Christian W. A. Pfirrmann, MD
Claudio Dora, MD
Marius R. Schmid, MD
Marco Zanetti, MD
Juerg Hodler, MD
Norbert Boos, MD

MR Image–based Grading of Lumbar Nerve Root Compromise due to Disk Herniation: Reliability Study with Surgical Correlation[1]

Index terms:
Nerves, roots, 339.499
Nerves, spinal
Spine, intervertebral disks, 336.783
Spine, MR, 33.12141

Published online before print
10.1148/radiol.2302021289
Radiology 2004; 230:583–588

Abbreviation:
SE = spin echo

[1] From the Departments of Radiology (C.W.A.P., M.R.S., M.Z., J.H.) and Spinal Surgery (C.D., N.B.), Orthopedic University Hospital Balgrist, Forchstrasse 340, CH-8008 Zurich, Switzerland. Supported in part by a grant from the Swiss National Science Foundation (32–52927.97). Received October 9, 2002; revision requested December 17; final revision received May 28, 2003; accepted June 25. **Address correspondence to** C.W.A.P. (e-mail: *christian@pfirrmann.ch*).

Author contributions:
Guarantor of integrity of entire study, C.W.A.P.; study concepts, C.W.A.P., N.B.; study design, N.B.; literature research, C.W.A.P.; clinical studies, C.D., M.R.S.; data acquisition, C.D., M.R.S.; data analysis/interpretation, C.W.A.P., C.D.; statistical analysis, C.W.A.P.; manuscript preparation, C.W.A.P.; manuscript definition of intellectual content, C.W.A.P., N.B., J.H.; manuscript editing, J.H., M.Z.; manuscript revision/review, N.B., J.H., M.Z.; manuscript final version approval, M.Z.

A system for grading lumbar nerve root compromise (no compromise, contact of disk material with nerve root, deviation of nerve root, and compression of nerve root) was tested in the interpretation of routine magnetic resonance images of 500 lumbar nerve roots in 250 symptomatic patients. Intra- and interobserver reliability was assessed for three independent observers. In the 94 nerve roots evaluated at surgery, surgical grading was correlated with image-based grading. κ statistics indicated substantial agreement between different readings by the same observer and between different observers (for intraobserver agreement, κ = 0.72–0.77; for interobserver agreement, κ = 0.62–0.67). Correlation of image-based grading with surgical grading was high (r = 0.86). The image-based grading system enabled reliable evaluation and reporting of nerve root compromise.

Disk herniations of the same size may be asymptomatic in one patient and lead to severe nerve root compromise in another patient (1). Magnetic resonance (MR) imaging reports usually focus on the morphology, location, and size of the herniated disk (2–4). The effect of disk herniation depends on the location and extent of the herniation relative to the diameter of the spinal canal (5,6). Therefore, a clinically relevant grading system for disk herniation must be based on the spatial relationship between herniated disk material and neural structures. Although this assumption seems obvious to both clinicians and radiologists, it is not yet reflected in radiologic classification systems of disk herniation (7). We believe that a standardized nomenclature in the assessment of abnormalities of the lumbar spine is a prerequisite for a comparison of data from different investigations (8). The reliability of the assessment has a crucial influence on the validity of the data.

The objective of this investigation was to describe a system for grading lumbar nerve root compromise depicted on routine MR images, to evaluate its reliability, and to correlate image-based grades with surgical grades.

Materials and Methods

Grading System

The system we used in grading compromise of the intraspinal extradural lumbar nerve root consists of four grade categories, summarized as follows. Grade 0 (normal): No compromise of the nerve root is seen. There is no evident contact of disk material with the nerve root, and the epidural fat layer between the nerve root and the disk material is preserved (Fig 1). Grade 1 (contact): There is visible contact of disk material with the nerve root, and the normal epidural fat layer between the two is not evident. The nerve root has a normal position, and there is no dorsal deviation (Fig 2). Grade 2 (deviation): The nerve root is displaced dorsally by disk material (Fig 3). Grade 3 (compression): The nerve root is compressed between disk material and the wall of the spinal canal; it may appear flattened or be indistinguishable from disk material (Fig 4).

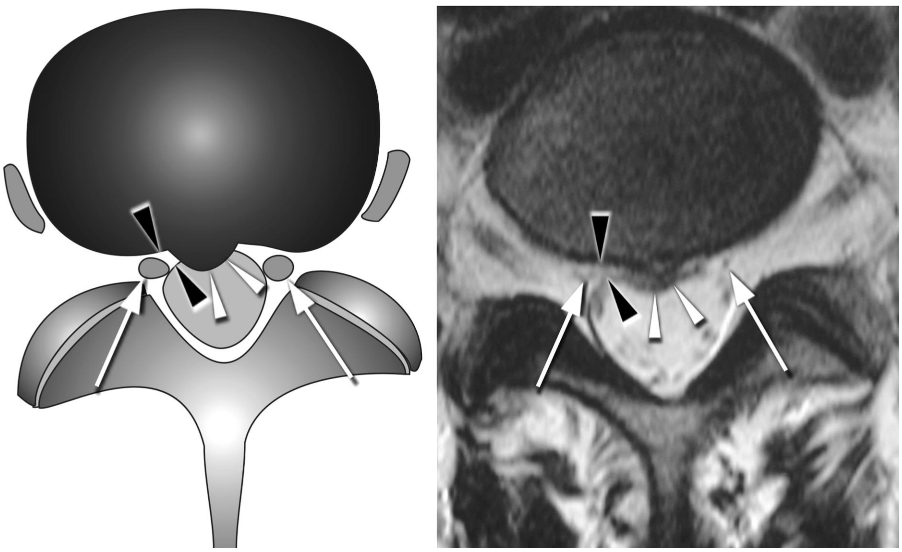

Figure 1. Diagram (left) and transverse T2-weighted fast spin-echo (SE) (repetition time msec/ echo time msec = 4,000/122) MR image (right) show no compromise of the nerve root. A normal epidural fat layer (black arrowheads) is visible between the nerve root (arrows) and the disk material (white arrowheads).

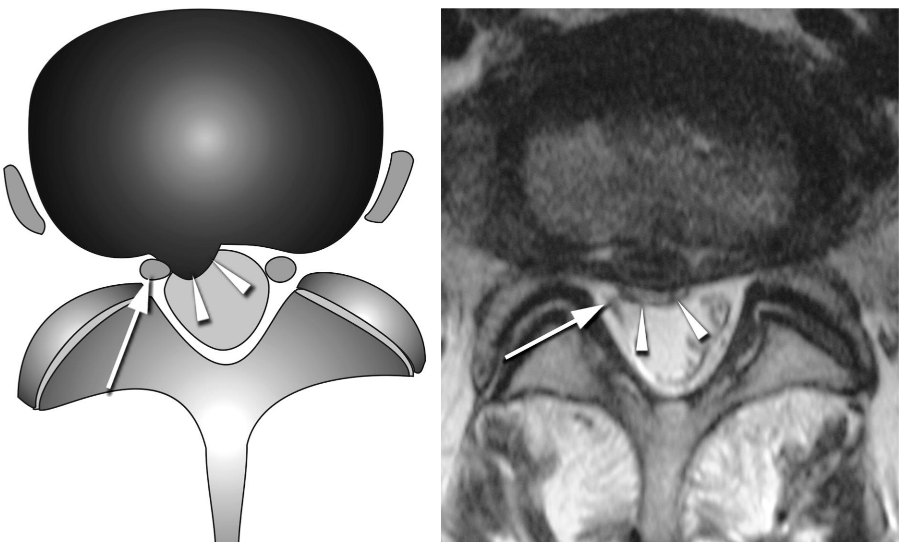

Figure 2. Diagram (left) and transverse T2-weighted fast SE (4,000/122) image (right) show contact of disk material (arrowheads) with the right nerve root (arrow). No epidural fat layer is visible between the nerve root and the disk material. The nerve root is in the normal position and is not dorsally deviated.

Study Population

The study population comprised 250 patients (139 men and 111 women; mean age, 46 years; age range, 19–84 years) in whom 500 nerve roots were retrospectively evaluated on MR images for compromise. In each patient, both nerve roots at a specific level were evaluated on images. In 80 consecutive patients (48 men with a mean age of 46.2 years and 32 women with a mean age of 48.5 years) who had undergone surgery after imaging, 94 nerve roots had been graded for compromise in the surgical report. The other 170 consecutive patients (91 men and 79 women; mean age, 48 years; age range, 16–84 years) had been referred from the outpatient spine clinic for imaging evaluation of sciatica. In the subgroup of 80 patients who underwent surgery, nerve root decompression had been performed in 68 patients and posterior lumbar intervertebral fusion had been performed in 12 patients. Our rationale for including patients who underwent spinal fusion was to ensure that cases with no neural compromise were included in the analysis. Only patients whose surgical reports provided an unequivocal description of the extent of nerve root compromise were included in the study.

The institutional review board did not require advance approval or informed consent for review of patients' records and images. Patients' rights are protected by a law that requires that patients be informed about the possibility of future review of their charts and radiographs for scientific purposes and be given the opportunity to forbid such use of their data. None of the patients included in our retrospective study had forbidden the use of their data.

MR Images

MR imaging of the lumbar spine was performed with a 1.5-T imager (Symphony; Siemens Medical Solutions, Erlangen, Germany) and a dedicated receive-only spine coil. All 250 images acquired in the 250 patients were obtained at the level of disk herniation or at the level of the disk believed to be responsible for the patient's symptoms. The imaging protocol included sagittal T1-weighted SE (700/12) and T2-weighted fast SE (5,000/130) sequences with the following parameters: matrix, 512 × 225; field of view, 300 mm; section thickness, 4 mm; intersection gap, 0.8 mm; number of signals acquired, four; echo train length, eight. Transverse T2-weighted fast SE (4,000/122) images also were acquired with the following parameters: matrix, 210 × 256; field of view, 150 mm; intersection gap, 0.8 mm; number of signals acquired, two; echo train length, eight. The imaging protocol conformed with the standards of the American College of Radiology (9).

Image Assessment

Three observers (C.D., M.R.S., C.W.A.P.)—a spinal radiology fellowship–trained orthopedic surgeon and two staff radiologists with 4 and 9 years of experience, respectively, in interpreting spinal MR images—independently graded each of the 500 lumbar nerve roots on two separate occasions. The three readers were blinded to the surgical reports and were unaware of the patients' medical history; cases that included surgical evaluation were intermixed with those that did not. On each MR image, the disk at the level most likely to be causing the symptoms or the level that had been surgically evaluated was marked by an in-

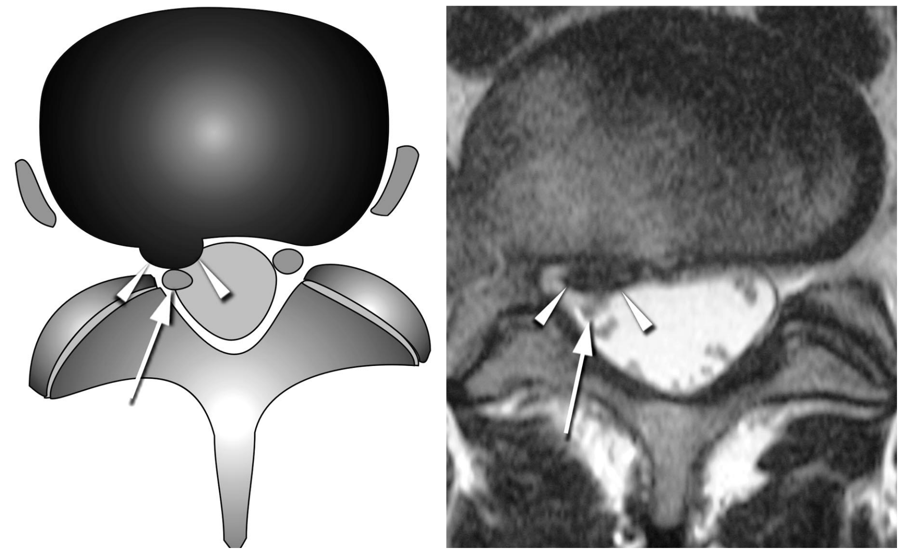

Figure 3. Diagram (left) and transverse T2-weighted fast SE (4,000/122) image (right) show dorsal deviation of the right nerve root (arrow), caused by contact with disk material (arrowheads).

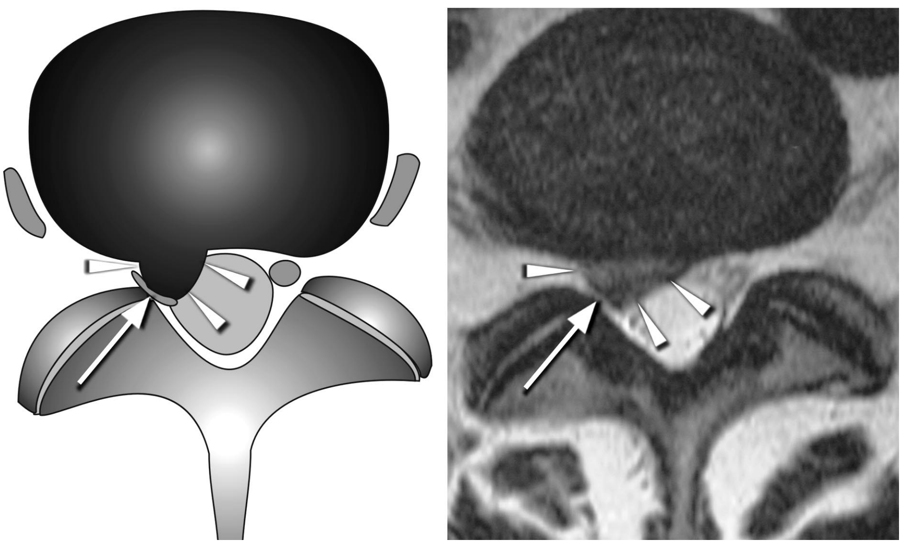

Figure 4. Diagram (left) and transverse T2-weighted fast SE (4,000/122) image (right) show compression of the right nerve root (arrow) between disk material (arrowheads) and the wall of the spinal canal. The nerve root appears flattened and is indistinguishable from disk material.

dependent radiologist (M.Z.) who was not involved in the image evaluation. Both nerve roots at the marked level were evaluated. Two evaluated nerve roots were at the L2 level, 12 were at the L3 level, 44 were at the L4 level, 178 were at the L5 level, and 264 were at the S1 level. To prevent a memory effect on the results of intraobserver variability assessement, an interval of at least 4 weeks was interposed between the first and second readings, and the cases were reviewed in a different order at the second reading than at the first. After the independent blinded readings, a consensus reading was performed for images about which there was grading disagreement. Data that resulted from the consensus reading were used for descriptive purposes only and were not included in the calculations of interobserver and intraobserver agreement.

Surgical Assessment

During surgery, the extent of nerve root compromise was assessed in terms similar to those used in MR image–based assessment. Nerve root deviation was defined as posterior dislocation of the nerve root by a bulging or herniated intervertebral disk. Nerve root compression was defined as posterior dislocation combined with deformation (eg, flattening) of the nerve root. Reliable intraoperative observation of contact between the nerve root and the intervertebral disk is not possible because of the very close spatial relationship of the posterior aspect of the annulus fibrosus and nerve root and because subtle differences in position cannot be assessed after the dorsal elements of the vertebra are removed during surgery. For the correlation of image-based grades with surgical grades, we therefore merged the first two grading categories used in MR image evaluation into a single category. Thus, three categories were considered: normal or contact, deviation, and compression.

Statistical Analysis

Differences in age and sex distribution in the study population were evaluated by using a nonpaired *t* test. Sex and distribution of nerve root compromise grades were assessed with a χ^2 test, and age and distribution of grades were assessed with analysis of variance. Intraobserver and interobserver reliability of MR image–based evaluations were estimated by calculating percentage agreement for evaluations in perfect agreement, in disagreement by one grade, or in disagreement by two grades. κ statistics with standard errors also were calculated for reliability (10). Following the system used by Landis and Koch (11), we described levels of agreement as follows: κ values of 0–0.20 indicated slight agreement, 0.21–0.40 indicated fair agreement, 0.41–0.60 indicated moderate agreement, 0.61–0.80 indicated substantial agreement, 0.81 or higher indicated excellent agreement, and 1.00 indicated absolute agreement. Frequency of disagreement was calculated for each grade. The Spearman coefficient was calculated for correlation between surgical grading and MR image–based grading.

Results

Demographic Assessment of the Study Population

There was no statistically significant difference in age distribution between the sexes (48 men with a mean age of 46.2 years and 32 women with a mean age of 48.5 years; $P = .580$ for nonpaired *t* test). No significant difference likewise was found between the sexes in grade of nerve root compromise (grade 0, 115 men and 94 women; grade 1, 57 men and 49 women; grade 2, 29 men and 21 women; grade 3, 77 men and 58 women; $P = .940$ for χ^2 test). Differences in mean patient age between grades of nerve root compromise also were not statistically significant (grade 0, 47.4 years; grade 1, 46.0 years; grade 2, 44.2 years; grade 3, 49.0 years; $P = .187$ for analysis of variance).

Grading of Nerve Root Compromise in the Study Population

The results of image-based grading of nerve root compromise by the three observers are summarized in Table 1. By consensus of all observers, 209 (42%) of 500 nerve roots evaluated on MR images were normal, 106 (21%) were in contact with disk material, 50 (10%) were deviated dorsally, and 135 (27%) were compressed. At surgery, 30 (32%) of 94 nerve roots were classified as normal or in contact with disk material, 14 (15%) were classified as deviated, and 50 (53%) were classified as compressed.

TABLE 1
Grading of Nerve Root Compromise by the Three Observers

Observer	Normal	Contact	Deviation	Compression
	First Reading			
A	222	90	56	132
B	204	118	50	128
C	216	102	63	119
	Second Reading			
A	235	84	43	138
B	211	102	55	132
C	208	118	62	112
	Consensus Reading			
A,B,C	209 (42)	106 (21)	50 (10)	135 (27)

Note.—Data are numbers of nerve roots, with percentages of total ($n = 500$) in parentheses.

Intraobserver Agreement

The results of the analysis of intraobserver reliability are summarized in Table 2. Perfect agreement was achieved in the grading of compromise in 401–423 (81%–85%) of 500 nerve roots. The κ values for intraobserver agreement were substantial (0.72–0.77). Disagreement by one grade occurred in grading of 76–98 (15%–20%) of the nerve roots. Disagreement by more than one grade occurred in grading of only one to three nerve roots.

TABLE 2
Intraobserver Reliability

Observer	κ Statistic	Standard Error	Agreement*	Disagreement by One Grade*	Disagreement by Two Grades*
A	0.77	0.023	423 (85)	76 (15)	1 (0.2)
B	0.72	0.025	401 (80)	98 (20)	1 (0.2)
C	0.73	0.024	406 (81)	91 (18)	3 (0.6)

* Data are numbers of nerve roots, with percentages of total ($n = 500$) in parentheses.

Interobserver Agreement

The κ statistics for interobserver agreement (0.62–0.67) were slightly lower than those for intraobserver agreement (0.72–0.77) (Table 3). Perfect agreement in grading of nerve root compromise was achieved in 368–386 (74%–77%) of the 500 nerve roots evaluated. In 109–127 (22%–25%) of the nerve roots, assessments of compromise diverged by one grade. Divergence by two grades occurred at assessment of between four and eight (0.8%–1.6%) of the nerve roots. No divergence by more than two grades occurred. There were no obvious differences in assessment of nerve root compromise between observers according to training background or level of experience.

TABLE 3
Interobserver Reliability

Observers	κ Statistic	Standard Error	Agreement*	Disagreement by One Grade*	Disagreement by Two Grades*
	First Reading				
A and B	0.65	0.027	378 (76)	118 (24)	4 (0.8)
A and C	0.67	0.026	386 (77)	109 (22)	5 (1.0)
B and C	0.62	0.027	368 (74)	127 (25)	5 (1.0)
	Second Reading				
A and B	0.67	0.026	386 (77)	110 (22)	4 (0.8)
A and C	0.64	0.027	374 (75)	120 (24)	6 (1.2)
B and C	0.63	0.027	370 (74)	122 (24)	8 (1.6)

* Data are numbers of nerve roots, with percentages of total ($n = 500$) in parentheses.

Analysis of Disagreement

The frequencies of intraobserver and interobserver disagreement for each of the three pairs of proximal grades are shown in Table 4. Disagreement was least frequent between the two highest grades of nerve root compromise (deviation and compression), which indicates that the proposed grading system has a good ability to characterize nerve root compromise. Disagreement was slightly more frequent between the two lowest grades (normal and contact) than between the other two grade pairs.

Correlation of MR Image–based Grades with Surgical Grades

The results of correlation of MR image–based grades with surgical grades are displayed in Table 5. The Spearman correlation coefficient was high ($r = 0.86$, $P < .001$). All but two of the nerve roots for which image-based grades were discrepant with surgical grades involved discrepancies of a single grade. In one nerve root that was graded on the MR image as in contact with the disk, surgery revealed compression. In another nerve root, in which compression was found on the MR image, only contact was found at surgery.

Discussion

The grading system described is a reliable method for evaluating and reporting

TABLE 4
Comparison of Frequencies of Intra- and Interobserver Disagreement by Proximal Grade Pairs

Grades between Which Disagreement Occurred	Intraobserver Disagreement	Interobserver Disagreement
Normal versus contact	44–55 (14–17)	58–71 (18–23)
Contact versus deviation	16–24 (10–15)	18–35 (12–22)
Deviation versus compression	16–23 (9–12)	19–37 (10–20)

Note.—Data are numbers of nerve roots, expressed as ranges, that were assigned a disputed grade in the specified proximal categories. Numbers in parentheses, also expressed as ranges, are percentages obtained by dividing the numbers of nerve roots assigned a disputed grade in the specified proximal categories by the total numbers of nerve roots assigned the same grade.

TABLE 5
Correlation of MR Image–based Grading with Surgical Grading

	MR Image–based Grade			
Surgical Grade	Normal or Contact	Deviation	Compression	Total
Normal or contact	27	2	1	30
Deviation	1	8	5	14
Compression	1	6	43	50
Total	29	16	49	94

Note.—Data are numbers of nerve roots. Spearman coefficient $r = 0.86$.

compromise of intraspinal extradural nerve roots in both clinical and research settings. The need for standardized nomenclature and classification of pathologic changes in the lumbar spine is emphasized in the recommendations of the combined task forces of the North American Spine Society, American Society of Spine Radiology, and American Society of Neuroradiology (2). Disk bulging and disk herniation (protrusion and extrusion, respectively) have been precisely defined, as have the descriptors for their anatomic location (zones in the transverse plane and levels in the craniocaudal direction) (2). However, none of these terms specify the relationship between the disk and the nerve root. Although lumbar disk herniation is a common finding in an asymptomatic population (12–14), nerve root compression is rare in asymptomatic volunteers (13).

Previous investigators have demonstrated that spinal canal dimensions are substantially smaller in symptomatic patients than in asymptomatic controls (5,15). The results of studies in asymptomatic volunteers who were matched according to age, sex, and occupational risk factors to patients with symptoms of disk herniation show a frequency of disk herniation substantially higher in the asymptomatic group (76%) (16) than the frequencies previously found in unmatched study populations (20%–36%) (13,14). Boos et al (16) showed that the only substantial morphologic difference between symptomatic patients and asymptomatic matched controls was the presence of neural compromise (83% vs 22%). Therefore, neural compromise may have a more important role in explaining pain than does morphologic extension of disk material beyond intervertebral space (15,17).

The pathophysiologic mechanisms that cause sciatica are still incompletely understood. Two mechanisms seem important: mechanical nerve root compromise and chemical irritation of the nerve root by nucleus pulposus tissue. Compression of nerve roots is known to correlate both with pain and with neural dysfunction in a segmental distribution of that specific nerve root (18). Compression per se may impair the transport of nutrients to the nerve tissue. On the other hand, Olmarker et al (19) have shown that nucleus pulposus tissue can induce neurophysiologic and histologic changes in porcine cauda equina nerve roots even without compression. Furthermore, nucleus pulposus appears to cause inflammation, as indicated by leukotaxis and an increase of vascular permeability (20). However, contact of disk material with the nerve root is infrequently associated with symptoms. In a series of MR imaging examinations of the lumbar spine in 60 asymptomatic subjects, contact of disk material with the nerve root was observed in 22%–23% of subjects (13).

The κ values for inter- and intraobserver variability in our study were substantial (0.62–0.77) and well within values reported for other grading systems of lumbar disk derangement. In a previous study (7) in which two nomenclatures were evaluated (normal, bulge, herniation; and normal, bulge, protrusion, extrusion), interobserver agreement was 80% (κ = 0.58) and intraobserver agreement was 86% (κ = 0.71–0.69). κ values from another previous study (21) of a different system for grading lumbar disk degeneration were slightly higher (for intraobserver variability, κ = 0.84–0.90; for interobserver variability, κ = 0.69–0.81). The analysis of observer disagreement in our study indicates that the system we tested is reliable in discriminating between higher grades of nerve root compromise but slightly less reliable in discriminating between a normal nerve root without compromise and a nerve root in contact with disk material.

The imaging protocol used in our study conformed with the standards established by the American College of Radiology (9). However, T1-weighted SE images could have been obtained either in addition to or instead of T2-weighted fast SE images, as is done in many institutions (9). T1-weighted images can help to clarify the relationship between the nerve root and epidural fat or thecal sac, and their use might have affected the performance of our grading system. A further limitation of our study is the retrospective nature of the analysis. Because all surgeries were performed after MR imaging, the surgeon's assessment might have been biased by the imaging studies. In addition, surgical assessment of nerve root compromise generally is limited by the close proximity of the intervertebral disk to the nerve root and the resultant difficulty in discerning contact between the two. It seems likely that reliable in vivo differentiation of contact from no contact is possible only with imaging. Therefore, these two grading categories (normal and contact) were not considered separately in the correlation of image-based grades with surgical grades. However, it is possible at surgery to reliably differentiate among compression, deviation, and normal or contact.

Another limitation of our study is the fact that two nerve roots at any given level were always analyzed together. Because the nerve root contralateral to a compromised nerve root has a greater chance of being normal than do nerve roots at other proximal levels, some bias may be inherent in our statistical calculations. In addition, the number of nerve roots at higher lumbar levels (L1–L2, L2–L3) in our study was small, compared with that of nerve roots at lower levels (L4–L5, L5–S1). However, we believe that the large overall number of nerve roots evaluated helps to minimize these limitations. Likewise, any bias that may have occurred in the assessments of observer reliability as a result of the memory effect should have been minimized by the delay of 4 weeks between the two readings.

In conclusion, the MR image–based grading system used in this study enables discrimination between grades of nerve root compromise in the lumbar spine with sufficient reliability for both research and clinical purposes.

References

1. Boos N, Dreier D, Hilfiker E, et al. Tissue characterization of symptomatic and asymptomatic disk herniations by quantitative magnetic resonance imaging. J Orthop Res 1997; 15:141–149.
2. Fardon DF, Milette PC. Nomenclature and classification of lumbar disk pathology: recommendations of the combined task forces of the North American Spine Society, American Society of Spine Radiology, and American Society of Neuroradiology. Spine 2001; 26:E93–E113.
3. Milette PC. The proper terminology for reporting lumbar intervertebral disk disorders. AJNR Am J Neuroradiol 1997; 18:1859–1866.
4. Modic MT, Masaryk TJ, Ross JS, Carter JR. Imaging of degenerative disk disease. Radiology 1988; 168:177–186.
5. Porter RW, Hibbert CS, Wicks M. The spinal canal in symptomatic lumbar disk lesions. J Bone Joint Surg Br 1978; 60-B:485–487.
6. Heliovaara M, Vanharanta H, Korpi J, Troup JD. Herniated lumbar disk syndrome and vertebral canals. Spine 1986; 11:433–435.
7. Brant-Zawadzki MN, Jensen MC, Obuchowski N, Ross JS, Modic MT. Interobserver and intraobserver variability in interpretation of lumbar disk abnormalities: a comparison of two nomenclatures. Spine 1995; 20:1257–1264.
8. Fardon DF. Nomenclature and classification of lumbar disk pathology. Spine 2001; 26:461–462.
9. American College of Radiology. Practice standard for the performance of magnetic resonance imaging (MRI) of the adult spine. Available online at: *www.acr.org/departments/stand_accred/pdf/mri_adult_spine.pdf*. Accessed September 1, 2002.
10. Cohen JA. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960; 20:37–46.
11. Landis RJ, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33:159–174.
12. Boden SD, Davis DO, Dina TS, Patronas NJ, Wiesel SW. Abnormal magnetic-resonance scans of the lumbar spine in asymptomatic subjects: a prospective investigation. J Bone Joint Surg Am 1990; 72:403–408.
13. Weishaupt D, Zanetti M, Hodler J, Boos N. MR imaging of the lumbar spine: prevalence of intervertebral disk extrusion and sequestration, nerve root compression, end plate abnormalities, and osteoarthritis of the facet joints in asymptomatic volunteers. Radiology 1998; 209:661–666.
14. Jensen MC, Brant-Zawadzki MN, Obuchowski N, Modic MT, Malkasian D, Ross JS. Magnetic resonance imaging of the lumbar spine in people without back pain. N Engl J Med 1994; 331:69–73.
15. Debois V, Herz R, Berghmans D, Hermans B, Herregodts P. Soft cervical disk herniation: influence of cervical spinal canal measurements on development of neurologic symptoms. Spine 1999; 24:1996–2002.
16. Boos N, Rieder R, Schade V, Spratt K, Semmer N, Aebi M. The diagnostic accuracy of magnetic resonance imaging, work perception and psychosocial factors in identifying symptomatic disk herniations. Spine 1995; 20:2613–2625.
17. Beattie PF, Meyers SP, Stratford P, Millard RW, Hollenberg GM. Associations between patient report of symptoms and anatomic impairment visible on lumbar magnetic resonance imaging. Spine 2000; 25:819–828.
18. Olmarker K, Rydevik B. Pathophysiology of sciatica. Orthop Clin North Am 1991; 22:223–234.
19. Olmarker K, Rydevik B, Nordborg C. Autologous nucleus pulposus induces neurophysiologic and histologic changes in porcine cauda equina nerve roots. Spine 1993; 18:1425–1432.
20. Olmarker K, Blomquist J, Stromberg J, Nannmark U, Thomsen P, Rydevik B. Inflammatogenic properties of nucleus pulposus. Spine 1995; 20:665–669.
21. Pfirrmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N. Magnetic resonance classification of lumbar intervertebral disk degeneration. Spine 2001; 26:1873–1878.