SPINE Volume 35, Number 21, pp 1919–1924
©2010, Lippincott Williams & Wilkins

Qualitative Grading of Severity of Lumbar Spinal Stenosis Based on the Morphology of the Dural Sac on Magnetic Resonance Images

Constantin Schizas, MD, PD, FRCS,* Nicolas Theumann, MD, PD,† Alexandre Burn, MD,* Rosamond Tansey, MD,* Douglas Wardlaw, FRCS,‡ Francis W. Smith, MD,§ and Gerit Kulik, PhD*

Study Design. Retrospective radiologic study on a prospective patient cohort.

Objective. To devise a qualitative grading of lumbar spinal stenosis (LSS), study its reliability and clinical relevance.

Summary of Background Data. Radiologic stenosis is assessed commonly by measuring dural sac cross-sectional area (DSCA). Great variation is observed though in surfaces recorded between symptomatic and asymptomatic individuals.

Methods. We describe a 7-grade classification based on the morphology of the dural sac as observed on T2 axial magnetic resonance images based on the rootlet/cerebrospinal fluid ratio. Grades A and B show cerebrospinal fluid presence while grades C and D show none at all. The grading was applied to magnetic resonance images of 95 subjects divided in 3 groups as follows: 37 symptomatic LSS surgically treated patients; 31 symptomatic LSS conservatively treated patients (average follow-up, 2.5 and 3.1 years); and 27 low back pain (LBP) sufferers. DSCA was also digitally measured. We studied intra- and interobserver reliability, distribution of grades, relation between morphologic grading and DSCA, as well relation between grades, DSCA, and Oswestry Disability Index.

Results. Average intra- and interobserver agreement was substantial and moderate, respectively (k = 0.65 and 0.44), whereas they were substantial for physicians working in the study originating unit. Surgical patients had the smallest DSCA. A larger proportion of C and D grades was observed in the surgical group. Surface measurements resulted in overdiagnosis of stenosis in 35 patients and under diagnosis in 12. No relation could be found between stenosis grade or DSCA and baseline Oswestry Disability Index or surgical result. C and D grade patients were more likely to fail conservative treatment, whereas grades A and B were less likely to warrant surgery.

Conclusion. The grading defines stenosis in different subjects than surface measurements alone. Since it mainly considers impingement of neural tissue it might be a more appropriate clinical and research tool as well as carrying a prognostic value.

Key words: lumbar spinal stenosis, magnetic resonance imaging, spine decompression, conservative treatment, classification. **Spine 2010;35:1919–1924**

Lumbar spinal stenosis (LSS) is diagnosed in an ever increasing number of patients referred to spinal surgeons for treatment as a result of the availability of magnetic resonance imaging (MRI) and ageing of the population. Radiologic LSS might not always be the cause of the presenting symptoms since it can to a certain degree be present in asymptomatic subjects.[1,2] Several parameters have been proposed in defining spinal stenosis, the prevailing being measurements of spinal canal or dural sac cross-sectional surface area (DSCA) in either computed tomography (CT) or axial MRI sequences taken at disc level. Surfaces measuring less than 100 mm^2 or 75 mm^2 represent respectively relative and absolute stenosis.[3] A significant variation in surface measurements is found though with overlapping of symptomatic and asymptomatic patient numerical values.[4] In an effort to counter the above limitations other parameters have been developed such as the stenosis ratio (SR).[5] In addition to showing poor correlation with clinical symptoms,[1,6,7] the above-mentioned quantitative parameters require accurate measurement using tools not always available in everyday clinical practice.

In an effort to improve the currently available radiologic LSS criteria, we designed a qualitative grading system based on the morphologic appearance of the dural sac as seen on T2-weighted axial images of the lumbar spine, taking into account the cerebrospinal fluid (CSF)/rootlet content.

Materials and Methods

This was a retrospective radiologic study coupled with data from a prospective LSS patient database. Institutional review board approval was granted for this study.

From the Departments of *Orthopedics and †Radiology, Centre Hospitalier Universitaire Vaudois (CHUV), University of Lausanne, Lausanne, Switzerland; ‡NHS Grampian, Orthopaedic Unit, Woodend Hospital, Aberdeen, Untied Kingdom; and §Centre for Spinal Research, Positional MRI Centre, Woodend Hospital, Aberdeen, United Kingdom.
Acknowledgment date: September 21, 2009. Revision date: December 19, 2010. Acceptance date: January 7, 2010.
The manuscript submitted does not contain information about medical device(s)/drug(s).
No funds were received in support of this work. Although one or more of the author(s) has/have received benefits for personal or professional use from a commercial party related directly or indirectly the subject of the manuscript, benefits will be directed solely to a research fund, foundation, educational institution, or other non-profit organization which the author(s) has/have been associated.
Supported by research funds (unrestricted grant) from DePuy Spine Switzerland (to the Department of Orthopedics).
Local Ethics committee approval granted (IRB).
Address correspondence and reprint requests to Constantin Schizas, MD, PD, FRCS, Orthopedic Department, Centre Hospitalier Universitaire Vaudois (CHUV), University of Lausanne, Avenue Pierre-Decker 4, CH-1011 Lausanne, Switzerland; E-mail: cschizas@hotmail.com

A total of 95 patients were included in this study. They were comprised of 3 groups as follows: surgically treated LSS patients (n = 37 patients), conservatively treated LSS patients (n = 31 patients), and low back pain sufferers (LBP group, n = 27 patients).

All surgical and conservative group patients (n = 68) were consecutive cases referred to our clinic with neurologic claudication and MRI studies available in the Pictures Archiving and Communication System of our institution demonstrating varying degrees of spinal stenosis. All surgical group patients failed at least 6 months of conservative treatment, including oral analgesia, physiotherapy, and epidural steroids. Conservative group patients were treated using a mixture of the above-mentioned methods. None of the conservative group patients required surgery during at least a 12-month period. Duration of symptoms before referral in our clinic was of 40.5 (standard deviation [SD], 41.5) months in the surgical group and 18.3 (SD, 27.8) months for the conservative group. Average follow-ups of the conservative and surgical groups were 3.1 years (SD, 1.4) and 2.5 years (SD, 0.8), respectively. LBP group patients were consecutive cases referred for spinal MRI in our institution and suffering from LBP either nonspecific, due to degenerative disc disease or to an inflammatory spondyloarthropathy. Suspected spinal stenosis cases were excluded from this group with none of those reporting radicular symptoms or neurologic claudication. The LBP group patients were not included in the clinical outcome part of the study, and therefore no follow-up was performed.

Average age was 71.5 years (SD, 9.3) in the surgical group, 68.4 years (SD, 12.8) in the conservative group and 51.5 years (SD, 17.4) in the LBP group of patients. Thirty-six were males and 59 were females. The female/male ratio was 0.94, 3.4, and 1.7 in the surgical, conservative, and LBP groups, respectively.

The proportion of patients presenting degenerative spondylolisthesis at any level was not significantly different between the surgical and conservative groups, although the surgical group had more subjects with L4–L5 spondylolisthesis (19 in the surgical group and 7 in the conservative group). No LBP patient presented a spondylolisthesis.

Surgery consisted of decompression in all cases, associated with instrumented fusion in 24 patients mainly due to the presence of degenerative spondylolisthesis.

Sagittal T1-weighted spin-echo, sagittal T2-weighted fast spin-echo, and axial T2-weighted fast spin-echo lumbar spine images were acquired by using magnets operating at field strength of 3.0 T, with a 3.0-mm section thickness and a 0.3-mm intersection gap. The field of view was 31 × 31 cm for sagittal images and 18 × 15 cm for axial images. The images were collected electronically and stored directly as Digital Imaging and Communications in Medicine files. All available MRI studies were accepted irrespective of the quality of the obtained images.

The grading is based on the CSF/rootlet ratio as seen axial T2 images and was conceived following observation of the different patterns according which the rootlets were disposed within the dural sac while the patient rested supine during MRI acquisition. Description of the grading is as follows (Figure 1):

Grade A stenosis: there is clearly CSF visible inside the dural sac, but its distribution is inhomogeneous:

A1: the rootlets lie dorsally and occupy less than half of the dural sac area.

A2: the rootlets lie dorsally, in contact with the dura but in a horseshoe configuration.

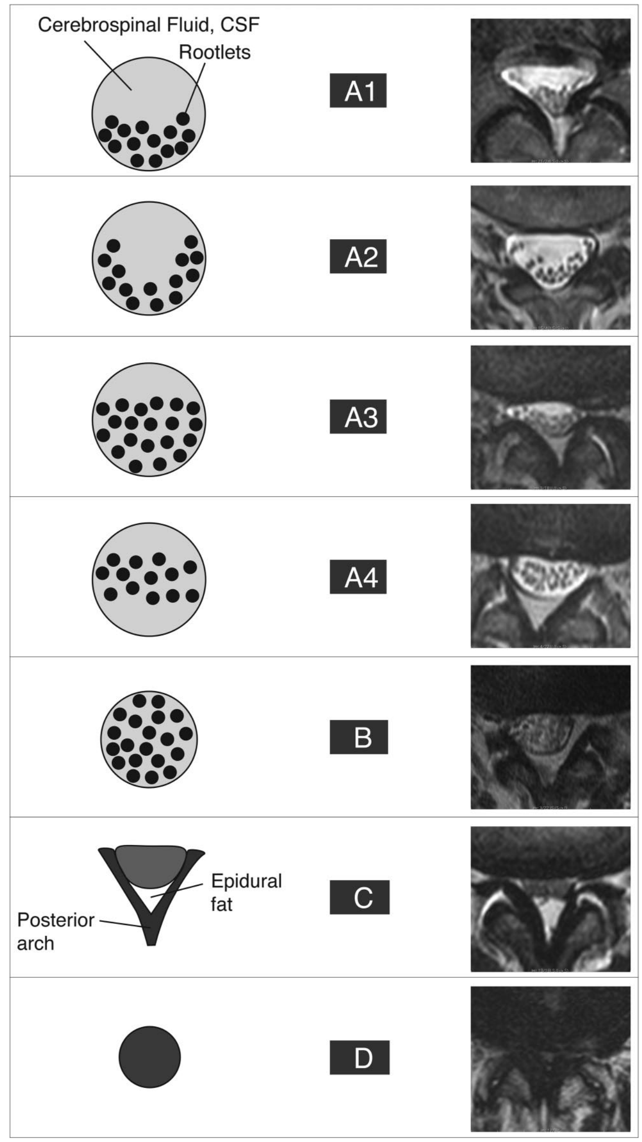

Figure 1. Description of the morphologic classification of spinal stenosis combining graphic and MRI examples.

A3: the rootlets lie dorsally and occupy more than half of the dural sac area.

A4: the rootlets lie centrally and occupy the majority of the dural sac area.

Grade B stenosis: the rootlets occupy the whole of the dural sac, but they can still be individualized. Some CSF is still present giving a grainy appearance to the sac.

Grade C stenosis: no rootlets can be recognized, the dural sac demonstrating a homogeneous gray signal with no CSF signal visible. There is epidural fat present posteriorly.

Grade D stenosis: in addition to no rootlets being recognizable there is no epidural fat posteriorly.

We defined grade A as no or minor stenosis, B as moderate stenosis, C as severe stenosis, and D as extreme stenosis.

The following parameters were also determined using OSIRIX imaging software: anteroposterior (AP) and transverse diameter of the dural sac measured at disc and pedicle levels for all levels through which axial acquisitions were available, surface area in the above-mentioned levels (measured surface) and surface area using the technique described by Hamanishi (calculated surface). The latter is calculated using the product of AP and transverse diameters of the dural sac multiplied by a ratio depending on its form.[8] Those measurements were performed twice within a 4-week interval by a senior spinal surgeon and senior radiologist in order to assess their intra- and interobserver reliability.

Intra- and Interobserver Reliability Assessment of the Morphologic Grading

A total of 57 axial disc level T2 images from the surgically treated group were randomly selected. Patient identification and disc level were blinded. The images were analyzed by a senior radiologist, senior spinal surgeon, and 2 orthopedic junior physicians working in the study originating institution. Orthopedic department trainees received oral tuition on the classification, supported by images encountered in clinical practice. Additionally those images were submitted to a senior spinal surgeon and radiologist practicing in a different country. The latter group received as sole training the above-mentioned description without oral briefing. Images were resubmitted 3 weeks later for repeat scoring following random rearrangement. Oswestry Disability Index (ODI)[9] was retrieved from a prospective LSS database at baseline for both surgical and conservative groups and at 12 months for the surgical group. Global outcome of surgery at 12 months was also assessed.[10]

Statistical Analysis

Intra- and interobserver reliability were assessed using weighted Cohen, kappa statistics for stenosis grades, and *t* test for surface and diameter measurements. Fisher exact test, linear regression, and odds ratio (OR) were used as appropriate.

■ Results

Inter- and Intraobserver Reliability

Stenosis morphologic grading: average inter- and intraobserver kappas were 0.44 (SD, 0.17) and 0.65 (SD, 0.14), respectively. For observers from the study originating unit those values were 0.67 (SD, 0.08) and 0.77 (SD, 0.06), while also demonstrating a learning effect between readings.

Numerical Values

Significant discrepancies were encountered in inter- and intraobserver DSCA direct measurements taken at pedicle level ($P = 0.015$ and 0.023, respectively) as well as interobserver calculated DSCA at disc level using the Hamanishi technique ($P < 0.001$). For the measured DSCA, the average intra- and interobserver variations were 6 and 9 mm^2, respectively.

Stenosis Grade distribution

At the L4–L5 disc level, a larger proportion of C and D grade stenosis were observed in the surgical group compared to both the conservative and LBP groups ($P < 0.01$ and $P < 0.001$, respectively). At all other levels, the distribution of stenosis grade was similar between surgical and conservative groups but significantly different in the LBP group ($P < 0.05$). There were nevertheless more A3 and A4 grades at the L3 and L4 pedicle level in the surgical group ($P < 0.05$).

The distribution of the stenosis grade among the 3 patient groups when the most stenotic level was taken into account can be seen in Table 1. No grade D was found in the LBP group, whereas grades A and B were the commonest in the conservative group and C and D in the surgical group. There was no significant difference in the number of patients who presented with multilevel stenosis (defined either according to morphologic grade or surface measurement) between the surgical and conservative groups.

Table 1. Distribution of Narrowest Grades per Patient per Group

	Grade A	Grade B	Grade C	Grade D
LBP group	23	1	3	0
Conservative group	16	6	6	3
Surgical group	1	3	20	13

LBP indicates low back pain.

Relation Between Surface Measurements and Morphologic Stenosis Grading

Even though some relation between morphologic grading and DSCA obtained using the direct measurement technique was observed, a significant overlap of grades for a given surface measurement could be observed (Figure 2). In particular, grade C stenosis (*i.e.*, severe) could be observed in 12 patients with levels measuring more than 100 mm^2. By contrast, 40 levels in 35 patients would be over diagnosed as stenotic if surface measurements alone were considered. An example is shown in Figure 3.

The narrowest level ranged from 48.7 mm^2 to 269.0 mm^2 in the LBP group, 18.5 mm^2 to 161.1 mm^2 in the conservative group, and 11.7 mm^2 to 94.3 mm^2 in the surgical group.

Even though the same proportion of patients in the conservative group exhibited at their narrowest level either a DSCA <75 mm (n = 9) or a grade C or D (n = 9), those criteria applied to different patients. More specifically, only 6 cases less than 75 mm^2 were C or D grades, whereas the 3 additional patients with C and D grades measured more than 75 mm.

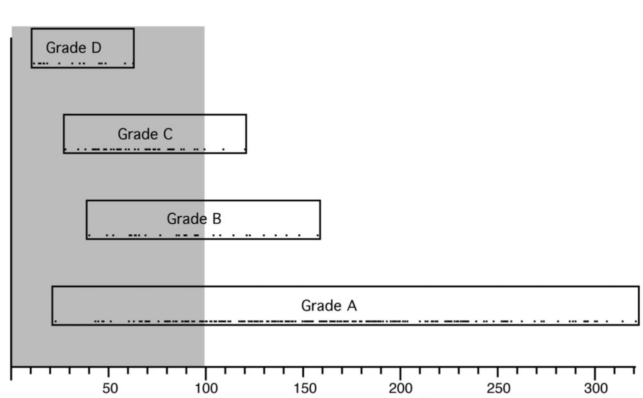

Figure 2. Relation between morphologic grading and dural sac surface area.

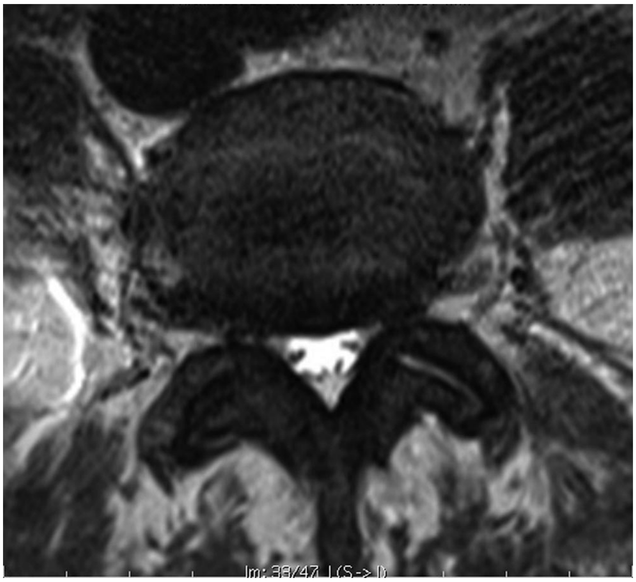

Figure 3. Axial T2 MRI image at L4–L5 level with dural sac surface area measuring 66 mm^2, but of grade A morphology without rootlet impingement within the dural sac.

Stenosis grades C and D or DSCA <75 mm^2 were both linked to an increased risk of failure of conservative treatment and need for surgery (OR, 29.8) even though as stated above those criteria did not apply to the same individuals. Other predictors of failure were spondylolisthesis (OR, 3.3), DSCA at pedicle level <160 mm^2 (OR, 3.4) and Laurencin SR <0.5 (OR, 4.5).

In the LBP group, asymptomatic stenosis was present in 7 cases when morphologic grade was considered (grades B, and C and D) but in 14 cases when surface alone was taken into account (<100 mm). Those values were respectively 4 and 11 if the definition was changed to C and D for grades and <75 mm for surface measurements.

In 9 patients, myelographic sequences were available. In 8 cases, myelographic blocs were observed corresponding to grade C and D stenosis (Figure 4).

An accessory finding of our study was that the 3 groups of patients were different as far as canal dimensions were concerned measured at pedicle (Figure 5) and

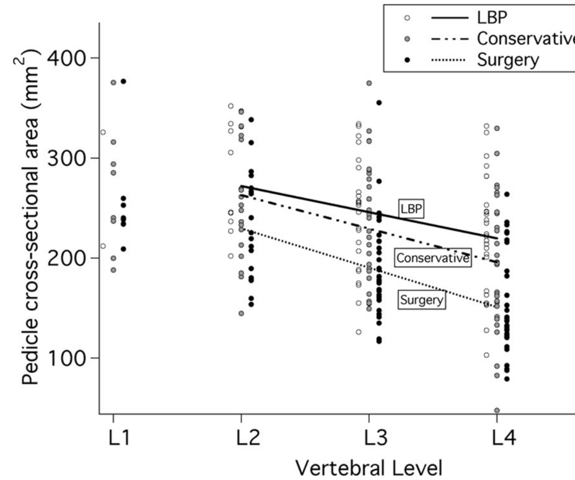

Figure 5. DSCA measured at pedicle level for all 3 groups, demonstrating the smaller surfaces of the surgical group.

disc level with surgical patients having smaller canals compared to the 2 other groups ($P < 0.0001$). Laurencin stenotic ratio was smaller in both the surgical and conservative group compared to the LBP group for all lumbar levels apart from the L4 to L5 level. The latter was even smaller in the surgical group ($P < 0.05$).

Clinical Result and Stenosis

Baseline ODI was similar in both surgical and conservative groups ($P > 0.5$). No correlation was found between baseline ODI and morphologic grade of stenosis or surface measurements (of the more stenotic level for each patient) in either the surgical or the conservative group.

Average ODI improved from 49% to 29% (SD, 19) after surgery in the surgical group. Global outcome was rated as "good" (operation helped or helped a lot) by 82% of surgical patients. No relation was found between ODI improvement and morphologic grade in the surgical group.

Discussion

Kappa Values Appraisal

The overall intraobserver agreement for the proposed morphologic grading was substantial including for ob-

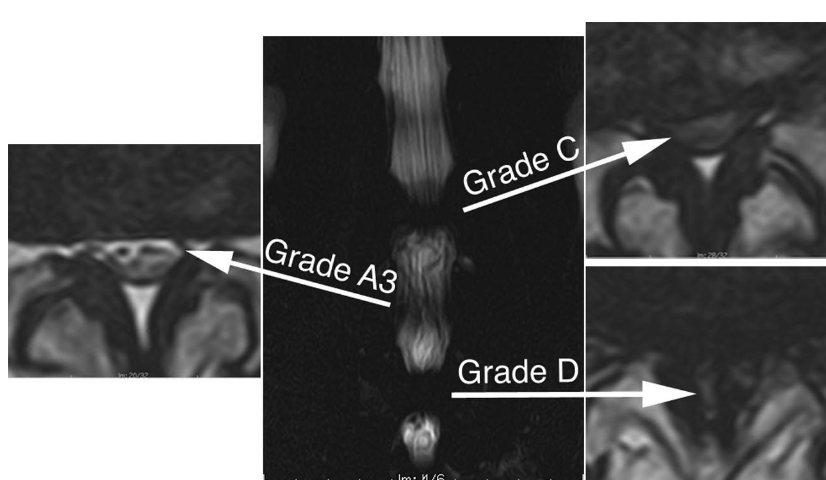

Figure 4. Myelographic sequence demonstrating myelographic bloc at 2 levels (grades C and D) associated with a grade A morphology.

servers working in different countries. Interobserver agreement was moderate, but it included a group that did not receive specific briefing on the present grading system. A similar effect was described by Lurie *et al* who studied reliability of MRI readings in a subpopulation of the Spine Patient Outcome Research Trial (SPORT) dealing with spinal stenosis.[11] Their qualitative assessment comprised only 3 grades (mild, moderate, and severe stenosis), while investigators received formal tuition. Despite the presence of only 3 grades as compared to 4 in our study, their kappa value was similar to the one observed within our unit. Other researchers found even poorer agreement (kappa = 0.26) when attempting to quantify severity of stenosis based on CT images.[12] The worst kappa values in our study compare favorably with that of other commonly used classifications in spinal surgery. Agreement in using the AO classification of spinal fractures was found to be moderate (kappa = 0.475),[13] whereas the Thoracolumbar Injury Classification and Severity Score showed an initial poor agreement (kappa = 0.189) improving over time (kappa = 0.509, 7 months later).[14] Interestingly, the load-sharing classification showed in 1 study to have a near perfect agreement and constitutes probably the only classification that reaches so high kappa values in spinal surgery.[15]

Relation Between Imaging and Clinical Symptoms

Only one previous study showed clear correlation between ODI and degree of stenosis, although the latter was assessed on myelograms measuring the AP canal diameter and correcting for radiologic magnification.[16] Amudsen *et al* found no relation between degree of stenosis (measured on myelography and CT) and clinical symptoms in 100 patients selected from a neurology department on the basis of clinical symptoms of LSS.[1] Lohman *et al* found no relation between cross sectional area of the canal measured on CT and clinical symptoms.[6] Sirvanci *et al* looked at the correlation between imaging and ODI[7] in 63 surgical candidates with LSS. They studied cross sectional area as well as subjective criteria of lateral recess and foraminal stenosis on axial MRI images but found no correlation between those parameters and ODI. Finally, Jonsson *et al* observed a nonsignificant tendency toward more walking disturbances in LSS patients with more pronounced constriction seen on myelography.[17]

None of the above-mentioned studies looked at morphologic characteristics and were limited in trying to establish a relation between measured parameters and symptoms or functional status.

The fact that we were unable to identify a relationship between severity of stenosis and clinical result could also be due to the fact that the majority of our surgical patients had a grade C or D stenosis. Furthermore, attempting to relate to severity of symptoms requires probably different outcome tools than the commonly used ODI or VAS. Other disease-specific tools such as the Zurich Claudication Questionnaire,[18] which has been shown to be more precise in LSS[19] might be more relevant.

This study is the first to propose a classification of spinal stenosis based on morphology and apply it to a clinical setting. If we consider complete absence of CSF as a sign of severe stenosis (such as seen in grades C and D) it becomes apparent that a morphologic grading would be more appropriate than simple area measurements. While it is true that C and D grades correspond to the narrower end of the surfaces measured, absence of compression (A grades) correspond to a much greater range of surfaces. Surface measurements alone fail to take into account degree of entrapment of neural tissue within the spinal canal and this fact could render a morphologic classification a more pertinent tool in clinical and research settings.

The MRI acquisitions followed the same protocol, but the orientation of the slices might differ from patient to patient. This nevertheless reflects the variability of MRI quality in everyday clinical practice as witnessed in larger scale studies.[11] Furthermore, this variability might have a larger influence on quantitative measures rather than qualitative ones such as the presence or absence of CSF.

In this study, we made no distinction between central and lateral stenosis due to the fact that the 2 are closely linked in the genesis of symptoms in patients with marked degenerative changes[20] and often coexist.[21]

In studying relation between surface and grading, we used the measured DSCA figures even though at pedicle level an average inter- and intraobserver variation of 9 and 6 mm^2 was observed. This variation was nevertheless up to 23 times smaller than the DSCA variation in a single grade.

Surgical candidates were selected by a single surgeon based on severity of symptoms, although this was not necessarily reflected by their ODI, the latter being probably not the best tool in assessing LSS patients with neurologic claudication. The selection was also based on the degree of stenosis, which was intuitively performed before the establishment of the grading system presented in this study. The main author did nevertheless consider patients to have more severe stenosis if no CSF could be identified in axial images and therefore selected mainly C and D grade patients. This selection proved useful since no patient from the conservative group had symptoms severe enough to warrant surgery at an average follow-up of 3.1 years. Most grade A and B patients with claudication did not warrant surgery for several years, whereas C and D grade patients had a significant risk of failing conservative measures giving this grading a prognostic value. Applying the present grading to other surgical series might give more information as to the relation between degree of stenosis and surgical outcome.

■ Conclusion

The present classification system taking into account the CSF content of the canal has the advantage of being easily applied in everyday clinical practice without the use of specific measuring tools while offering moderate to substantial agreement between observers on its initial implementation. It appears to reflect better the definition of stenosis than numerical values of surface measurements alone since it takes into account the degree of impingement of the contents of the dural sac. Further validation of the present grading by its use in other patient groups is necessary in order to judge on its transferability to different settings.

■ Key Points

- Grading stenosis based on morphology rather than surface measurements defines stenosis in different subjects.
- Patients with Grades C and D stenosis are more likely to fail conservative treatment.

References

1. Amundsen T, Weber H, Lilleas F, et al. Lumbar spinal stenosis. Clinical and radiologic features. *Spine* 1995;20:1178–86.
2. Boden SD, McCowin PR, Davis DO, et al. Abnormal magnetic-resonance scans of the cervical spine in asymptomatic subjects. A prospective investigation. *J Bone Joint Surg Am* 1990;72:1178–84.
3. Schonstrom N, Lindahl S, Willen J, et al. Dynamic changes in the dimensions of the lumbar spinal canal: an experimental study in vitro. *J Orthop Res* 1989;7:115–21.
4. Penning L, Wilmink JT. Posture-dependent bilateral compression of L4 or L5 nerve roots in facet hypertrophy. A dynamic CT-myelographic study. *Spine* 1987;12:488–500.
5. Laurencin CT, Lipson SJ, Senatus P, et al. The stenosis ratio: a new tool for the diagnosis of degenerative spinal stenosis. *Int J Surg Investig* 1999; 1:127–31.
6. Lohman CM, Tallroth K, Kettunen JA, et al. Comparison of radiologic signs and clinical symptoms of spinal stenosis. *Spine* 2006;31:1834–40.
7. Sirvanci M, Bhatia M, Ganiyusufoglu KA, et al. Degenerative lumbar spinal stenosis: correlation with Oswestry Disability Index and MR imaging. *Eur Spine J* 2008;17:679–85.
8. Hamanishi C, Matukura N, Fujita M, et al. Cross-sectional area of the stenotic lumbar dural tube measured from the transverse views of magnetic resonance imaging. *J Spinal Disord* 1994;7:388–93.
9. Fairbank JC, Couper J, Davies JB, et al. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980;66:271–3.
10. Grob D, Mannion AF. The patient's perspective on complications after spine surgery. *Eur Spine J* 2009;18(suppl 3):380–5.
11. Lurie JD, Tosteson AN, Tosteson TD, et al. Reliability of readings of magnetic resonance imaging features of lumbar spinal stenosis. *Spine* 2008;33: 1605–10.
12. Drew R, Bhandari M, Kulkarni AV, et al. Reliability in grading the severity of lumbar spinal stenosis. *J Spinal Disord* 2000;13:253–8.
13. Wood KB, Khanna G, Vaccaro AR, et al. Assessment of two thoracolumbar fracture classification systems as used by multiple surgeons. *J Bone Joint Surg Am* 2005;87:1423–9.
14. Patel AA, Vaccaro AR, Albert TJ, et al. The adoption of a new classification system: time-dependent variation in interobserver reliability of the thoracolumbar injury severity score classification system. *Spine* 2007;32:E105–10.
15. Dai LY, Jin WJ. Interobserver and intraobserver reliability in the load sharing classification of the assessment of thoracolumbar burst fractures. *Spine* 2005;30:354–8.
16. Hurri H, Slatis P, Soini J, et al. Lumbar spinal stenosis: assessment of long-term outcome 12 years after operative and conservative treatment. *J Spinal Disord* 1998;11:110–5.
17. Jonsson B, Annertz M, Sjoberg C, et al. A prospective and consecutive study of surgically treated lumbar spinal stenosis. Part I: clinical features related to radiographic findings. *Spine* 1997;22:2932–7.
18. Stucki G, Daltroy L, Liang MH, et al. Measurement properties of a self-administered outcome measure in lumbar spinal stenosis. *Spine* 1996;21: 796–803.
19. Pratt RK, Fairbank JC, Virr A. The reliability of the Shuttle Walking Test, the Swiss Spinal Stenosis Questionnaire, the Oxford Spinal Stenosis Score, and the Oswestry Disability Index in the assessment of patients with lumbar spinal stenosis. *Spine* 2002;27:84–91.
20. Szpalski M, Gunzburg R. Lumbar spinal stenosis in the elderly: an overview. *Eur Spine J* 2003;12(suppl 2):S170–5.
21. Lee CK, Rauschning W, Glenn W. Lateral lumbar spinal canal stenosis: classification, pathologic anatomy and surgical decompression. *Spine* 1988; 13:313–20.