SPINE Volume 31, Number 10, pp 1168–1176
©2006, Lippincott Williams & Wilkins, Inc.

Diagnosis of Lumbar Spinal Stenosis

A Systematic Review of the Accuracy of Diagnostic Tests

Irene de Graaf, MSc,* Anneloes Prak, MSc,* Sita Bierma-Zeinstra, PhD,* Siep Thomas, PhD,* Wilco Peul, MD,† and Bart Koes, PhD*

Study Design. Systematic review of diagnostic studies.

Objective. To investigate the diagnostic performance of tests used to detect lumbar spinal stenosis.

Summary of Background Data. Little is known about the diagnostic accuracy of tests in detecting lumbar spinal stenosis. A systematic review will provide more insight in this topic.

Methods. We performed a literature search in Medline (PubMed) and Embase for original diagnostic studies on lumbar spinal stenosis, in which one or more different tests were evaluated with a reference standard, and diagnostic values were reported or could be calculated. Two reviewers independently checked all abstracts and full text articles for inclusion criteria. Included articles were assessed for their quality using the Quadas tool. Study features and diagnostic values were extracted from the included studies.

Results. Of the 24 articles included in this review, 15 concerned imaging tests, 7 evaluated "clinical tests," and 2 reported on other diagnostic tests. The overall quality was poor; only 5 studies scored positive on more than 50% of the quality items. Estimates of the diagnostic value of the tests differed considerably. The imaging studies showed no superior accuracy for myelography compared with CT or MRI. Overall, there is considerable variation in the clinical tests; some studies show high sensitivity, whereas others show high specificity.

Conclusions. Because of heterogeneity and overall poor quality, no firm conclusions about the diagnostic performance of the different tests can be drawn. Better-designed studies exploring the accuracy of diagnostic tests are needed to improve the diagnostic policy.

Key words: lumbar spinal stenosis, diagnostic performance, systematic review. **Spine 2006;31:1168–1176**

Diagnoses are made to predict prognosis of symptoms and are meant to guide decisions about treatment. In the ideal situation, there is a clear diagnostic entity with an agreed gold standard to prove its existence as well as knowledge about natural course and effectiveness of treatments. In cases where the gold standard test is not directly available, other diagnostic tests (proxies) are developed to come close to this gold standard. At the present stage of development of the diagnostic entity lumbar spinal stenosis, this entity is defined as any type of narrowing of the lumbar spinal canal, causing compression of the content of the spinal canal, due to a conflict between the available space in the canal and its content. Symptoms are thought to be caused by direct mechanical compression or indirect vascular compression of the nerve roots or the cauda equina.[1,2] Symptoms are reported to be diverse; patients often complain of pain in the legs and the classic neurogenic claudication that is characterized by pain during walking, numbness, tingling, weakness, and radiating pain down to ankles. Symptoms are provoked by standing and relieved when sitting. Taking a flexed posture also reduces symptoms because in that way the available space in the lumbar spinal canal increases.[1,3]

Lumbar spinal stenosis is a result of degenerative, developmental, or congenital disorders. The degenerative type is often due to arthrotic changes of intervertebral discs, facet joints, and ligaments surrounding the vertebral canal. The degenerative type occurs most often, especially in those 50 to 60 years of age.[1,4–6] People with the congenital type may complain earlier in life; their stenosis is a result of congenitally anatomic changes or malformations, *e.g.*, an excessive scoliosis or excessive lordotic curves.[1,4,6] Developmental spinal stenosis is a condition in which the narrow spinal canal is caused by a growth disturbance of the posterior elements in the spinal canal.[7]

Lumbar spinal stenosis may occur at different localizations in the spinal canal, sometimes at more than one location at the same time. In central canal stenosis, nerve roots in the cauda equina may be compressed. Lateral recess stenosis and foraminal stenosis may cause compression of the nerve roots leaving the spine.[4,8]

Various tests are currently used to diagnose lumbar spinal stenosis. It is important to know the value of these diagnostic tests because a false-positive test result may lead to unnecessary surgery (*e.g.*, laminectomy) and/or expensive or invasive additional diagnostic interventions. Reduction of false-negative test results is also important because maybe patients with this chronic disease could be successfully treated (either with surgery or conservative therapy).

Little is known about the diagnostic accuracy of different tests available in detecting lumbar spinal stenosis. In 1992, Kent *et al* performed a systematic review of the diagnostic performance of MRI, CT, and myelography.[9] However, other diagnostic studies have since been intro-

From the *Department of General Practice, Erasmus MC, Rotterdam, The Netherlands; and †Department of Neurosurgery, LUMC, Leiden, The Netherlands. Drs. de Graaf and Prak contributed equally to this study.
Acknowledgment date: September 28, 2004. First revision date: March 2, 2005. Second revision date: June 8, 2005. Acceptance date: June 23, 2005.
The manuscript submitted does not contain information about medical device(s)/drug(s).
No funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.
Address correspondence and reprint requests to Sita Bierma-Zeinstra, PhD, Department of General Practice, Erasmus MC, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, E-mail: s.bierma-zeinstra@erasmusmc.nl

Item 1*: Was the spectrum of patients representative of the patients who will receive the test in practice?
Item 2: Were selection criteria clearly described?
Item 3*: Is the reference standard likely to correctly classify the target condition?
Item 4*: Is the time period between the reference standard and index test short enough to be reasonably sure that the target condition did not change in the time between the two tests?
Item 5: Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?
Item 6: Did patients receive the same reference standard regardless of the index test result?
Item 7: Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?
Item 8: Was the execution of the index test described in sufficient detail to permit replication of the test?
Item 9: Was the execution of the reference standard described in sufficient detail to permit its replication?
Item 10: Were the index test results interpreted without knowledge of the results of the reference standard?
Item 11: Were the reference standard results interpreted without knowledge of the results of the index test?
Item 12: Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?
Item 13*: Were uninterpretable/ intermediate test results reported?
Item 14: Were withdrawals from the study explained?

*Item 1: All people have symptoms of lumbar spinal stenosis. People should not have confirmed lumbar spinal stenosis at the start of the study.
*Item 3: Surgical confirmation (stenosis confirmed by the surgeon, either visually or by measurement) or a combination of clinical diagnoses and imaging methods were used as a reference standard.
*Item 4: The time period is less than one year.
*Item 13: Uninterpretables of both reference test and index test.

Figure 1. The separate items of the Quadas tool for methodologic assessment of diagnostic studies.[11]

duced, possibly with increasing quality on study designs. Moreover, it is important to explore the accuracy of all the available tests, including clinical tests.

Therefore, we performed a systematic review of studies evaluating the diagnostic value of imaging, clinical and other tests used to detect lumbar spinal stenosis. Although the clinical decision tree in lumbar spinal stenosis shows gaps in diagnostic knowledge as well as in prognostic and therapeutic knowledge, our scientific inventory focuses on the diagnostic level of the decision tree.

Methods

Search Methods. An optimal search strategy for identifying diagnostic accuracy studies was recently published.[10] We used this strategy and added disease-specific items relevant for the present question. A search was conducted in Medline (PubMed) and Embase for relevant articles up to March 2004 using the following strategy:

1. sensitivity OR specificity OR screening OR "false positive" OR "false negative" OR accuracy OR "predictive value" OR "predictive values" OR "reference value" OR "reference values" OR "reference standard*" OR roc OR likelihood
2. stenos* OR verbiest OR "neurogen* claudicat*" OR "claudicat* spinalis" OR LSS
3. low back OR "low back" OR lumbar OR spinal OR lumbosacral
4. diagn* #1 AND #2 AND #3 AND #4

Selection. The following selection criteria were used:

1) The study investigated the diagnostic accuracy of imaging, clinical, and other tests in detecting lumbar spinal stenosis, in an adult study population.
2) One or more different diagnostic tests as well as a reference test should be included within the design.
3) Diagnostic values (sensitivity, specificity, predictive values, and accuracy) are reported or can be calculated.
4) If the results concern a subgroup of patients with lumbar spinal stenosis, these should be analyzed separately in the same article.
5) The article should be written in English, German, French, or Dutch.

Two reviewers read all abstracts, independently of each other. Full text was retrieved of articles that could not be excluded based on title and abstract. These full text articles were read and checked for inclusion by two persons independently. If there was no agreement, a third reviewer was asked to make the final decision.

Reference lists of all included articles and one meta-analysis were reviewed to search for relevant references that were not found by our search strategy.

Quality. To assess the articles for their quality, we used the "Quality assessment for diagnostic accuracy studies" (Quadas) tool.[11] The quality items are shown in Figure 1. Two reviewers independently scored the included articles; a final score was made by agreement. The kappa value was calculated to assess

Table 1. Study Characteristics of the Included Diagnostic Studies on Imaging Tests

	N	Participants on Whom Results Are Based (N)*	Mean Age (range)	% Male	Characteristics of Participants	Setting and Design§	Type Stenosis
Ilkko (1988)[13]	116	116	44 (10–79)	58	Indication for radiologic examination, because of: sciatica (n = 69), suspected spinal stenosis (n = 18), chronic lumbar pain (n = 15), spondylosis and spondylolisthesis (n = 4), suspected fracture, pain associated with degeneration, piriformis syndrome, discitis, polyneuropathy and MS (n = 10)	Setting: tertiary; design: prospective	Central spinal stenosis, including bony and soft tissue (disc protrusion, ligamentum flavum)
Rankine (1997)[14]	79	79	48.8	42	Indication for MRI; symptoms: low back pain, with or without sciatica	Setting: unknown; design: retrospective	Central canal stenosis
Tervonen (1889)[15]	76	76	42 (21–63)	45	Indication for examination by CT or myelography; symptoms: unknown	Setting: tertiary; design: prospective	Central spinal stenosis
Eberhardt (1994)[16]	65	65	51.8 (23–80)	42	Surgery for lumbar complaints; symptoms: radicular complaints	Setting: unknown; design: prospective	Spinal stenosis: osseous narrowing of the spinal canal, with liquor disturbance, subgroup is spondylolisthesis
Engel (1985)[17]	67	19	Unknown	Unknown	Symptomatic spine patients, not all patients had symptomatology warranting myelogram or surgery; symptoms: low back pain, with or without radiation down one or both legs	Setting: tertiary; design: unknown	Focal stenosis isolated or superimposed on diffuse (including herniated disc)
Donmez (1990)[18]	50	50	(21–61)	56	Patients were preoperatively evaluated by CT; symptoms: low back pain, leg pain, paresthesia, neurogenic claudication	Setting: unknown; design: prospective	Lateral recess and central spinal stenosis
Freund (1997)[19]	25	25	44	64	Clinical indication for myelography, suspected for spinal canal stenosis; symptoms: unknown	Setting: unknown; design: prospective	Degenerative osseous spinal canal stenosis
Modic (1986)[20]	60	48	46 (19–73)	Unknown	Clinical history and physical exam that indicated a strong probability of disc herniation or canal stenosis with a likelihood of required surgery; symptoms: unknown	Setting: unknown; design: prospective	Neural foramina, lateral recess, central canal stenosis as one group
Bischoff (1993)[21]	57	28	(20–79)¶	51¶	Surgically explored for suspected disc herniation (n = 47) or spinal stenosis (n = 28); symptoms: unknown	Setting: tertiary; design: retrospective	Spinal stenosis including nerve root compression due to facet joint arthritis, foraminal, lateral recess stenosis as one group
Feldmeyer (1982)[22]	38	7	(16–81)¶	55¶	Suspected disc herniation (n = 18), or canal stenosis (n = 13), suspected for other spinal diseases (n = 7) with indication of preoperative myelography; symptoms: bilateral radiation, sciatica, neurogenic claudication and absence of Lasègue	Setting: unknown; design: prospective	Canal stenosis including narrow canal only, or associated with herniated disc or bony protrusions
Arrault (1987)[23]	60	60	Unknown	58	Surgically confirmed central and lateral spinal stenosis; symptoms: unknown	Setting: tertiary; design: retrospective	Central stenosis, lateral stenosis (only osseous or associated with herniated disc)
Bell (1984)[24]	122	46	Unknown	Unknown	Surgically confirmed disc herniation (n = 76), spinal stenosis (n = 46) or both; symptoms: unknown	Setting: unknown; design: retrospective	Spinal stenosis including nerve compression due to facet joint abnormalities and lateral recess stenosis
Bolender (1985)[25]	55	24	(36–85)¶	Unknown¶	Surgery for suspected central, lateral and foraminal stenosis; symptoms: pain, neurogenic claudication, sensory changes, weakness and absence of reflexes in the lower extremities	Setting: tertiary; design: retrospective	Central stenosis
Herkowitz (1982)[26]	30	18	59 (21–95)	50	Surgically confirmed disc herniation (n = 12) or stenosis (n = 18); symptoms: sciatica	Setting: unknown; design: retrospective	Spinal stenosis

(*Continued*)

Table 1. Continued

	N	Participants on Whom Results Are Based (N)*	Mean Age (range)	% Male	Characteristics of Participants	Setting and Design§	Type Stenosis
Jia (1991)[27]	78	27	41 (25–67)¶	65¶	Surgically confirmed disc herniation (n = 65), including 10 with concurrent osseous lateral recess stenosis and 4 with hypertrophic ligamentum flavum; nerve root canal or lateral recess stenosis only (n = 8) and central stenosis (n = 5); symptoms: unknown	Setting: unknown; design: unknown	Central canal, nerve root canal, lateral recess stenosis as one group

*No. of patients with (suspected) lumbar spinal stenosis, or no. of patients receiving reference, in case results are only based on this subgroup.
†When it was possible to report the mean age of the participants on who results are based, this age was reported. When it was not possible, the mean age of the total group was reported.
‡When it was possible to report the percentage male on who results were based, this percentage was reported. When it was not possible, percentage male of the total group was reported.
§Retrospective or prospective design.
¶Based on total N, not given on no. of participants on whom results are based.

the interobserver variation of the initial assessment of both reviewers.

Linear regression was performed to investigate whether the quality of the studies increased over time. The dependent variable was the quality score, and the independent variable was year of publication, adjusted for type of study (imaging tests, clinical tests, and other tests).

Data Extraction. Study characteristics of the included studies were extracted; in order to gain insight in the diagnostic accuracy, we focused on the sensitivity, specificity, and the negative and positive likelihood ratio of the test at issue.

All reported calculations and results in the studies were checked. When the diagnostic outcomes were not reported, we calculated them if sufficient data were presented. The confidence intervals (CI) of the sensitivity and specificity were also calculated. When the sensitivity and specificity were not reported and could not be calculated, we extracted other values such as accuracy, positive predictive value (PPV), and negative predictive value (NPV).

When possible, diagnostic values of the tests are statistically pooled; otherwise, the results are summarized in a qualitative manner.

■ Results

Search and Selection

Our search strategy in Medline (PubMed) resulted in 1,514 references and Embase yielded an additional 17 references. Reviewing of the reference lists resulted in 42 additional abstracts. In total, 138 articles were retrieved in full text; one article could not be retrieved.[12] Finally, 24 articles were included in this review. Main reasons for exclusion were: lack of a reference standard, a description of symptoms and signs in lumbar spinal stenosis in a case series only, no separate diagnostic outcomes for lumbar spinal stenosis were reported, only known cases of lumbar spinal stenosis were evaluated, only cervical or thoracal spinal stenosis were diagnosed, and the study was a case report, or an intervention study. Some articles were excluded for more than one reason.

Type of Studies

Fifteen articles concerned imaging tests (*e.g.*, CT, MRI, myelography, plain radiography, and ultrasound) for diagnosing lumbar spinal stenosis,[13–27] and 7 articles evaluated "clinical tests",[28–34] *e.g.*, standardized history, physical examination, and pain drawings (drawings in which patients indicate pain patterns). Two articles reported other diagnostic tests: dermatomal somatosensory-evoked potentials and selective lumbar root sheath infiltration (the selective inactivation of nerve roots, with evaluation of the effect on the experienced pain).[35,36]

The characteristics of the included studies on imaging tests are shown in Table 1, and the characteristics of the included studies on clinical tests and other tests are shown in Table 2.

Quality

Results of the quality assessment are shown in Appendix 1. The overall quality was poor; only five studies scored positive on more than 50% of the quality items; these studies were considered to be high-quality studies.[13,14,15,28,29] Most of the imaging studies[13–27] scored negative on the following items: not all people or a random selection received the same reference standard, not all people received the same reference standard regardless of the index test result, and the reference test was not interpreted without knowledge of the index test result. In almost all studies, selection criteria were not clearly described, the execution of the reference standard was not described in sufficient detail to allow replication of the test, uninterpretables and withdrawals often were not mentioned. The time interval between the index test and reference test (item 4) was often not reported; only four studies reported this time interval.[13–16]

The kappa value for interobserver variation in the initial quality assessment of the reviewers was 0.73.

Table 2. Study Characteristics of the Included Diagnostic Studies on Clinical Tests and Other Tests

	N	Participants on Whom Results Are Based (N)*	Mean Age (range)¶	% Male¶	Characteristics of Participants	Setting and Design§	Type Stenosis
Clinical tests							
Roach (1997)[28]	106	99	55 (19–88)	51	Symptoms: recurrent or chronic low back pain	Setting: tertiary; design: prospective	Spinal stenosis: narrowing of the spinal canal or foramina of developmental or degenerative origin
Jensen (1989)[29]	23	23	55 (23–72)	57	Indication for myelography; symptoms: neurogenic claudication, unilateral in 14 cases, bilateral in 9 cases	Setting: unknown; design: prospective	Spinal stenosis
Katz (1995)[30]	93	75	65 (40–91)	31	Symptoms: low back pain with or without radiation to the lower extremities	Setting: tertiary; design: prospective	Spinal stenosis: compression of nerve roots by narrowing of the spinal canal or neural foramina
Mann (1991/1992)[31–33]	250	(**A**) 25 (randomly selected); (**B** + **C**) 250	Unknown	Unknown	Five categories of patients were selected, with 50 patients in each category: benign back pain, herniated nucleus pulposus, spinal stenosis, serious underlying disorders or psychogenic regional pain disturbance; symptoms: unknown	Setting: tertiary design: retrospective	Spinal stenosis: narrowing of the spinal canal and foramina of congenital, developmental or degenerative origin
Fritz (1997)[34]	45	45	58	Unknown	All patients had previously undergone MRI or CT; symptoms: low back and lower extremity pain and self-reported limitations in walking tolerance	Setting: tertiary; design: prospective	Spinal stenosis: any narrowing of the spinal canal and/or nerve root canals
Other tests							
Snowden (1992)[35]	58	(**A**) 40 (**B**) 18 (**C**) 58	64 (23–85)	69%	High index of suspicion for spinal stenosis by history and physical examination, indication for DSEP, previously examined by MRI and/or CT. symptoms: unknown	Setting: tertiary; design: retrospective	Central, lateral, recess or foraminal stenosis, as one group
Castro (1991)[36]	30	30	53	Unknown	Suspicion of nerve root compression by degenerative changes; symptoms: upper leg pain, aggravating when standing, typical neurogenic claudication (n = 9), segmental pain in a dermatome (n = 12), long history of back complaints	Setting: unknown; design: prospective	Degenerative stenosis of the nerve root canal or lateral recess, as one group

*No. of patients with (suspected) lumbar spinal stenosis, or no. of patients receiving reference, in case results are only based on this subgroup.
†When it was possible to report the mean age of the participants on who results are based, this age was reported. When it was not possible, the mean age of the total group was reported
‡When it was possible to report the percentage male on who results were based, this percentage was reported. When it was not possible, percentage male of the total group was reported.
§Retrospective or prospective design.
¶Based on total N, not given on no. of participants on whom results are based.

The linear regression showed a positive association between year of publication and quality of the study (β = 1.36, P = 0.034).

Data Extraction

Because of the heterogeneity of the tests, study population, and reference standards, statistical pooling was not possible. Therefore, the results are summarized in a qualitative manner.

Imaging Tests, Surgery as a Reference Standard

Table 3 presents data on the diagnostic performance of the imaging tests.

In 12 studies, surgical confirmation was used as a reference standard.[16–27] None of these studies had a "high quality."

Of these 12 studies, 6 reported the sensitivity for CT,[18,20,22–25] which ranged from 21%[25] to 100%.[22] When the study with the lowest quality (only 14% of the items scored positive)[25] was not taken into consideration, the sensitivity ranged from 74% to 100%.[18,20,22–24]

Ten articles reported the sensitivity for conventional myelography,[16,19,20–27] which ranged from 54%[20] to 100%.[19]

One study reported the sensitivity for CT myelography, which was 87%.[21]

Three showed diagnostic values for MRI[20,21,27] with a sensitivity ranging from 77%[20] to 87%.[21] Two studies reported that the sensitivity for three-dimensional magnetic resonance myelography (three-dimensional MRM) was 100%.[16,19]

Of the 9 studies that investigated myelography as well as CT, MRI, or three-dimensional MRM, 5 studies showed a higher sensitivity for MRI, three-dimensional MRM, or CT than for myelography,[16,20–23] and 4 studies showed the same sensitivity for myelography as for CT or MRI.[19,21,23,25] A higher sensitivity for myelography was reported in 3 studies.[24,25,27] The specificity of myelography is slightly higher than that of CT and MRI, in the two studies that reported the specificity of these tests.[20,21]

Ultrasound was evaluated in one of the studies that used surgery as a reference standard; in this study a sensitivity of 95% was reported.[17]

The specificity of most imaging studies that used surgical confirmation as a reference standard could not be estimated; only in 3 studies was this possible.[18,20,21] In these cases, sensitivity and specificity were estimated based on involved levels or nerve roots.

Overall, the studies showed no superior accuracy for myelography compared to CT, MRI, or three-dimensional MRM, with three-dimensional MRM showing the highest sensitivity.

Imaging Tests, Other Reference Standards

In three studies evaluating imaging tests, other reference standards were used. These three studies had a "high quality."

In a study that used three-sequence MRI as a reference standard, a sensitivity of 60% was shown for a single-sequence MRI[14]; the specificity was 95%.

For ultrasound, a sensitivity of 90% and a specificity of 96% were found in a study in which myelography or CT was used as a reference standard.[15]

The sensitivity and specificity of plain radiography, as shown by one study in which CT was used as a reference standard,[13] were 66% and 93%, respectively.

In summary, three-sequence MRI appeared to be more sensitive than single-sequence MRI, and the accuracy of ultrasound appeared to be almost equal to that of CT or myelography.

Clinical Tests

Table 4 presents data on the diagnostic performance of the clinical tests.

Two of the studies that investigated clinical tests had a "high quality."[28,29]

The "clinical tests" consisted of questionnaires,[28,34] treadmill tests evaluated in two different ways,[29,34] standardized history and physical examination,[30] and pain drawings analyzed in three different ways but in the same study population.[31–33] The reference standards were: a combination of clinical, radiologic, and other diagnostic tests,[28,31–33] myelography,[29] expert opinion,[30] and MRI or CT.[34] As shown in Tables 3 and 4 (available online through Article Plus), there is considerable variation in tests and in test results.

Only twice were almost comparable items investigated in two separate studies. In the study by Fritz *et al*, "pain in the legs relieved when sitting" had a sensitivity of 81% and a specificity of 16%, in which all participants had previously undergone MRI or CT (that also served as the reference standard) for low back pain and lower extremity pain, with self-reported limitations in walking tolerance.[34] However, in the study by Katz *et al*, in which an expert opinion served as a reference standard, "no pain when seated" had a sensitivity of 46% and a specificity of 93%; included in this study were patients with low back pain with or without radiation to the lower extremities.[30] Both of these studies were performed in a tertiary setting.

In the "high-quality" study by Roach *et al*, in which a combination of clinical, radiologic, and imaging tests are used as a reference standard, again in a tertiary setting, using a population with chronic or recurrent low back pain, pseudoclaudication had a sensitivity of 63% and a specificity of 71%.[28] In the study by Katz *et al*, "worse when walking" had a sensitivity of 71% and a specificity of 30%, and used an expert opinion as a reference standard.[30]

Two studies with a "high quality" showed sensitivity above 75% in the following items and tests: "radiating leg pain" and the "Pain Response to Activity and Positioning questionnaire" in patients with disc disease with spinal stenosis,[28] and "any change of neurologic status."[29] Other items with a sensitivity above 75% in studies with a low quality were: "age >65 years," "pain below buttocks," "no pain with flexion,"[30] "prolonged recovery after level walking," "pain in legs relieved by sitting," and "best/worst posture with regard to symptoms."[34]

In "low-quality" studies, items and tests with a specificity above 75 are: "expert evaluation of pain drawings,"[32] "statistical analysis of pain drawings,"[31] "no pain when seated," "symptoms improve when seated," "wide-based gait," "abnormal Romberg," "pinprick deficit," "weakness," "vibration deficit," "absent Achilles reflex,"[30] "earlier onset of symptoms with level walking," and "longer total walking time during inclined walking."[34]

Only the model based on a discriminate analysis (with the variables "time to onset of symptoms" and "recovery time"), investigated in a study with low quality, showed both a sensitivity and specificity above 75%, respectively, 77% and 95%.[34]

Overall, there is considerable variation in the clinical tests. Some findings show high sensitivity, while others show high specificity. Only in one low-quality study were both sensitivity and specificity above 75%, and this concerned a modeled combination of symptoms.[34]

Other Tests

Table 4 presents data on the diagnostic performance of the other tests.

The 2 studies investigating other tests were both of low quality. One study investigated the diagnostic performance of dermatomal somatosensory-evoked potentials with CT or MRI as a reference standard; this study showed a sensitivity of 94%.[35] The other study investigated the diagnostic performance of the selective lumbar root sheath infiltration with successful outcome of surgery as a reference standard; this study did not report a sensitivity or specificity but showed a positive predictive value of 95%.[36]

Discussion

The present study investigated the value of imaging, clinical, and other tests in diagnosing lumbar spinal stenosis. There is no indication suggesting that invasive imaging (myelography) is more accurate than noninvasive imaging. Three-sequence MRI appeared to be more sensitive than single-sequence MRI. In clinical tests, a high sensitivity is often accompanied by a low specificity and *vice versa*. Only one low-quality study, with a modeled combination of symptoms, reported both sensitivity and specificity above 75%. Selective lumbar root sheath infiltration shows a positive predictive value of 95% in one study; however, based on the invasiveness of the test and the low quality of that study, this test should not yet be recommended.

Because of the heterogeneity, the overall poor quality, and the small sample size of the studies, it is not possible to draw definite conclusions about which tests are best for diagnosing lumbar spinal stenosis.

Heterogeneity

In the investigated studies, there is much heterogeneity in study design, diagnostic test of interest, test characteristics, patient characteristics, reference standard, and definition of lumbar spinal stenosis. In many studies, it remains unclear what is meant by lumbar spinal stenosis because it is not further specified. In some articles, results are only concerned with central canal stenosis; whereas in others, central, lateral, and foraminal stenosis are taken together as one group.

Although we acknowledged the fact that it would introduce more variation in the findings, we did not want to include one single cause (degenerative or congenital or developmental) or only one location (central or lateral or foraminal) of stenosis in our review. This choice was made because in clinical practice one may be confronted with different locations and causes of stenosis in one and the same patient.

Quality

Besides the incomparability of the findings, the overall quality of the studies was poor, although the quality of the studies did increase slightly over time. The poor quality is due to defects in study design, resulting in biases like partial or differential verification bias. Partial verification bias occurs when people with positive index test results are more likely to receive the reference test. Differential verification bias occurs when people with a positive index test receive another, often more invasive reference test.[11] Sometimes the specificity cannot be calculated at all when all people with negative index tests do not receive the reference test and are therefore excluded from the study. It should be mentioned that sometimes it is not possible to avoid verification bias, *e.g.*, in case of an invasive reference standard such as surgery, or in case of retrospective studies. Verification bias and other shortcomings in design, data collection, and reporting (*e.g.*, evaluating tests in a diseased population and a control group, interpretation of the reference test with knowledge of the index test result) affect the estimates of diagnostic accuracy, mostly resulting in an overestimation.[37] In the review by Kent *et al* investigating the accuracy of CT, MRI, and myelography in the diagnosis of lumbar spinal stenosis, the estimates of sensitivity and specificity were also considered to be imprecise because they were based on small studies with potential biases.[9]

Sensitivity, specificity, and likelihood ratios show the capacity of a test to distinguish the diseased from the nondiseased. In many studies, specificity was not reported or could not be estimated, thus leaving us with only sensitivity. However, a high sensitivity by itself does not tell much about the diagnostic performance of a certain test because this test could also be positive in nondiseased people.

Because specificity based on patients could not always be calculated, some authors calculated the sensitivity and specificity based on the involved spinal levels or nerve roots. We did not consider this to be a valid estimation because a test must distinguish diseased patients from nondiseased patients.

Gold Standard

Another problem in diagnostic accuracy studies on lumbar spinal stenosis is the fact that there is no consensus about the gold standard. Determining the true sensitivity and specificity is hampered because the reference test is automatically considered to be better than the index test. Thus, in cases of a false-negative reference test, a positive index test result will wrongfully be counted as a false positive. The reported values therefore should be considered to be the sensitivity and specificity in diagnosing "stenosis according to the reference test." In our review, we decided that surgical confirmation and a combination of clinical and radiologic methods are (up to now) the best available reference standards. However, the surgical findings depend on positioning of the patient, and the clinical observation of the anatomy is a gray scale. To conclude that there is a stenosis or not is very difficult, except for lateral recess stenosis. Because one has to open the spinal canal by removing a part or the whole lamina, one cannot measure the diameter. Only in case of severe stenosis is there no doubt. Besides, when surgical confirmation is used as a reference standard, there are other problems that affect the design of the study, *e.g.*, problems with blinding or the fact that not every patient will undergo an operation, as can be seen by the poor scores items 5, 6, and 11 of the Quadas tool (Appendix 1).

The study on nerve root blocks,[36] described in "Other Tests" differs from the other studies in the way that they took a successful outcome of surgery as gold standard and thereby they were the only study who tried to link outcome of a diagnostic test with outcome of treatment. However, only the people with a positive test were offered surgery, and 95% of those who eventually went to surgery had a successful outcome. This figure is very high, possibly by the selection of people with this positive test. To get a selection of people where the positive predictive value is high, it is exactly what you are looking for to decide for surgical intervention. However, this figure alone is not very informative; we also should know how this compares with successful outcome of surgery without doing this specific diagnostic test or doing another diagnostic test. Further, nothing is known about outcome in people with a negative test. There are a lot of missed opportunities in this study, which also is reflected in its low-quality score.

Clinical and Radiologic Diagnoses

It should be considered what the diagnosis of lumbar spinal stenosis actually stands for. In diagnosing this condition, there is a tendency to focus on imaging studies. However, it has been reported that in about 30% of asymptomatic subjects lumbar spinal abnormalities can be seen on imaging studies.[38–40] Even higher percentages of spinal abnormalities can be seen in a study by Jensen *et al*.[41] In a study by Boden *et al*, lumbar spinal stenosis is seen on MRI in 21% of asymptomatic individuals aged 60 years or older.[39] Moreover, there are discrepancies between clinical symptoms and imaging findings in lumbar spinal stenosis.[7,42] Hence, there is a difference between the clinical and radiologic diagnosis of lumbar spinal stenosis. Therefore, clinical and radiologic findings should be considered together when diagnosing this disease.

Future Research

Further diagnostic research on lumbar spinal stenosis is essential. Because of the lack of an adequate reference standard, follow-up studies (preferably randomized) to quantify the effect of a diagnostic technology on patient outcome are needed. Evaluating a test on patient outcome comprises the evaluation of diagnostic tests plus all possible administered therapies combined.[43] In such studies, the problem of verification bias is eliminated, and the clinical decision tree from diagnostic test to effect of treatment is scientifically evaluated in its entirety. In this context, the most promising tests retrieved from this systematic review should be evaluated first.

Because of the poor quality and the incomparability of the investigated studies, we cannot draw firm conclusions about the diagnostic accuracy of imaging, clinical, and other tests in diagnosing lumbar spinal stenosis. Given this situation, at present the most promising tests are CT or MRI (preferably more than one sequence), avoiding myelography because of its invasiveness and lack of superior accuracy. These tests could be combined with clinical tests with high LR+ (>10) such as a combination of time to onset of symptoms and recovery time in treadmill walking, or wide-based gait during physical examination. However, also clinical tests with moderate $\text{LR}+^{5-10}$ can be very useful when the prior chance on presence of the disease in the study population is between about 20% and 80%[44]; this was the case for longer walking time during inclined treadmill walking, or symptoms improve when seating.

■ Key Points

- The overall quality of studies that investigate the diagnostic accuracy of tests used to detect lumbar spinal stenosis is poor.
- There is no indication that any of the imaging tests has a superior accuracy.
- In clinical tests, a high sensitivity is often accompanied by a low specificity and *vice versa*.
- More prospective studies are needed, which prevent the occurrence of verification bias.

tables

References

1. Epstein NE, Maldonado VC, Cusick JF. Symptomatic lumbar spinal stenosis. *Surg Neurol* 1998;50:3–10.
2. van Woerkom TCAM, Tavy DLJ. Aandoeningen van ruggemerg, zenuwwortels en cauda equina. In: Hijdra A, Koudstaal PJ, Roos RAC, eds. *Neurologie*. Maarssen: Elsevier Gezondheidszorg, 1998:429–32.
3. Hall S, Bartleson JD, Onofrio BM, et al. Lumbar spinal stenosis: clinical features, diagnostic procedures, and results of surgical treatment in 68 patients. *Ann Intern Med* 1985;103:271–5.
4. Alvarez JA, Hardy RH Jr. Lumbar spine stenosis: a common cause of back and leg pain. *Am Fam Physician* 1998;57:1825–40.
5. Postacchini F. The diagnosis of lumbar spinal stenosis: analysis of clinical and radiographic findings in 43 cases. *Ital J Orthop Traumatol* 1985;11:5–21.
6. Tan SB. Spinal canal stenosis. *Singapore Med J* 2003;44:168–9.
7. Amundsen T, Weber H, Lilleås F, et al. Lumbar spinal stenosis. Clinical and radiologic features. *Spine* 1995;20:1178–86.
8. Szpalski M, Gunzburg R. Lumbar spinal stenosis in the elderly: an overview. *Eur Spine J* 2003;12(suppl):170–5.
9. Kent DL, Haynor DR, Larson EB, et al. Diagnosis of lumbar spinal stenosis

in adults: a metaanalysis of the accuracy of CT, MR, and myelography. *AJR Am J Roentgenol* 1992;158:1135–44.

10. Devillé WL, Buntinx F, Bouter LM, et al. Conducting systematic reviews of diagnostic studies: didactic guidelines. BMC Medical Research Methodology [serial online]. 2002;2:2–9. Available at: http://www.biomedcentral.com/1471-22882/2/9. Accessed July 3, 2002.
11. Whiting P, Rutjes AWS, Reitsma JB, et al. The development of Quadas: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Medical Research Methodology [serial online] 2003;3:3–25. Available at: http://www.biomedcentral.com/1471-2288/3/25. Accessed November 10, 2003.
12. Laurencin CT, Lipson SJ, Senatus P, et al. The stenosis ratio: a new tool for the diagnosis of degenerative spinal stenosis. *Int J Surg Invest* 1999;1:127–31.
13. Ilkko E. Diagnosis of lumbar central spinal stenosis by plain radiography. *J Med Imag* 1989;3:91–101.
14. Rankine JJ, Hutchinson CE, Hughes DG. MRI of lumbar spondylosis: a comparison of sagittal T2 weighted and three sequence examinations. *Br J Radiol* 1997;70:1112–21.
15. Tervonen O, Koivukangas J. Transabdominal ultrasound measurement of the lumbar spinal canal: its value for evaluation of lumbar spinal stenosis. *Spine* 1989;14:232–5.
16. Eberhardt KE, Hollenbach HP, Huk WJ. 3D-MR Myelographie (3D-MRM) zur diagnose von lumbalen Nervenwurzelkompressionssyndromen. Vergleichsstudie zu konventioneller Myelographie. *Aktuelle Radiol* 1994;4:313–7.
17. Engel JM, Engel GM, Gunn DR. Ultrasound of the spine in focal stenosis and disc disease. *Spine* 1985;10:928–31.
18. Donmez T, Caner H, Cila A, et al. Diagnostic value of computed tomography in spinal and lateral recess stenosis, preoperatively and for long-term follow-up: a prospective study in 50 cases. *Radiat Med* 1990;8:111–5.
19. Freund M, Hutzelmann A, Steffens JC, et al. MR-Myelographie bei spinalkanalstenosen. *Rofo Fortschr Geb Rontgenstr Neuen Bildgeb Verfahr* 1997;167:474–8.
20. Modic MT, Masaryk T, Boumphrey F, et al. Lumbar herniated disk disease and canal stenosis: prospective evaluation by surface coil MR, CT and myelography. *AJR Am J Roentgenol* 1986;147:757–65.
21. Bischoff RJ, Rodriguez RP, Gupta K, et al. A comparison of computed tomography-myelography, magnetic resonance imaging, and myelography in the diagnosis of herniated nucleus pulposus and spinal stenosis. *J Spinal Disord* 1993;6:289–95.
22. Feldmeyer JJ, Schnyder P, Regli F, et al. Apport de la tomodensitométrie au diagnostic de la hernie discale et du canal étroit lombaires. *Schweiz Med Wochenshr* 1982;112:599–604.
23. Arrault I, Benoist M, Rocolle J, et al. Valuer diagnostique de la saccoradiculographie et du scanner dans les stenoses lombaires. *Rev Rhum Mal Osteoartic* 1987;54:627–31.
24. Bell GR, Rothman RH, Booth RE, et al. A study of computer-assisted tomography: II. Comparison of metrizamide myelography and computed tomography in the diagnosis of herniated lumbar disc and spinal stenosis. *Spine* 1984;9:552–6.
25. Bolender NF, Schonstrom NS, Spengler DM. Role of computed tomography and myelography in the diagnosis of central spinal stenosis. *J Bone Joint Surg Am* 1985;67:240–6.
26. Herkowitz HN, Wiesel SW, Booth RE Jr, et al. Metrizamide myelography and epidural venography: their role in the diagnosis of lumbar disc herniation and spinal stenosis. *Spine* 1982;7:55–64.
27. Jia LS, Shi ZR. MRI and myelography in the diagnosis of lumbar canal stenosis and disc herniation: a comparative study. *Chin Med J (Engl)* 1991;104:303–6.
28. Roach KE, Brown MD, Albin RD, et al. The sensitivity and specificity of pain response to activity and position in categorizing patients with low back pain. *Phys Ther* 1997;77:730–8.
29. Jensen OH, Schmidt-Olsen S. A new functional test in the diagnostic evaluation of neurogenic intermittent claudication. *Clin Rheumatol* 1989;8:363–7.
30. Katz JN, Dalgas M, Stucki G, et al. Degenerative lumbar spinal stenosis: diagnostic value of the history and physical examination. *Arthritis Rheum* 1995;38:1236–41.
31. Mann NH 3rd, Brown MD, Enger I. Statistical diagnosis of lumbar spine disorder using computerized patient pain drawings. *Comput Biol Med* 1991;21:383–97.
32. Mann NH 3rd, Brown MD, Enger I. Expert performance in low-back disorder recognition using patient pain drawings. *J Spinal Disord* 1992;5:254–9.
33. Mann NH 3rd, Brown MD, Hertz DB, et al. Initial-impression diagnosis using low-back pain patient pain drawings. *Spine* 1993;18:41–53.
34. Fritz JM, Erhard RE, Delitto A, et al. Preliminary results of the use of a two-stage treadmill test as a clinical diagnostic tool in the differential diagnosis of lumbar spinal stenosis. *J Spinal Disord* 1997;10:410–6.
35. Snowden ML, Haselkorn JK, Kraft GH, et al. Dermatomal somatosensory evoked potentials in the diagnosis of lumbosacral spinal stenosis: comparison with imaging studies. *Muscle Nerve* 1992;15:1036–44.
36. Castro WH, van Akkerveeken PF. Der diagnostische Wert der selektiven lumbalen Nervenwurzelblockade. *Z Orthop Ihre Grenzgeb* 1991;129:374–9.
37. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061–6.
38. Hitselberger WE, Witten RM. Abnormal myelograms in asymptomatic patients. *J Neurosurg* 1968;28:204–6.
39. Boden SD, Davis DO, Dina TS, et al. Abnormal magnetic-resonance scans of the lumbar spine in asymptomatic subjects: a prospective investigation. *J Bone Joint Surg Am* 1990;72:403–8.
40. Wiesel SW, Tsourmas N, Feffer HL, et al. A study of computer-assisted tomography: 1. The incidence of positive CAT scans in an asymptomatic group of patients. *Spine* 1984;9:549–51.
41. Jensen MC, Brant-Zawadski MN, Obuchowski N, et al. Magnetic resonance imaging of the lumbar spine in people without back pain. *N Engl J Med* 1994;331:69–73.
42. Beattie PF, Meyers SP, Stratford P, et al. Associations between patient report of symptoms and anatomic impairment visible on lumbar magnetic resonance imaging. *Spine* 2000;25:819–28.
43. Moons KGM, Grobbee DE. Diagnostic studies as multivariable, prediction research. *J Epidemiol Community Health* 2002;56:337–8.
44. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature: III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994;271:703–7.