Appendix

Supplementary Methodology Details

Subject recruitment: Reports on serial lumbar magnetic resonance imaging performed with the university scanner were screened to ensure that the person was between the ages of fifty-five and eighty years old and that he or she did not meet any exclusion criteria (previous surgery, tumor, etc.). The radiologists' reports were supplemented by a review of all images by a study physician to select persons with a "preliminary diagnosis of stenosis." The university's computerized medical records of persons with no apparent stenosis on magnetic resonance imaging were further reviewed to exclude persons with pain radiating below the knee. The subsequent group was labeled "preliminary diagnosis of no stenosis." All potential subjects were then screened by telephone to make sure that they did not meet exclusion criteria, including known polyneuropathy, diabetes, heavy alcohol use, previous lumbar surgery, or relative contraindications to magnetic resonance imaging or electrodiagnostic testing. Subjects who planned to have surgery were also excluded, as one of the project's long-term goals was to follow symptoms and signs of spinal stenosis over eighteen months.

Magnetic resonance imaging: All subjects underwent a non-contrast lumbosacral spinal magnetic resonance imaging scan (GE Signa Horizon LX; GE Medical Systems, Milwaukee, Wisconsin) including sagittal T2-weighted scans (field of view, 30 cm; scan thickness, 3.0 mm; interscan spacing, 0.5 mm; matrix, 384 × 192; repetition time, 3000 msec; echo time, 102 msec; pulse, fast spin echo), sagittal T1-weighted scans (field of view, 30 cm; scan thickness, 3 mm; interscan spacing, 0.5 mm; matrix, 256 × 192; repetition time, 400 to 700 msec; echo time, min full; pulse, spin echo), and axial T2-weighted scans (field of view, 20 cm; scan thickness, 4 mm; interscan spacing, 5 mm, five slices through each disc space from T12-L1 through L5-S1; matrix, 256 × 256; repetition time, 3000 to 5000 msec; echo time, 102 msec; pulse, fast spin echo).

Images were reviewed by a neuroradiologist, for whom clinical and electrodiagnostic findings were masked, at a workstation (Windows Advantage Workstation; GE Medical Systems). Numerous anatomic measurements were made with use of an electronic cursor. Previous work suggested that symptoms arise because vascular compromise occurs when the vasa nervorum are compressed at two levels[54]. Therefore, composite scores based on the average of the smallest two canal diameters, smallest two thecal sac diameters, etc., were developed. The test-retest reliability of the radiologists was excellent. With a mean of 23.6 ± 5.5 months between the measurements performed by the same rater on the initial magnetic resonance images of thirty-five subjects, all but one of the thirty measures were found to have intraclass correlation coefficient of ≥0.90 and all of the obtained results were significant at the $p < 0.01$ level.

Electrodiagnostic testing: With all of the above information masked, an electrodiagnostic medicine specialist performed a detailed electrodiagnostic study. The masking procedure included having the physiatrist who recorded the history and performed the physical examination instruct the subject, position the subject on the table, palpate markings for the paraspinal examination, and acclimatize the subject to the test by performing a very small exploration of a randomly chosen location in the paraspinals. An assistant was in the room at all times to remind the subject not to give any hints about his or her status. Because masking during electrodiagnostic tests had not previously been validated, an analysis was performed and it showed that <6% of electrodiagnostic tests had any potential for clinically relevant unmasking[43].

Electrodiagnostic testing was performed with a Nicolet Viking II electromyography machine (Nicolet Biomedical, Madison, Wisconsin) with use of a 50 to 75-mm monopolar needle. As recommended by Dillingham et al.[51], the testing included exploration of five muscles with overlapping root innervation. The muscles include the tensor fasciae latae (L4, L5, and S1 innervated), vastus medialis (L2, L3, and L4), tibialis anterior (L4 and L5), extensor hallucis longus (L5 and S1), and medial gastrocnemius (S1 and S2) in either the more symptomatic limb or, if symptoms were absent or symmetrical, in a limb chosen by the assistant ahead of time by a coin toss. In each muscle, spontaneous and insertional activities were scored as 0 to 4+ after six insertions in four different directions. Ten motor units per muscle were sampled. Informal measurements of typical motor-unit amplitude, number of polyphasic motor units, and motor unit recruitment (firing rate of the first motor unit when a second motor unit was recruited) were recorded. Sural sensory response and peroneal motor responses from the ankle, fibular head, and popliteal space were {tested}. Bilateral H waves and peroneal F waves (ten supramaximal stimulations per limb) were performed by a technician and interpreted by the electromyographer. Skin temperature was monitored, and heat was applied with use of a hydrocollator when necessary to keep skin temperature above 32°C.

Paraspinal mapping needle electrodiagnostic testing of the paraspinal muscles was performed bilaterally. In this study, the abbreviated paraspinal mapping technique (sometimes called “MiniPM”) was used rather than the extended original technique, because mathematical modeling has shown that its sensitivity and specificity are similar to those of the original method, with drastic reduction in the number of needle insertions. The paraspinal mapping technique is described in detail in a publication-quality course handout for the American Association of Neuromuscular and Electrodiagnostic Medicine, available at www.aanem.org, and a slightly older version is in text reference[38]. Briefly, the technique includes palpation of the inferior border of the three lowest lumbar spinous processes and the midpoint between the posterior superior iliac spines, measuring 2.5 cm lateral and (for the L3, L4, and L5 spinous processes) 1 cm cranial. From each of these four locations, a 50 to 75-mm monopolar electrodiagnostic testing needle is inserted at a 45° to 60° angle to the surface in three different directions—cranial 45°, directly across to the spinous process, and caudal 45°—and advanced through the muscle in 5-mm movements to detect abnormal muscle membrane instability. Any instability found (positive waves or fibrillations) must last more than one second and be reproducible. Scores for the medialmost 1 cm are calculated separately from those for the more lateral components of the insertion. Depending on the severity of the findings, scores ranged from 0 to 4 in any of twenty-four locations. A total score for the side (number of +’s) is determined. Paraspinal mapping was performed bilaterally, but only the data for the side on which limb electrodiagnostic testing was performed were used in the current paper.

Statistical methods: Data were initially entered in a Microsoft Excel database where errors were checked and corrected. SPSS version 12.0 was used for statistical analysis[62]. First, group differences in the various clinical measures were examined with use of either one-way analysis of variance (for three-group comparisons) or a t test (for two-group comparisons) for continuous measures and a chi-square test of independence for categorical measures. When analysis of variance was used, post hoc tests with use of the Tukey honestly significant difference as the criterion were conducted to determine how each of the three groups differed. The sensitivity and specificity of canal size were initially examined by determining the canal size values from the normative sample and calculating the cutoff as one standard deviation below the mean.

Discriminant function analyses were used to examine the ability of the canal size and electrodiagnostic testing measures to predict the diagnosis given by the physiatrist. First, the abilities of canal size, paraspinal mapping, and the presence or absence of an abnormal result of the limb examination to predict the group to which the patient belonged (stenosis versus back pain and stenosis versus asymptomatic) were assessed separately. Variables were entered directly. In addition, to determine whether the classifications obtained were significantly better than those expected by chance, the significance of the observed and expected hit rates was determined with use of the maximum chance criterion recommended by Huberty[63]. Prior probability of group membership in the discriminant analyses was set to be equal. This approach is conservative, as the classification scheme obtained is not biased toward assigning subjects to a group with a higher base rate when the initial group sizes are unequal. In addition, this approach is recommended when the base rates of a particular disorder in a given population are not well known[64]. Following these analyses, a stepwise discriminant analysis was performed to statistically determine the best combination of predictors. The criterion for entry was based on the Wilk method[64]. Significance was accepted at $p < 0.05$.

Additional statistical analyses: The ability of each of the measures to distinguish subjects with stenosis from those with back pain is presented in Table E-1. None of the variables alone significantly contributed to the prediction of group membership, although both electrodiagnostic measures were marginally significant ({chi square} (1) = 2.83, p = 0.09 for paraspinal mapping and chi square (1) = 3.05, p = 0.08 for the electrodiagnostic limb testing). The classifications derived from the measurements of canal size and the findings of the electrodiagnostic limb examination were not significantly better than those expected by chance. However, the classifications derived from the paraspinal mapping scores were significantly better than what would be expected from chance alone ($z = 1.85$, $p < 0.05$). Overall, the classification scheme derived from paraspinal mapping accurately grouped 63.2% of the patients. For this classification, a paraspinal mapping score of ≤4 was deemed normal, whereas a score of >4 was indicative of stenosis. This classification scheme had a sensitivity of 45.1% and a specificity of 84.1%.

The ability of the measures to discriminate subjects with stenosis from asymptomatic subjects is also presented in Table E-1. Canal size significantly contributed to the prediction of group membership (chi square (1) = 7.56, p = 0.006). However, the classifications were not significantly more accurate than those expected by chance. Paraspinal mapping scores also significantly contributed to the prediction of group membership (chi square (1) = 6.44, p = 0.011), but again the classifications were not significantly better than those expected by chance. Finally, the presence or absence of an abnormal finding on limb testing also significantly contributed to the prediction of group membership (chi square (1) = 9.56, p = 0.002) with the classifications not significantly better than those expected by chance.

Stepwise discriminant function analyses were conducted to determine if a combination of canal size and electrodiagnostic testing measures significantly predicted group membership. The criterion for entry was based on the Wilk method[64]. Separate analyses were conducted to compare the subjects who had clinical stenosis with those who had back pain, with asymptomatic subjects, and with the latter two groups combined. In the comparison of subjects who had clinical stenosis with the subjects who had back pain, none of the variables satisfied the entry criteria. In the comparison of subjects who had clinical stenosis with asymptomatic subjects, both canal size and the results of electrodiagnostic limb testing fulfilled the entry criteria. The resulting discriminant function significantly predicted group membership (chi square (1) = 18.50,

$p < 0.001$). Inspection of the standardized canonical discriminate function coefficients indicated that the electrodiagnostic limb testing had the greatest contribution to the prediction of group membership (0.796), followed by canal size (–0.725). The discriminant function accurately classified 65.9% of the original cases; specificity was 60.8%, while sensitivity was 74.2%. However, the classification scheme did not predict group membership better than that expected by chance alone.

In the comparison of subjects who had clinical stenosis with the combined group of asymptomatic persons and persons with back pain, again both the results of electrodiagnostic limb testing and the canal size fulfilled the entry criteria, and the resulting discriminant function significantly predicted group membership (chi square (1) = 14.23, $p = 0.002$). The standardized canonical discriminate function coefficients indicated that the limb examination again made the greatest contribution to the prediction of group membership (0.814), followed by canal size (–0.646). The discriminant function accurately classified 61.9% of the original cases; specificity was 49.0%, while sensitivity was 70.7%. However, the classification rate obtained was not significantly better than that expected by chance.

Table E1. Magnetic resonance imaging measurements in the different clinically defined patient populations. "sac" = thecal sac.

Magnetic Resonance Imaging Measurements	Clinical Spinal Stenosis n=51 Mean (SD)	Mechanical Low Back Pain n=44 Mean (SD)	Asymptomatic n=31 Mean (SD)	F Statistic	Significance (p value)
Minimum canal diameter[a]	13.13 (2.92)	13.83 (2.70)	15.05 (3.10)	4.261*	0.016
Smallest two canal diameters[b]	14.35 (2.70)	14.78 (2.33)	16.19 (2.71)	5.040**	0.008
Minimum sac diameter	8.39 (2.54)	8.43 (2.62)	9.52 (2.25)	2.301	0.104
Smallest two sac diameters	9.17 (2.43)	9.37 (2.35)	10.25 (2.18)	2.159	0.120
Minimum canal area	269.53 (65.86)	272.77 (55.89)	278.10 (69.11)	0.176	0.839
Smallest two canal areas	286.93 (62.70)	287.17 (47.67)	301.37 (53.92)	0.774	0.463
Minimum sac area	96.18 (50.97)	96.59 (46.35)	117.00 (38.49)	2.295	0.105
Smallest two sac areas	109.98 (48.42)	109.85 (44.07)	131.26 (37.35)	2.698	0.071
Minimum lateral recess, symptomatic side	6.13 (1.10)	6.19 (1.14)	5.92 (1.16)	0.554	0.576
Smallest two lateral recesses, symptomatic side	6.55 (1.01)	6.48 (1.02)	6.31 (0.86)	0.600	0.550
Minimum interfacet distance from the osseous margin	15.77 (2.34)	15.40 (1.75)	16.10 (2.80)	0.875	0.420
Smallest two interfacet distances from the osseous margin	16.45 (2.24)	16.06 (1.68)	16.78 (2.57)	1.038	0.357
Minimum interfacet distance from the inner ligamentous margin	6.65 (2.88)	6.23 (2.53)	7.14 (3.03)	0.958	0.386
Smallest two interfacet distances from the inner ligamentous margin	7.41 (2.86)	6.97 (2.38)	7.83 (2.83)	0.952	0.389

* $p < 0.05$.

* *$p < 0.01$.

[a]The mean difference between stenosis and asymptomatic, −1.92, is significant at the 0.05 level.

[b]The mean difference between stenosis and asymptomatic, −1.84, is significant at the 0.05 level.