudeau 2001; Stankovic 1999; Vroomen 1998; Vroomen 2002 (prim care)). Studies also gave little information about the extent to which findings at either imaging or surgery matched with relevant clinical findings (e.g. with respect to suspected level or side of herniation).

Clinical follow-up after surgery has been suggested to be the optimal reference standard: if symptoms disappear after appropriate surgery, the cause of the problem must have been disc herniation with nerve root compression. However, results regarding the benefits of surgical intervention are not unequivocal (Gibson 2007; Peul 2007). There can be a strong placebo effect, and spontaneous recoverymay occur, casting strong doubt on the validity of followup after surgery as a reference standard. Disc herniation found at surgery plus immediate post-surgery leg pain relief may be the optimal standard to identify lumbar disc herniation, but is only feasible in the selective population of surgical patients. Other, less invasive diagnostic procedures that were not addressed in this review (e.g. nerve conduction studies (Yagci 2009)), could be part of the diagnostic work-up and aid in the diagnosis of disc herniation as a cause of radiculopathy. Such procedures could form part of the reference standard in primary diagnostic research, but the diagnostic performance of these methods still needs to be firmly established.

Index tests

The limited performance (in particular sensitivity) of some commonly used tests, such as reflex impairments, may partly be explained by the effects of ageing. Most of our studies included working age populations, but the effect of age on the results of neurological testing is important for clinical practice. It is not uncommon for older adults to lose the Achilles reflex, although this usually is a bilateral change. Absent reflexes are found in 30% of those aged 61-70, and evenmore often in older subjects (Bowditch 1996). Spangfort also demonstrated an increasing occurrence of reflex impairments with age (Spangfort 1972). Consequently, in older patients, a missing ankle reflex does not necessarily indicate the presence of a herniated disc (false positives), which reduces the specificity of this test. Asymmetrical loss of the tendon reflexes is a more meaningful definition of a positive test result, but reflex asymmetry was described in only two studies (Albeck 1996; Gurdjian 1961).

Reliability

This review focused on the diagnostic performance (i.e. validity) of physical examination in patients with back pain, and a systematic search and synthesis of evidence on reliability was outside the scope of this

review. However, adequate reliability (interand intra-observer agreement) is a prerequisite for good performance of diagnostic tests. Our review showed that the procedures for physical examination were often poorly described, and it was often unclear whether or not tests were standardised, observers were trained, or what thresholds were used to define positive test results. Only four studies provided some information on reliability.

Vucetic 1996 only reported data on inter-observer variation of measuring spinal mobility, with coefficients of variation ranging between 5% and 7%. Poiraudeau 2001 reported good intraobserver reliability for the Bell test, hyperextension test, SLR and XSLR, with kappa ranging between 0.76 and 0.96. However, inter-observer agreement varied between pairs of observers and between tests, being fair for the Bell test (kappa 0.58 to 0.64), and weak to fair for hyperextension, SLR and XSLR (most kappa lower than 0.5). Vroomen et al. published a separate study on the reliability of physical examination tests in their cohort (Vroomen 2000), reporting good inter-observer agreement for muscle weakness and sensory deficits (kappa 0.57 to 0.82), and fair agreement for impaired reflexes (kappa 0.42 to 0.53). They found that assessments of the SLR and XSLR were most consistent (all kappa > 0.66), which seems to agree with results reported by Kosteljanetz 1988, who indicated that differences between observers in the angle at which pain was reported during the SLR were smaller than 10° in most patients. Other studies, reporting on observer agreement in other populations, have also reported variable results, for example, on the reliability of visual inspection of scoliosis (Clare 2005; Donahue 1996). The results of these studies indicate that reliability of tests performed during physical examination can be far from optimal, which will partly explain the poor diagnostic performance of most tests included in this review.

The importance of reliability in diagnostic test evaluation holds not only for index tests, but also for the reference standard. The reliability of imaging techniques in the identification of lumbar disc herniation has been reported to be moderate to good (Lurie 2008; Vroomen 2002 (prim care)), but evidence on the reliability of surgical findings is scarce. In many studies, information on surgical findings may have been derived from surgeons' operative notes, and we found no standards or criteria to guide surgeons when documenting their surgical findings. Poor reliability of surgical findings may have affected the results of diagnostic performance of the physical examination tests reported in this review, but given the lack of evidence, the extent of this effect is difficult to estimate.

Strengths and weaknesses of the review

Even though recent studies have recommended against the use of a methodological filter (Doust 2005; Leeflang 2006) we decided to employ a broad, sensitive filter to identify diagnostic accuracy studies. Searching for eligible publications was not easy: many relevant studies were published before 1985, were poorly indexed in electronic databases, and were often not specifically designed as a diagnostic accuracy study. Several of the (older) publications were finally identified through reference checking. In order to study the consequences of this decision we carried out a sensitivity analysis, repeating the search without the use of amethodological filter, and studying all additional citations from PubMed and EMBASE. As the methodological filter had only been applied to some elements of the electronic search (see Appendix 1), the number of additional citations was limited to approximately 450. However, none of these citations turned out to be relevant to our review. In about half of the citations, the study was clearly not a diagnostic accuracy study, but most other studies were also excluded, for other reasons. Therefore, in this review and for this particular topic, the application of a sensitive methodological filter in some parts of the search strategy had no implications for the identification of relevant studies.

Poor reporting in the original publications affected assessment of quality (risk of bias), and was one of the reasons for disagreements on some QUADAS items. The older studies in particular, and those studies not specifically designed as diagnostic accuracy studies, provided little information on important aspects of study design. The introduction and implementation of the STARDguidelines may improve reporting of diagnostic studies in future (Bossuyt 2003; Bossuyt 2003a; Smidt 2006). Assessment of quality in diagnostic reviews is further facilitated by defining clear guidelines for review authors on how to score individual items, and by piloting the procedures.

The number of studies for each index test was often small, and therefore a thorough analysis of the influence of potential sources of heterogeneity was only possible for the SLR. However, the results for most

other tests were clustered around the diagonal in ROCspace, indicating poor diagnostic performance. Further analysis of heterogeneity in these subsets of studies is unlikely to affect these results.

Applicability of findings to clinical practice and policy

Especially in primary care, physical examination is often used to distinguish between patients at low or high probability of lumbar disc herniation, and to decide which patients should be referred for imaging or are likely candidates for surgery. Therefore, the applicability of the results of our review is limited by the fact that most studies concerned surgical populations, and do not adequately represent patients with back and sciatica in other health care settings.

Equally important is the fact that most studies only presented the diagnostic value of individual aspects of physical examination.

When carried out in isolation, the diagnostic performance ofmost tests was poor, but in clinical practice, the combination of several elements of diagnostic information, including demographic data and information from patient history, together will contribute to estimating the likelihood of nerve root impingement. Investigating diagnostic performance of individual test results may underestimate the diagnostic performance of the process of physical examination.

This is confirmed by the relatively good performance of multivariable diagnostic model presented by Vucetic 1999. However, the models by Vroomen 2002 (prim care) showed that physical examination may have little to add to the results of a clinical history (area under the ROC curve 0.83 versus 0.80). Therefore, future diagnostic studies should focus on the relative contribution of information from patient history, physical examination and diagnostic imaging in order to develop diagnostic strategies that distinguish better between patients with or without radiculopathy due to lumbar disc herniation. These studies should be carried out in primary care populations. Another important contribution would be to use alternative outcomes, such as treatment decisions or recovery from symptoms, in order to study the role of physical examination in the management of patients with back pain and sciatica, and explore the consequences of positive and negative test results.

AUTHORS' CONCLUSIONS

Implications for practice

Available evidence suggests that when used in isolation, several aspects of physical examination (scoliosis, paresis or muscle weakness, muscle wasting, impaired reflexes, sensory deficits) do not accurately distinguish between low-back pain patients with or without lumbar radiculopathy due to disc herniation. For other tests (forward flexion, hyper extension test, and slump test), there was insufficient evidence to provide recommendations regarding their diagnostic accuracy or usefulness. In surgical populations, the SLR showed high sensitivity (and variable specificity), whereas the XSLR showed high specificity (coupled with low sensitivity).

However, these results were found in populations with a very high prevalence of disc herniation (mostly above 75%) and likely a severe spectrumof disease, and cannot be generalised to other populations.

The diagnostic performance of physical examination tests in primary care populations and other general, unselected patient groups, is still unclear as evidence from these settings is scarce.

An overview of the results of all tests is given in a summary table (Summary of results). Clear implications for practice are difficult to formulate, but the available evidence indicates that in patients with low-back pain and sciatica, a diagnosis of lumbar disc herniation should not be based on the results of one single physical examination test. Better performancemay be obtained when combinations of tests are evaluated, including information from both patient history and physical examination, but this requires further study.

Implications for research

There is a strong need for good quality prospective cohort studies that are carried out in general populations of patients in primary care presenting with low-back pain with radiating leg symptoms.

Preferably, these studies should evaluate the performance of combinations of diagnostic information in order to derive a diagnostic algorithm based on patient history and physical examination.

The performance of such diagnostic models can be tested against imaging in a consecutive series of patients with back and sciatica.

Clear definitions should be given for positive results of both index tests and reference standard outcome. Subsequent research may investigate the impact of applying a diagnostic model on decisions regarding referral and treatment, and on patient outcomes.

A C K N OWL E D G E M E N T S

We would like to thank Gustavo Zanoli and Cesar Hincapie for their help with papers published in Italian and Spanish, and Petra Macaskill, Rob Scholten and Vicki Pennick for their useful comments on this review. We are grateful to Vicki Pennick and Irina Schelkanova for their assistance with entering the review in RevMan 5.

FIGURE 1

Figure 1. Results of the search for studies evaluating the diagnostic performance of physical examination in the diagnosis of lumbar disc herniation

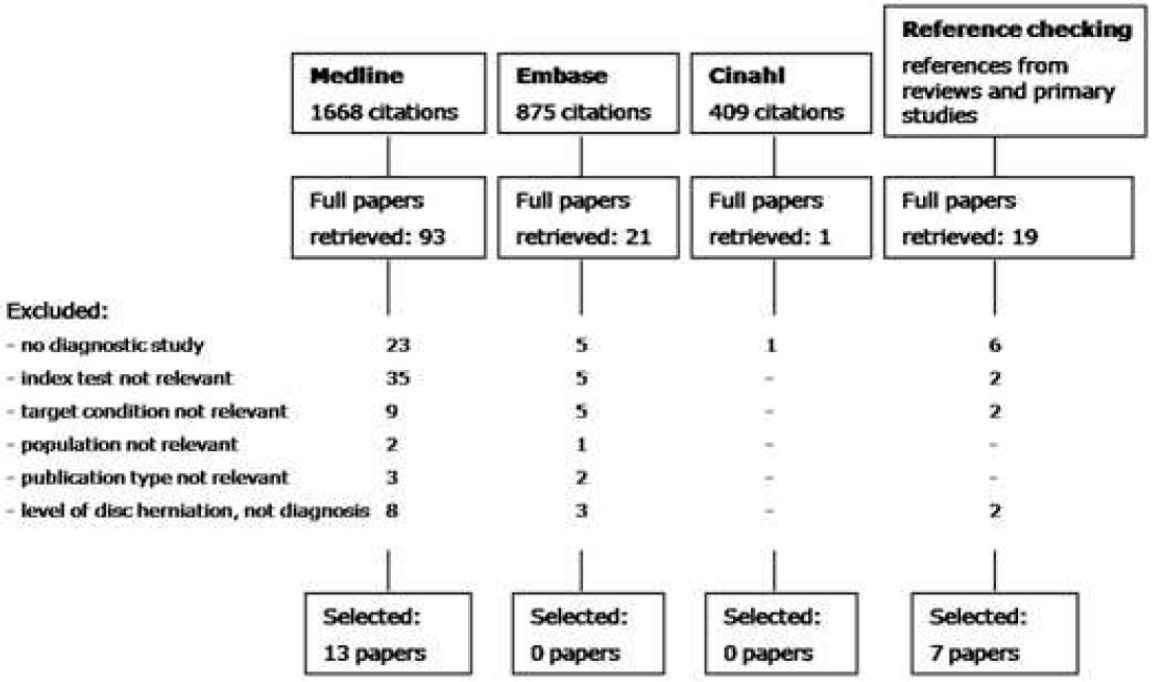


FIGURE 2

Quality assessment summary: review authors' judgements about each risk of bias item for each included study.

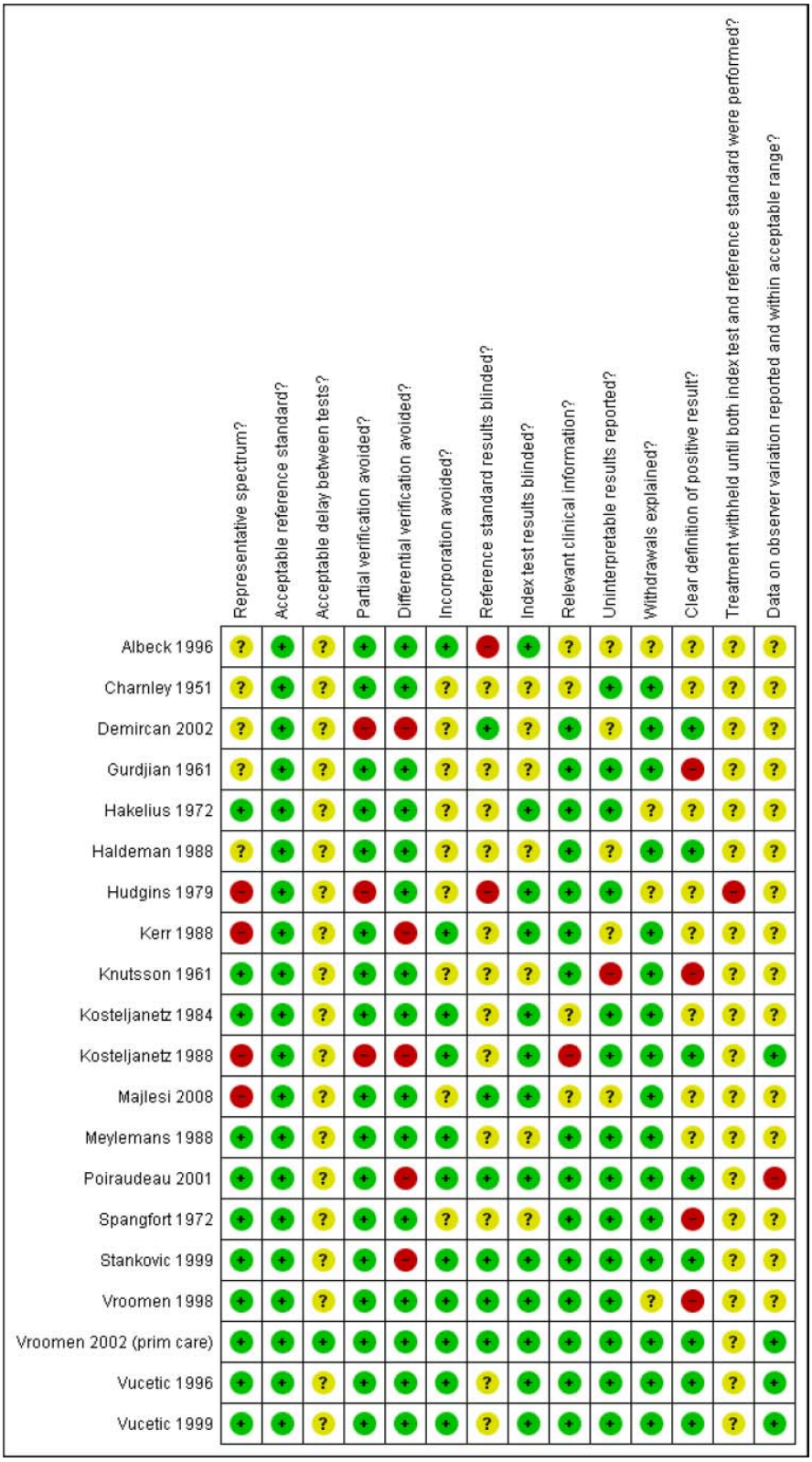


FIGURE 3

Figure 3. Forest plot: SLR (leg pain at any angle) - reference test: imaging

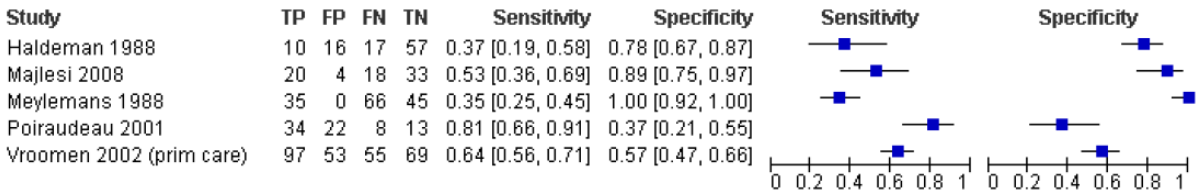


FIGURE 4

Figure 4. Forest plot: SLR (leg pain at any angle) - reference test surgery

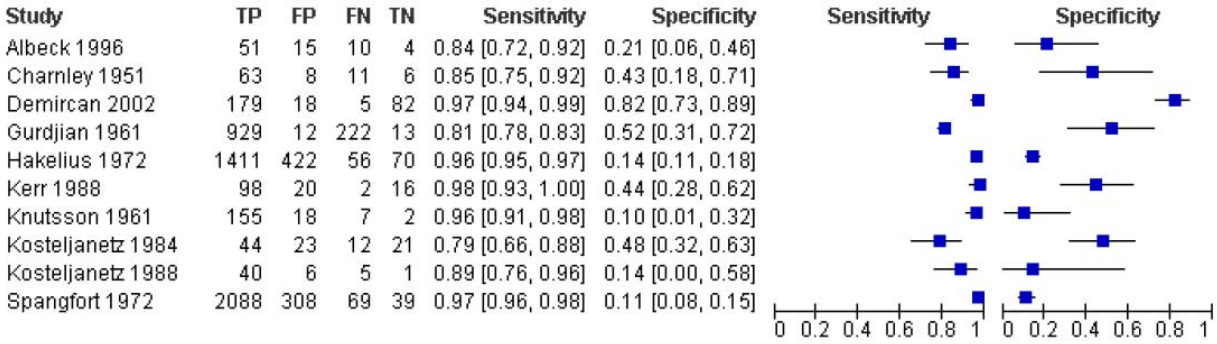


FIGURE 5

Figure 5. Summary ROC Plo: SLR (leg pain at any angle) - reference test surgery *Note: the case-control s by Demircan et al (2002 - upper left in ROC-space) showed clear review bias and was excluded from the poo analysis.*

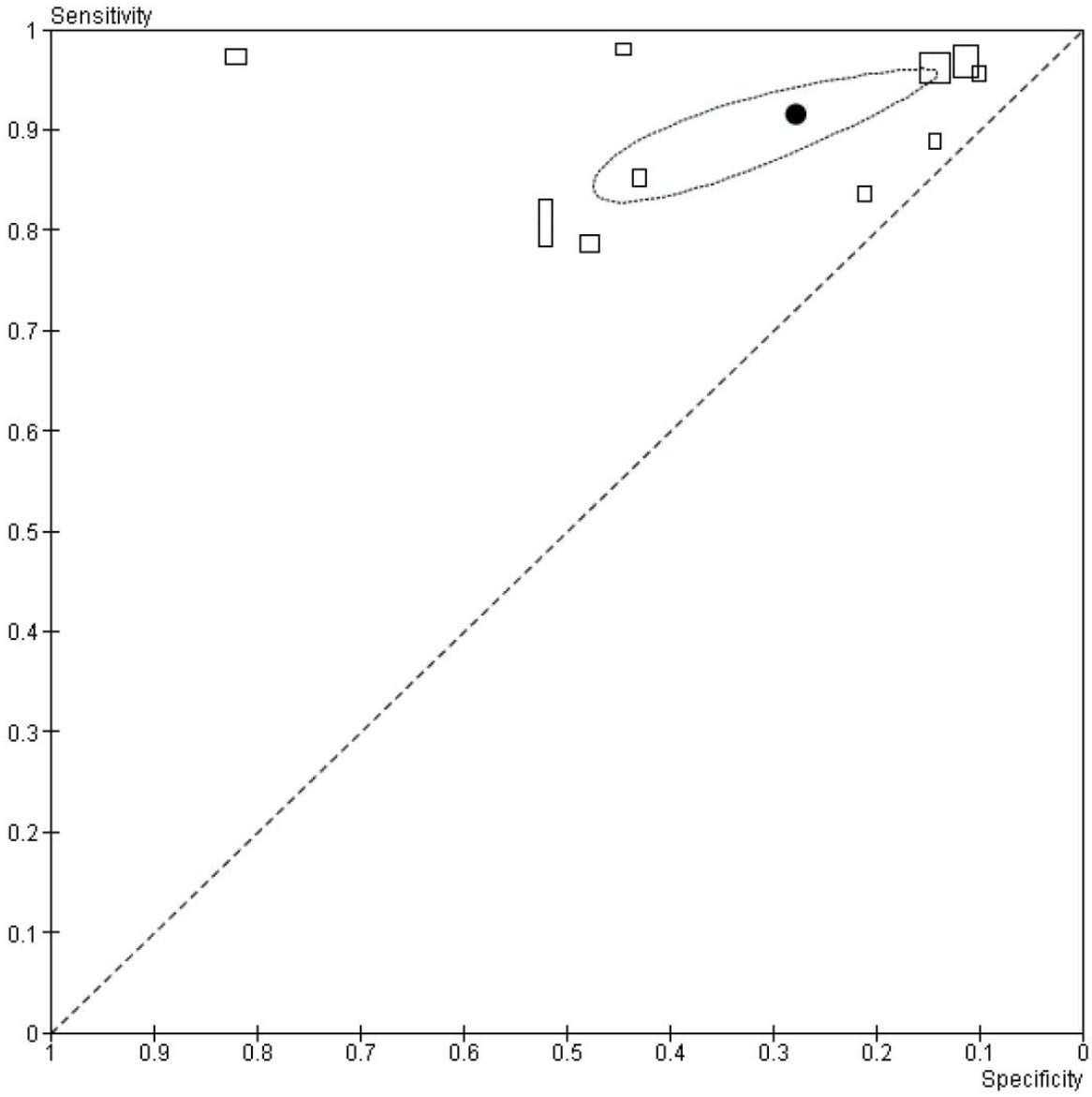


FIGURE 6

Figure 6. Forest plot:XSLR (leg pain at any angle) - reference test surgery or imaging (study by Poiraudeau et al.)

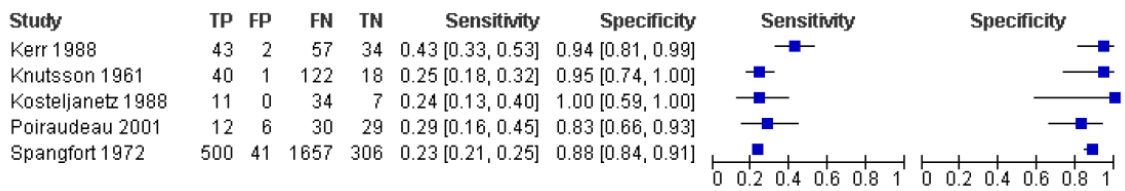


FIGURE 7

Figure 7. Summary ROC Plot of X-SLR (leg pain at any angle) - reference test surgery or imaging. *Note: T study by Kerr et al. (1988 - highest sensitivity among the studies) was found to have a high risk of verification b*

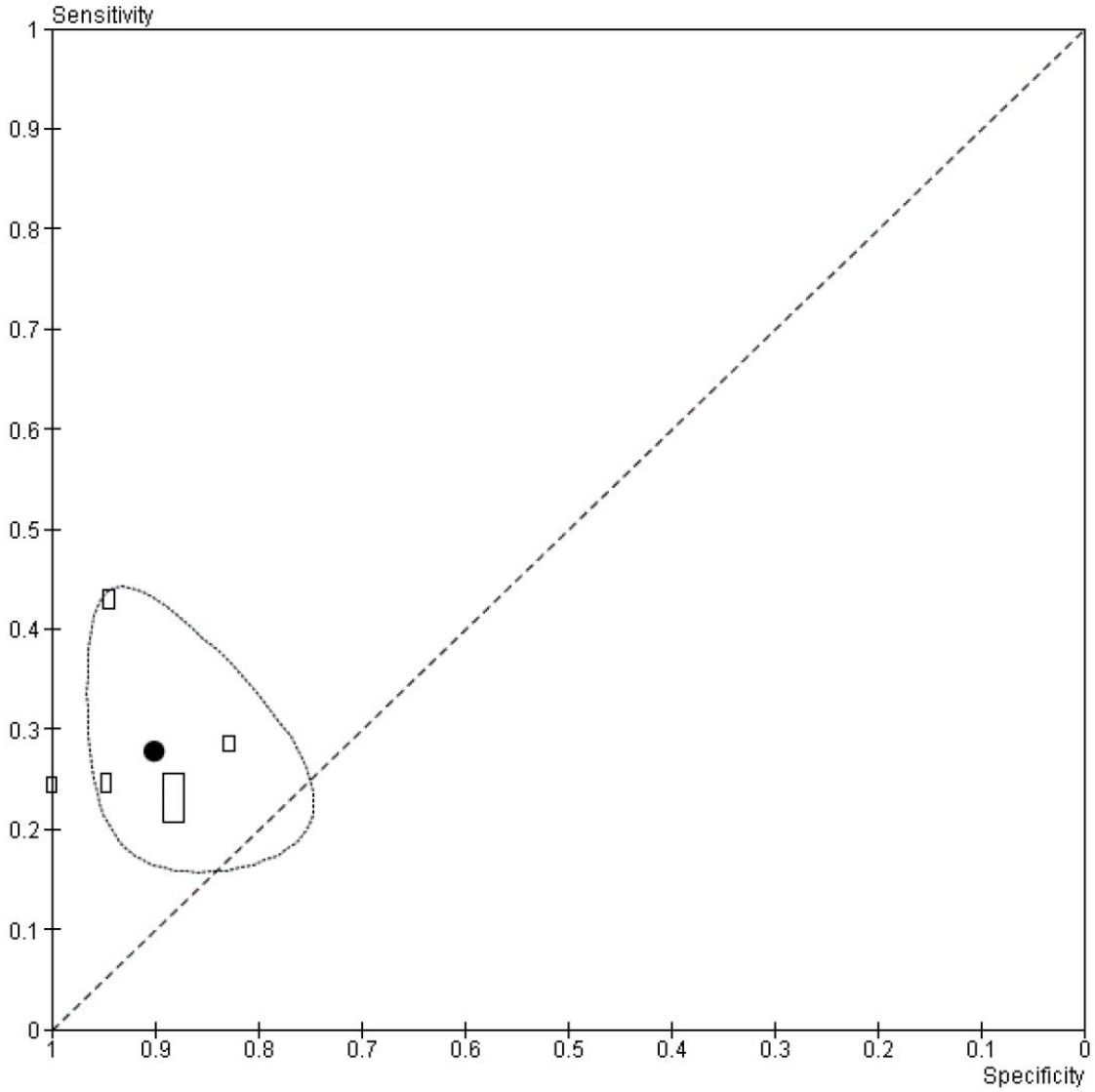


FIGURE 8

Figure 8. Forest plot: Scoliosis (visual inspection) - reference test surgery

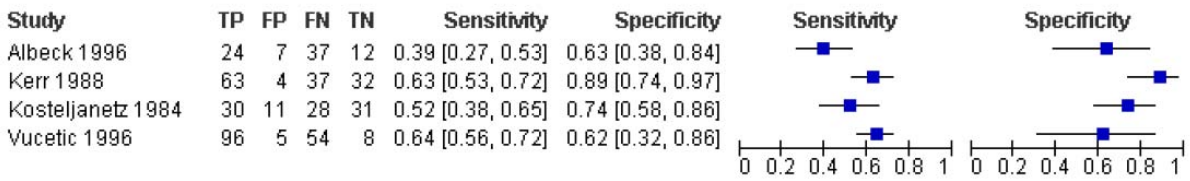


FIGURE 9

Figure 9. Forest plot: Paresis (dorsiflexion toe/ankle) - reference test surgical findings or imaging (study by Vroomen et al.)

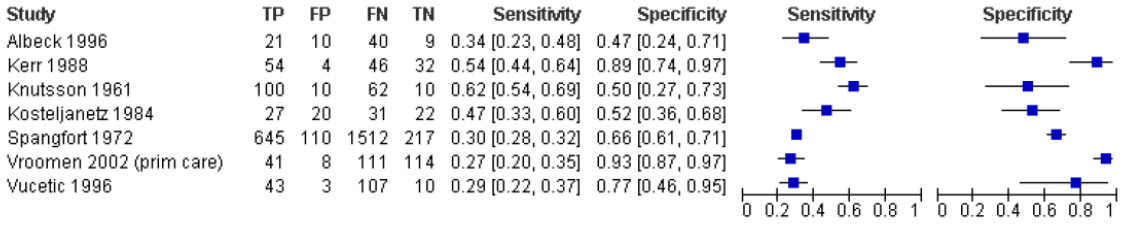


FIGURE 10

Figure 10. Forest plot: Muscle wasting - reference test surgical findings

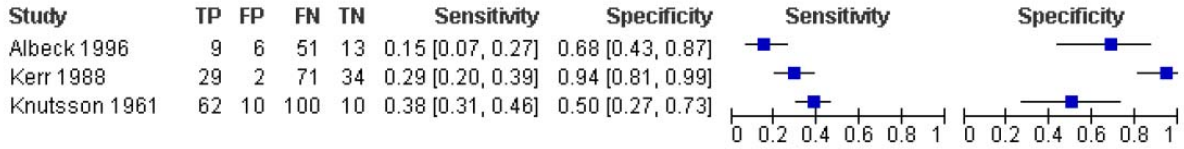


FIGURE 11

Figure 11. Forest plot: Impaired reflexes (Achilles tendon) - reference test: surgical findings or imaging (study by Vroomen et al).

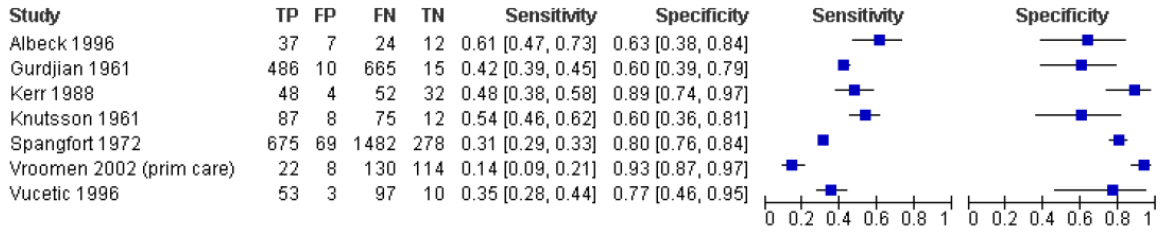


FIGURE 12

Figure 12. Forest plot: Sensory deficits - reference test: surgical findings or imaging (study by Vroomen et al.)

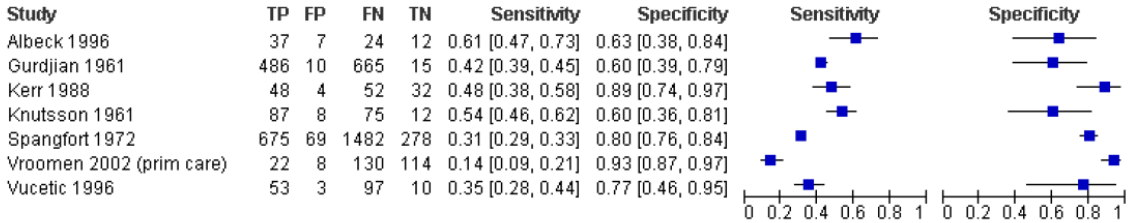


FIGURE 13

Figure 13. Forest plot of Forward flexion - reference test surgery or imaging (study by Vroomen et al.)

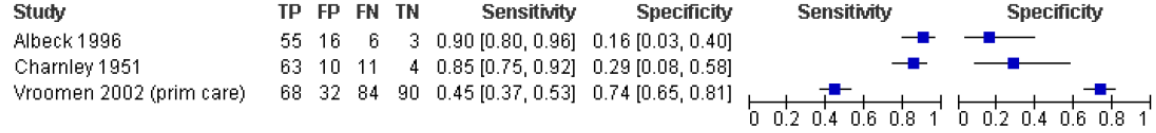


TABLE 1

Summary of results. Summary Table: performance of aspects of physical examination in the diagnosis of lumbar radiculopathy due to disc herniation in patients with low-back pain

What is the performance of tests performed during physical examination to identify radiculopathy due to lower lumbar disc herniation in patients with low-back pain and sciatica?
Population: Patients with LBP and sciatica in primary or secondary care
Prior testing:Variable or not described, imaging (X-ray or myelography was reported for some surgical populations
Index tests: All relevant physical examination tests, including the straight leg raising test, crossed straight leg raising test, paresis or muscle weakness, sensory deficits, and impaired reflexes
Target condition: radiculopathy due to lumbar disc herniation
Reference standard: Diagnostic imaging (MRI, CT, myelography) or findings at surgery.
Studies: Cohort studies (16) or case-control studies (n=3)

Diagnostic test	Setting / reference standard	Pooled estimate for sensitivity (95% CI) *or range of estimates*	Pooled estimate for specificity (95% CI) *or range estimates*	Mean prevalence disc herniation (range)**	Summary estimate LR of positive test result (95% CI)	Summary estimate LR negative test result (95% CI)

Summary of results. Summary Table: performance of aspects of physical examination in the diagnosis of lumbar radiculopathy due to disc herniation in patients with low-back pain (*Continued*)

Population: patients with LBP in primary or secondary care **Index test:** Straight leg raising test (SLR) 5 studies, no pooled analysis	Diagnostic imaging: 1 primary care / MRI: 4 sec care / CT-MTI:	0.64 (0.56 to 0.71)* range: 0.35 to 0.81	0.57 (0.47 to 0.66)* range: 0.37 to 1.00	55% 40% (26 to 55%)		
Population: patients with LBP referred for surgery **Index test**: Straight leg raising test (SLR) pooled estimate of 9 studies	Surgical findings	0.92 (0.87 to 0.95)	0.28 (0.18 to 0.40)	82% (58 to 98%)	1.3 (1.1 to 1.4)	0.30 (0.24 to 0.39)
Population: patients with LBP referred for surgery **Index test**: Crossed straight leg raising test (XSLR) pooled estimate of 5 studies	Surgical findings	0.28 (0.22 to 0.35)	0.90 (0.85 to 0.94)	77% (55 to 89%)	2.1 (1.6 to 2.8)	0.86 (0.83 to 0.89)
Population: patients with LBP referred for surgery **Index test**: Scoliosis 5 studies, no pooled analysis	Surgical findings	range: 0.39 to 0.68	range: 0.62 to 0.89	66% (58 to 76%)		
Population: patients with LBP in primary care or referred for surgery **Index test**: Paresis or muscle weakness 7 studies, no	1 primary care / MRI: 6 surgical findings:	0.27 (0.20 to 0.37)* range: 0.29 to 0.62	0.93 (0.88 to 0.97)* range: 0.50 to 0.89	55% 74% (58 to 89%)		

Summary of results. Summary Table: performance of aspects of physical examination in the diagnosis of lumbar radiculopathy due to disc herniation in patients with low-back pain (*Continued*)

pooled analysis						
Population: patients with LBP referred for surgery **Index test**: Muscle wasting 3 studies, no pooled analysis	surgical findings	range: 0.15 to 0.38	range: 0.50 to 0.94	83% (76 to 89%)		
Population: patients with LBP in primary care or referred for surgery **Index test**: Impaired reflexes 7 studies, no pooled analysis	1 primary care / MRI: 6 surgical findings:	0.15 (0.09 to 0.21)* range: 0.31 to 0.62	0.93 (0.88 to 0.97)* range: 0.60 ? 0.89	55% 82% (63 to 98%)		
Population: patients with LBP in primary care or referred for surgery **Index test**: Sensory deficits 6 studies, no pooled analysis	1 primary care / MRI: 5 surgical findings:	0.28 (0.21 to 0.36)* range: 0.26 to 0.67	0.66 (0.56 to 0.74)* range: 0.42 to 0.69	55% 72% (58 to 89%)		
Population: patients with LBP in primary or referred for surgery **Index test**: Forward flexion 3 studies, no pooled analysis	1 primary care / MRI: 2 surgical findings:	0.45 (0.37 to 0.53)* range: 0.85 and 0.90	0.74 (0.65 to 0.81)* range: 0.16 and 0.29	55% range: 76 and 84%		
Population: patients with LBP in secondary care **Index test**: Extension test 2 studies, no pooled analysis	secondary care / MRI-CT	range: 0.13 to 0.90, depending on cut-point	range: 0.17 to 0.94, depending on cut-point	range: 55 and 88%		

Summary of results. Summary Table: performance of aspects of physical examination in the diagnosis of lumbar radiculopathy due to disc herniation in patients with low-back pain (*Continued*)

Population: patients with LBP in secondary care **Index test:** Slump test 2 studies, no pooled analysis	secondary care / MRI-CT	range: 0.44 to 0.87, depending on cut-point	range: 0.23 to 0.63, depending on cut-point	88% and unclear (case control study)		
Population: patients with LBP in secondary care **Indext test**: Bell test 1 study	secondary care / MRI-CT	0.49 (0.33 to 0.65)	0.63 (0.45 to 0.79)	55%		

* Results from the single primary care study are presented separately. For surgical studies, range of estimates is given (no pooled analysis because of heterogeneity)

** Not weighted for sample size

REFERENCES

References to studies included in this review

Albeck 1996 {published data only}
_ Albeck MJ. A critical assessment of clinical diagnosis of disc herniation in patients with monoradicular sciatica. Acta Neurochir
(Wien) 1996;138:40–4.
Charnley 1951 {published data only}
Charnley J. Orthopaedic signs in the diagnosis of disc protrusion.
With special reference to the straight leg raising test. Lancet 1951; 260:186–92.
Demircan 2002 {published data only}
Demircan MN, Colak A, Kutlay M, Kibici K, Topuz K. Cramp finding: can it be used as a new diagnostic and prognostic factor in lumbar disc surgery?. Eur Spine J 2002;11:47–51.
Gurdjian 1961 {published data only}
Gurdjian ES,Webster JE, Ostrowski AZ, HardyWG, Lindner DW, Thomas LM. Herniated lumbar intervertebral discs: An analysis of 1176 operated cases. J Trauma 1961;1:158–76.
Hakelius 1972 {published data only}
Hakelius A, Hindmarsh J. The significance of neurological signs and myelographic findings in the diagnosis of lumbar root compression. Acta Orthop Scand 1972;43:239–46.
Haldeman 1988 {published data only}
Haldeman S, Shouka M, Robboy S. Computed tomography, electrodiagnostic and clinical findings in chronic workers' compensation patients with back and leg pain. Spine 1988;13: 345–50.
Hudgins 1979 {published data only}
Hudgins WR. The crossed straight leg raising test: a diagnostic sign of herniated disc. J OccupMed 1972;21:407–8.
Kerr 1988 {published data only}
Kerr RS, Cadoux-Hudson TA, Adams CB. The value of accurate clinical assessment in the surgical management of the lumbar disc protrusion. J Neurol Neurosurg Psychiatry 1988;51:169–73.
Knutsson 1961 {published data only}
Knutsson B. Comparative value of electromyographic, myelographic and clinical neurological examinations in diagnosis of lumbar root compression syndrome. Acta Orthop Scand Suppl
1961;49:1–134.
Kosteljanetz 1984 {published data only}
Kosteljanetz M, Espersen JO, Halaburt H, Miletic T. Predictive value of clinical and surgical findings in patients with lumbagosciatica.
A prospective study (Part I). Acta Neurochir (Wien) 1984; 73:67–76.
Kosteljanetz 1988 {published data only}
Kosteljanetz M, Bang F, Schmidt-Olsen S. The clinical significance of straight-leg raising (Lasègue's sign) in the diagnosis of prolapsed lumbar disc. Interobserver variation and correlation with surgical finding. Spine 1988;13:393–5.
Majlesi 2008 {published data only}
Majlesi J, Togay H, Unalan H, Toprak S. The sensitivity and specificity of the Slump and the Straight Leg Raising tests in patients with lumbar disc herniation. J Clin Rheumatol 2008;14:87–91.
Meylemans 1988 {published data only}
Meylemans L, Vancraeynest T, Bruyninckx F, Rosselle N. A comparative study of EMG and CAT scan in the lumbo-ischial syndrome. Acta Belg Med Phys 1988;11:35–42.
Poiraudeau 2001 {published data only}
Poiraudeau S, Foltz V, Drape JL, Fermanian J, Lefevre-Colau MM, Mayoux-Benhamou MA, et al.Value of the bell test and the hyperextension test for diagnosis in sciatica associated with disc herniation: comparison with Lasègue's sign and the crossed Lasègue's sign. Rheumatology (Oxford) 2001;40:460–6.
Spangfort 1972 {published data only}
Spangfort EV. The lumbar disc herniation. A computer-aided analysis of 2,504 operations. Acta Orthop Scand Suppl 1972;142: 1–95.
Stankovic 1999 {published data only}
Stankovic R, Johnell O, Maly P, Willner S. Use of lumbar extension, slump test, physical and neurological examination in the evaluation of patients with suspected herniated nucleus pulposus. A prospective clinical study. Man Ther 1999;4:25–32.
Vroomen 1998 {published data only}

Vroomen PC, Van Hapert SJ, van Acker RE, Beuls EA, Kessels AG, Wilmink JT. The clinical significance of gadolinium enhancement of lumbar disc herniations and nerve roots on preoperative MRI.
Neuroradiology 1998;40:800–6.
Vroomen 2002 (prim care) {published data only}
Vroomen PC, de Krom MC, Wilmink JT, Kester AD, Knottnerus JA. Diagnostic value of history and physical examination in patients suspected of lumbosacral nerve root compression. J Neurol
Neurosurg Psychiatry 2002;72:630–4.
Vucetic 1996 {published data only}
Vucetic N, Svensson O. Physical signs in lumbar disc hernia. Clin
Orthop Relat Res 1996;333:192–201.
Vucetic 1999 {published data only}
Vucetic N, Astrand P, Guntner P, Svensson O. Diagnosis and prognosis in lumbar disc herniation. Clin Orthop Relat Res 1999; 361:116–22.
References to studies excluded from this review
Albert 1993 {published data only}
Albert TJ, Balderston RA, Heller JG, Herkowitz HN, Garfin SR, Tomany K, et al.Upper lumbar disc herniations. Journal of Spinal
Disorders 1993;6(4):351–9.
Brugnoni 1998 {published data only}
Brugnoni G. Semeiotics of manual medicine in lumbar discal herniation. Riabilitazione 1998;31(4):163–72.
Edgar 1974 {published data only}
Edgar MA, Park WM. Induced pain patterns on passive straight-leg raising in lower lumbar disc protrusion. A prospective clinical, myelographic and operative study in fifty patients. J Bone Joint Surg
Br 1974;56-B(4):658–67.
Jensen 1987 {published data only}
Jensen OH. The level-diagnosis of a lower lumbar disc herniation: the value of sensibility and motor testing. Clin Rheumatol 1987;6
(4):564–9.
Kortelainen 1985 {published data only}
Kortelainen P, Puranen J, Koivisto E, Lahde S. Symptoms and signs of sciatica and their relation to the localization of the lumbar disc herniation. Spine 1985;10(1):88–92.
Lansche 1960 {published data only}
Lansche WE. Ford LT [Correlation of the myelogram with clinical and operative findings in lumbar disc lesions]. J Bone Jopint Surg
[Am] 1960;42:193–206.
Portnoy 1972 {published data only}
Portnoy HD, Ahmad M. Value of the neurological examination, electromyography and myelography in herniated lumbar disc.
Mich.Med 1972;71(14):429–34.
Rainville 2003 {published data only}
Rainville J, Jouve C, Finno M, Limke J. Comparison of four tests of quadriceps strength in L3 or L4 radiculopathies. Spine 2003;28
(21):2466–71.
Reihani-Kermani 2003 {published data only}
Reihani-Kermani H. Level-diagnosis of lumbar disc herniation.
Iranian Journal of Medical Sciences 2003;28(3):135–8.
Reihani-Kermani 2004 {published data only}
Reihani-Kermani H. Correlation of clinical presentation with intraoperative level diagnosis in lower lumbar disc herniation. Ann
Saudi Med 2004;24(0256-4947, 4):273–5.
Supik 1994 {published data only}
Supik LF, Broom MJ. Sciatic tension signs and lumbar disc herniation. Spine 1994;19(9):1066–9.
Thelander 1992 {published data only}
Thelander U, Fagerlund M, Friberg S, Larsson S. Straight leg raising test versus radiologic size, shape, and position of lumbar disc hernias. Spine 1992;17(4):395–9.
Xin 1987 {published data only}
Xin SQ, Zhang QZ, Fan DH. Significance of the straight-legraising test in the diagnosis and clinical evaluation of lower lumbar intervertebral-disc protrusion. J Bone Joint Surg Am 1987;69(4): 517–22.

Additional references

Bachmann 2002
Bachmann LM, Coray R, Estermann P, Ter Riet G. Identifying diagnostic studies in MEDLINE: reducing the number needed to read. Journal of the American Medical Informatics Association 2002;9 (6):653–8.
Bachmann 2003
Bachmann LM, Estermann P, Kronenberg C, ter Riet G.
Identifying diagnostic accuracy studies in EMBASE. Journal of the
Medical Library Association 2003;91(3):341–6.
Boden 1990
Boden SD, Davis DO, Dina TS, Patronas NJ, Wiesel SW.
Abnormal magnetic-resonance scans of the lumbar spine in asymptomatic subjects. A prospective investigation. Journal of Bone
and Joint Surgery 1990;72(3):403–8.
Boden 1996
Boden SD, Wiesel SW. Lumbar spine imaging: role in clinical decision making. J AM Acad Orthop Surg 1996;4:238–48.
Borenstein 2001
Borenstein DG, OMara JW, Bodwen SD, Lauerman WC, Jacobson A, Platenberg C, et al.The value of magnetic resonance imaging of the lumbar spine to predict low-back pain in asymptomatic subjects : a seven year follow-up study. J Bone Joint Surg 2001;83-A: 1306–11.
Bossuyt 2003
Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al.Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. Ann Intern Med 2003;138:40–4.
Bossuyt 2003a
Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Moher D, Rennie D, de Vet HC, Lijmer JG, Standards for Reporting of Diagnostic Accuracy. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Ann Intern Med 2003;138:W1–12.
Bowditch 1996
Bowditch MG, Sanderson P, Livesey JP. The significance of an absent ankle reflex. J Bone Joint Surg 1996;78-Br:276–9.
Clare 2005
Clare HA, Adams R, Maher CG. Reliability of detection of lumbar lateral shift. J Manipulative Physiol Ther 2005;26:476–80.
Devillé 2000
Devillé WL, van der Windt DA, Dzaferagic A, Bezemer PD, Bouter LM. The test of Lasègue: systematic review of the accuracy in diagnosing herniated discs. Spine 2000;25(9):1140–7.
Devillé 2000a
Devillé WL, Bezemer PD, Bouter LM. Publications on diagnostic test evaluation in family medicine journals: an optimal search strategy. Journal of Clinical Epidemiology 2000;53(1):65–9.
Deyo 1990
Deyo RA, Loeser JD, Bigos SJ. Herniated lumbar intervertebral disk. Annals of Internal Medicine 1990;112(8):598–603.
Deyo 1992
Deyo RA, Rainville J, Kent DL. What can the history and physical examination tell us about low back pain?. JAMA 1992;268(6): 760–5.
Donahue 1996
Donahue MS, Riddle D, Sullivan MS. Intertester reliability of a modified version of McKenzie's lateral shift assessments obtained on patients with low back pain. Phys Ther 1996;76:706–26.
Doust 2005
Doust JA, Pietrzak E, Sanders S, Glasziou PP. Identifying studies for systematic reviews of diagnostic tests was difficult due to the poor sensitivity and precision of methodologic filters and the lack of information in the abstract. J Clin Epidemiol 2005;58:444–9.
Gibson 2007
Gibson JNA, Waddell G. Surgical interventions for lumbar disc prolapse. Cochrane Database of Systematic Reviews 2007, Issue 2.
[DOI: 10.1002/14651858.CD001350.pub4] Handbook 2005

Smidt N, Deeks J, Moore T (Eds). Cochrane Handbook for Cochrane Reviews of Diagnostic Test Accuracy. The Cochrane Collaboration July 2005, issue version 0.3.
Jarvik 2005
Jarvik JG, Hollingworth W, Heagerty PJ, Boyko EJ, Deyo RA.
Three-year incidence of low back pain in an initially asymptomatic cohort: clinical and imaging risk factors. Spine 2005;30:1541–8.
Kim 2002
Kim YS, Chin DK, Yoon DH, Jin BH, Cho YE. Predictors of successful outcome for lumbar chemonucleolysis: analysis of 3000 cases during the past 14 years. Neurosurgery 2002;51(5 Suppl): S123–8.
Kohlboeck 2004
Kohlboeck G, Greimel KV, Piotrowski WP, Leibetseder M, Krombholz-Reindl M, Neuhofer R, et al.Prognosis of multifactorial outcome in lumbar discectomy: a prospective longitudinal study investigating patients with disc prolapse. Clinical Journal of Pain
2004;20(6):455–61.
Leeflang 2006
Leeflang MM, Scholten RJ, Rutjes AW, Reitsma JB, Bossuyt PM.
Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. J Clin
Epidemiol 2006;59:234–40.
Lijmer 1999
Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, et al.Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999;282(11):1061–6.
Lurie 2008
Lurie JD, Tosteson AN, Tosteson TD, Carragee E, Carrino JA, Kaiser J, et al.Reliability of magnetic resonance imaging readings for lumbar disc herniation in the Spine Patient Outcomes Research Trial (SPORT). Spine 2008;33:991–8.
Maitland 1985
Maitland GD. The slump test: examination and treatment. Austr J
Physiother 1985;31:215.
Peul 2007
Peul WC, van Houwelingen HC, van den Hout WB, Brand R, Eekhof JA, Tans JT, et al: The Hague Spine Intervention Prognostic Study Group. Surgery versus prolonged conservative treatment for sciatica. N Engl J Med 2007;356:2245–56.
Rebain 2002
Rebain R, Baxter GD, McDonough S. A systematic review of the passive straight leg raising test as a diagnostic aid for low back pain (1989 to 2000). Spine 2002;27(17):E388–95.
Rebain 2003
Rebain R, Baxter GD, McDonough S. The passive straight leg raising test in the diagnosis and treatment of lumbar disc herniation: a survey of United kingdom osteopathic opinion and clinical practice. Spine 2003;28(15):1717–24.
Reitsma 2005
Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews.
Journal of Clinical Epidemiology 2005;58(10):982–90.
Rutjes 2006
Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. CMAJ 2006;174(4):469–76.
Smidt 2006
Smidt N, Rutjes AW, van der Windt DA, Ostelo RW, Bossuyt PM, Reitsma JB, et al.The quality of diagnostic accuracy studies since the STARD statement: has it improved?. Neurology 2006;67:792–7.
van den Hoogen 1995
van den Hoogen HM, Koes BW, van Eijk JT, Bouter LM. On the accuracy of history, physical examination, and erythrocyte sedimentation rate in diagnosing low back pain in general practice.
A criteria-based review of the literature. Spine 1995;20(3):318–27.
Vroomen 1999
Vroomen PC, de Krom MC, Knottnerus JA. Diagnostic value of history and physical examination in patients suspected of sciatica due to disc herniation: a systematic review. Journal of Neurology

1999;246(10):899–906.
Vroomen 2000
Vroomen PC, De Krom MC, Knottnerus AJ. Consistency of history taking and physical examination in patients with suspected lumbar nerve root involvement. Spine 2000;25:91–7.
Weinstein 2006
Weinstein JN, Tosteson TD, Lurie JD, Tosteson AN, Hanscom B, Skinner JS, et al.Surgical vs nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT): a randomized trial. JAMA 2006;296(20):2441–50.
Whiting 2004
Whiting P, Rutjes AW, Dinnes J, Reitsma J, Bossuyt PM, Kleijnen J. Development and validation of methods for assessing the quality of diagnostic accuracy studies. Health Technology Assessment 2004;8 (25):iii, 1-234.
Yagci 2009
Yagci I, Gunduz OH, Ekinci G, Diracoglu D, Us O, Akyuz G. The Utility of Lumbar Paraspinal Mapping in the Diagnosis of Lumbar Spinal Stenosis. Am J Phys Med Rehabil 2009;Aug 15:E-pub ahead of print.
_ Indicates the major publication for the study

CHARACTERISTICSOFSTUDIES

Characteristics of included studies *[ordered by study ID]*

Albeck 1996

Clinical features and settings	Secondary care, Denmark - sampling unclear - first surgery - diagnostic imaging before surgery, results not reported
Participants	80 patients with monoradicular pain, who failed conservative treatment: 60% male, median age 40 (21 to 59) years
Study design	Prospective cohort
Target condition and reference standard(s)	Surgical findings: Extruded nucleus pulposus tissue
Index and comparator tests	Segmental spasm, trunk list, mobility (finger to floor), SLR, sensory deficits, impaired reflexes
Follow-up	
Notes	prevalence disc herniation: 76%

Table of Methodological Quality

Item	Authors' judgement	Description
Representative spectrum? All tests	Unclear	unclear from text: consecutive series?
Acceptable reference standard? All tests	Yes	findings at surgery
Acceptable delay between tests? All tests	Unclear	unclear from text
Partial verification avoided? All tests	Yes	all patients received same reference standard
Differential verification avoided? All tests	Yes	all patients received same reference standard
Incorporation avoided? All tests	Yes	diagnosis based on surgical findings only
Reference standard results blinded? All tests	No	surgeon not blind to results of phys ex

Albeck 1996 (*Continued*)

Index test results blinded? All tests	Yes	phys ex before surgery
Relevant clinical information? All tests	Unclear	unclear from text: unclear who performed phys ex, and if this person was aware of other info
Uninterpretable results reported? All tests	Unclear	unclear from text
Withdrawals explained? All tests	Unclear	unclear from text
Clear definition of positive result? All tests	Unclear	unclear from text: execution of test unclear
Treatment withheld until both index test and reference standard were performed? All tests	Unclear	no information
Data on observer variation reported and within acceptable range? All tests	Unclear	no data presented

Charnley 1951

Clinical features and settings	Secondary care, USA, 1946-56 - sampling method unclear - no information on previous surgery - no information on pre-surgery tests
Participants	88 patients with sciatica: selection criteria, age and sex not reported.
Study design	Historical cohort
Target condition and reference standard(s)	Surgical findings: lumbar disc protrusion
Index and comparator tests	SLR, SLR in combination with forward flexion
Follow-up	
Notes	Prevalence disc herniation: 84%

Table of Methodological Quality

Item	Authors' judgement	Description

Charnley 1951 (*Continued*)

Representative spectrum? All tests	Unclear	No information on selection
Acceptable reference standard? All tests	Yes	findings at surgery
Acceptable delay between tests? All tests	Unclear	unclear from text
Partial verification avoided? All tests	Yes	all patients received same reference standard
Differential verification avoided? All tests	Yes	all patients received same reference standard
Incorporation avoided? All tests	Unclear	unclear if surgeon performed tests
Reference standard results blinded? All tests	Unclear	unclear if surgeon was blind
Index test results blinded? All tests	Unclear	unclear from text
Relevant clinical information? All tests	Unclear	unclear from text
Uninterpretable results reported? All tests	Yes	all test results reported
Withdrawals explained? All tests	Yes	no withdrawals
Clear definition of positive result? All tests	Unclear	no information on execution of tests
Treatment withheld until both index test and reference standard were performed? All tests	Unclear	unclear from text
Data on observer variation reported and within acceptable range? All tests	Unclear	no data presented

Demircan 2002

Clinical features and settings	Secondary care, Turkey, 1997-90 - sampling method unclear - no information on previous surgery - no information on pre-surgery tests
Participants	100 surgical patients (A) 100 patients with LBP and sciatica, no need for surgery (B) 100 healthy controls (C) Male: A 76%, B 64%, C 66% Mean age: A 33 (20-42), B 36 (20-45), C 36 (20-45) years.
Study design	Case control
Target condition and reference standard(s)	Surgical findings (A) MRI (A, B, C): Lumbar disc protrusion or sequestration.
Index and comparator tests	SLR
Follow-up	
Notes	

Table of Methodological Quality

Item	Authors' judgement	Description
Representative spectrum? All tests	Unclear	patient sampling procedures unclear
Acceptable reference standard? All tests	Yes	MRI or surgical findings
Acceptable delay between tests? All tests	Unclear	unclear from text
Partial verification avoided? All tests	No	only test positives received surgery
Differential verification avoided? All tests	No	patients received either MRI or surgery as reference standard
Incorporation avoided? All tests	Unclear	unclear if PE was performed independent from reference test
Reference standard results blinded? All tests	Yes	surgeon / radiologist not informed of results of cramp test

Demircan 2002 (*Continued*)

Index test results blinded? All tests	Unclear	case control study, unclear if index test was carried out before or after surgery
Relevant clinical information? All tests	Yes	similar as in usual care
Uninterpretable results reported? All tests	Unclear	unclear from text
Withdrawals explained? All tests	Yes	no withdrawals
Clear definition of positive result? All tests	Yes	clear definition of cramp test given
Treatment withheld until both index test and reference standard were performed? All tests	Unclear	unclear from text
Data on observer variation reported and within acceptable range? All tests	Unclear	no data presented

Gurdjian 1961

Clinical features and settings	Secondary care, United Kingdom - sampling method unclear - no information on previous surgery - myelography in all before surgery, positive in > 80%
Participants	1176 surgical patients with sciatica: 65% men, 52% older than 40 years.
Study design	Historical cohort
Target condition and reference standard(s)	Surgical findings: lumbar disc protrusion or rupture
Index and comparator tests	SLR, loss of Achilles and patellar tendon reflex
Follow-up	
Notes	Prevalence disc herniation: 98%

Table of Methodological Quality

Item	Authors' judgement	Description

Gurdjian 1961 (*Continued*)

Representative spectrum? All tests	Unclear	methods for patient sampling unclear
Acceptable reference standard? All tests	Yes	findings at surgery
Acceptable delay between tests? All tests	Unclear	unclear from text
Partial verification avoided? All tests	Yes	all patients received same reference standard
Differential verification avoided? All tests	Yes	all patients received same reference standard
Incorporation avoided? All tests	Unclear	unclear who performed phys ex
Reference standard results blinded? All tests	Unclear	unclear from text
Index test results blinded? All tests	Unclear	unclear who performed tests
Relevant clinical information? All tests	Yes	as in usual care
Uninterpretable results reported? All tests	Yes	all test results reported
Withdrawals explained? All tests	Yes	no withdrawals
Clear definition of positive result? All tests	No	no information on criteria or on execution of test
Treatment withheld until both index test and reference standard were performed? All tests	Unclear	unclear from text
Data on observer variation reported and within acceptable range? All tests	Unclear	no data

Hakelius 1972

Clinical features and settings	Secondary care, Sweden, 1939-64 - sampling: all surgical patients - first surgery - myelography in all patients before surgery, results unclear
Participants	1986 surgical patients with neurological signs of DH: age and sex not reported.
Study design	Historical cohort
Target condition and reference standard(s)	Surgical findings: Lumbar disc protrusion or sequestration exerting pressure on the nerve root
Index and comparator tests	Lasègue's sign = SLR
Follow-up	
Notes	Prevalence disc herniation: 75% (based on 1959 patients)

Table of Methodological Quality

Item	Authors' judgement	Description
Representative spectrum? All tests	Yes	consecutive series of patients
Acceptable reference standard? All tests	Yes	findings at surgery
Acceptable delay between tests? All tests	Unclear	unclear from text
Partial verification avoided? All tests	Yes	all patients received same reference standard
Differential verification avoided? All tests	Yes	all patients received same reference standard
Incorporation avoided? All tests	Unclear	unclear who carried out index tests or reference test
Reference standard results blinded? All tests	Unclear	unclear who carried out test
Index test results blinded? All tests	Yes	phys ex carried out before surgery
Relevant clinical information? All tests	Yes	as in usual care

Hakelius 1972 (*Continued*)

Uninterpretable results reported? All tests	Yes	all test results reported
Withdrawals explained? All tests	Unclear	unexplained number did not receive reference test
Clear definition of positive result? All tests	Unclear	unclear from text
Treatment withheld until both index test and reference standard were performed? All tests	Unclear	unclear from text
Data on observer variation reported and within acceptable range? All tests	Unclear	no data presented

Haldeman 1988

Clinical features and settings	Secondary care, USA - sampling method unclear - 84% first spinal surgery
Participants	100 workers > 6 months LBP; leg pain; received conservative treatment but unlikely to undergo surgery: 48% older than 40 years
Study design	Cohort: prospective?
Target condition and reference standard(s)	CT: spinal stenosis (> 50% occlusion); or large disc bulges (> 5 mm); or herniation with/without sequestration
Index and comparator tests	SLR / Lasègue's test, neural examination (radiculopathy)
Follow-up	
Notes	Prevalence disc herniation: 26%

Table of Methodological Quality

Item	Authors' judgement	Description
Representative spectrum? All tests	Unclear	sampling procedures unclear
Acceptable reference standard? All tests	Yes	findings at surgery

Haldeman 1988 (*Continued*)

Acceptable delay between tests? All tests	Unclear	unclear from text
Partial verification avoided? All tests	Yes	all patients received same reference standard
Differential verification avoided? All tests	Yes	all patients received same reference standard
Incorporation avoided? All tests	Unclear	unclear if diagnosis was based on CT only or also phys ex
Reference standard results blinded? All tests	Unclear	unclear if radiology reports were assessed by the same person as phys ex
Index test results blinded? All tests	Unclear	unclear from text
Relevant clinical information? All tests	Yes	as in usual care
Uninterpretable results reported? All tests	Unclear	unclear from text
Withdrawals explained? All tests	Yes	no withdrawals reported
Clear definition of positive result? All tests	Yes	criteria for positive test results described and execution described
Treatment withheld until both index test and reference standard were performed? All tests	Unclear	unclear from text
Data on observer variation reported and within acceptable range? All tests	Unclear	no information

Hudgins 1979

Clinical features and settings	Secondary care, USA - consecutive sampling - first surgery - no information on pre-surgery tests
Participants	351 patients with LBP and/or leg pain referred to neurosurgical service, positive SLR (work-up bias): age and sex unclear.

Hudgins 1979 (*Continued*)

Study design	Cohort: historical or prospective?
Target condition and reference standard(s)	Surgical findings: lumbar disc protrusion Clinical follow-up: patients not responding to conservative Rx, pain after 6 months, or surgery elsewhere.
Index and comparator tests	XSLR in combination with positive SLR
Follow-up	
Notes	Prevalence disc herniation: 70%

Table of Methodological Quality

Item	Authors' judgement	Description
Representative spectrum? All tests	No	Only patients with a positive straight leg raising test included
Acceptable reference standard? All tests	Yes	scored positively for 274 surgical patients, but unclear for all 351 herniated disc suspects (which included 77 patients for whom the reference standard consisted of clinical follow-up).
Acceptable delay between tests? All tests	Unclear	unclear from text
Partial verification avoided? All tests	No	only those with positive test results receive surgery
Differential verification avoided? All tests	Yes	scored positively for 274 surgical patients, but unclear for all 351 herniated disc suspects (which included 77 patients for whom the reference standard consisted of clinical follow-up).
Incorporation avoided? All tests	Unclear	unclear from text: surgeon interpreted findings, but did he/she use results of phys ex?
Reference standard results blinded? All tests	No	Index and reference test by surgeon
Index test results blinded? All tests	Yes	physical examination before surgery
Relevant clinical information? All tests	Yes	as in usual care

Hudgins 1979 (*Continued*)

Uninterpretable results reported? All tests	Yes	all test results reported
Withdrawals explained? All tests	Unclear	About 20% withdrawals, impact on performance unclear
Clear definition of positive result? All tests	Unclear	unclear from text
Treatment withheld until both index test and reference standard were performed? All tests	No	patients received non-operative management
Data on observer variation reported and within acceptable range? All tests	Unclear	no data

Kerr 1988

Clinical features and settings	Secondary care, UK - consecutive sampling for cases - sampling unclear for controls - no information on previous surgery - myelography in all before surgery: positive
Participants	100 patients with protruded lumbar disc, back pain & sciatica: 55% male, mean age 40 years 36 controls with back pain and sciatica but normal myelogram (work-up bias): mean age 41 years.
Study design	Case control
Target condition and reference standard(s)	Cases: myelography & surgical findings: lumbar disc protrusion or sequestration with distortion of nerve root Controls: normal myelogram
Index and comparator tests	SLR, XSLR, scoliosis, calf muscle wasting, motor weakness, Achilles tendon reflex abnormality, sensory deficits
Follow-up	
Notes	

Table of Methodological Quality

Item	Authors' judgement	Description

Kerr 1988 *(Continued)*

Representative spectrum? All tests	No	only patients with positive myelograms receive surgery and are selected
Acceptable reference standard? All tests	Yes	findings at surgery
Acceptable delay between tests? All tests	Unclear	unclear from text
Partial verification avoided? All tests	Yes	all patients received same reference standard
Differential verification avoided? All tests	No	all patients received same reference standard
Incorporation avoided? All tests	Yes	phys ex not specifically included in diagnosis of HD
Reference standard results blinded? All tests	Unclear	Data collected from medical records, and unclear if surgeon was blinded during original data collection
Index test results blinded? All tests	Yes	phys ex before surgery
Relevant clinical information? All tests	Yes	as in usual care
Uninterpretable results reported? All tests	Unclear	unclear from text
Withdrawals explained? All tests	Yes	no withdrawals reported
Clear definition of positive result? All tests	Unclear	Some definitions given, but little information on execution of tests
Treatment withheld until both index test and reference standard were performed? All tests	Unclear	unclear from text
Data on observer variation reported and within acceptable range? All tests	Unclear	No data presented

Knutsson 1961

Clinical features and settings	Secondary care, Sweden, 1958-59 - sampling method unclear - 89% first spinal surgery - myelography in 205 patients, positive in approximately 80%
Participants	205 patients operated upon for DH: 61% male, 58% 40 years or older.
Study design	Historical cohort
Target condition and reference standard(s)	Surgical findings: lumbar disc herniation or protrusion
Index and comparator tests	SLR / Lasègue's test?, Achilles or patellar tendon reflex abnormality, motor weakness, impaired sensibility, muscular atrophy
Follow-up	
Notes	Prevalence disc herniation: 89% (first time surgery) 67% (previous surgery)

Table of Methodological Quality

Item	Authors' judgement	Description
Representative spectrum? All tests	Yes	consecutive series of patients
Acceptable reference standard? All tests	Yes	findings at surgery
Acceptable delay between tests? All tests	Unclear	unclear from text
Partial verification avoided? All tests	Yes	all patients received same reference standard
Differential verification avoided? All tests	Yes	all patients received same reference standard
Incorporation avoided? All tests	Unclear	unclear who performed index tests
Reference standard results blinded? All tests	Unclear	unclear who performed index tests
Index test results blinded? All tests	Unclear	unclear who performed index tests
Relevant clinical information? All tests	Yes	as in usual care

Knutsson 1961 (*Continued*)

Uninterpretable results reported? All tests	No	all test results reported
Withdrawals explained? All tests	Yes	no withdrawals reported
Clear definition of positive result? All tests	No	No information on either cut points or ways tests were performed
Treatment withheld until both index test and reference standard were performed? All tests	Unclear	unclear from text
Data on observer variation reported and within acceptable range? All tests	Unclear	no data

Kosteljanetz 1984

Clinical features and settings	Secondary care, Denmark, 1978-80 - consecutive sampling - first surgery - X-ray in all before surgery to exclude malignancy
Participants	107 patients with LBP and symptoms & signs suggesting root compression, >3 weeks unsuccessful conservative treatment: 51% male.
Study design	Prospective cohort
Target condition and reference standard(s)	Surgical findings: complete or incomplete lumbar disc herniation and evidence of nerve root compression
Index and comparator tests	Lasègue's sign = SLR, paraesthesia, scoliosis, sensory loss, paresis
Follow-up	
Notes	Prevalence disc herniation: 58% (based on 100 patients)

Table of Methodological Quality

Item	**Authors' judgement**	**Description**
Representative spectrum? All tests	Yes	consecutive series of patients
Acceptable reference standard? All tests	Yes	findings at surgery

Kosteljanetz 1984 (*Continued*)

Acceptable delay between tests? All tests	Unclear	unclear from text
Partial verification avoided? All tests	Yes	all patients received same reference standard
Differential verification avoided? All tests	Yes	all patients received same reference standard
Incorporation avoided? All tests	Yes	phys ex not clearly included in diagnosis
Reference standard results blinded? All tests	Unclear	surgeon also interpreted results of phys ex?
Index test results blinded? All tests	Yes	phys ex before surgery
Relevant clinical information? All tests	Unclear	unclear from text
Uninterpretable results reported? All tests	Yes	all test results reported
Withdrawals explained? All tests	Yes	no withdrawals reported
Clear definition of positive result? All tests	Unclear	Cut points described but no information on execution
Treatment withheld until both index test and reference standard were performed? All tests	Unclear	unclear from text
Data on observer variation reported and within acceptable range? All tests	Unclear	no data presented

Kosteljanetz 1988

Clinical features and settings	Secondary care, Denmark, 1986-87 - Sampling based on positive myelography - first surgery
Participants	55 patients with suspected DH, unilateral sciatica, positive myelogram: 60% male, median age 45 (18-73) years.
Study design	Prospective cohort

Kosteljanetz 1988 (*Continued*)

Target condition and reference standard(s)	Surgical findings: prolapsed lumbar disc
Index and comparator tests	Lasègue's sign = SLR, crossed Lasègue's sign = XSLR
Follow-up	
Notes	Prevalence disc herniation: 87% (based on 52 patients)

Table of Methodological Quality

Item	Authors' judgement	Description
Representative spectrum? All tests	No	selection for surgery and inclusion in study based on positive myelography
Acceptable reference standard? All tests	Yes	findings at surgery
Acceptable delay between tests? All tests	Unclear	unclear from text
Partial verification avoided? All tests	No	only those with positive myelography received surgery
Differential verification avoided? All tests	No	only those with positive myelography received surgery
Incorporation avoided? All tests	Yes	phys ex not clearly included in diagnosis
Reference standard results blinded? All tests	Unclear	Surgeon aware of results of phys ex?
Index test results blinded? All tests	Yes	phys ex before surgery
Relevant clinical information? All tests	No	surgeon specifically blinded to phys ex => not as in usual care
Uninterpretable results reported? All tests	Yes	all test results reported
Withdrawals explained? All tests	Yes	no withdrawals reported
Clear definition of positive result? All tests	Yes	clear description of definition and execution

Kosteljanetz 1988 (*Continued*)

Treatment withheld until both index test and reference standard were performed? All tests	Unclear	unclear from text
Data on observer variation reported and within acceptable range? All tests	Yes	reliability investigated and acceptable

Majlesi 2008

Clinical features and settings	Secondary care, Turkey, 2005 - sampling method unclear - no information on previous surgery
Participants	38 patients with bulging, protruded, or extruded disc, back and/or leg pain: 79% male, mean age 38 years. 37 patients, no abnormalities on MRI, back and/or leg pain: 68% male, 40 years.
Study design	Case control (nested cohort design?)
Target condition and reference standard(s)	MRI: with bulging, protruded, or extruded disc
Index and comparator tests	SLR, slump test
Follow-up	
Notes	

Table of Methodological Quality

Item	Authors' judgement	Description
Representative spectrum? All tests	No	case control, sampling unclear, different exclusion criteria cases / controls (e.g. comorbidity)
Acceptable reference standard? All tests	Yes	MRI
Acceptable delay between tests? All tests	Unclear	unclear from text
Partial verification avoided? All tests	Yes	all patients received same reference standard
Differential verification avoided? All tests	Yes	all patients received same reference standard

Majlesi 2008 (*Continued*)

Incorporation avoided? All tests	Unclear	unclear if diagnosis was only based on imaging or also on phys ex
Reference standard results blinded? All tests	Yes	radiologist blinded to results phys ex
Index test results blinded? All tests	Yes	phys ex before surgery
Relevant clinical information? All tests	Unclear	unclear from text
Uninterpretable results reported? All tests	Unclear	unclear from text
Withdrawals explained? All tests	Yes	no withdrawals reported
Clear definition of positive result? All tests	Unclear	No information on both execution and cut-offs
Treatment withheld until both index test and reference standard were performed? All tests	Unclear	unclear from text
Data on observer variation reported and within acceptable range? All tests	Unclear	no data presented

Meylemans 1988

Clinical features and settings	Setting unclear, Belgium, 1985-86 - consecutive sampling - no information on previous surgery
Participants	146 patients with LBP and leg pain, duration < two months, no previous radicular symptoms: 58% men, majority 30-50 years.
Study design	Prospective (?) cohort
Target condition and reference standard(s)	CT: "radiculopathy", unclear if disc protrusion was found in all cases
Index and comparator tests	SLR, neurological examination, combination of both
Follow-up	
Notes	Prevalence disc herniation: 40%

Meylemans 1988 (*Continued*)

Table of Methodological Quality

Item	Authors' judgement	Description
Representative spectrum? All tests	Yes	consecutive series of patients
Acceptable reference standard? All tests	Yes	CT
Acceptable delay between tests? All tests	Unclear	unclear from text
Partial verification avoided? All tests	Yes	all patients received same reference standard
Differential verification avoided? All tests	Yes	all patients received same reference standard
Incorporation avoided? All tests	Yes	phys ex not explicitly part of diagnosis
Reference standard results blinded? All tests	Unclear	unclear from text
Index test results blinded? All tests	Unclear	unclear who carried out index tests
Relevant clinical information? All tests	Yes	as in usual care
Uninterpretable results reported? All tests	Yes	all results reported
Withdrawals explained? All tests	Yes	no withdrawals reported
Clear definition of positive result? All tests	Unclear	information on cut-off, not on execution
Treatment withheld until both index test and reference standard were performed? All tests	Unclear	unclear from text
Data on observer variation reported and within acceptable range? All tests	Unclear	no data presented

Poiraudeau 2001

Clinical features and settings	Secondary care, France - consecutive sampling - first surgery
Participants	78 patients with lumbosacral pain + pain below the knee (L5/S1) or neurological impairment: 42% male, mean age 50 (SD 16) years.
Study design	Prospective cohort
Target condition and reference standard(s)	MRI, CT or myelography: compression of lumbar nerve root by disc herniation
Index and comparator tests	Bell test, hyper extension test, Lasègue's sign = SLR crossed Lasègue's sign = XSLR, Combinations of tests
Follow-up	
Notes	Prevalence disc herniation: 55%

Table of Methodological Quality

Item	Authors' judgement	Description
Representative spectrum? All tests	Yes	consecutive series of patients
Acceptable reference standard? All tests	Yes	diagnostic imaging
Acceptable delay between tests? All tests	Unclear	unclear from text
Partial verification avoided? All tests	Yes	all patients received one of the reference standards
Differential verification avoided? All tests	No	patients received different types of imaging as reference standard
Incorporation avoided? All tests	Yes	phys ex not clearly included in diagnosis
Reference standard results blinded? All tests	Yes	radiologists blinded
Index test results blinded? All tests	Yes	phys ex before reference standard
Relevant clinical information? All tests	Yes	as in usual care

Poiraudeau 2001 (*Continued*)

Uninterpretable results reported? All tests	Yes	all test results reported
Withdrawals explained? All tests	Yes	no withdrawals
Clear definition of positive result? All tests	Yes	criteria and execution described
Treatment withheld until both index test and reference standard were performed? All tests	Unclear	unclear from text
Data on observer variation reported and within acceptable range? All tests	No	Data reported, but observer variation is quite high

Spangfort 1972

Clinical features and settings	Secondary care, Sweden, 1951-66 - all surgical cases - no information on previous surgery - myelography in 80%, positive in approximately 80%
Participants	2504 Patients with suspected lumbar disc herniation: 70% male, 40.8 (15-74) years.
Study design	Historical cohort
Target condition and reference standard(s)	Surgical findings: complete or incomplete disc herniation or bulging disc, assumed to cause pressure on nerve root
Index and comparator tests	Lasègue's sign = SLR, crossed Lasègue's sign = XSLR, impaired ankle reflex, paresis
Follow-up	
Notes	Prevalence disc herniation: 86% (78% when excluding bulging disc)

Table of Methodological Quality

Item	**Authors' judgement**	**Description**
Representative spectrum? All tests	Yes	consecutive series of patients
Acceptable reference standard? All tests	Yes	findings at surgery

Spangfort 1972 (*Continued*)

Acceptable delay between tests? All tests	Unclear	unclear from text
Partial verification avoided? All tests	Yes	all patients received same reference standard
Differential verification avoided? All tests	Yes	all patients received same reference standard
Incorporation avoided? All tests	Unclear	unclear from text
Reference standard results blinded? All tests	Unclear	author performed record review, but also interpreted surgical findings?
Index test results blinded? All tests	Unclear	author performed record review, although index test was carried out before reference test
Relevant clinical information? All tests	Yes	as in usual care
Uninterpretable results reported? All tests	Yes	all test results reported
Withdrawals explained? All tests	Yes	no withdrawals
Clear definition of positive result? All tests	No	no description of positivity criteria or of execution of test
Treatment withheld until both index test and reference standard were performed? All tests	Unclear	unclear from text
Data on observer variation reported and within acceptable range? All tests	Unclear	no data presented

Stankovic 1999

Clinical features and settings	Secondary care, Sweden - consecutive sampling - first surgery
Participants	105 patients with LBP and/or radiating pain in the leg, eligible for surgery: 66% male, 42.7 (SD 9.8) years.
Study design	Prospective cohort

Stankovic 1999 (*Continued*)

Target condition and reference standard(s)	CT and/or MRI: bulging disc, or herniated disc assumed to compromise nerve root
Index and comparator tests	Slump test, lumbar extension test
Follow-up	
Notes	Prevalence disc herniation: 88% (50% when excluding bulging disc)

Table of Methodological Quality

Item	Authors' judgement	Description
Representative spectrum? All tests	Yes	consecutive series of patients
Acceptable reference standard? All tests	Yes	diagnostic imaging
Acceptable delay between tests? All tests	Unclear	unclear from text
Partial verification avoided? All tests	Yes	all patients received reference standard
Differential verification avoided? All tests	No	patients received CT and/or MRI
Incorporation avoided? All tests	Yes	phys ex not explicitly included in diagnosis
Reference standard results blinded? All tests	Yes	radiologist blinded
Index test results blinded? All tests	Yes	phys ex before imaging
Relevant clinical information? All tests	Yes	as in usual care
Uninterpretable results reported? All tests	Yes	all test results reported
Withdrawals explained? All tests	Yes	no withdrawals reported
Clear definition of positive result? All tests	Yes	description given of execution of tests and of positivity criteria

Stankovic 1999 (*Continued*)

Treatment withheld until both index test and reference standard were performed? All tests	Unclear	unclear from text
Data on observer variation reported and within acceptable range? All tests	Unclear	no data presented

Vroomen 1998

Clinical features and settings	Setting unclear, Netherlands - consecutive sampling - first surgery
Participants	71 patients (18-60 years), symptoms & signs & radiological findings consistent with single-level unilateral DH, indication surgery: 65% male, age 39 (SD 9.1) years.
Study design	Prospective cohort
Target condition and reference standard(s)	MRI: annular rupture, migration, nerve root compression by disc material
Index and comparator tests	Paresis, finger-floor distance, ankle/knee tendon reflex, SLR, sensory loss (ORs, insufficient data for 2x2 table)
Follow-up	
Notes	Prevalence disc herniation: 83% (annular rupture)

Table of Methodological Quality

Item	**Authors' judgement**	**Description**
Representative spectrum? All tests	Yes	consecutive series of patients
Acceptable reference standard? All tests	Yes	diagnostic imaging
Acceptable delay between tests? All tests	Unclear	unclear from text
Partial verification avoided? All tests	Yes	all patients received same reference standard
Differential verification avoided? All tests	Yes	all patients received same reference standard

Vroomen 1998 (*Continued*)

Incorporation avoided? All tests	Yes	phys ex not included in diagnosis
Reference standard results blinded? All tests	Yes	radiologist blinded
Index test results blinded? All tests	Yes	phys ex before imaging
Relevant clinical information? All tests	Yes	as in usual care
Uninterpretable results reported? All tests	Yes	all test results reported
Withdrawals explained? All tests	Unclear	Not entirely clear if all 71 participants received both index and reference tests
Clear definition of positive result? All tests	No	no description of execution of test or positivity criteria
Treatment withheld until both index test and reference standard were performed? All tests	Unclear	unclear from text
Data on observer variation reported and within acceptable range? All tests	Unclear	no data presented

Vroomen 2002 (prim care)

Clinical features and settings	Primary care, Netherlands - referred by GPs, sampling unclear - first surgery
Participants	338 patients with a new episode of LBP and leg pain sufficiently severe to warrant action: 51% male, age 46 (SD 12) years.
Study design	Prospective cohort
Target condition and reference standard(s)	MRI: lumbosacral nerve root compression
Index and comparator tests	SLR, paresis, finger-floor distance, loss of ankle/knee tendon reflex, sensory loss. Combinations of tests
Follow-up	
Notes	Prevalence disc herniation: 55%

Vroomen 2002 (prim care) (*Continued*)

Table of Methodological Quality		
Item	**Authors' judgement**	**Description**
Representative spectrum? All tests	Yes	seems to be a consecutive series of patients, clear criteria
Acceptable reference standard? All tests	Yes	diagnostic imaging
Acceptable delay between tests? All tests	Yes	less than 24 hours
Partial verification avoided? All tests	Yes	all patients received same reference standard
Differential verification avoided? All tests	Yes	all patients received same reference standard
Incorporation avoided? All tests	Yes	phys ex not included in diagnosis
Reference standard results blinded? All tests	Yes	radiologist blinded
Index test results blinded? All tests	Yes	phys ex before MRI
Relevant clinical information? All tests	Yes	as in usual care
Uninterpretable results reported? All tests	Yes	all test results reported
Withdrawals explained? All tests	Yes	no withdrawals
Clear definition of positive result? All tests	Yes	information on test execution and positivity criteria in separate paper
Treatment withheld until both index test and reference standard were performed? All tests	Unclear	unclear from text
Data on observer variation reported and within acceptable range? All tests	Yes	inter-observer variation reported and acceptable

Vucetic 1996

Clinical features and settings	Secondary care, Sweden - consecutive sampling - no information on previous surgery - myelography in all, results unclear
Participants	163 surgical patients with clinical and radiographic signs of single-level lumbar disc herniation: 53% male, age 43 (SD 10.2) years.
Study design	Prospective cohort
Target condition and reference standard(s)	Surgical findings: lumbar disc protrusion, extrusion, or sequestration
Index and comparator tests	Decreased sensibility, paresis, loss of reflexes, scoliosis. Combinations of tests
Follow-up	
Notes	Prevalence disc herniation: 92% (63% when excluding protruding disc)

Table of Methodological Quality

Item	Authors' judgement	Description
Representative spectrum? All tests	Yes	consecutive series of patients
Acceptable reference standard? All tests	Yes	findings at surgery
Acceptable delay between tests? All tests	Unclear	unclear from text
Partial verification avoided? All tests	Yes	all patients received same reference standard
Differential verification avoided? All tests	Yes	all patients received same reference standard
Incorporation avoided? All tests	Yes	phys ex not explicitly included in diagnosis
Reference standard results blinded? All tests	Unclear	unclear who carried out index and reference tests
Index test results blinded? All tests	Yes	index test performed before reference test
Relevant clinical information? All tests	Yes	as in usual care

Vucetic 1996 (*Continued*)

Uninterpretable results reported? All tests	Yes	all test results reported
Withdrawals explained? All tests	Yes	no withdrawals reported
Clear definition of positive result? All tests	Yes	clear description of test execution and definition of positivity criteria
Treatment withheld until both index test and reference standard were performed? All tests	Unclear	unclear from text
Data on observer variation reported and within acceptable range? All tests	Yes	inter-observer variation reported and acceptable

Vucetic 1999

Clinical features and settings	Secondary care, Sweden, 1981-84 - consecutive sampling - first surgery - myelography in all, results unclear (same cohort as Vucetic 1996)
Participants	160 patients with clinical and radiographic signs of lumbar disc herniation; 53% male, age 43 (SD 10).
Study design	Prospective (?) cohort
Target condition and reference standard(s)	Surgical findings: lumbar disc protrusion, ruptured annulus, or sequestration
Index and comparator tests	Crossed Lasègue's sign = XSLR Combinations of tests (ORs, insufficient data for 2x2 table).
Follow-up	
Notes	Prevalence disc herniation: 61%

Table of Methodological Quality

Item	Authors' judgement	Description
Representative spectrum? All tests	Yes	consecutive series of patients
Acceptable reference standard? All tests	Yes	findings at surgery

Vucetic 1999 (*Continued*)

Acceptable delay between tests? All tests	Unclear	unclear from text
Partial verification avoided? All tests	Yes	all patients received same reference standard
Differential verification avoided? All tests	Yes	all patients received same reference standard
Incorporation avoided? All tests	Yes	phys ex not explicitly included in diagnosis standard
Reference standard results blinded? All tests	Unclear	unclear who performed tests
Index test results blinded? All tests	Yes	index test carried out before reference test
Relevant clinical information? All tests	Yes	as in usual care
Uninterpretable results reported? All tests	Yes	all test results reported
Withdrawals explained? All tests	Yes	no withdrawals reported
Clear definition of positive result? All tests	Yes	clear description of test execution and of positivity criteria
Treatment withheld until both index test and reference standard were performed? All tests	Unclear	unclear from text
Data on observer variation reported and within acceptable range? All tests	Yes	inter-observer variation examined and acceptable

phys ex = physical examination; MRI = magnetic resonance imaging; CT = computed tomography; HD = herniated disc

Characteristics of excluded studies *[ordered by study ID]*

Albert 1993	Population: not low back pain
Brugnoni 1998	Index test: not physical examination
Edgar 1974	Target condition: level of disc herniation, no comparison with absence of DH
Jensen 1987	Target condition: level of disc herniation, no comparison with absence of DH
Kortelainen 1985	Target condition: level of disc herniation, no comparison with absence of DH
Lansche 1960	Target condition: level of disc herniation, no comparison with absence of DH
Portnoy 1972	Target condition: level of disc herniation, no comparison with absence of DH
Rainville 2003	Target condition: level of disc herniation, no comparison with absence of DH
Reihani-Kermani 2003	Target condition: level of disc herniation, no comparison with absence of DH
Reihani-Kermani 2004	Target condition: level of disc herniation, no comparison with absence of DH
Supik 1994	Target condition: level of disc herniation, no comparison with absence of DH
Thelander 1992	Target condition: level of disc herniation, no comparison with absence of DH
Xin 1987	Target condition: level of disc herniation, no comparison with absence of DH

DATA

Presented below are all the data for all of the tests entered into the review.

Tests. Data tables by test

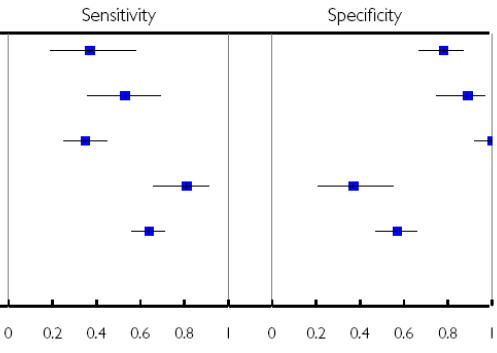
Test	No. of studies	No. of participants
1 SLR (leg pain at any angle) - reference test: imaging	5	672
2 SLR (leg pain at any angle) - reference test surgery	10	6561
3 X-SLR (leg pain at any angle) - reference test surgery or imaging	5	2950
4 Scoliosis (visual inspection) - reference test: surgical findings	4	479
5 Paresis (dorsiflexion toe/ankle) - reference test:surgical findings or imaging	7	3419
6 Muscle wasting - reference test: surgical findings	3	397
7 Impaired reflexes (Achilles tendon) - reference test: surgical findings or imaging	7	4515
8 Sensory deficits - reference test surgical findings or imaging	6	935
9 Forward flexion - reference test surgery or imaging	3	442

Test 1. SLR (leg pain at any angle) - reference test: imaging.

Review: Physical examination for lumbar radiculopathy due to disc herniation in patients with low-back pain

Test: 1 SLR (leg pain at any angle) - reference test: imaging

Study	TP	FP	FN	TN	Sensitivity	Specificity
Haldeman 1988	10	16	17	57	0.37 [0.19, 0.58]	0.78 [0.67, 0.87]
Majlesi 2008	20	4	18	33	0.53 [0.36, 0.69]	0.89 [0.75, 0.97]
Meylemans 1988	35	0	66	45	0.35 [0.25, 0.45]	1.00 [0.92, 1.00]
Poiraudeau 2001	34	22	8	13	0.81 [0.66, 0.91]	0.37 [0.21, 0.55]
Vroomen 2002 (prim care)	97	53	55	69	0.64 [0.56, 0.71]	0.57 [0.47, 0.66]

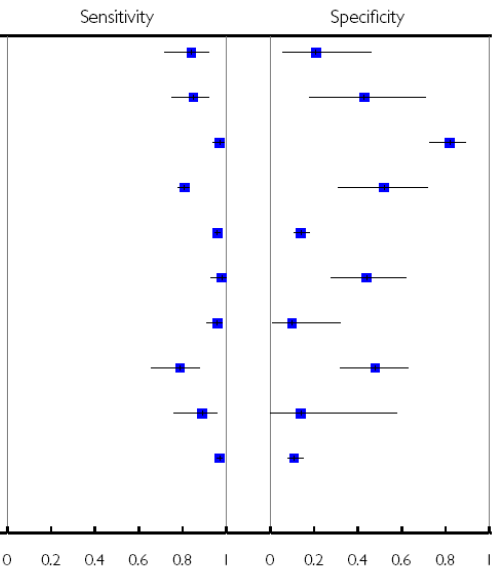


Test 2. SLR (leg pain at any angle) - reference test surgery.

Review: Physical examination for lumbar radiculopathy due to disc herniation in patients with low-back pain

Test: 2 SLR (leg pain at any angle) - reference test surgery

Study	TP	FP	FN	TN	Sensitivity	Specificity
Albeck 1996	51	15	10	4	0.84 [0.72, 0.92]	0.21 [0.06, 0.46]
Charnley 1951	63	8	11	6	0.85 [0.75, 0.92]	0.43 [0.18, 0.71]
Demircan 2002	179	18	5	82	0.97 [0.94, 0.99]	0.82 [0.73, 0.89]
Gurdjian 1961	929	12	222	13	0.81 [0.78, 0.83]	0.52 [0.31, 0.72]
Hakelius 1972	1411	422	56	70	0.96 [0.95, 0.97]	0.14 [0.11, 0.18]
Kerr 1988	98	20	2	16	0.98 [0.93, 1.00]	0.44 [0.28, 0.62]
Knutsson 1961	155	18	7	2	0.96 [0.91, 0.98]	0.10 [0.01, 0.32]
Kosteljanetz 1984	44	23	12	21	0.79 [0.66, 0.88]	0.48 [0.32, 0.63]
Kosteljanetz 1988	40	6	5	1	0.89 [0.76, 0.96]	0.14 [0.00, 0.58]
Spangfort 1972	2088	308	69	39	0.97 [0.96, 0.98]	0.11 [0.08, 0.15]

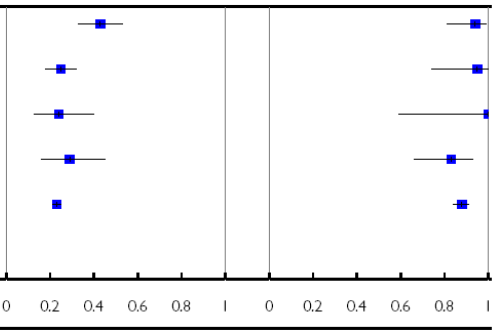


Test 3. X-SLR (leg pain at any angle) - reference test surgery or imaging.

Review: Physical examination for lumbar radiculopathy due to disc herniation in patients with low-back pain

Test: 3 X-SLR (leg pain at any angle) - reference test surgery or imaging

Study	TP	FP	FN	TN	Sensitivity	Specificity
Kerr 1988	43	2	57	34	0.43 [0.33, 0.53]	0.94 [0.81, 0.99]
Knutsson 1961	40	1	122	18	0.25 [0.18, 0.32]	0.95 [0.74, 1.00]
Kosteljanetz 1988	11	0	34	7	0.24 [0.13, 0.40]	1.00 [0.59, 1.00]
Poiraudeau 2001	12	6	30	29	0.29 [0.16, 0.45]	0.83 [0.66, 0.93]
Spangfort 1972	500	41	1657	306	0.23 [0.21, 0.25]	0.88 [0.84, 0.91]

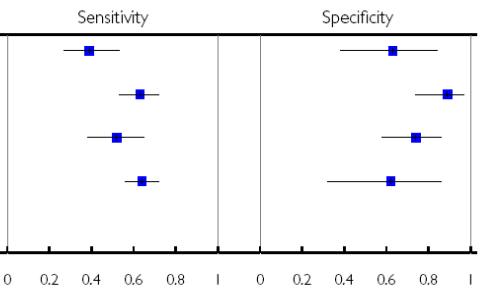


Test 4. Scoliosis (visual inspection) - reference test: surgical findings.

Review: Physical examination for lumbar radiculopathy due to disc herniation in patients with low-back pain

Test: 4 Scoliosis (visual inspection) - reference test: surgical findings

Study	TP	FP	FN	TN	Sensitivity	Specificity
Albeck 1996	24	7	37	12	0.39 [0.27, 0.53]	0.63 [0.38, 0.84]
Kerr 1988	63	4	37	32	0.63 [0.53, 0.72]	0.89 [0.74, 0.97]
Kosteljanetz 1984	30	11	28	31	0.52 [0.38, 0.65]	0.74 [0.58, 0.86]
Vucetic 1996	96	5	54	8	0.64 [0.56, 0.72]	0.62 [0.32, 0.86]

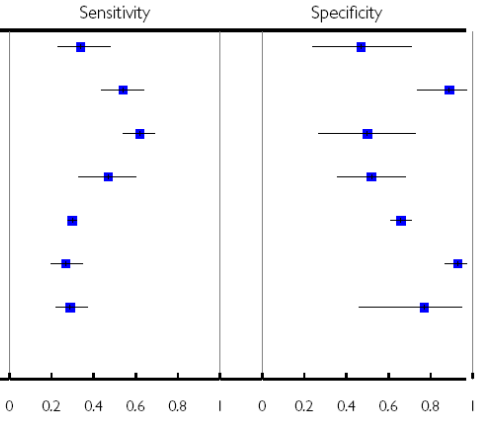


Test 5. Paresis (dorsiflexion toe/ankle) - reference test:surgical findings or imaging.

Review: Physical examination for lumbar radiculopathy due to disc herniation in patients with low-back pain

Test: 5 Paresis (dorsiflexion toe/ankle) - reference test:surgical findings or imaging

Study	TP	FP	FN	TN	Sensitivity	Specificity
Albeck 1996	21	10	40	9	0.34 [0.23, 0.48]	0.47 [0.24, 0.71]
Kerr 1988	54	4	46	32	0.54 [0.44, 0.64]	0.89 [0.74, 0.97]
Knutsson 1961	100	10	62	10	0.62 [0.54, 0.69]	0.50 [0.27, 0.73]
Kosteljanetz 1984	27	20	31	22	0.47 [0.33, 0.60]	0.52 [0.36, 0.68]
Spangfort 1972	645	110	1512	217	0.30 [0.28, 0.32]	0.66 [0.61, 0.71]
Vroomen 2002 (prim care)	41	8	111	114	0.27 [0.20, 0.35]	0.93 [0.87, 0.97]
Vucetic 1996	43	3	107	10	0.29 [0.22, 0.37]	0.77 [0.46, 0.95]

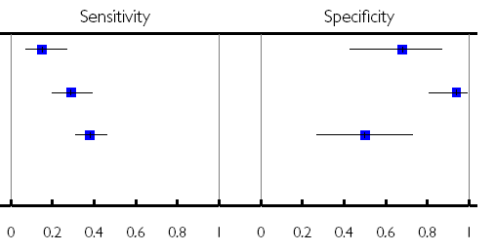


Test 6. Muscle wasting - reference test: surgical findings.

Review: Physical examination for lumbar radiculopathy due to disc herniation in patients with low-back pain

Test: 6 Muscle wasting - reference test: surgical findings

Study	TP	FP	FN	TN	Sensitivity	Specificity
Albeck 1996	9	6	51	13	0.15 [0.07, 0.27]	0.68 [0.43, 0.87]
Kerr 1988	29	2	71	34	0.29 [0.20, 0.39]	0.94 [0.81, 0.99]
Knutsson 1961	62	10	100	10	0.38 [0.31, 0.46]	0.50 [0.27, 0.73]

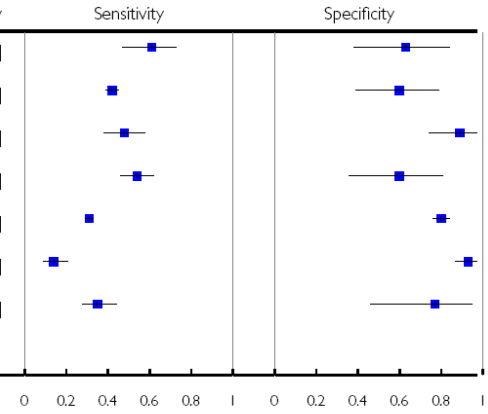


Test 7. Impaired reflexes (Achilles tendon) - reference test: surgical findings or imaging.

Review: Physical examination for lumbar radiculopathy due to disc herniation in patients with low-back pain

Test: 7 Impaired reflexes (Achilles tendon) - reference test: surgical findings or imaging

Study	TP	FP	FN	TN	Sensitivity	Specificity
Albeck 1996	37	7	24	12	0.61 [0.47, 0.73]	0.63 [0.38, 0.84]
Gurdjian 1961	486	10	665	15	0.42 [0.39, 0.45]	0.60 [0.39, 0.79]
Kerr 1988	48	4	52	32	0.48 [0.38, 0.58]	0.89 [0.74, 0.97]
Knutsson 1961	87	8	75	12	0.54 [0.46, 0.62]	0.60 [0.36, 0.81]
Spangfort 1972	675	69	1482	278	0.31 [0.29, 0.33]	0.80 [0.76, 0.84]
Vroomen 2002 (prim care)	22	8	130	114	0.14 [0.09, 0.21]	0.93 [0.87, 0.97]
Vucetic 1996	53	3	97	10	0.35 [0.28, 0.44]	0.77 [0.46, 0.95]

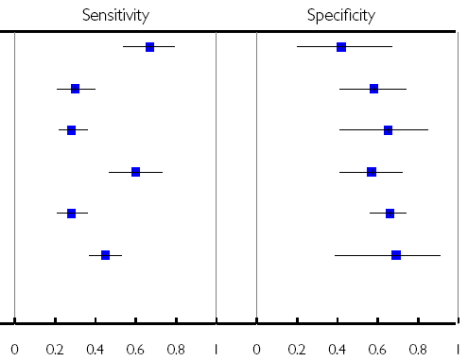


Test 8. Sensory deficits - reference test surgical findings or imaging.

Review: Physical examination for lumbar radiculopathy due to disc herniation in patients with low-back pain

Test: 8 Sensory deficits - reference test surgical findings or imaging

Study	TP	FP	FN	TN	Sensitivity	Specificity
Albeck 1996	41	11	20	8	0.67 [0.54, 0.79]	0.42 [0.20, 0.67]
Kerr 1988	30	15	70	21	0.30 [0.21, 0.40]	0.58 [0.41, 0.74]
Knutsson 1961	46	7	116	13	0.28 [0.22, 0.36]	0.65 [0.41, 0.85]
Kosteljanetz 1984	35	18	23	24	0.60 [0.47, 0.73]	0.57 [0.41, 0.72]
Vroomen 2002 (prim care)	43	42	109	80	0.28 [0.21, 0.36]	0.66 [0.56, 0.74]
Vucetic 1996	67	4	83	9	0.45 [0.37, 0.53]	0.69 [0.39, 0.91]

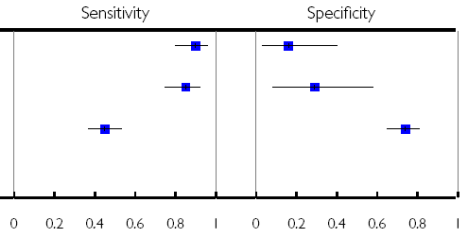


Test 9. Forward flexion - reference test surgery or imaging.

Review: Physical examination for lumbar radiculopathy due to disc herniation in patients with low-back pain

Test: 9 Forward flexion - reference test surgery or imaging

Study	TP	FP	FN	TN	Sensitivity	Specificity
Albeck 1996	55	16	6	3	0.90 [0.80, 0.96]	0.16 [0.03, 0.40]
Charnley 1951	63	10	11	4	0.85 [0.75, 0.92]	0.29 [0.08, 0.58]
Vroomen 2002 (prim care)	68	32	84	90	0.45 [0.37, 0.53]	0.74 [0.65, 0.81]



APPENDICES

Appendix 1. MEDLINE search strategy

1 Index test: tests performed during physical examination

1a

"straight leg raising"[tw] OR lasegue[tw] OR (provocation[tw] AND "intra abdominal pressure"[tw]) OR "bell test"[tw] OR "hyper-extension test"[tw] OR "femoral nerve stretch test"[tw] OR (achilles[tw] AND (areflexia[tw] OR reflex*[tw])) OR (knee[tw] AND (extens*[tw] OR reflex[tw])) OR "Reflex, stretch"[mesh] OR (dermatom*[tw] AND (somatosensory[tw] OR sensibility[tw])) OR slump[tw] OR ("muscle strength"[tw] AND leg[tw] AND (test[tw] OR tests[tw] OR testing[tw] OR sign[tw])) OR ((Bragard*[tw] OR Naffziger*[tw]) AND (test[tw] OR tests[tw] OR testing[tw] OR sign[tw])) OR (measur*[tw] AND "calf wasting"[tw]) OR (im-pair*[tw] AND "ankle reflex"[tw]) OR (weakness[tw] AND dorsiflexion[tw] AND foot[tw])

1b

Physical examination[mesh] OR "physical examination" OR "function test" OR "physical test" OR (clinical[tw] AND (diagnosis[tw] OR sign[tw] OR signs[tw] OR significance[tw] OR symptom*[tw] OR parameter*[tw] OR assessment[tw] OR finding*))

2 Population: low-back pain and anatomical location

2a

back pain[mesh] OR sciatica[mesh] OR "back ache"[tw] OR backache[tw] OR "back pain"[tw] OR dorsalgia[tw] OR lumbago[tw] OR sciatica[tw] OR ischias[tw] OR ischialgia[tw] OR lumboischialgia[tw] OR radiculalgia[tw] OR ((Pain[mesh] OR pain[tw] OR ache*[tw] OR aching[tw] OR complaint*[tw] OR dysfunction*[tw] OR disabil*[tw] OR neuralgia[tw]) AND (Back[mesh] OR spine[mesh] OR back[ti] OR lowback[tw] OR lumbar[tw] OR lumbal[tw] OR lumbo*[tw] OR sciatic[tw] OR spine[tw] OR spinal[tw] OR radicular[tw] OR "nerve root"[tw] OR "nerve roots"[tw] OR disk[tw] OR disc[tw] OR disks[tw] OR discs[tw] OR vertebra*[tw] OR intervertebra*[tw] OR sacroilia*[tw] OR Sacroiliac-joint[mesh]))

2b

low[tw] OR lower[tw] OR lowback[tw] OR sciatic*[tw] OR ischia*[tw] OR lumbo*[tw] OR lumba*[tw] OR sacroilia*[tw]

3 Target condition: lumbar radiculopathy

Intervertebral disk displacement[mesh] OR polyradiculopathy[mesh] OR radiculopath* OR radiculiti* OR ((disc OR discs OR disk OR disks) AND (displacement OR hernia* OR protru* OR avulsion*)) OR (("nerve root" OR "nerve roots") AND (compress* OR entrap* OR inflammat* OR disorder*)) OR (nerve compression syndromes[mesh] AND (Back[mesh] OR spine[mesh] OR back[ti] OR lowback[tw] OR lumbar[tw] OR lumbal[tw] OR lumbo*[tw] OR sciatic[tw] OR spine[tw] OR spinal[tw] OR radicular[tw] OR "nerve root"[tw] OR "nerve roots"[tw] OR disk[tw] OR disc[tw] OR disks[tw] OR discs[tw] OR vertebra*[tw] OR intervertebra*[tw] OR sacroilia*[tw] OR Sacroiliac-joint[mesh]))

4 Methodological filter (primary diagnostic studies)

4a

diagnosis[sh] OR pathophysiology[sh] OR etiology[sh]

4b

diagnosis[sh] OR diagnosis[mesh:noexp]

4c

Diagnostic errors[mesh] OR "Diagnosis, differential"[mesh] OR "Reproducibility of results"[mesh] OR Reference standards[mesh] OR "Sensitivity and specificity"[mesh] OR Comparative study[pt] OR "Evaluation Studies as Topic"[Mesh] OR Evaluation studies[pt] OR Longitudinal studies[mesh] OR sensitivit* OR specificit* OR accura* OR likelihood ratio* OR predict* OR index test OR reference test OR (false[tw] AND (positive[tw] OR negative[tw])) OR pretest[tw] OR pre-test[tw] OR posttest[tw] OR post-test[tw] OR "gold standard" OR roc[tw] OR odds[tw] OR validity OR validation OR validate* OR validation studies[pt] OR verif*[ti] OR evaluat*[ti] OR value*[ti] OR reference values[mesh] OR cutoff OR cut-off OR repeatability OR reproducibility OR efficacy OR reliability OR error*[tw] OR suitability[tw] OR utility[tw]

5. Exclusion criteria: children, reviews, case reports, animal studies

((child[mesh] OR infant[mesh]) NOT (adult[mesh] OR adolescent[mesh])) OR Review[pt] OR case reports[pt] OR (animals[mesh] NOT humans[mesh])

Searches (combinations)

A. 1a and (2a or 3) and 2b not 5

B. 1a and ((2a and 4a) or (3 and 4b)) not 5

C. 1b and 2a and 2b and 3 and (4a or 4b) not 5

D. 1b and 2b and 3 and 4b and 4c not 5

Final search: A or B or C or D

Appendix 2. EMBASE search strategy

1 Index test: tests performed during physical examination

1a

straight-leg-raising OR lasegue OR (provocation AND intra-abdominal-pressure) OR bell-test OR hyperextension-test OR femoral-nerve-stretch-test OR (achilles AND (areflexia OR reflex*)) OR (knee AND (extens* OR reflex)) OR Tendon-reflex/exp OR (dermatom* AND (somatosensory OR sensibility)) OR slump OR (Muscle-strength/exp AND leg/exp AND (test OR tests OR testing OR sign OR signs)) OR ((Bragard* OR Naffziger*) AND (test OR tests OR testing OR sign OR signs)) OR (measur* AND calf-wasting) OR (impair* AND ankle-reflex) OR (weakness AND dorsiflexion AND foot)

1b

clinical-examination/de OR clinical-feature/de OR clinical-observation/de OR physical-examination/exp OR functional-assessment/de OR function-test/exp OR provocation-test/de OR physical-examination OR function-test OR physical-test OR 'clinical *2 diagnosis' OR 'clinical *2 sign' OR 'clinical 2 signs' OR 'clinical 2 significance' OR 'clinical 2 symptom' OR 'clinical 2 symptoms' OR 'clinical 2 parameter' OR 'clinical 2 parameters' OR clinical-assessment OR ((symptom OR symptoms) AND (sign OR signs)) OR 'clinical *2 finding' OR 'clinical *2 findings'

1c

diagnosis OR etiology OR pathophysiology

1d

diagnosis

2 Population: low-back pain and anatomical location

2a

backache/exp OR backache OR back-ache OR back-pain OR dorsalgia OR lumbago OR sciatica OR ischias OR ischialgia OR lumboischialgia OR radiculalgia OR ((Pain/exp OR pain OR ache* OR aching OR complaint* OR dysfunction* OR disabil* OR neuralgia) AND (Back/exp OR back:ti OR lowback OR lumbar OR lumbal OR lumbo* OR sciatic OR spine OR spinal OR radicular OR nerve-root OR nerve-roots OR disk* OR disc* OR vertebra* OR intervertebra* OR sacroilia* OR Intervertebral-articulation/de OR Sacroiliac-joint/de))

2b

low OR lower OR lowback OR sciatic* OR ischia* OR lumbo* OR lumba* OR sacroilia*

3 Target condition: lumbar radiculopathy

Intervertebral-disk-disease/exp OR radiculopathy/exp OR radiculopath* OR radiculiti* OR ((disc OR discs OR disk OR disks) AND (displacement OR hernia* OR protru* OR avulsion*)) OR ((nerve-root OR nerve-roots) AND (compress* OR entrap* OR inflammat* OR disorder*))

4 Methodological 'filter' (primary diagnostic studies)

accuracy/de OR clinical-study/de OR comparative-study/de OR correlation-analysis/de OR correlation-coefficient/de OR diagnostic-accuracy/de OR diagnostic-error/de OR diagnostic-value/de OR differential-diagnosis/de OR evaluation/de OR factorial-analysis/de OR follow-up/de OR (major-clinical-study/de AND methodology/de) OR Medical-decision-making/de OR observer-variation/de OR prediction-and-forecasting/exp OR preoperative-evaluation/de OR prospective-study/de OR reliability/de OR receiver-operating-characteristic/de OR reproducibility/de OR sensitivity-analysis/de OR sensitivity-and-specificity/exp OR standard/de OR utilization-review/de OR validation-process/de OR sensitivit* OR specificit* OR accura* OR likelihood-ratio OR predict*

5 Exclusion criteria: children, reviews, case reports, animal studies

(child/exp NOT adult/exp) OR Review:it OR review/de OR case-report/de OR ((animal/exp OR animal-experiment/exp OR animal-model/exp OR animal-disease/exp) NOT human/exp)

Term combinations applied:

A 1a and (2a or 3) and 2b not 5

B 1a and ((2a and 1c) or (3 and 1d)) not 5

C 1b and 2a and 2b and 3 and 1c not 5

D 1b and 1d and 2b and 3 and 4 not 5

Total = A or B or C or D

Appendix 3. CINAHL search strategy

1 Index test: tests performed during physical examination

1a

straight leg raising OR lasegue OR (provocation AND intra abdominal pressure) OR bell test OR hyperextension test OR femoral nerve stretch test OR (achilles AND (areflexia OR reflex*)) OR (knee AND (extens* OR reflex)) OR MH "Reflex, stretch" OR (dermatom* AND (somatosensory OR sensibility)) OR slump OR (muscle strength AND leg AND (test OR tests OR testing OR sign)) OR ((TX Bragard* OR TX Naffziger*) AND (test OR tests OR testing OR sign)) OR (measur* AND calf wasting) OR (impair* AND ankle reflex) OR (weakness AND dorsiflexion AND foot)

1b

MH "Physical examination+" OR physical examination OR function test OR physical test OR (clinical AND (diagnosis OR sign OR signs OR significance OR symptom* OR parameter* OR assessment OR finding*))

2 Population: low-back pain and anatomical location

MH "Back pain+" OR MH sciatica OR back ache OR backache OR back pain OR dorsalgia OR lumbago OR sciatica OR ischias OR ischialgia OR lumboischialgia OR radiculalgia OR ((MH Pain OR pain OR ache* OR aching OR complaint* OR dysfunction* OR disabil* OR neuralgia) AND (MH Back OR MH Spine OR TI back OR lowback OR lumbar OR lumbal OR lumbo* OR sciatic OR spine OR spinal OR radicular OR nerve root* OR disk OR disc OR disks OR discs OR vertebra* OR intervertebra* OR sacroilia* OR MH "Sacroiliac joint"))

3 Target condition: lumbar radiculopathy

MH "Intervertebral disk displacement" OR MH "Polyradiculopathy+" OR radiculopath* OR radiculiti* OR ((disc OR discs OR disk OR disks) AND (displacement OR hernia* OR protru* OR avulsion*)) OR ((nerve root*) AND (compress* OR entrap* OR inflammat* OR disorder*)) OR (MH Nerve compression syndromes AND (MH Back OR MH Spine OR TI back OR lowback OR lumbar OR lumbal OR lumbo* OR sciatic OR spine OR spinal OR radicular OR (nerve root*) OR disk OR disc OR disks OR discs OR vertebra* OR intervertebra* OR sacroilia* OR MH "Sacroiliac joint"))

Exclusion criteria: publication types

(ZT "Case study") OR (ZT "Editorial") OR (ZT "Letter") OR (ZT "Review")

Term combinations applied:

A 1a and (2 or 3) not 4

B 1b and 3 not 4

Total = A or B

Appendix 4. Criteria for Quality Assessement (QUADAS)

Item and Guide to classification
1. Was the spectrum of patients representative of the patients who will receive the test in practice? Is it a selective sample of patients? Differences in demographic or clinical features between the study population and the source population may lead to selection bias or spectrum variation. In this item we will focus on selection bias: is a selective sample of patients included? • **Classify as 'yes'** if a consecutive series of patients or a random sample has been selected. Information should be given about setting, in- and exclusion criteria, and preferably number of patients eligible and excluded. If a mixed population of primary and secondary care patients is used: the number of participants from each setting is presented. • **Classify as 'no'** if healthy controls are used. Score also 'no' if non-response is high and selective, or there is clear evidence of selective sampling. Score also 'no' if a population is selected that is otherwise unsuitable, for example, patients are known to have other specific causes of LBP (severe OA, malignancies, etc). • **Classify as 'unclear'** if insufficient information is given on the setting, selection criteria, or selection procedure to make a judgment.

(*Continued*)

2. Is the reference standard likely to classify the target condition correctly?

Estimates of test performance are based on the assumption that the reference standard will identify nerve root compression due to disc herniation with 100% sensitivity and 100% specificity. Such reference standards are rare. Errors due to an imperfect reference standard may bias the estimation of diagnostic performance. For this review acceptable reference standards are: 1) findings at surgery demonstrating nerve root compression or irritation due to disc herniation; and 2) myelography indicating nerve root compression; and 3) although probably of lower quality, CT/MRI findings indicating nerve root compression;

- **Classify as 'yes'** if one of these procedures is used as reference standards.
- **Classify as 'no'** if you seriously question the methods used, if consensus among observers, or a combination of aspects of physical examination and history ('clinical judgement') is used as reference standard. *(Use of imaging/surgery is actually a selection criterion, so the latter may not occur)*
- **Classify as 'unclear'** if insufficient information is given on the reference standard.
- **Classify as 'not able'** if you consider yourself not capable to assess this item. If you have doubts, for example, regarding the quality of MRI-procedures but feel not competent to make an adequate assessment, we can consult a radiologist.

3. Is the time period between the reference standard and the index test short enough to be reasonably sure that the target condition did not change between the two tests?

The index tests and reference standard should ideally be carried out at the same time. If there is a considerable delay, misclassification (due to spontaneous recovery or worsening of the condition) may occur.

- **Classify as 'yes'** if the time period between physical examination and the reference standard is one week or less.
- **Classify as 'no'** if the time period between physical examination and the reference standard is longer than one week.
- **Classify as 'unclear'** if there is insufficient information on the time period between index tests and reference standard.

4. Did the whole sample or a random selection of the sample receive verification using a reference standard of diagnosis?

When not all of the study patients receive confirmation of their diagnosis by a reference standard, partial verification bias may occur. Bias is very likely if the results of the index test influence the decision to perform the reference standard. Random allocation of patients to the reference standard should in theory not affect diagnostic performance. [Verification bias is also known as work-up bias or sequential ordering bias].

- **Classify as 'yes'** if it is clear that all patients who received the index test went on to receive a reference standard, even if the reference standard is not the same for all patients.
- **Classify as 'no'** if not all patients who received the index test received verification by a reference standard.
- **Classify as 'unclear'** if insufficient information is provided to assess this item.

5. Did patients receive the same reference standard regardless of the index test result?

Differential verification bias occurs when the results of the index tests are verified by different reference standards. This is not unlikely in this review: some patients may be referred for surgery following physical examination, whereas others only go on to receive diagnostic imaging. Bias is likely to occur when this decision depends on the results of the index test.

- **Classify as 'yes'** if it is clear that all patients receiving the index test are subjected to the same reference standard.
- **Classify as 'no'** if different reference standards are used.
- **Classify as 'unclear'** if insufficient information is provided to assess this item.

6. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?

It is not unlikely that the results of the physical examination are used when establishing the final diagnosis. In this case incorporation bias may occur (overestimating diagnostic accuracy). Knowledge of the results of the index test does not necessarily mean that these results are incorporated in the reference standard. For example, if the reference standard consists of MRI-results only (regardless of knowledge of the results of the straight leg raising test), the index test is *not* part of the reference standard. However, if the final diagnosis is based on the results of both MRI-findings *and* a positive straight leg raising test, incorporation bias will occur.

- **Score 'yes'** if the index is no part of the reference standard.
- **Score 'no'** if the index test is clearly part of the reference standard.
- **Score 'unclear'** if insufficient information is provided to assess this item.

(*Continued*)

7. Were the reference standard results interpreted without knowledge of the results of the index test?

Interpretation of the results of physical examination may be influenced by knowledge of the results of the reference standard, and vice versa. This is known as reviewer bias, and may lead to over-estimation of diagnostic accuracy. In our review the risk of bias may be substantial as both index test and reference standard often involve a subjective assessment of results. If the index test always precedes the reference standard, interpretation of the results of the index test will usually be without knowledge of the results of the reference standard. The reverse may also be true, although surgery is unlikely to precede physical examination.

- **Classify as 'yes'** if the results of the reference standard are interpreted blind to the results of the index tests. Also score **'yes'** if the sequence of testing is always the same and, consequently, the reference standard is interpreted blind of the index test.
- **Classify as 'no'** if the assessor is aware of the results of the index test.
- **Classify as 'unclear'** if insufficient information is given on independent or blind assessment of the index test.

8. ***Were the index test results interpreted without knowledge of the results of the reference standard?***

Interpretation of the results of physical examination may be influenced by knowledge of the results of the reference standard, and vice versa. This is known as reviewer bias, and may lead to over-estimation of diagnostic accuracy. In our review the risk of bias may be substantial as both index test and reference standard often involve a subjective assessment of results. If the index test always precedes the reference standard, interpretation of the results of the index test will usually be without knowledge of the results of the reference standard. The reverse may also be true, although surgery is unlikely to precede physical examination.

- **Classify as 'yes'** if the results of the index test are interpreted blind to the results of the reference test. Score also **'yes'** if the sequence of testing is always the same and, consequently, the index test is interpreted blind of the reference standard.
- **Classify as 'no'** if the assessor is aware of the results of the reference standard.
- **Classify as 'unclear'** if insufficient information is given on independent or blind assessment of the reference standard.

9. Were the same clinical data available when the index test results were interpreted as would be available when the test is used in practice?

The knowledge of demographic and clinical data, such as age, gender, symptoms, history of low-back pain, previous treatments, or other aspects of physical examination may influence the interpretation of test results. The way this item is scored depends on the objective of the index test. If an aspect of physical examination is intended to replace other tests, these clinical data should *not* be available. However, if in practice clinical data are usually available when interpreting the results of the index test, this information should be available to the assessors of the index test.

- **Classify as 'yes'** if clinical data (i.e. patient history, other physical tests) would normally be available when the test results are interpreted and similar data are available in the study.
- **Classify as 'yes'** if clinical data would normally not be available when the test results are interpreted and these data are also not available in the study.
- **Classify as 'no'** if this is not the case, e.g. if other test results are available that can not be regarded as part of routine care.
- **Classify as 'unclear'** if the paper does not explain which clinical information was available at the time of assessment.

10. Were uninterpretable / intermediate test results reported?

Uninterpretable or intermediate test results are often not reported in diagnostic studies. Authors may simply remove these results from the analysis, which may lead to biased results of diagnostic performance. If uninterpretable or intermediate test results occur randomly and are not related to disease status, bias is unlikely. Whatever the cause of uninterpretable results they should be reported in order to estimate their potential influence on diagnostic performance.

- **Classify as 'yes'** if all test results are reported for all patients, including uninterpretable, indeterminate or intermediate results.
- **Classify as 'yes'** if the authors do not report any uninterpretable, indeterminate or intermediate results AND the results are reported for all patients who were described as having been entered into the study.
- **Classify as 'no'** if you think that such results occurred, but have not been reported.
- **Classify as 'unclear'** if it is unclear whether all results have been reported.

(*Continued*)

11. Were withdrawals from the study explained?

Patients may withdraw from the study before the results of both index test and reference standard are known. If withdrawals systematically differ from patients remaining in the study, then estimates of diagnostic test performance may be biased. A flow chart is sometimes provided (in more recently published papers) which may help to score this item.

- **Classify as 'yes'** if it is clear what happens to all patients who entered the study (all patients are accounted for, preferably in a flow chart).
- **Classify as 'yes'** if the authors do not report any withdrawals AND if the results are available for all patients who were reported to have been entered in the study.
- **Classify as 'no'** if it is clear that not all patients who were entered completed the study (received both index test and reference standard), and not all patients are accounted for.
- **Classify as 'unclear'** when the paper does not clearly describe whether or not all patients completed all tests, and are included in the analysis.

Note: In many diagnostic studies one may doubt whether or not all eligible patients have been entered in the study and are described in the paper. This issue is more strongly related to selection bias and will be scored under item 1.

Additional QUADAS items

12. Did the study provide a clear definition of what was considered to be a "positive" result of the index test?

Aspects of physical examination, for example the straight leg raising test, require a subjective judgement. Furthermore, several methods of performing the test have been described, and several cut-offs have been proposed. Consequently, it is essential that an adequate description is given of the methods used to carry out (aspects of) physical examination, and how a positive result is defined.

- **Classify as 'yes'** if the paper provides a clear description of the way the index test is performed, including a definition of a positive test result.
- **Classify as 'no'** if no description is given of the way the index test is performed, and no definition is given of a positive test result.
- **Classify as 'unclear'** if the methods of the index test are described, but no clear definition of a positive result has been provided, or vice versa.

13. Was treatment withheld until both index test and reference standard were performed?

If index tests and reference standard are not performed on the same day, some type of intervention may be initiated in between index test and reference standard. This might lead to misclassification (if some recovery of symptoms occurs).

- **Classify as 'yes'** if no treatment is given in the time period between physical examination and the reference standard.
- **Classify as 'no'** if an intervention is given that in your opinion could possibly influence the prognosis of low-back pain due to nerve root compression / irritation.
- **Classify as 'unclear'** if there is insufficient information regarding treatment between index test and reference standard.

14. Were data on observer variation reported and within acceptable range?

Studies on the reproducibility of physical examination in patients with musculoskeletal pain show that there may be considerable inter-observer variation. This may strongly influence the diagnostic performance of the index test. It is difficult to give minimal cut-off scores for inter-observer agreement. A kappa or ICC of 0.70 is often considered to be acceptable, but this is certainly an arbitrary definition.

- **Classify as 'yes'** if the paper provides information on inter-observer variation, and the results are acceptable.
- **Classify as 'no'** if information is given on inter-observer variation, and the results demonstrate poor agreement.
- **Classify as 'unclear'** if there is insufficient information is provided regarding inter-observer variation

Appendix 5. Diagnostic performance of Straight leg raising test

Reference	Index test	Reference standard	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Vroomen 2002 Primary care	**SLR Leg pain**	**MRI: nerve root compression**	**97**	**53**	**55**	**69**	**0.64 (0.56 to 0.71)**	**0.57 (0.47 to 0.66)**
Albeck 1996	**SLR Leg pain any angle**	**S: extrusion**	**51**	**15**	**10**	**4**	**0.84 (0.72 to 0.92)**	**0.21 (0.06 to 0.46)**
Charnley 1951	**SLR Leg pain < 60°**	**S: protrusion**	**63**	**8**	**11**	**6**	**0.85 (9.75 to 0.92)**	**0.43 (0.18 to 0.71)**
Charnely 1951	SLR Leg pain < 40°	S: protrusion	58	5	16	9	0.78 (0.67 to 0.87)	0.64 (0.35 to 0.87)
Demircan 2002	SLR: pain < 30°	S: protrusion / sequestration	129	5	55	95	0.70 (0.63 to 0.77)	0.95 (0.89 to 0.98)
Demircan 2002	SLR: pain < 60°	S: protrusion / sequestration	172	8	12	92	0.94 (0.89 to 0.97)	0.92 (0.85 to 0.97)
Demircan 2002	**SLR: pain < 90°**	**S: protrusion / sequestration**	**179**	**18**	**5**	**82**	**0.97 (0.94 to 0.99)**	**0.82 (0.73 to 0.89)**
Gurdjian 1961	**SLR (cut off not given)**	**S: protrusion**	**929**	**12**	**222**	**13**	**0.81 (0.78 to 0.83)**	**0.52 (0.31 to 0.72)**
Hakelius 1972	SLR: pain < 30°	S: protrusion / sequestration	600	107	867	385	0.41 (0.38 to 0.44)	0.78 (0.74 to 0.82)
Hakelius 1972	SLR: pain < 60°	S: protrusion / sequestration	1229	313	238	179	0.84 (0.82 to 0.86)	0.36 (0.32 to 0.41)

(*Continued*)

Hakelius 1972	**SLR: pain any angle**	**S: protrusion / sequestration**	**1411**	**422**	**56**	**70**	**0.96 (0.95 to 0.97)**	**0.14 (0.11 to 0.18)**
Haldeman 1988	**SLR: leg pain < 60°**	**CT: bulge / stenosis**	**10**	**16**	**17**	**57**	**0.37 (0.19 to 0.58)**	**0.78 (0.67 to 0.87)**
Haldeman et al. 1988	SLR: any leg pain or back pain at <60°	CT: bulge / stenosis	17	39	10	34	0.62 (0.43 to 0.81)	0.47 (0.35 to 0.59)
Kerr 1988	**SLR: leg pain any angle**	**S: protrusion / sequestration**	**98**	**20**	**2**	**16**	**0.98 (0.93 to 0.99)**	**0.44 (0.28 to 0.62)**
Knutsson 1961 Subgroup first surgery	**SLR (cut off not given)**	**S: herniation / protrusion**	**155**	**18**	**7**	**2**	**0.96 (0.91 to 0.98)**	**0.10 (0.01 to 0.32)**
Kosteljanetz 1984	SLR Leg pain < 30°	S: complete / incomplete herniation	8	3	48	41	0.14 (0.06 to 0.26)	0.93 (0.81 to 0.99)
Kosteljanetz 1984	SLR Leg pain < 50°	S: complete / incomplete herniation	29	12	27	32	0.52 (0.38 to 0.65)	0.73 (0.57 to 0.85)
Kosteljanetz 1984	**SLR Leg pain any angle**	**S: complete / incomplete herniation**	**44**	**23**	**12**	**21**	**0.79 (0.66 to 0.88)**	**0.48 (0.33 to 0.63)**
Kosteljanetz 1984	SLR: leg pain / back pain any angle	S: complete / incomplete herniation	51	35	5	9	0.91 (0.80 to 0.97)	0.21 (0.10 to 0.35)
Kosteljanetz 1988	**SLR: Leg pain at any angle**	**S: prolapse**	**40**	**6**	**5**	**1**	**0.89 (0.76 to 0.96)**	**0.14 (0.004 to 0.58)**
Kosteljanetz 1988	SLR: leg pain / back pain at any angle	S: prolapse	43	6	2	1	0.96 (0.85 to 0.99)	0.14 (0.004 to 0.58)

(*Continued*)

Majlesi 2008	SLR Pain < 70°	MRI: bulging, protruding, extruding	20	4	18	33	0.52 (0.42 to 0.58)	0.89 (0.79 to 0.95)
Meylemans 1988	SLR pain < 45°	CT: radiculopathy	35	0	66	45	0.35 (0.26 to 0.45)	1.00 (0.92 to 1.00)
Poireaudeau 2001	SLR: Leg pain at any angle	MRI, CT or myelography: herniation	34	22	8	13	0.81 (0.66 to 0.91)	0.37 (0.22 to 0.55)
Spangfort 1972	SLR: Leg pain at any angle	S: bulging, (in)complete herniation	2088	308	69	39	0.97 (0.96 to 0.98)	0.11 (0.08 to 0.15)

Abbreviations: S = surgery, CT = computed tomography, MRI = magnetic resonance imaging, TP = true positives, FP = false positives, FN = false negatives, TN = true negatives, CI = confidence interval. Shaded comparisons are used for further analyses

Appendix 6. Diagnostic performance of Crossed straight leg raising test

Reference	Index test	Reference standard	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Kerr 1988	XSLR: contralateral leg pain, any angle	S: protrusion / sequestration	43	2	57	34	0.43 (0.33 to 0.53)	0.94 (0.81 to 0.99)
Knutsson 1961 Subgroup first surgery	XSLR (cut off not given)	S: herniation / protrusion	40	1	122	18	0.25 (0.18 to 0.32)	0.95 (0.74 to 0.99)
Kostel-janetz 1988	XSLR: contralateral leg pain any angle	S: prolapse	11	0	34	7	0.24 (0.13 to 0.40)	1.00 (0.59 to 1.00)
Kosteljanetz 1988	XSLR: contra-lateral leg pain	S: prolapse	19	1	26	6	0.42 (0.28 to 0.58)	0.86 (0.42 to 0.99)

(*Continued*)

	or back pain any angle							
Poireaudeau 2001	**XSLR: contralateral leg pain any angle**	**MRI, CT or myelography: herniation**	**12**	**6**	**30**	**29**	**0.29 (0.16 to 0.45)**	**0.83 (0.66 to 0.93)**
Spangfort 1972	**XSLR: contralateral leg pain at any angle**	**S: bulging (in)complete herniation**	**500**	**41**	**1657**	**306**	**0.23 (0.21 to 0.25)**	**0.88 (0.84 to 0.91)**

Abbreviations: S = surgery, CT = computed tomography, MRI = magnetic resonance imaging, TP = true positives, FP = false positives, FN = false negatives, TN = true negatives, CI = confidence interval. Shaded comparisons are used for further analyses

Appendix 7. Diagnostic performance of physical examination for Scoliosis

Reference	Index test	Reference standard	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Albeck 1996	**Scoliosis (visual inspection)**	**S: extrusion**	**24**	**7**	**37**	**12**	**0.39 (0.27 to 0.53)**	**0.63 (0.38 to 0.84)**
Kerr 1988	**Scoliosis: visual inspection**	**S: protrusion/ sequestration**	**63**	**4**	**37**	**32**	**0.63 (0.53 to 0.72)**	**0.89 (0.74 to 0.97)**
Kostel-janetz 1984	**Scoliosis ?**	**S: (in)complete herniation**	**30**	**11**	**28**	**31**	**0.52 (0.38 to 0.65)**	**0.74 (0.58 to 0.86)**
Vucetic 1996	Scoliosis Visual inspection	S: extrusion, sequestration	70	31	33	29	0.68 (0.58 to 0.77)	0.48 (0.35 to 0.62)
Vucetic 1996	**Scoliosis Visual inspection**	**S: sequestration, extrusion, or protrusion**	**96**	**5**	**54**	**8**	**0.64 (0.56 to 0.72)**	**0.62 (0.32 to 0.86)**

Abbreviations: S = surgery, CT = computed tomography, MRI = magnetic resonance imaging, TP = true positives, FP = false positives, FN = false negatives, TN = true negatives, CI = confidence interval. Shaded comparisons are used for further analyses

Appendix 8. Diagnostic performance of physical examination for Paresis or muscle weakness

Reference	Index test	Reference standard	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Vroomen 2002 Primary care	**Paresis**	**MRI: nerve root compression**	**41**	**8**	**111**	**114**	**0.27 (0.20 to 0.37)**	**0.93 (0.88 ? 0.97)**
Albeck 1996	**Paresis**	**S: extrusion**	**21**	**10**	**40**	**9**	**0.34 (0.23 to 0.47)**	**0.47 (0.24 to 0.71)**
Kerr 1988	**Reduced power dorsiflexion**	**S: protrusion / sequestration**	**54**	**4**	**46**	**32**	**0.54 (0.44 to 0.64)**	**0.89 (0.74 to 0.97)**
Knutsson 1961 All	Weakness paralysis big toe	S: herniation / protrusion	112	14	66	14	0.63 (0.55 to 0.70)	0.50 (0.31 to 0.69)
Knutsson 1961 Subgroup first surgery	**Weakness or paralysis big toe**	**S: herniation / protrusion**	**100**	**10**	**62**	**10**	**0.62 (0.54 to 0.69)**	**0.50 (0.27 to 0.73)**
Knutsson 1961 Subgr previous surgery	Weakness or paralysis big toe	S: herniation / protrusion	12	4	4	4	0.75 (0.48 to 0.93)	0.50 (0.16 to 0.84)
Kostel-janetz 1984	**Paresis, muscle weakness**	**S: (in)complete herniation**	**27**	**20**	**31**	**22**	**0.47 (0.33 to 0.61)**	**0.52 (0.36 to 0.68)**
Spangfort 1972	Paresis dorsiflexion	S: (in)complete herniation	585	170	1357	392	0.30 (0.28 to 0.32)	0.70 (0.66 to 0.74)
Spangfort 1972	**Paresis dorsiflexion**	**S: (in)complete herniation / bulging disc**	**645**	**110**	**1512**	**217**	**0.30 (0.28 to 0.32)**	**0.66 (0.61 to 0.72)**
Vucetic 1996	Extensor paresis big toe or ankle	S: extrusion, sequestration	27	19	76	41	0.26 (0.18 to 0.36)	0.68 (0.55 to 0.80)

(*Continued*)

Vucetic 1996	**Extensor paresis big toe or ankle**	**S: sequestration, extrusion, or protrusion**	**43**	**3**	**107**	**10**	**0.29 (0.22 to 0.37)**	**0.77 (0.46 to 0.95)**

Abbreviations: S = surgery, CT = computed tomography, MRI = magnetic resonance imaging, TP = true positives, FP = false positives, FN = false negatives, TN = true negatives, CI = confidence interval. Shaded comparisons are used for further analyses

Appendix 9. Diagnostic performance of physical test for Muscle wasting

Reference	Index test	Reference standard	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Albeck 1996	Muscle wasting	S: extrusion	9	6	51	13	0.15 (0.07 to 0.27)	0.68 (0.43 to 0.87)
Kerr 1988	Muscle wasting: 1 cm diff calf circumference	S: protrusion / sequestration	29	2	71	34	0.29 (0.20 to 0.39)	0.94 (0.81 to 0.99)
Knuttson 1961 All	Muscle wasting	S: herniation / protrusion	69	13	109	15	0.39 (0.32 to 0.46)	0.54 (0.34 to 0.73)
Knutsson 1961 Subgroup first surgery	Muscle wasting	S: herniation / protrusion	62	10	100	10	0.38 (0.31 to 0.46)	0.50 (0.27 to 0.73)
Knutsson 1961 Subgr previous surgery	Muscle wasting	S: herniation / protrusion	7	3	9	5	0.44 (0.20 to 0.70)	0.63 (0.25 to 0.92)

Abbreviations: S = surgery, CT = computed tomography, MRI = magnetic resonance imaging, TP = true positives, FP = false positives, FN = false negatives, TN = true negatives, CI = confidence interval.

Appendix 10. Diagnostic performance of physical test for Impaired reflexes

Reference	Index test	Reference standard	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Vroomen 2002 Primary care	**Absence tendon reflexes**	**MRI: nerve root compression**	**22**	**8**	**130**	**114**	**0.15 (0.09 to 0.21)**	**0.93 (0.88 to 0.97)**
Albeck 1996	**Impaired reflex (asymmetric)**	**S: extrusion**	**37**	**7**	**24**	**12**	**0.61 (0.47 to 0.73)**	**0.63 (0.38 to 0.84)**
Gurdjian 1961	**Impaired Achilles tendon reflex (9 cases bilateral, all other unilateral)**	**S: protrusion**	**486**	**10**	**665**	**15**	**0.42 (0.39 to 0.45)**	**0.60 (0.39 to 0.79)**
Gurdjian 1961	Impaired patellar tendon reflex	S: protrusion	84	3	1067	22	0.07 (0.06 to 0.09)	0.88 (0.69 to 0.98)
Kerr 1988	**Abnormal ankle tendon reflex**	**S: protrusion / sequestration**	**48**	**4**	**52**	**32**	**0.48 (0.38 to 0.58)**	**0.89 (0.74 to 0.97)**
Knutsson 1961 All	Weakness Achilles tendon reflex	S: herniation / protrusion	100	13	78	15	0.56 (0.49 to 0.64)	0.54 (0.34 to 0.73)
Knutsson 1961 Subgroup first surgery	**Weakness Achilles tendon reflex**	**S: herniation / protrusion**	**87**	**8**	**75**	**12**	**0.54 (0.46 to 0.62)**	**0.60 (0.36 to 0.81)**
Knutsson 1961 Subgr previous surgery	Weakness Achilles tendon reflex	S: herniation / protrusion	13	5	3	3	0.81 (0.54 to 0.96)	0.38 (0.09 to 0.76)
Knutsson 1961 All	Weakness patellar tendon reflex	S: herniation / protrusion	27	7	151	21	0.15 (0.10 to 0.21)	0.75 (0.55 to 0.89)

(*Continued*)

Knutsson 1961 Subgroup first surgery	Weakness patellar tendon reflex	S: herniation / protrusion	23	7	139	13	0.14 (0.09 to 0.21)	0.65 (0.41 to 0.85)
Knutsson 1961 Subgr previous surgery	Weakness patellar tendon reflex	S: herniation / protrusion	4	0	12	8	0.25 (0.07 to 0.52)	1.00 (0.63 to 1.00)
Spangfort 1972	Impairment ankle reflex	S: (in)complete herniation	618	126	1324	436	0.32 (0.30 to 0.34)	0.78 (0.74 to 0.81)
Spangfort, 1972	**Impairment ankle reflex**	**S: (in)complete herniation or bulging disc**	**675**	**69**	**1482**	**278**	**0.31 (0.29 to 0.33)**	**0.80 (0.76 to 0.84)**
Vucetic 1996	Areflexia : ≥1 tendon reflexes absent	S: extrusion, sequestration	44	12	59	48	0.43 (0.33 to 0.53)	0.80 (0.68 to 0.89)
Vucetic 1996	**Areflexia : ≥1 tendon reflexes absent**	**S: sequestration, extrusion, or protrusion**	**53**	**3**	**97**	**10**	**0.35 (0.28 to 0.44)**	**0.77 (0.46 to 0.95)**

Abbreviations: S = surgery, CT = computed tomography, MRI = magnetic resonance imaging, TP = true positives, FP = false positives, FN = false negatives, TN = true negatives, CI = confidence interval. Shaded comparisons are used for further analyses

Appendix 11. Diagnostic performance of physical test for Sensory deficits

Reference	Index test	Reference standard	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Vroomen 2002 Primary care	**Sensory loss: hypaesthesia**	**MRI: nerve root compression**	**43**	**42**	**109**	**80**	**0.28 (0.21 to 0.36)**	**0.66 (0.56 to 0.74)**
Vroomen 2002	Sensory loss: hypalgesia	MRI: nerve root	26	19	126	103	0.17 (0.12 to 0.24)	0.84 (0.77 to 0.90)

(*Continued*)

Primary care		compression						
Albeck 1996	**Sensory deficits**	**S: extrusion**	**41**	**11**	**20**	**8**	**0.67 (0.54 to 0.79)**	**0.42 (0.20 to 0.67)**
Kerr 1988	**Tingling, numbness**	**S: protrusion / sequestration**	**30**	**15**	**70**	**21**	**0.30 (0.21 to 0.40)**	**0.58 (0.41 to 0.75)**
Knuttson 1961 all	Impaired sensibility	S: herniation / protrusion	51	10	127	18	0.29 (0.22 to 0.36)	0.64 (0.44 to 0.81)
Knutsson 1961 Subgroup first surgery	**Impaired sensibility**	**S: herniation / protrusion**	**46**	**7**	**116**	**13**	**0.28 (0.21 to 0.36)**	**0.65 (0.41 to 0.85)**
Knutsson 1961 Sub previous surgery	Impaired sensibility	S: herniation / protrusion	5	3	11	5	0.31 (0.11 to 0.59)	0.63 (0.25 to 0.92)
Kostel-janetz 1984	**Sensory loss**	**S: (in)complete / herniation**	**35**	**18**	**23**	**24**	**0.60 (0.47 to 0.73)**	**0.57 (0.41 to 0.72)**
Vucetic 1996	Decreased sensibility	S: extrusion, sequestration	43	28	60	32	0.42 (0.32 to 0.52)	0.53 (0.40 to 0.66)
Vucetic 1996	**Decreased sensibility**	**S: sequestration, extrusion, or protrusion**	**67**	**4**	**83**	**9**	**0.45 (0.37 to 0.53)**	**0.69 (0.39 to 0.91)**

Abbreviations: S = surgery, CT = computed tomography, MRI = magnetic resonance imaging, TP = true positives, FP = false positives, FN = false negatives, TN = true negatives, CI = confidence interval. Shaded comparisons are used for further analyses

Appendix 12. Diagnostic performance of Forward Flexion

Reference	Index test	Reference standard	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Vroomen 2002 Primary care	Finger-floor >25 cm	MRI: nerve root compression	68	32	84	90	0.45 (0.37 to 0.53)	0.74 (0.65 to 0.81)
Albeck 1996	Forward flexion (knee or higher)	S: extrusion	55	16	6	3	0.90 (0.80 to 0.96)	0.16 (0.03 to 0.40)
Charnley 1951	Forward flexion (<3 inches below knee)	S: protrusion	63	10	11	4	0.85 (0.75 to 0.92)	0.29 (0.08 to 0.58)

Abbreviations: S = surgery, CT = computed tomography, MRI = magnetic resonance imaging, TP = true positives, FP = false positives, FN = false negatives, TN = true negatives, CI = confidence interval.

Appendix 13. Diagnostic performance of Extenstion Test

Reference	Index test	Reference standard	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Poiraudeau 2001	Hyper extension	MRI, CT or myelography: herniation	19	11	23	24	0.45 (0.30 to 0.61)	0.69 (0.51 to 0.83)
Stankovic 1999	Extension test: major loss	CT/MRI: herniation	10	3	42	50	0.19 (0.10 to 0.33)	0.94 (0.84 to 0.99)
Stankovic 1999	Extension test: major loss	CT/MRI: bulging / herniation	12	1	81	11	0.13 (0.07 to 0.22)	0.92 (0.62 to 0.99)
Stankovic 1999	Extension test: moderate loss	CT/MRI: herniation	24	24	28	29	0.46 (0.32 to 0.61)	0.55 (0.40 to 0.68)
Stankovic 1999	Extension test: moder-	CT/MRI: bulging /	42	6	51	6	0.45 (0.35 to 0.56)	0.50 (0.21 to 0.79)

(*Continued*)

	ate loss	herniation						
Stankovic 1999	Extension test: any loss	CT/MRI: herniation	47	44	5	9	0.90 (0.79 to 0.97)	0.17 (0.08 to 0.30)
Stankovic 1999	Extension test: any loss	CT/MRI: bulging / herniation	82	9	11	3	0.88 (0.80 to 0.94)	0.25 (0.06 to 0.57)

Abbreviations: S = surgery, CT = computed tomography, MRI = magnetic resonance imaging, TP = true positives, FP = false positives, FN = false negatives, TN = true negatives, CI = confidence interval

Appendix 14. Diagnostic performance of other tests

Reference	Index test	Reference standard	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Majlesi 2008	Slump test: Any pain?	MRI: bulging, protruding, extruding disc	32	6	6	31	0.84 (0.74 to 0.90)	0.83 (0.73 to 0.90)
Stankovic 1999	Slump test: Pain below knee	CT/MRI: herniation	25	21	27	32	0.48 (0.34 to 0.62)	0.60 (0.46 to 0.74)
Stankovic 1999	Slump test: Pain below knee	CT/MRI: bulging / herniation	41	5	52	7	0.44 (0.34 to 0.55)	0.58 (0.28 to 0.85)
Stankovic 1999	Slump test: Pain in buttock / leg	CT/MRI: herniation	40	34	12	19	0.77 (0.63 to 0.88)	0.36 (0.23 to 0.50)
Stankovic 1999	Slump test: Pain in buttock / leg	CT/MRI: bulging / herniation	66	8	27	4	0.71 (0.61 to 0.80)	0.33 (0.10 to 0.65)
Stankovic 1999	Slump test: Pain anywhere	CT/MRI: herniation	49	41	3	12	0.94 (0.84 to 0.99)	0.23 (0.12 to 0.36)

(*Continued*)

Stankovic 1999	Slump test: Pain anywhere	CT/MRI: bulging / herniation	81	9	12	3	0.87 (0.79 to 0.93)	0.25 (0.06 to 0.57)
Poiraudeau 2001	Bell test	MRI, CT or myelography: herniation	20	13	21	22	0.49 (0.33 to 0.65)	0.63 (0.45 to 0.79)

Abbreviations: S = surgery, CT = computed tomography, MRI = magnetic resonance imaging, TP = true positives, FP = false positives, FN = false negatives, TN = true negatives, CI = confidence interval

Appendix 15. Diagnostic performance of combinations of tests

Reference	Index test	Reference standard	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Charnley 1951	SLR + finger-to floor <60° + <3 inches below knee	S: protrusion	57	6	17	8	0.77 (0.66 to 0.86)	0.57 (0.29 to 0.82)
Charnely 1951	SLR + finger to floor <40° + <3 inches below knee	S: protrusion	54	5	20	9	0.73 (0.61 to 0.83)	0.64 (0.35 to 0.87)
Hudgins 1979 All patients	SLR+ and XSLR+: contralateral leg pain, any angle	S: protrusion or clinical follow-up	56	2	188	105	0.23 (0.18 to 0.29)	0.98 (0.93 to 0.99)
Hudgins 1979 Subgroup surgery	SLR+ and XSLR+: contralateral leg pain, any angle	S: protrusion	54	2	173	45	0.24 (0.18 to 0.30)	0.96 (0.86 to 0.99)
Hudgins 1979 Subgroup conservative	SLR+ and XSLR+: contralateral leg pain, any	clinical follow-up	2	0	15	60	0.12 (0.02 to 0.36)	1.00 (0.94 to 1.00)

(*Continued*)

	angle							
Meylemans 1988	Neurological signs (sensibility, reflexes)	CT: radiculopathy	40	6	61	39	0.40 (0.30 to 0.50)	0.87 (0.73 to 0.95)
Meylemans 1988	Neurological signs + SLR (<45°)	CT: radiculopathy	22	0	79	45	0.22 (0.14 to 0.31)	1.00 (0.92 to 1.00)
Poiraudeau 2001 **	SLR + Bell test	MRI, CT or myelography: herniation	18	12	25	23	0.42 (0.27 to 0.58)	0.66 (0.48 to 0.81)
Poiraudeau 2001 **	SLR + hyper extension	MRI, CT or myelography: herniation	18	10	25	25	0.42 (0.27 to 0.58)	0.71 (0.54 to 0.85)
Poiraudeau 2001 **	XSLR + Bell test	MRI, CT or myelography: herniation	9	4	34	31	0.21 (0.10 to 0.36)	0.89 (0.73 to 0.97)
Poiraudeau 2001 **	XSLR + hyper extension	MRI, CT or myelography: herniation	9	2	34	33	0.21 (0.10 to 0.36)	0.94 (0.81 to 0.99)
Poiraudeau et al. 2001 **	SLR + XSLR	MRI, CT or myelography: herniation	12	9	31	26	0.28 (0.15 to 0.44)	0.74 (0.57 to 0.88)
Poiraudeau 2001**	SLR + XSLR + Bell test + hyper extension	MRI, CT or myelography: herniation	7	3	36	32	0.16 (0.07 to 0.31)	0.91 (0.77 to 0.98)

Abbreviations: S = surgery, CT = computed tomography, MRI = magnetic resonance imaging, TP = true positives, FP = false positives, FN = false negatives, TN = true negatives, CI = confidence interval. ; ** average results of three observers