SPINE Volume 26, Number 17, pp 1873–1878

Magnetic Resonance Classification of Lumbar Intervertebral Disc Degeneration

Christian W. A. Pfirrmann, MD,* Alexander Metzdorf, MD,† Marco Zanetti, MD,* Juerg Hodler, MD,* and Norbert Boos, MD†

Study Design. A reliability study was conducted.

Objectives. To develop a classification system for lumbar disc degeneration based on routine magnetic resonance imaging, to investigate the applicability of a simple algorithm, and to assess the reliability of this classification system.

Summary of Background Data. A standardized nomenclature in the assessment of disc abnormalities is a prerequisite for a comparison of data from different investigations. The reliability of the assessment has a crucial influence on the validity of the data. Grading systems of disc degeneration based on state of the art magnetic resonance imaging and corresponding reproducibility studies currently are sparse.

Methods. A grading system for lumbar disc degeneration was developed on the basis of the literature. An algorithm to assess the grading was developed and optimized by reviewing lumbar magnetic resonance examinations. The reliability of the algorithm in depicting intervertebral disc alterations was tested on the magnetic resonance images of 300 lumbar intervertebral discs in 60 patients (33 men and 27 women) with a mean age of 40 years (range, 10–83 years). All scans were analyzed independently by three observers. Intra- and interobserver reliabilities were assessed by calculating kappa statistics.

Results. There were 14 Grade I, 82 Grade II, 72 Grade III, 68 Grade IV, and 64 Grade V discs. The kappa coefficients for intra- and interobserver agreement were substantial to excellent: intraobserver (kappa range, 0.84–0.90) and interobserver (kappa range, 0.69–0.81). Complete agreement was obtained, on the average, in 83.8% of all the discs. A difference of one grade occurred in 15.9% and a difference of two or more grades in 1.3% of all the cases.

Conclusion. Disc degeneration can be graded reliably on routine T2-weighted magnetic resonance images using the grading system and algorithm presented in this investigation. [Key words: disc degeneration, intervertebral disc, magnetic resonance imaging, reliability] **Spine 2001;26:1873–1878**

Magnetic resonance imaging (MRI) is the most important method for the clinical assessment of intervertebral disc pathology. The signal characteristics of the disc in T2-weighted MRIs reflect changes caused by aging or degeneration.[14,16,20]

From the *Division of Musculoskeletal Radiology, and the †Department of Orthopedic Surgery, Orthopedic University Hospital, Balgrist Zurich, Switzerland.
Supported by grant 32-52927.97 from the Swiss National Science Foundation.
Acknowledgment date: August 16, 2000.
First revision date: November 27, 2000.
Acceptance date: January 1, 2001.
Device status category: 11.
Conflict of interest category: 14.

A standardized nomenclature in the assessment of disc alterations is a prerequisite for comparison of data from different investigations.[3] A morphologic grading system relating to the pathologic changes in the disc is needed.[9] The reliability (interobserver and intraobserver reproducibility) of the assessment has a crucial influence on the validity of the data. However, reproducibility studies for the assessment of intervertebral disc degeneration currently are sparse despite their clinical importance.

The MRI technique is developing continuously. Fast spin-echo (FSE) imaging was introduced in spinal MRI during the early 1990s[12] and now is widely used. As compared with the conventional spin-echo technique, FSE imaging offers a significant reduction in scanning time, a better signal-to-noise ratio, and fewer motion artifacts.[1] However, conventional spin-echo images cannot be compared directly with FSE images, and disc disease may not have exactly the same signal characteristics on these two types of sequences. On FSE images, the contrast between fat- and fluid-containing structures is less than on spin-echo images,[1,5] and the normal intervertebral disc has lower signal intensity on FSE images. Most prior classification systems of degenerative disc disease were developed and tested on spin-echo MRIs.[3,17]

The objective of this investigation was to develop a classification system for lumbar disc degeneration observed on MRIs, to investigate the applicability of a simple algorithm, and to assess the reliability of this classification system.

Methods

Grading System and Algorithm for the Lumbar Intervertebral Discs. A comprehensive grading system for lumbar disc degeneration (Table 1, Figure 1) was developed by the senior author on the basis of the literature and previously published work.[2,3,9,13,23] The feasibility of an algorithm to assess the grades of disc degeneration was tested, and the algorithm was optimized (Figure 2) by reviews of lumbar MRI examinations during routine clinical work.

Participants. The study involved lumbar MRIs of 60 patients (33 men and 27 women) with a mean age of 40 years (range, 10–83 years). Over a period of 3 weeks, 40 routine MRI scans of the lumbar spine were collected consecutively. To ensure an adequate number of nondegenerated and adolescent discs, 20 examinations of patients between the ages of 10 and 20 years were randomly selected from a database of examinations at the same institution in 1999. All the patients included in this study presented initially to the outpatient spine clinic, then were re-

Table 1. Classification of Disc Degeneration*

Grade	Structure	Distinction of Nucleus and Anulus	Signal Intensity	Height of Intervertebral Disc
I	Homogeneous, bright white	Clear	Hyperintense, isointense to cerebrospinal fluid	Normal
II	Inhomogeneous with or without horizontal bands	Clear	Hyperintense, isointense to cerebrospinal fluid	Normal
III	Inhomogeneous, gray	Unclear	Intermediate	Normal to slightly decreased
IV	Inhomogeneous, gray to black	Lost	Intermediate to hypointense	Normal to moderately decreased
V	Inhomogeneous, black	Lost	Hypointense	Collapsed disc space

* Modified from Pearce (cited by Eyre et al[9]).

ferred for a lumbar MRI. None of these examinations were used for the development and optimization of the algorithm.

Imaging Technique. The MRIs of the lumbar spine were performed on a 1-T scanner (Siemens Impact Expert; Siemens Medical Systems, Erlangen, Germany) using a dedicated receive only spine coil. The imaging protocol included sagittal T1-weighted spin-echo (repetition time [TR] 700 msec/echo time [TE] 12 msec) and T2-weighted FSE (TR 5000 msec/TE 130 msec) images with the following parameters: matrix, 512 × 225; field of view, 225 × 300 mm; slice thickness, 4 mm; interslice gap, 0.8 mm; number of excitations, 4; echo train length (ETL), 15 (the first echo of this sequence is discarded), and axial T2-weighted axial FSE scans (TR 5000 msec/TE 72 msec; matrix, 210 × 256; field of view, 150 × 150 mm; interslice gap, 0.8 mm; number of excitations, 2; echo train length, 7). All the sequences were acquired without fat saturation.

Image Assessment. Three observers with different levels of experience analyzing spinal MRIs (*i.e.*, an orthopedic surgeon, a fellowship-trained musculoskeletal radiologist, and a musculoskeletal senior staff radiologist) graded each of the 300 lumbar intervertebral discs on the T2-weighted sagittal images. The observers were not involved in the development of the grading system and the algorithm. In all, 60 selected MRIs were randomly ordered in three sets of 20 MRIs and interpreted independently by the three observers. Each observer was allowed to review only one set per day to avoid rater fatigue. The choice of 20 scans represented the usual number of spinal MRIs read on a daily basis. All the MRIs were analyzed by the observers on a separate occasion, with a minimum interval of 1 week. The observers were asked to follow the algorithm strictly as given. A handout of the classification system (Table 1), the algorithm (Figure 2), and a set of sample MRIs (Figure 1) was available to the raters during the image review. To obtain a reference grade for each disc, a consensus readout was performed after all the data were collected.

Data Analysis. The reliability of the MRI evaluations was estimated using agreement percentage and kappa statistics within raters (intraobserver reliability) and between raters (interobserver reliability).[4] According to Landis and Koch,[11] the agreement was rated as follows: kappa 0 to 0.2 indicated slight agreement, 0.21 to 0.4 fair agreement, 0.41 to 0.60 moderate agreement, 0.61 to 0.8 substantial agreement, and 0.81 upward excellent agreement. With this rating, absolute agreement would be 1. Frequency of disagreement was calculated for each grade.

■ Results

Grades of Disc Degeneration in the Study Population

Altogether, 300 lumbar discs were analyzed in a study population of 60 individuals. The number of disc degeneration grades assessed by each reader are summarized in Table 2. The consensus reading resulted in 14 Grade I discs (5%), 82 Grade II discs (27%), 72 Grade III discs (24%), 68 Grade IV (23%), and 64 Grade V discs (21%).

Intraobserver Agreement

The results of the intraobserver agreement are summarized in Table 3. Intraobserver agreement was "excellent" for all three readers, with kappa values ranging from 0.84 to 0.90. Complete intraobserver agreement was achieved in a range from 264 (88%) to 277 (92.3%) of 300 discs. All but one disagreement were within one grade of difference, with a range of 23 (7.7%) to 35 (11.7%) discs.

Interobserver Agreement

As expected, interobserver agreement was somewhat lower than intraobserver agreement (Table 3). Nevertheless, the agreement ranged from substantial to excellent, with kappa values ranging from 0.69 to 0.81. Complete agreement was achieved in a range from 233 (77.7%) to 257 (85.7%) of all 300 discs. A difference of one grade occurred in 43 (14.3%) to 71 (23.7%) assessments of the discs, a difference of two grades in two discs (0.7%), and a difference of three grades in one disc (0.3%).

Evaluation of Distinction Between Grades and Analysis of Disagreement

The relation between the frequency of the different disc degeneration grades in the study population and the frequency of disagreement is displayed in Table 4. The relative disagreement shows a fairly even distribution among grades, indicating that the proposed grading system has good discrimination ability. However, disagreement was more frequent between Grades I and II in terms of inter- and intraobserver agreement, and between Grades III and IV in terms of interobserver agreement. The cases with disagreements of two and three grades referred to disc spaces with marked narrowing of the disc height and normal to slightly decreased signal of the nucleus. The case with a disagreement of three grades referred to a transitional vertebra at the lumbosacral junction. The fifth lum-

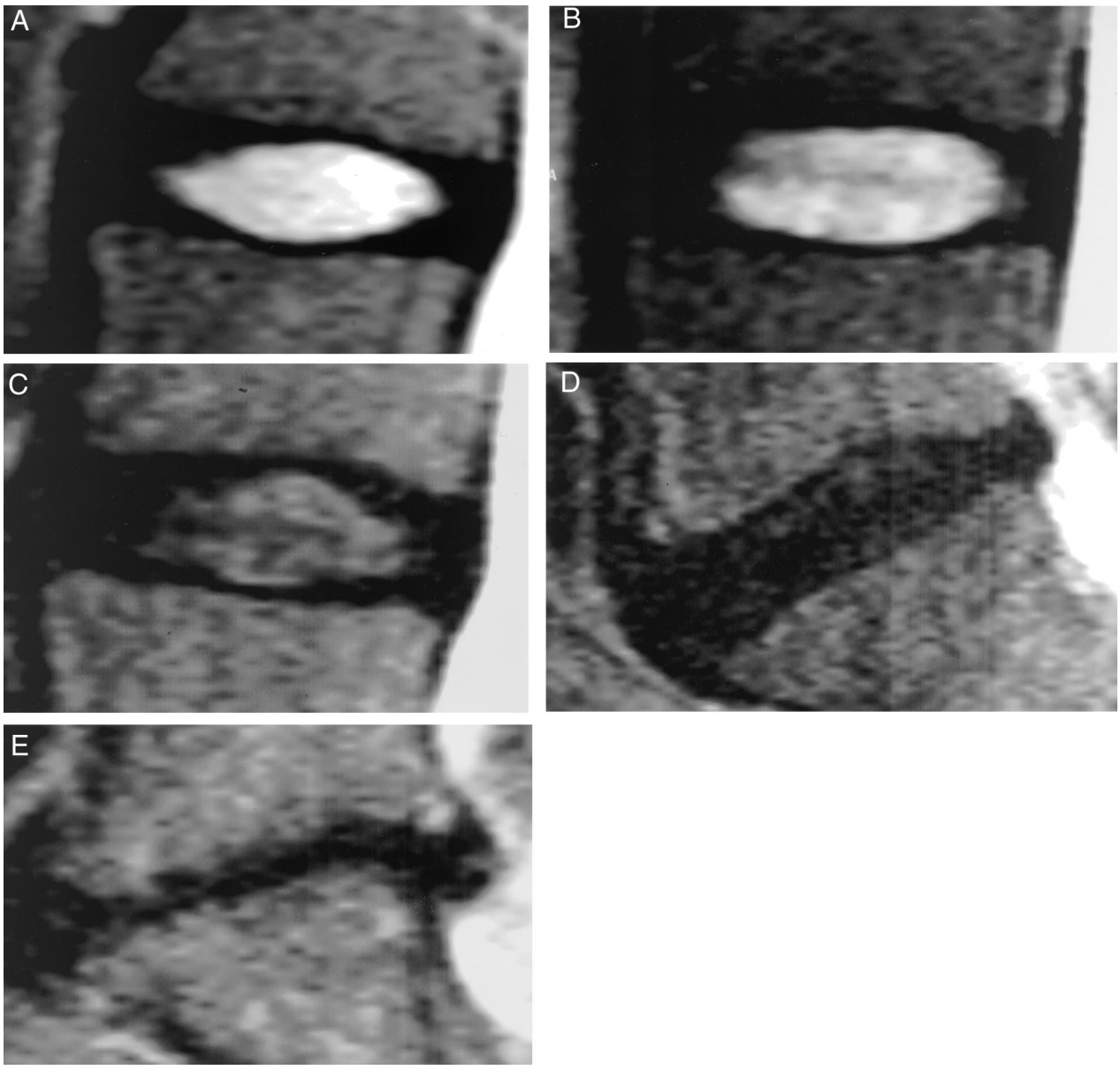

Figure 1. **A–E**, Grading system for the assessment of lumbar disc degeneration. Grade I: The structure of the disc is homogeneous, with a bright hyperintense white signal intensity and a normal disc height. Grade II: The structure of the disc is inhomogeneous, with a hyperintense white signal. The distinction between nucleus and anulus is clear, and the disc height is normal, with or without horizontal gray bands. Grade III: The structure of the disc is inhomogeneous, with an intermediate gray signal intensity. The distinction between nucleus and anulus is unclear, and the disc height is normal or slightly decreased. Grade IV: The structure of the disc is inhomogeneous, with an hypointense dark gray signal intensity. The distinction between nucleus and anulus is lost, and the disc height is normal or moderately decreased. Grade V: The structure of the disc is inhomogeneous, with a hypointense black signal intensity. The distinction between nucleus and anulus is lost, and the disc space is collapsed. Grading is performed on T2-weighted midsagittal (repetition time 5000 msec/echo time 130 msec) fast spin-echo images.

bar vertebral body was partially sacralized with a hypoplastic L5–S1 disc. Despite the signal characteristics of a Grade II disc, it was read as Grade V by two readers because of the markedly smaller disc space.

■ Discussion

A number of morphologic grading systems for lumbar disc degeneration have been proposed.[10] Most previous classification systems and reliability studies of lumbar disc abnormalities on MRI have focused on the posterior aspect of the disc, distinguishing among bulging, protrusion, and extrusion.[3] Studies focusing on the MRI characteristics of the disc structure are rare.[17] The current grading system and algorithm are based on MRI signal intensity, disc structure, distinction between nucleus and anulus, and disc height. The latter is important for distinguishing between Grades IV and V discs. For Grades III and IV, the disc height is not a discriminative feature. Changes in the vertebral body marrow adjacent to the intervertebral disc were not included in this algorithm because they can be well characterized by the three types that Modic et al suggested.[15] The current authors suggest

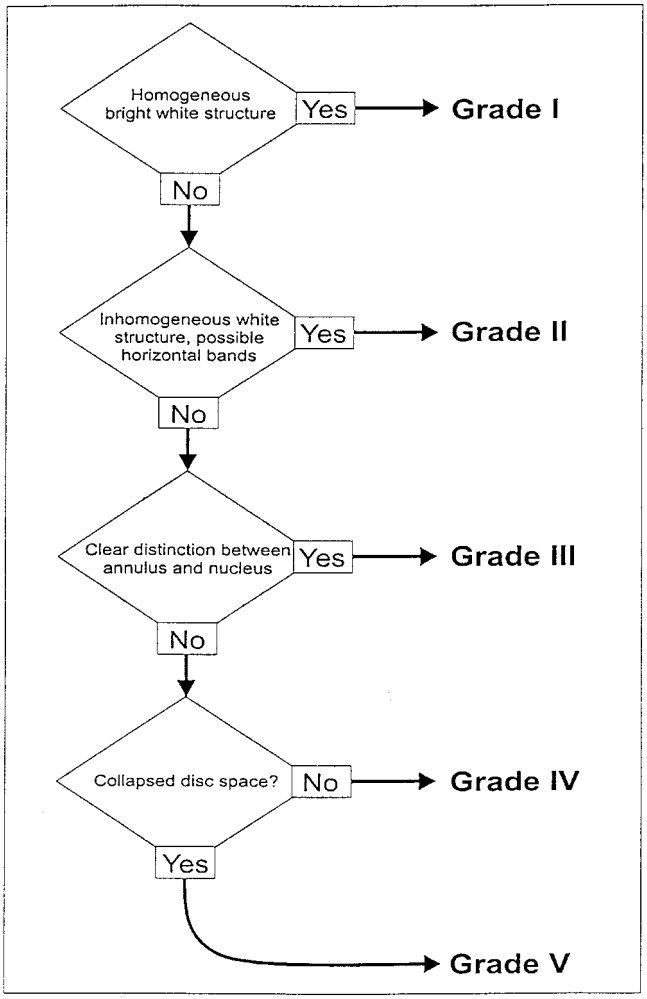

Figure 2. Algorithm for the grading system and for the assessment of the lumbar disc degeneration grade.

adding the Modic classification (Types 1–3) to their grading system for further specification of a degenerated intervertebral disc in the case of concomitant bone marrow changes.

The signal intensity of the disc in relation to chemical composition and histologic changes has been studied.[18,21] The signal loss of the disc on T2-weighted MRIs correlates with the progressive degenerative changes of the intervertebral disc.[14] The brightness of the nucleus has been shown to correlate directly with the proteoglycan concentration, but not with the water or collagen content.[16] Variations in disc hydration or composition can be detected noninvasively by quantitative MRI in studies of cohorts with sufficient accuracy.[2] Recent studies have used MRI to assess disc degeneration *in vivo* with regard to potential risk factors. At this writing, no formally tested classification is available from the literature. Brant-Zawadzki et al[3] emphasized the need for a comprehensive standardized terminology.

Table 2. Assigned and Consensus Grades for Three Observers

	Grades				
Observer	I	II	III	IV	V
A1	18	85	74	66	57
A2	17	78	84	68	53
B1	15	74	76	71	64
B2	15	75	81	64	65
C1	11	84	63	67	75
C2	7	94	57	70	72
Consensus:	14 (5%)	82 (27%)	72 (24%)	68 (23%)	64 (21%)

Table 3. Intra- and Interobserver Reliability*

		Agreement		Disagreement					
				1 Grade		2 Grades		3 Grades	
Observers	Kappa	n	%	n	%	n	%	n	%
Intraobserver									
A (1–2)	0.90	277	92.3	23	7.7	0	0.0	0	0
B (1–2)	0.90	276	92.0	24	8.0	0	0.0	0	0
C (1–2)	0.84	264	88.0	35	11.7	1	0.3	0	0
Interobserver									
A1–B1	0.79	252	84.0	47	15.7	0	0.0	1	0.3
A1–C1	0.74	240	80.0	57	19.0	2	0.7	1	0.3
B1–C1	0.81	255	85.0	44	14.7	1	0.3	0	0.0

* Interobserver reliability only between the first readings of all observers is displayed.

In this study, the authors proposed a classification system for gross morphology of the lumbar intervertebral disc. This grading system is an extension of the preliminary work by Pearce et al.[9,16,23] The MRI grading system is based on a gross morphology grading scheme. Previous evaluation reported it to be reliable for assessing gross (macroscopic) disc morphology.[23] The concordance of this classification system with morphologic and MRI changes also was reported to be high.[9] Analysis of the biochemical changes on the same specimens has suggested that MRI reflects the proteoglycan content of the nucleus more closely than gross tissue morphology.[22]

Observer variability of lumbar degenerative disease has been studied with various imaging techniques.[6–8] With MRI, values of agreement on disc degeneration show a high variation in the literature depending on the variable investigated. Brant-Zawadzki et al[3] reported kappa values of 0.58 to 0.71 for determining the degree of "disc extension beyond the interspace" depending on two different nomenclatures. In a study by Raininko et al,[17] the observer variability for the signal intensity of the nuclear complex ranged from 0.65 to 0.94. In the study of Weishaupt et al[24] investigating asymptomatic volunteers, the kappa value for grading disc morphology was 0.84 and 0.79. The intra- and interobserver reliability in the current study (kappa, 0.69–0.90) is well in line with the reliability found for grading gross morphology (kappa, 0.67–0.94).[23] No obvious differences were seen between the three readers despite their different backgrounds and levels of experience.

The frequency of disagreement in discriminating grades was fairly uniform among the different levels and almost limited to one degree. This indicates the absence of a weak spot for the grading system and the algorithm. However, a slightly higher frequency of disagreement

Table 4. Influence of Category (Grades I to IV) on Intra- and Interobserver Reliability

Grades	Consensus* n	Intraobserver Disagreement Range		Interobserver Disagreement Range	
		n	Consensus (%)	n	Consensus (%)
Disagreement in one grade					
Grades I and II	96	3–12	3–13	10–16	10–17
Grades II and III	154	7–8	5–5	7–17	5–11
Grades III and IV	140	7–8	5–6	7–24	5–17
Grades IV and V	132	4–8	3–6	7–9	5–7
Disagreement in two grades					
Grades III and V	136	0–1	0–1	0–2	0–1
Disagreement in three grades					
Grades II and V	146	0	0–1	0–1	0–1

* Total number of discs assigned in the consensus reading for the two grades compared.

was observed between Grades I and II, and between Grades III and IV. This can be explained by the main discriminating features between these grades (homogeneous *versus* inhomogeneous bright nucleus for Grades I and II and the possibility of differentiating the anulus and nucleus for Grades III and IV), which are subject to a larger scope of interpretation than for the other grades. The difficulty grading a hypoplastic intervertebral disc in a case with a partially sacralized L5 vertebral body, which resulted in a Grade V assessment because of the marked smaller disc space despite the signal characteristics of a Grade II disc, could have been eliminated with the strict use of the algorithm. Other reasons for a high signal in the intervertebral disc, such as a localized fluid collection in a fissure or the vacuum phenomenon,[19] could be rare but potential pitfalls for the assessment of the disc grade. These rare pitfalls can be avoided in equivocal cases when the characteristics of disc structure, the MRI signal intensity, the distinction between nucleus and anulus, and the disc height are considered simultaneously.

In conclusion, this study describes a grading system for the assessment of lumbar disc degeneration in form of a simple algorithm using contemporary MRI technique. The system is comprehensive, with intra- and interobserver reliability sufficient to discriminate between the different grades. This grading system therefore provides a standardized and reliable assessment of MRI disc morphology for research and clinical purposes.

■ Key Points

- This study suggests a grading system for the assessment of lumbar disc degeneration in the form of a simple algorithm using contemporary MRI technique.
- The system is comprehensive, with intra- and interobserver reliability sufficient to discriminate the different grades.
- This grading system therefore provides a standardized and reliable assessment of MRI disc morphology for research and clinical purposes.

References

1. Atlas SW, Hackney DB, Listerud J. Fast spin-echo imaging of the brain and spine. Magn Reson Q 1993;9:61–83.
2. Boos N, Wallin A, Schmucker T, et al. Quantitative MR imaging of lumbar intervertebral disc and vertebral bodies: methodology, reproducibility, and preliminary results. Magn Reson Imaging 1994;12:577–87.
3. Brant-Zawadzki MN, Jensen MC, Obuchowski N, et al. Interobserver and intraobserver variability in interpretation of lumbar disc abnormalities: a comparison of two nomenclatures. Spine 1995;20:1257–63, discussion 64.
4. Cohen JA. A coefficient of agreement for nominal scales. Educ Psychol Measurement 1960;20:37–46.
5. Constable RT, Anderson AW, Zhong J, et al. Factors influencing contrast in fast spin-echo MR imaging. Magn Reson Imaging 1992;10:497–511.
6. Coste J, Judet O, Barre O, et al. Inter- and intraobserver variability in the interpretation of computed tomography of the lumbar spine. J Clin Epidemiol 1994;47:375–81.
7. Coste J, Paolaggi JB, Spira A. Reliability of interpretation of plain lumbar spine radiographs in benign, mechanical low back pain. Spine 1991;16:426–8.
8. Deyo RA, McNiesh LM, Cone RO. Observer variability in the interpretation of lumbar spine radiographs. Arthritis Rheum 1985;28:1066–70.
9. Eyre D, Benya P, Buckwalter J, et al. Intervertebral disk: Part B. Basic science perspectives. In: Frymoyer JW, Gordon SL, eds. New Perspectives on Low Back Pain. Park Ridge, IL: American Academy of Orthopaedic Surgeons, 1989.
10. Friberg S, Hirsch C. Anatomical and clinical studies on lumbar disc degeneration. Acta Orthop Scand 1949;19:222–42.
11. Landis RJ, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
12. Listerud J, Einstein S, Outwater E, et al. First principles of fast spin echo. Magn Reson Q 1992;8:199–244.
13. Masaryk TJ, Ross JS, Modic MT, et al. High-resolution MR imaging of sequestered lumbar intervertebral disks. AJR Am J Roentgenol 1988;150:1155–62.
14. Modic MT, Masaryk TJ, Ross JS, et al. Imaging of degenerative disk disease. Radiology 1988;168:177–86.
15. Modic MT, Steinberg PM, Ross JS, et al. Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging. Radiology 1988;166:193–9.
16. Pearce RH, Thompson JP, Bebault GM, et al. Magnetic resonance imaging reflects the chemical changes of aging degeneration in the human intervertebral disk. J Rheumatol Suppl 1991;27:42–3.
17. Raininko R, Manninen H, Battie MC, et al. Observer variability in the assessment of disc degeneration on magnetic resonance images of the lumbar and thoracic spine. Spine 1995;20:1029–35.
18. Schiebler ML, Camerino VJ, Fallon, MD, et al. *In vivo* and *ex vivo* magnetic resonance imaging evaluation of early disc degeneration with histopathologic correlation. Spine 1991;16:635–40.
19. Schweitzer ME, el-Noueam KI. Vacuum disc: frequency of high signal intensity on T2-weighted MR images. Skeletal Radiol 1998;27:83–6.
20. Sether LA, Yu S, Haughton VM, et al. Intervertebral disk: normal age-related changes in MR signal intensity. Radiology 1990;177:385–8.

21. Tertti M, Paajanen H, Laato M, et al. Disc degeneration in magnetic resonance imaging: a comparative biochemical, histologic, and radiologic study in cadaver spines. Spine 1991;16:629–34.
22. Thompson JP, Pearce RH, Ho B. Correlation of gross morphology and chemical composition with magnetic resonance images of human lumbar intervertebral discs. Trans Orthop Res Soc 1988;13:276.
23. Thompson JP, Pearce RH, Schechter MT, et al. Preliminary evaluation of a scheme for grading the gross morphology of the human intervertebral disc. Spine 1990;15:411–5.
24. Weishaupt D, Zanetti M, Hodler J, et al. MR imaging of the lumbar spine: prevalence of intervertebral disk extrusion and sequestration, nerve root compression, endplate abnormalities, and osteoarthritis of the facet joints in asymptomatic volunteers. Radiology 1998;209:661–6.

Address reprint requests to

Norbert Boos, MD
Orthopedic University Hospital
Balgrist
Forchstrasse 340
CH-8008 Zurich
Switzerland
E-Mail: nboos@balgrist.unizh.ch