Diagnosis of lumbar spinal stenosis: a systematic review of the accuracy of diagnostic tests

de Graaf I, Prak A, Bierma-Zeinstra S, Thomas S, Peul W, Koes B

CRD summary
This review assessed the accuracy of various tests for the diagnosis of lumbar spinal stenosis. The authors concluded that there was insufficient evidence from the included diverse poor-quality studies to draw firm conclusions, and that further research is required. The review was generally well conducted and the conclusions correctly reflect the limited evidence from the included studies.

Authors' objectives
To review the accuracy of tests used to detect lumbar spinal stenosis.

Searching
MEDLINE and EMBASE were searched up to May 2004; the search terms were reported and included a diagnostic filter. The reference lists of included studies and one meta-analysis were reviewed to identify additional relevant studies. Studies were included if they were published in English, German, French or Dutch.

Study selection
Study designs of evaluations included in the review
Diagnostic accuracy studies that compared the index test with a reference standard were eligible for inclusion. Both prospective and retrospective studies were included.

Specific interventions included in the review
Studies of imaging, clinical or other tests used to detect lumbar spinal stenosis were eligible for inclusion. The imaging tests evaluated in the included studies were computed tomography (CT), conventional myelography, CT myelography, magnetic resonance imaging (MRI), three-dimensional magnetic resonance myelography (MRM) and ultrasound. The clinical tests included questionnaires, treadmill tests, standardised history and physical examination, and pain drawings. Other tests evaluated were dermatomal somatosensory-evoked potentials and selective lumbar root sheath infiltration. Some studies only sought to identify central stenosis; other studies identified central, lateral or foraminal stenosis as a group.

Reference standard test against which the new test was compared
The studies had to include a reference standard to be included in the review, but no inclusion criteria relating to the reference standard were specified. The reference standards used were surgical confirmation, MRI, myelography, CT, expert opinion, and a combination of clinical, radiological and other diagnostic tests.

Participants included in the review
Studies of adults being evaluated for lumbar spinal stenosis were eligible for inclusion. Studies that included a subgroup of patients with lumbar spinal stenosis had to analyse these patients separately. Where reported, the patients in the included studies were aged from 10 to 95 years (mean age: 42 to 65) and 31 to 67% were men. The patients were being evaluated for a variety of medical conditions: sciatica, suspected spinal stenosis, chronic lumbar pain, spondylolysis and spondylolisthesis, suspected fracture, pain associated with degeneration piriformis syndrome, discitis, polyneuropathy, low back pain, leg pain, parasthesia, neurogenic claudication, suspected spinal canal stenosis, disc herniation, suspected central lateral and foraminal stenosis, herniated nucleus pulposus, serious underlying disorders or psychogenic regional pain disturbance, and suspicion of nerve root compression.

Outcomes assessed in the review
The studies had to report measures of diagnostic accuracy (sensitivity, specificity, predictive values and accuracy) or sufficient data to calculate these measures.
How were decisions on the relevance of primary studies made?
Two reviewers independently screened all abstracts. Full-text articles were obtained if articles could not be excluded on the basis of information in the title and abstract. Full articles were also screened independently by two reviewers. Any disagreements were resolved through referral to a third reviewer.

Assessment of study quality
The studies were assessed for methodological quality using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool (see Other Publications of Related Interest). Studies that met more than 50% of the quality criteria were considered to be high quality. Two reviewers independently scored the articles and agreed on a final score.

Data extraction
The authors did not state how the data were extracted for the review, or how many reviewers performed the data extraction.

Where possible, data on sensitivity, specificity and positive and negative likelihood ratios were extracted from the individual studies. The reviewers checked all calculations and results of the primary studies. When these measures were not reported they were calculated if sufficient data were presented. The confidence intervals (CIs) of sensitivity and specificity were also calculated. When sensitivity and specificity were not reported, other measures of accuracy such as test accuracy and positive and negative predictive values were extracted.

Methods of synthesis
How were the studies combined?
The studies were grouped by index and reference test and combined in a narrative, with accompanying tables. The authors stated that statistical pooling of the studies was not possible because of differences between the studies with respect to index tests, reference standards and participants.

How were differences between studies investigated?
 Differences between the studies were discussed in the text. Linear regression was performed to investigate whether the quality of studies increased over time (adjusted for index test).

Results of the review
Twenty-four studies were included: 15 studies evaluated imaging tests (n=978, results from 688 were used in the review), 7 studies evaluated clinical tests (n=517) and 2 studies evaluated other tests (n=88).

The quality of the included studies was poor, with only 5 studies scoring positive on more than 50% of items.

Imaging studies (15 studies).
Estimates of sensitivity varied considerably between studies, and in most studies it was not possible to calculate the specificity. There was no difference in accuracy between myelography and CT or MRI.

Clinical studies (7 studies).
The clinical tests evaluated were very different between the studies. Estimates of sensitivity and specificity also showed considerable variation. Some clinical findings showed high sensitivity whereas others showed high specificity. Only one low-quality study, which evaluated a model based on a discriminate analysis, showed both sensitivity and specificity above 75%.

Other tests (2 studies).
Both studies were of low quality. One evaluated dermatomal somatosensory-evoked potentials and showed a sensitivity of 94%. The other investigated selective lumbar root sheath infiltration and did not report on the sensitivity or
specificity, although it reported a positive predictive value of 95%.

**Authors' conclusions**
Due to limitations in the quality of the primary studies and differences between them, it was not possible to draw firm conclusions about the diagnostic performance of the different tests. High-quality studies that evaluate the accuracy of diagnostic tests are required to improve the diagnostic policy.

**CRD commentary**
This was generally a well conducted and reported review. It addressed a focused question supported by clearly defined inclusion criteria. The literature search was limited to two databases and included a diagnostic filter, thus some relevant studies might have been missed. Attempts were made to reduce language bias but no attempts were made to locate unpublished studies, thus the review may be subject to publication bias. Methods were used to minimise reviewer errors and bias in the study selection and validity assessment processes, but it was unclear whether similar steps were taken when extracting the data. Adequate details of the included studies were reported in tables obtained directly from the authors. A detailed quality assessment was undertaken and the results were incorporated into a narrative synthesis of the results. In view of the differences amongst the studies, a narrative synthesis was appropriate. The authors' conclusions about the limitations of the data and the need for further research appear appropriate given the deficiencies of the primary studies.

**Implications of the review for practice and research**
Practice: The authors stated that the diagnosis of lumbar stenosis should be based on both clinical and radiological findings.

Research: The authors stated that further prospective diagnostic accuracy studies that are not subject to verification bias are required. In addition, since there is no adequate reference standard for the diagnosis of lumbar stenosis, follow-up studies (preferably randomised) are needed to determine the effect of diagnostic tests (plus treatments given) on patient outcomes.

**Bibliographic details**

**PubMedID**
16648755

**DOI**
10.1097/01.brs.0000216463.32136.7b

**Other publications of related interest**

**Indexing Status**
Subject indexing assigned by NLM

**MeSH**
Diagnostic Tests, Routine; Humans; Lumbar Vertebrae /pathology; MEDLINE; Reproducibility of Results; Sensitivity and Specificity; Spinal Stenosis /diagnosis

**AccessionNumber**
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.