

# Lumbar Spine: Reliability of MR Imaging Findings<sup>1</sup>

John A. Carrino, MD, MPH  
Jon D. Lurie, MD, MS  
Anna N. A. Tosteson, ScD  
Tor D. Tosteson, ScD  
Eugene J. Carragee, MD  
Jay Kaiser, MD  
Margaret R. Grove, MS  
Emily Blood, MS  
Loretta H. Pearson, MPhil  
James N. Weinstein, DO, MS  
Richard Herzog, MD

<sup>1</sup> From the Section of Musculoskeletal Radiology, Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, 601 N Caroline St, JHOC 5165, Baltimore, MD 21287 (J.A.C.); Departments of Medicine (J.D.L., A.N.A.T.), Community and Family Medicine (A.N.A.T., T.D.T., M.R.G., E.B., L.H.P.), and Orthopaedics (J.N.W.), Dartmouth Medical School, Lebanon, NH; Department of Orthopaedic Surgery, Stanford School of Medicine, Stanford, Calif (E.J.C.); National Orthopedic Imaging Associates, Greenbrae, Calif (J.K.); and Division of Teleradiology, Hospital for Special Surgery, New York, NY (R.H.). Received November 15, 2007; revision requested January 16, 2008; revision received June 12; final version accepted June 26. Supported by National Institutes of Health grant P60 AR048094-01A1, Multidisciplinary Clinical Research Center in Musculoskeletal Diseases, Project 1: Prognostic Value of Lumbar Spine MR Imaging. Address correspondence to J.A.C. (e-mail: [jcarrin2@jhmi.edu](mailto:jcarrin2@jhmi.edu)).

© RSNA, 2008

## Purpose:

To characterize the inter- and intraobserver variability of qualitative, non-disk contour degenerative findings of the lumbar spine at magnetic resonance (MR) imaging.

## Materials and Methods:

The case accrual method used to perform this institutional review board-approved, HIPAA-compliant retrospective study was the random selection of 111 interpretable MR examination cases of subjects from the Spine Patient Outcomes Research Trial. The subjects were aged 18–87 years (mean, 53 years  $\pm$  16 [standard deviation]). Four independent readers rated the cases according to defined criteria. A subsample of 40 MR examination cases was selected for reevaluation at least 1 month later. The following findings were assessed: spondylolisthesis, disk degeneration, marrow endplate abnormality (Modic changes), posterior anular hyperintense zone (HIZ), and facet arthropathy. Inter- and intraobserver agreement in rating the data was summarized by using weighted  $\kappa$  statistics.

## Results:

Interobserver agreement was good ( $\kappa = 0.66$ ) in rating disk degeneration and moderate in rating spondylolisthesis ( $\kappa = 0.55$ ), Modic changes ( $\kappa = 0.59$ ), facet arthropathy ( $\kappa = 0.54$ ), and posterior HIZ ( $\kappa = 0.44$ ). Interobserver agreement in rating the extent of Modic changes was moderate:  $\kappa$  values were 0.43 for determining superior anteroposterior extent, 0.47 for determining superior craniocaudal extent, 0.57 for determining inferior anteroposterior extent, and 0.48 for determining inferior craniocaudal extent. Intraobserver agreement was good in rating spondylolisthesis ( $\kappa = 0.66$ ), disk degeneration ( $\kappa = 0.74$ ), Modic changes ( $\kappa = 0.64$ ), facet arthropathy ( $\kappa = 0.69$ ), and posterior HIZ ( $\kappa = 0.67$ ). Intraobserver agreement in rating the extent of Modic changes was moderate, with  $\kappa$  values of 0.54 for superior anteroposterior, 0.60 for inferior anteroposterior, 0.50 for superior craniocaudal, and 0.60 for inferior craniocaudal extent determinations.

## Conclusion:

The interpretation of general lumbar spine MR characteristics has sufficient reliability to warrant the further evaluation of these features as potential prognostic indicators.

© RSNA, 2008

Supplemental material: <http://radiology.rsna.org/cgi/content/full/2493071999/DC1>

Low back pain is highly prevalent and has substantial socioeconomic implications (1). Radiating leg pain is often caused by spinal abnormalities. Magnetic resonance (MR) imaging is frequently used to examine patients who have low back pain with or without leg pain. The relationships between anatomic abnormalities of the lumbar spine detected at MR imaging, clinical history, and patient outcome are controversial. Previous work has revealed a high prevalence of spine abnormalities in asymptomatic patients (2–7). MR imaging of the spine can depict alterations in both the anatomy (eg, herniation, stenosis) and tissue properties (eg, disk desiccation, reactive marrow changes), which then need to be considered within a clinical context.

In addition to disk contour abnormalities (eg, bulge, herniation), many spine MR–depicted degenerative findings involving the intervertebral disks, bone marrow, neuroforamina, spinal canal, and facet joints exist and may be overlooked or poorly understood by those treating patients with spine conditions (8). These structures contain no-

ciceptors and thus may be pain generators that contribute to a clinical condition (eg, lumbago, neurogenic claudication, or sciatica) and have the potential to help predict a good or poor outcome of treatments (eg, surgery). Variations in rates of surgery may be related in part to substantial variability among physicians in interpreting the abnormalities identified with advanced radiologic techniques (9).

One objective of the Spine Patient Outcomes Research Trial (SPORT) is to improve treatment recommendations and health outcomes for patients with chronic low back pain by assessing the role of lumbar spine MR imaging findings as indicators of the prognosis. Some imaging features of the lumbar spine have been shown to be useful predictors of diagnoses. The MR imaging findings of intervertebral disk anular hyperintense zone (HIZ) (10) and degenerative marrow changes (11) are associated with positive concordant diskography results. Bone scintigraphy with single photon emission computed tomography is used to predict response to facet injections (12). We considered that MR imaging findings—either alone or in combination with clinical parameters—might have the potential to serve as predictors of outcome. In medical prediction models, clinical parameters often dominate test findings, and this may reflect a lack of granularity in rating diagnostic tests such as imaging. The feature-rating data reported by radiologists may be merged into a prediction model to improve diagnostic accuracy, make a prognosis, or quantitate risk. Methodology standards should be followed to generate high-quality validated and robust prediction models (13). To effectively use the presence of a finding as a predictor of outcome, those interpreting the images must reliably assess the finding. One reason that a model

might lose its predictive power is the incorrect assessment of features (predictors), which causes the inputs in the prediction model to be faulty.

It is well known in radiology that observer performance can be an important source of variability in imaging-based diagnoses. Some prior work regarding observer performance in the interpretation of lumbosacral spine MR imaging data has been done in a variety of settings involving intervertebral disk and other abnormalities (14–19). Results of these prior investigations suggest that the reliability of characterizing non-disk contour lumbar spine MR imaging findings is reasonable, and we considered whether these might serve as predictors of outcome. However, these prior studies were limited by their focus on only one finding (14,16–19), small sample size of fewer than 25 MR examination cases (15), and having fewer than four readers (14,17–19). These investigations were also focused primarily on clinical diagnostic work, whereas we were interested in the effectiveness of MR imaging findings as potential predictors of outcome. Thus, our purpose in this investigation was to characterize the inter- and intraobserver variability of qualitative, non-disk contour degenerative findings of the lumbar spine at MR imaging.

#### Advances in Knowledge

- With use of a sample of 111 baseline MR examination cases collected to evaluate the reliability of non-disk contour degenerative findings at MR imaging, interobserver agreement was good in rating disk degeneration and moderate in rating spondylolisthesis, Modic-type degenerative changes, facet arthropathy, and posterior hyperintense zone.
- The intraobserver agreement based on 40 MR examination cases was good for rating spondylolisthesis, disk degeneration, Modic-type degenerative changes, facet arthropathy, and posterior HIZ.
- Inter- and intraobserver agreement was moderate for rating the superior anteroposterior, inferior anteroposterior, superior cranio-caudal, and inferior cranio-caudal extents of Modic-type changes.

#### Implication for Patient Care

- Non-disk contour degenerative findings of the spine at MR imaging have sufficient reliability to be further evaluated as potential prognostic indicators.

#### Published online before print

10.1148/radiol.2493071999

Radiology 2009; 250:161–170

#### Abbreviations:

HIZ = hyperintense zone

SPORT = Spine Patient Outcomes Research Trial

#### Author contributions:

Guarantors of integrity of entire study, J.A.C., J.D.L., A.N.A.T.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; manuscript final version approval, all authors; literature research, J.A.C., J.D.L., E.J.C., J.N.W.; clinical studies, J.A.C., J.D.L., E.J.C., J.K., J.N.W., R.H.; statistical analysis, A.N.A.T., T.D.T., M.R.G., E.B.; and manuscript editing, J.A.C., J.D.L., A.N.A.T., E.J.C., M.R.G., L.H.P., J.N.W., R.H.

Authors stated no financial relationship to disclose.

See also the editorial by Jarvik and Deyo in this issue.

## Materials and Methods

This retrospective study was approved by the Committee for Protection of Human Subjects at Dartmouth College and was Health Insurance Portability and Accountability Act compliant. Informed subject consent was waived for this investigation, in which deidentified data from the SPORT were used.

### Case Accrual and Image Acquisition

The SPORT is an ongoing multicenter randomized clinical trial funded by the National Institute of Arthritis and Musculoskeletal and Skin Diseases, with a goal of comparing the outcomes between patients who undergo surgery or nonsurgical treatment for the most common spine diagnoses: intervertebral disk herniation, spinal stenosis, and degenerative spondylolisthesis. These abnormalities comprise the diagnostic subgroups of SPORT. The nonsurgical treatment protocol involved the usual care recommended, which included, at least, active physical therapy, education and/or counseling with home exercise instruction, and nonsteroidal antiinflammatory drugs if they could be tolerated. Nonsurgical treatments were individualized for each patient and were tracked prospectively. Thirteen spine centers participated in the research study. SPORT-eligible participants had symptoms and confirmatory signs of lumbar radiculopathy that persisted for at least 6 weeks—with disk herniation at a corresponding level and side at imaging—or they had neurogenic claudication or radicular leg pain for 12 weeks, with imaging revealing spinal stenosis with or without degenerative spondylolisthesis. All patients were considered surgical candidates by the enrolling physician.

Exclusion criteria included prior lumbar surgery, cauda equina syndrome, scoliosis involving a vertebral column curvature of greater than 15°, vertebral fractures, spinal infection or tumor, inflammatory spondyloarthropathy, pregnancy, comorbid conditions contraindicating surgery, or inability or unwillingness to undergo surgery within 6 months. The inclusion and exclusion

criteria are also described in detail elsewhere (20).

The baseline MR imaging examination cases of subjects enrolled in the SPORT were used in this investigation to evaluate finding reliability. The cases were randomly sampled from all patients in the intervertebral disk herniation, spinal stenosis, and degenerative spondylolisthesis diagnosis subgroups of the SPORT who had available images by using pseudorandom numbers generated by statistical software (SPLUS; Insightful, Seattle, Wash). A total of 578 complete MR examination cases were available for this study. The random sampling was performed by the study statistician (T.D.T.). The number of MR examination cases to be reviewed was determined by using a sample size calculation based on the standard error of the  $\kappa$  statistics. For each diagnostic subgroup, it was determined, by using the multireader variance formula of Banerjee (21), that for an imaging characteristic with a prevalence of 44%, an interobserver  $\kappa$  value based on 55 readings by the four readers (J.A.C., E.J.C., R.H., J.K.) had an approximate standard error of 0.028, or a coefficient of variation of less than 5%, for  $\kappa$  values of 0.60 or higher. To accommodate possible losses due to unevaluable images, a total of 120 MR examination cases were selected initially, and 111 of these were usable. The subsamples of 20 herniated disk cases and 20 spondylolisthesis or stenosis cases were selected randomly to study intraobserver agreement by using the same procedure.

The subjects' demographic data and diagnoses are listed in Table 1, and specific imaging findings in groups of patients based on the original interpretations of the enrolling clinicians are listed in Table 2. The mean age of the subjects was 53 years, about half of them were women, most of them were white and non-Hispanic, most had nerve root tension signs and neurologic deficits, and their mean Oswestry Disability score at baseline was 46. These characteristics were generally similar to those of the subjects in the overall SPORT group; thus, the sample was believed to be rep-

resentative of the entire SPORT population.

Initially, each observer judged image quality to be good, fair, or inadequate for interpretation. These definitions were subjective, without standardization, and based on the experienced observers' assessments. The mean time to reinterpretation was 122 days (range, 54–372 days) for reader A, 105 days (range, 24–299 days) for reader B, 103 days (range, 42–301 days) for reader C, and 107 days (range, 57–324 days) for reader D.

Lumbar spine MR images were acquired from the clinical practices of each trial site; as a result, a variety of MR imaging techniques were performed. Sagittal T1-weighted spin-echo (400–600/8–14 [repetition time msec/echo time msec]), sagittal T2-weighted fast spin-echo or turbo spin-echo (3372–5300/60–120 [repetition time msec/effective echo time msec], with [ $n = 37$ ] or without [ $n = 71$ ] fat suppression, echo train length of 10–18), and axial T2-weighted fast spin-echo (2400–3500/60–120 [repetition time msec/effective echo time msec], echo train length of eight to 16) lumbar spine images were acquired by using magnets operating at a field strength of 1.5 T. The following parameters were commonly used to perform these sequences: 90° flip angle, two to four acquired signals, 256 × 192 matrix, and 3.5–4.0-mm section thickness with a 0.5–1.0-mm intersection gap. The field of view was 24–32 cm for sagittal images and 20–24 cm for axial images. The images were collected electronically and stored directly as DICOM (Digital Imaging and Communications in Medicine) files or collected as printed film hard copies and then digitized by using a high-definition scanner and stored in the DICOM format. All images were deidentified for patient confidentiality.

### Observers

The observers for the qualitative and semiquantitative evaluations were four physicians experienced in spine MR image interpretation: three musculoskeletal radiologists and one orthopedic spine surgeon. Two of the musculoskel-

etal radiologists (R.H., J.K.) and the orthopedic spine surgeon (E.J.C.) had more than 25 years experience reading MR images, and one musculoskeletal radiologist (J.A.C.) had 12 years experience reading them. Each observer received training and a handbook containing standardized definitions of imaging characteristics, with pictorial and diagrammatic examples (Appendix E1, <http://radiology.rsnajnl.org/cgi/content/full/2493071999/DC1>). Definitions were derived from the literature or in consensus. Prior to study initiation, the readers evaluated a sample set of images and held an in-person meeting to review them and refine the standardized definitions.

### Image Evaluation

The non-disk contour degenerative spine MR findings assessed were spondylolisthesis, disk degeneration, marrow endplate abnormality (Modic changes), intervertebral disk posterior annular HIZ, and facet arthropathy (ie, osteoarthritis). All MR images were

presented to the observers on compact disks created by using eFilm Lite software (Merge Technologies, Milwaukee, Wis). The types and numbers of display monitors used were not standardized across the readers. The image interpretations were recorded by using a standardized data collection form on which the reader was prompted to select from multiple-choice lists of imaging findings at each intervertebral disk level. The MR examination cases to be reviewed were prepared in monthly batches of approximately 12 cases and included cases from each diagnostic subgroup of the SPORT (ie, intervertebral disk herniation, spinal stenosis, and degenerative spondylolisthesis). The MR imaging findings reviewed are described in the following paragraphs. The training materials from the MR imaging review handbook for our study that were used for this investigation are also provided in Appendix E1 (<http://radiology.rsnajnl.org/cgi/content/full/2493071999/DC1>).

Spondylolisthesis was considered to

be present when a displacement of greater than 1 mm was identified at the intervertebral disk level and was classified as anterolisthesis (forward slip) or retrolisthesis (backward slip) on the basis of the position of the upper (cephalic) vertebra with respect to the lower (caudal) vertebra.

Disk degeneration was graded on a five-point ordinal scale, as described by Pfirrmann et al (14). In grade I disk degeneration, the nucleus pulposus is homogeneous and has high signal intensity (bright white), there is clear distinction of the nucleus and the anulus, and the height of the intervertebral disk is normal. Grade II degeneration has the same features as grade I degeneration, with the exception that the nucleus pulposus is inhomogeneous, with or without horizontal bands. In grade III degeneration, the nucleus pulposus is inhomogeneous and the distinction of the nucleus and the anulus remains clear; however, the nucleus pulposus signal intensity is intermediate and the height of the intervertebral disk is normal to slightly decreased. In grade IV degeneration, the nucleus pulposus is inhomogeneous with low to intermediate signal intensity, distinction of the nucleus and the anulus is lost, and the height of the intervertebral disk is normal to moderately decreased. In grade V degeneration, the nucleus pulposus is hypointense (black) and either inhomogeneous or homogeneous, distinction of the nucleus and the anulus is lost, and the height of the intervertebral disk indicates a collapsed disk space.

Endplate marrow abnormality was designated by using the Modic classification system (22). Modic 1 refers to edema-like signal intensity, Modic 2 refers to fatlike signal intensity, and Modic 3 refers to sclerosis-like signal intensity. Each vertebral level was scored by using the dominant Modic change. Semiquantitative assessment of the extent of alteration in endplate marrow signal intensity was performed for each endplate involved above (inferior) or below (superior) a specified intervertebral disk level. Each observer assessed the degree of signal intensity alteration from the anteroposterior

**Table 1**

#### Baseline Demographic and Clinical Characteristics

Characteristic	All Patients (n = 111)	Patients in Intraobserver Rating Analysis (n = 40)
Age (y)*	53.5 ± 16.2	51.8 ± 16.7
Female	57 (51)	22 (55)
Non-Hispanic	106 (96)	37 (92)
White	93 (84)	31 (78)
Body mass index*	28.5 ± 5.2	29.1 ± 5.5
Diagnosis		
IDH	72 (65)	20 (50)
SPS	57 (51)	13 (32)
DS	25 (22)	7 (18)
Time since recent back pain episode, <6 months	29 (26)	24 (60)
Straight leg raise test, ipsilateral side	42 (38)	15 (38)
Straight leg raise test, contralateral or both sides	8 (7)	4 (10)
Any neurologic deficit	72 (65)	24 (60)
Any pseudoclaudication	54 (49)	20 (50)
Straight leg raising or femoral tension	53 (48)	20 (50)
Any pain radiation	100 (90)	36 (90)
Any neurologic deficit	72 (65)	24 (60)
Body pain score*	29.4 ± 17.5	29.6 ± 16.9
Oswestry Disability score*	45.8 ± 21.2	46.9 ± 21
Stenosis or sciatica bothersomeness score*	14.3 ± 5.9	13.9 ± 6.5

Note.—Unless otherwise noted, data are numbers of patients, with percentages in parentheses. DS = degenerative spondylolisthesis, IDH = intervertebral disk herniation, SPS = spinal stenosis.

\* Data are means ± standard deviations.

( $\leq 50\%$  or  $>50\%$ ) and craniocaudal ( $<25\%$ ,  $25\%$ – $50\%$ , or  $>50\%$ ) extents.

HIZ was defined, on the basis of the original description, as an area of high signal intensity in the posterior anulus that was brighter than the nucleus pulposus depicted on the T2-weighted images (23). In addition to determining whether an HIZ was present or absent, we also noted the location of the zone—if present; therefore, multiple HIZs could be recorded. The location of an HIZ was designated as central, posterolateral, foraminal, or multiple. The presence of an anterior HIZ was not assessed. The location of the HIZ was determined by using a combination of the axial and sagittal planes.

The degree of facet osteoarthritis was rated on an ordinal scale of normal, mild, moderate, or severe by using criteria developed by means of group consensus, sample illustrative images, and descriptions in the literature. Instead of using formal definitions from the existing literature, we developed an example atlas for our MR imaging review handbook (Appendix E1, <http://radiology.rsnajnl.org/cgi/content/full/2493071999/DC1>).

### Statistical Analyses

Frequency distributions of the assessed imaging characteristics were calculated according to reader. These distributions of rating categories were compared for systematic differences in the use of particular categories among readers by using a conditional logistic regression that was based on dichotomized classifications, matching according to case, and the inclusion of reader as a classification variable. *P* values were calculated by using Wald statistics for the reader variable.

$\kappa$  Statistics (24) were used to summarize the intra- and interobserver reliability in rating the MR imaging data and were calculated by using linear weights to give more importance to disagreements that were farther apart on an ordinal scale. Intrareader  $\kappa$  values were calculated individually for each reader, and interreader  $\kappa$  values were calculated for each reader pair. Disk level was used as the unit of analysis. To

**Table 2**

#### Baseline Imaging Parameters Based on Original Interpretations of Enrolling Clinicians

Parameter	All Patients ( <i>n</i> = 111)	Patients in Intraobserver Rating Analysis ( <i>n</i> = 40)
<b>Disk level of spondylolisthesis</b>		
L3-4	5 (4)	0 (0)
L4-5	24 (22)	7 (18)
<b>Disk level of herniation</b>		
L2-3 or L3-4	5 (4)	0 (0)
L4-5	24 (22)	6 (15)
L5-S1	28 (25)	14 (35)
<b>Herniation type</b>		
Protruding	14 (13)	5 (12)
Extruding	39 (35)	15 (38)
Sequestered	4 (4)	0 (0)
Posterolateral herniation	46 (41)	19 (48)
<b>Disk level of stenosis</b>		
L2-3	6 (5)	2 (5)
L3-4	27 (24)	13 (32)
L4-5	52 (47)	19 (48)
L5-S1	8 (7)	4 (10)
<b>No. of levels with moderate or severe stenosis</b>		
None	1 (1)	0 (0)
One	29 (26)	9 (22)
Two	17 (15)	8 (20)
Three or more	7 (6)	3 (8)
<b>Stenosis location</b>		
Central	43 (39)	16 (40)
Lateral recess	50 (45)	18 (45)
Neuroforamen	21 (19)	7 (18)
<b>Stenosis severity</b>		
Mild	1 (1)	0
Moderate	21 (19)	7 (18)
Severe	32 (29)	13 (32)

Note.—Data are numbers of patients, with percentages in parentheses.

accommodate the use of multiple levels per individual, overall interreader  $\kappa$  values and 95% confidence intervals were calculated by using the bootstrap technique with 1000 samples each, with 111 MR examinations taken with replacement from the individual image records including all levels.

A weighted average of the pairwise  $\kappa$  values was calculated by using weights based on the estimated standard errors. The mean of the bootstrapped weighted averages was used as the estimate, and confidence intervals were determined by using the quantiles of the bootstrap distribution. For intrareader  $\kappa$  analysis, the images used included those that were evaluable at the first and second readings. The bootstrap procedure was

implemented by using 1000 samples each, with 40 MR examinations taken with replacement from the individual image records used in the reliability study. A stratified estimate of the overall weighted intrareader  $\kappa$  value was calculated at each bootstrap iteration. The strength of agreement was interpreted on the basis of the  $\kappa$  values suggested by Altman (25), as adapted from the method of Landis and Koch (24):  $\kappa$  Values of 0.81–1.00 indicated very good agreement; 0.61–0.80, good agreement; 0.41–0.60, moderate agreement; 0.21–0.40, fair agreement; and 0.20 or lower, poor agreement.

Because there were variations in both image acquisition technique and image format, which may contribute to

image variance to the extent that they affect reader interpretations, we derived agreement statistics ( $\kappa$  values) that were stratified according to a key acquisition parameter (sagittal T2-weighted fat-suppressed vs non-fat-suppressed images) and a key image format (digitized vs native digital) for inter- and intraobserver reliability.

### Results

For interobserver analysis, 111 MR examination cases were rated by all four readers. Forty of these cases were read twice for assessment of intraobserver agreement, with varying numbers of images reported as being evaluable by the reader: 32 images were evaluable for reader A; 35 images, for reader B; 38 images, for reader C; and 36 images, for reader D.

The frequency distributions of the four observers' ratings of the lumbar spine MR findings spondylolisthesis, intervertebral disk posterior annular HIZ (ie, posterior HIZ), disk degeneration, marrow endplate abnormality (Modic changes), and facet osteoarthritis are shown in the Figure. There were similar distributions across the readers for posterior HIZ and Modic changes. At Wald testing, the distributions were significantly different across readers for spondylolisthesis ( $P = .009$ ) and facet arthropathy ( $P = .006$ ), borderline significantly different for disk degeneration ( $P = .055$ ), and not significantly different for posterior HIZ ( $P = .22$ ) and Modic changes ( $P = .52$ ).

### Interobserver Reliability

$\kappa$  Values for both overall interobserver agreement and interobserver agreement for each reader, with corresponding 95% confidence intervals, are shown in Table 3. Overall, there was good agreement in rating disk degeneration ( $\kappa = 0.66$ ) and moderate agreement in rating spondylolisthesis ( $\kappa = 0.55$ ), Modic changes ( $\kappa = 0.59$ ), facet arthropathy ( $\kappa = 0.54$ ), and posterior HIZ ( $\kappa = 0.44$ ). Regarding the extent of Modic changes, overall agreement was moderate in determining the superior anteroposterior ( $\kappa = 0.43$ ), inferior

anteroposterior ( $\kappa = 0.57$ ), superior craniocaudal ( $\kappa = 0.47$ ), and inferior craniocaudal ( $\kappa = 0.48$ ) extents. For ordinal findings (ie, disk degeneration and facet arthropathy), there was typically a difference of only one grade between the observers.

### Intraobserver Reliability

$\kappa$  Values for both overall intraobserver agreement and intraobserver agreement for each reader, with corresponding 95% confidence intervals, are shown in Table 4. Overall, there was good agreement in the rating of most MR findings: spondylolisthesis ( $\kappa = 0.66$ ), disk degeneration ( $\kappa = 0.74$ ), Modic changes ( $\kappa = 0.64$ ), facet arthropathy ( $\kappa = 0.69$ ), and posterior HIZ ( $\kappa = 0.67$ ). Regarding the extent of Modic changes, overall agreement was moderate for determining the superior anteroposterior ( $\kappa = 0.54$ ), inferior anteroposterior ( $\kappa = 0.60$ ), superior craniocaudal ( $\kappa = 0.50$ ), and inferior craniocaudal ( $\kappa = 0.60$ ) extents.

Stratified analyses performed according to specific parameter subsets (sagittal T2-weighted image type and image format type) revealed no qualitative differences between inter- and intraobserver reliability data and overall agreement  $\kappa$  values (results not shown). Therefore, no substantial variability was introduced, regardless of whether fat-suppressed sagittal T2-weighted images were acquired or whether the images were stored as digitized film hard copies.

### Discussion

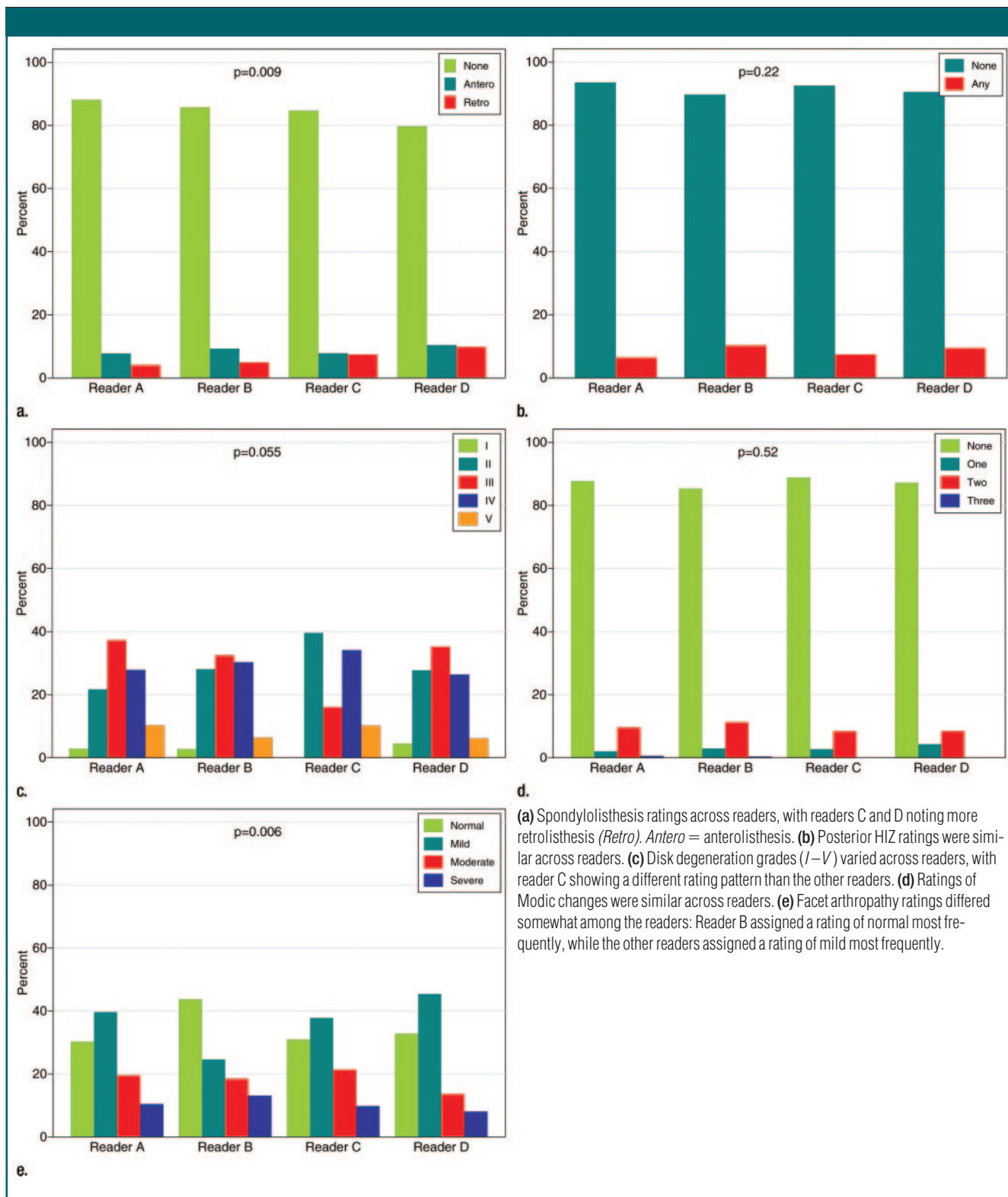
In this study, we used a sample of baseline MR examination cases of subjects enrolled in the SPORT. The images were rated by four independent readers according to defined criteria for non-disk contour-related degenerative MR findings, and a subgroup of the images were rerated by the same readers. We found that typically, intraobserver reliability was good and interobserver reliability was moderate. Thus, some variability existed between readers despite standardized definitions and reader training. For

the ordinal findings (ie, disk degeneration and facet arthropathy), there was typically a difference of only one grade between observers.

Several prior studies of observer variability have been performed. Mulconrey et al achieved good overall interobserver agreement ( $\kappa = 0.728$ ) among four readers in rating 17 lumbar spine MR examination cases (at 80 levels) for the presence of spondylolisthesis (15), which was substantially better than our moderate agreement. The system that we used to grade disk degeneration was initially described and tested with three readers who reviewed 60 MR examination cases (14). In that study, the  $\kappa$  coefficients were moderate to good ( $\kappa = 0.69$ – $0.81$ ) for interobserver agreement and good to very good ( $\kappa = 0.84$ – $0.90$ ) for intraobserver agreement. By using a simplified dichotomous definition of disk degeneration, Mulconrey and colleagues (15) achieved good agreement ( $\kappa = 0.773$ ) overall among four observers. These values are similar to the good interobserver and intraobserver agreement that we achieved in grading disk degeneration.

There have been several investigations of observer performance in the evaluation of marrow endplate abnormalities (Modic changes). In a study by Jones et al (16), five independent observers of differing experience read 50 MR examination cases (one level evaluated per MR image) and found overall interobserver agreement to be very good ( $\kappa = 0.85$ ) and intraobserver agreement to be good or very good ( $\kappa = 0.71$ – $1.00$ ). These values are higher than those that we derived. However, a selection bias could have been introduced in that study because the patients were preselected on the basis of the MR imaging features with only two normal levels present. In addition, only one interobserver agreement value was reported, without mention of pairwise comparisons; thus, it is unclear what this value represented.

In another study, Peterson et al (17) achieved moderate interobserver agreement ( $\kappa = 0.52$ ) overall and good intraobserver agreement ( $\kappa = 0.71$  and  $0.87$ ) overall for two readers in rating



(a) Spondylolisthesis ratings across readers, with readers C and D noting more retrolisthesis (*Retro*). *Antero* = anterolisthesis. (b) Posterior HIZ ratings were similar across readers. (c) Disk degeneration grades (I–V) varied across readers, with reader C showing a different rating pattern than the other readers. (d) Ratings of Modic changes were similar across readers. (e) Facet arthropathy ratings differed somewhat among the readers: Reader B assigned a rating of normal most frequently, while the other readers assigned a rating of mild most frequently.

51 spine MR examination cases (at 255 levels); these results are similar to our results with four readers. The investigation of Mulconrey et al (15), involving the rating of 17 lumbar spine MR examination cases (at 80 levels) by four readers, revealed good interobserver agreement ( $\kappa = 0.669$ ). Our results for interobserver reliability were lower and borderline good overall ( $\kappa = 0.59$ ). However, Mulconrey et al used a dichotomized classification system (present vs absent), which would be expected to improve concordance. These investigators did not evaluate intraobserver agreement. We extended the work done on Modic changes by also evaluating the extent of these alterations and

achieved moderate inter- and intraobserver agreement.

The reliability of the diagnosis of HIZ has also been evaluated. In the study by Smith et al (18), who evaluated the intervertebral disk posterior annular HIZ and correlated this observation with provocative diskography findings, two readers reviewed 175 MR examination cases without repeat readings. Interobserver agreement in rating the high-signal-intensity zone in given disks was moderate ( $\kappa = 0.57$ ), with a 95% confidence interval for  $\kappa$  values of 0.44 (fair) to 0.70 (good), which was lower than the interobserver agreement between two observers originally reported by Aprill

and Bogduck (23): 98% (66/67). Our interobserver agreement (moderate) and intraobserver agreement (good) results are similar to those of Smith et al (18).

In a study of imaging facet arthropathy by Weishaupt et al (19), two musculoskeletal radiologists independently graded the severity of osteoarthritis in 308 lumbar facet joints at MR imaging. These investigators used well-defined criteria. In their study, the normal facet joints showed a uniform joint space of 2–4 mm, without osteophytosis or subchondral bone reaction. Mild (grade 1) osteoarthritis was characterized by narrowing of the facet joint space (<2 mm) and/or small osteophytes, and/or mild

**Table 3**

**Interobserver Variability Measured by Using Weighted  $\kappa$  Statistics and Based on 111 MR Examination Cases**

Finding	Readers A and B	Readers A and C	Readers A and D	Readers B and C	Readers B and D	Readers C and D	Overall
Spondylolisthesis	0.68 (0.56, 0.78)	0.56 (0.45, 0.68)	0.43 (0.32, 0.55)	0.60 (0.49, 0.71)	0.55 (0.45, 0.65)	0.47 (0.36, 0.57)	0.55 (0.48, 0.62)
Disk degeneration grade	0.64 (0.59, 0.70)	0.62 (0.56, 0.69)	0.67 (0.61, 0.72)	0.70 (0.65, 0.75)	0.70 (0.65, 0.76)	0.63 (0.58, 0.69)	0.66 (0.63, 0.70)
Extent of Modic changes*	0.51 (0.37, 0.63)	0.59 (0.44, 0.72)	0.58 (0.45, 0.70)	0.57 (0.42, 0.69)	0.56 (0.42, 0.67)	0.74 (0.63, 0.83)	0.59 (0.49, 0.68)
Superior anteroposterior	0.26 (−0.01, 0.54)	0.44 (0.22, 0.70)	0.25 (0.01, 0.53)	0.54 (0.30, 0.79)	0.56 (0.28, 0.82)	0.64 (0.39, 0.84)	0.43 (0.27, 0.61)
Superior craniocaudal	0.33 (0.11, 0.54)	0.39 (0.20, 0.60)	0.54 (0.32, 0.74)	0.43 (0.20, 0.65)	0.51 (0.23, 0.73)	0.72 (0.53, 0.88)	0.47 (0.33, 0.62)
Inferior anteroposterior	0.60 (0.39, 0.80)	0.71 (0.47, 0.94)	0.49 (0.17, 0.77)	0.42 (0.16, 0.67)	0.54 (0.30, 0.77)	0.67 (0.47, 0.86)	0.57 (0.43, 0.69)
Inferior craniocaudal	0.41 (0.18, 0.61)	0.47 (0.27, 0.64)	0.48 (0.27, 0.67)	0.53 (0.29, 0.78)	0.43 (0.17, 0.66)	0.62 (0.42, 0.80)	0.48 (0.36, 0.60)
No vs any posterior HIZ	0.32 (0.17, 0.48)	0.38 (0.22, 0.55)	0.41 (0.24, 0.57)	0.44 (0.31, 0.57)	0.52 (0.38, 0.64)	0.58 (0.43, 0.71)	0.44 (0.34, 0.55)
Facet arthropathy	0.53 (0.47, 0.59)	0.40 (0.33, 0.47)	0.55 (0.48, 0.61)	0.57 (0.52, 0.62)	0.63 (0.58, 0.68)	0.53 (0.48, 0.59)	0.54 (0.50, 0.57)

Note.—Data are  $\kappa$  statistics for interobserver agreement in rating the given findings, computed by using 1000 bootstrapped samples. Numbers in parentheses are 95% confidence intervals based on 1000 bootstrapped samples.

\* Modic changes extent is contingent on Modic changes being present.

**Table 4**

**Intraobserver Variability Measured by Using Weighted  $\kappa$  Statistics and Based on 40 Repeated Readings**

Finding	Reader A	Reader B	Reader C	Reader D	Overall
Spondylolisthesis	0.74 (0.57, 0.88)	0.72 (0.57, 0.84)	0.54 (0.39, 0.70)	0.67 (0.53, 0.80)	0.66 (0.58, 0.74)
Disk degeneration grade	0.75 (0.65, 0.83)	0.67 (0.57, 0.76)	0.82 (0.76, 0.88)	0.72 (0.65, 0.80)	0.74 (0.68, 0.79)
Extent of Modic changes*	0.56 (0.35, 0.76)	0.59 (0.40, 0.76)	0.67 (0.49, 0.80)	0.73 (0.60, 0.84)	0.64 (0.54, 0.74)
Superior anteroposterior	−0.17 (−0.50, 0)	0.62 (0.21, 1.00)	0.45 (0.02, 0.88)	0.92 (0.75, 1.00)	0.54 (0.33, 0.70)
Superior craniocaudal	0.5 (0, 1.00)	0.43 (0, 0.79)	0.57 (−0.10, 1.00)	0.42 (0, 0.84)	0.50 (0.23, 0.72)
Inferior anteroposterior	0.31 (−0.27, 1.00)	0.64 (0.23, 0.92)	0.59 (0, 1.00)	0.67 (0.27, 0.94)	0.60 (0.31, 0.82)
Inferior craniocaudal	0.49 (−0.09, 1.00)	0.54 (0.19, 0.82)	0.77 (0.44, 1.00)	0.56 (0.17, 0.81)	0.60 (0.40, 0.77)
None vs any posterior HIZ	0.58 (0.20, 0.88)	0.59 (0.40, 0.76)	0.71 (0.49, 0.88)	0.79 (0.58, 0.95)	0.67 (0.53, 0.78)
Facet arthropathy	0.64 (0.56, 0.72)	0.66 (0.57, 0.74)	0.71 (0.63, 0.79)	0.73 (0.66, 0.80)	0.69 (0.64, 0.74)

Note.—Data are  $\kappa$  statistics for intraobserver agreement in rating the given findings, computed by using 1000 bootstrapped samples. Numbers in parentheses are 95% confidence intervals based on 1000 bootstrapped samples.

\* Modic changes extent is contingent on Modic changes being present.

articular process hypertrophy. Moderate (grade 2) osteoarthritis was characterized by narrowing of the facet joint space (<2 mm) and/or moderate osteophytes, and/or moderate articular process hypertrophy and/or mild subarticular bone reaction (erosions). Severe (grade 3) osteoarthritis was characterized by narrowing of the facet joint space (<2 mm) and/or large osteophytes, and/or large articular process hypertrophy and/or severe subarticular bone reaction (erosions and/or cysts). These investigators achieved fair interobserver agreement ( $\kappa = 0.41$ ) and good intraobserver agreement ( $\kappa = 0.70$  and  $0.76$ ). Our results are slightly better: moderate interobserver agreement ( $\kappa = 0.54$ ) and good intraobserver agreement ( $\kappa = 0.69$ ). These results are consistent with our consensus-based grading of normal, mild, moderate, and severe osteoarthritis, which roughly corresponds to the grading scheme used by Weishaupt et al (19).

The observer performance issues identified in this investigation have implications for research, clinical, and training activities. For research activities that involve the use of MR imaging findings in prediction models for diagnostic or prognostic purposes, multiple expert readers often are needed to expedite the data acquisition for inputs into the model. Having a reliable way to assess a predictive parameter across different readers is important for the efficiency of clinical trials. Once tests are out of the research arena, they often lose some performance capability in nonideal nonlaboratory settings. Therefore, reliable assessment of findings is important for clinical effectiveness. The degenerative MR imaging findings of the spine that we investigated will render sufficient reader agreement for use in clinical practice if they are deemed to be predictive of outcome on the basis of future SPORT analysis results. We believe that training, nomenclature (lexicon), image representations (teaching files), and defined relationships (ontology) will promote more uniform detection, characterization, and reporting of important spine MR imaging features. We recommend that a just-in-time

learning apparatus be used by providing imaging galleries of representative cases and illustrative examples that cover the spectrum of abnormalities, as described elsewhere (26). This apparatus could be implemented as an online teaching file and as a decision support tool for clinical work.

This study had a number of limitations. The variability in image acquisition methods introduced some heterogeneity; however, this variability reflects that commonly encountered in routine clinical practice. The nonuniformity of MR imaging protocols also could have introduced some variability; however, all readers had experience interpreting fat-suppressed as well as non-fat-suppressed images. While there was no standardization of the setting or process for performing the readings themselves, we did not see any systematic differences in interreader agreement between reader pairs. It is important to note that our results may be overestimations of the reliability that might be achieved by readers in routine clinical practice due to the methods we used to standardize terminology, including a prestudy face-to-face meeting and a detailed handbook of definitions and examples. We suggest that groups in practice should take particular efforts to clearly standardize the terminology and definitions used to report data.

Given the several cases in which we achieved only moderate agreement in this methodologically rigorous investigation, use of the spine MR imaging features described herein as prognosticators of outcome will be difficult to generalize to routine clinical practice unless decision support tools are used. In addition, the readers were aware that all of the images were from patients who had disk herniation or spinal stenosis severe enough to qualify them as surgical candidates and for entry into the SPORT. The lack of healthy (in terms of spinal status) subjects also may have affected the readings; however, an abundant number of normal as well as abnormal spinal levels were evaluated.

In conclusion, degenerative lumbar spine MR findings have sufficient reliability to potentially be used as predictors of

clinical prognoses and outcomes if a rigorous training paradigm is used. Agreement in rating the majority of these findings was not perfect, even with training and reference images. Thus, for clinical practice, decision support tools such as image galleries and teaching atlases may need to be used to maintain a high level of reliability—that is, to minimize intra- and interobserver variability.

## References

1. Katz JN. Lumbar disc disorders and low-back pain: socioeconomic factors and consequences. *J Bone Joint Surg Am* 2006; 88(suppl 2):21–24.
2. 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(2): 69–73.
3. 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(3):661–666.
4. Stadnik TW, Lee RR, Coen HL, Neirynek EC, Buisseret TS, Osteaux MJ. Annular tears and disk herniation: prevalence and contrast enhancement on MR images in the absence of low back pain or sciatica. *Radiology* 1998; 206(1):49–55.
5. 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(3):403–408.
6. Carragee EJ, Paragioudakis SJ, Khurana S. 2000 Volvo Award winner in clinical studies: lumbar high-intensity zone and discography in subjects without low back problems. *Spine* 2000;25(23):2987–2992.
7. Chung CB, Vande Berg BC, Tavernier T, et al. End plate marrow changes in the asymptomatic lumbosacral spine: frequency, distribution and correlation with age and degenerative changes. *Skeletal Radiol* 2004; 33(7):399–404.
8. Jinkins JR. Acquired degenerative changes of the intervertebral segments at and supra-adjacent to the lumbosacral junction: a radio-anatomic analysis of the nondiscal structures of the spinal column and perispinal soft tissues. *Eur J Radiol* 2004;50(2):134–158.
9. Weinstein JN, Lurie JD, Olson PR, Bronner KK, Fisher ES. United States' trends and

- regional variations in lumbar spine surgery: 1992–2003. *Spine* 2006;31(23):2707–2714.
10. Schellhas KP, Pollei SR, Gundry CR, Heithoff KB. Lumbar disc high-intensity zone: correlation of magnetic resonance imaging and discography. *Spine* 1996;21(1):79–86.
  11. Weishaupt D, Zanetti M, Hodler J, et al. Painful lumbar disk derangement: relevance of endplate abnormalities at MR imaging. *Radiology* 2001;218(2):420–427.
  12. Pneumaticos SG, Chatziioannou SN, Hipp JA, Moore WH, Esses SI. Low back pain: prediction of short-term outcome of facet joint injection with bone scintigraphy. *Radiology* 2006;238(2):693–698.
  13. Carrino JA, Ohno-Machado L. Development of radiology prediction models using feature analysis. *Acad Radiol* 2005;12(4):415–421.
  14. Pfirrmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N. Magnetic resonance classification of lumbar intervertebral disc degeneration. *Spine* 2001;26(17):1873–1878.
  15. Mulconrey DS, Knight RQ, Bramble JD, Paknikar S, Harty PA. Interobserver reliability in the interpretation of diagnostic lumbar MRI and nuclear imaging. *Spine J* 2006;6(2):177–184.
  16. Jones A, Clarke A, Freeman BJ, Lam KS, Grevitt MP. The Modic classification: inter- and intraobserver error in clinical practice. *Spine* 2005;30(16):1867–1869.
  17. Peterson CK, Gatterman B, Carter JC, Humphreys BK, Weibel A. Inter- and intraexaminer reliability in identifying and classifying degenerative marrow (Modic) changes on lumbar spine magnetic resonance scans. *J Manipulative Physiol Ther* 2007;30(2):85–90.
  18. Smith BM, Hurwitz EL, Solsberg D, et al. Interobserver reliability of detecting lumbar intervertebral disc high-intensity zone on magnetic resonance imaging and association of high-intensity zone with pain and anular disruption. *Spine* 1998;23(19):2074–2080.
  19. Weishaupt D, Zanetti M, Boos N, Hodler J. MR imaging and CT in osteoarthritis of the lumbar facet joints. *Skeletal Radiol* 1999;28(4):215–219.
  20. Birkmeyer NJ, Weinstein JN, Tosteson AN, et al. Design of the Spine Patient Outcomes Research Trial (SPORT). *Spine* 2002;27(12):1361–1372.
  21. Banerjee M. Beyond kappa: a review of interrater agreement measures. *Can J Stat* 1999;27(1):3–23.
  22. Modic MT, Steinberg PM, Ross JS, Masaryk TJ, Carter JR. Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging. *Radiology* 1988;166(1 pt 1):193–199.
  23. Aprill C, Bogduk N. High-intensity zone: a diagnostic sign of painful lumbar disk on magnetic resonance imaging. *Br J Radiol* 1992;65(773):361–369.
  24. Landis JR, Koch GG. Measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159–174.
  25. Altman D. *Practical statistics for medical research*. Boca Raton, Fla: Chapman & Hall-CRC, 1991.
  26. Swett HA, Fisher PR, Cohn AI, Miller PL, Mutalik PG. Expert system-controlled image display. *Radiology* 1989;172(2):487–493.

# Radiology 2008

## This is your reprint order form or pro forma invoice

(Please keep a copy of this document for your records.)

Reprint order forms and purchase orders or prepayments must be received 72 hours after receipt of form either by mail or by fax at 410-820-9765. It is the policy of Cadmus Reprints to issue one invoice per order.

**Please print clearly.**

Author Name \_\_\_\_\_  
Title of Article \_\_\_\_\_  
Issue of Journal \_\_\_\_\_ Reprint # \_\_\_\_\_ Publication Date \_\_\_\_\_  
Number of Pages \_\_\_\_\_ KB # \_\_\_\_\_ Symbol Radiology  
Color in Article? Yes / No (Please Circle)

**Please include the journal name and reprint number or manuscript number on your purchase order or other correspondence.**

### Order and Shipping Information

#### Reprint Costs (Please see page 2 of 2 for reprint costs/fees.)

\_\_\_\_\_ Number of reprints ordered \$ \_\_\_\_\_  
\_\_\_\_\_ Number of color reprints ordered \$ \_\_\_\_\_  
\_\_\_\_\_ Number of covers ordered \$ \_\_\_\_\_  
**Subtotal** \$ \_\_\_\_\_  
Taxes \$ \_\_\_\_\_

*(Add appropriate sales tax for Virginia, Maryland, Pennsylvania, and the District of Columbia or Canadian GST to the reprints if your order is to be shipped to these locations.)*

First address included, add \$32 for  
each additional shipping address \$ \_\_\_\_\_

**TOTAL** \$ \_\_\_\_\_

#### Shipping Address (cannot ship to a P.O. Box) Please Print Clearly

Name \_\_\_\_\_  
Institution \_\_\_\_\_  
Street \_\_\_\_\_  
City \_\_\_\_\_ State \_\_\_\_\_ Zip \_\_\_\_\_  
Country \_\_\_\_\_  
Quantity \_\_\_\_\_ Fax \_\_\_\_\_  
Phone: Day \_\_\_\_\_ Evening \_\_\_\_\_  
E-mail Address \_\_\_\_\_

#### Additional Shipping Address\* (cannot ship to a P.O. Box)

Name \_\_\_\_\_  
Institution \_\_\_\_\_  
Street \_\_\_\_\_  
City \_\_\_\_\_ State \_\_\_\_\_ Zip \_\_\_\_\_  
Country \_\_\_\_\_  
Quantity \_\_\_\_\_ Fax \_\_\_\_\_  
Phone: Day \_\_\_\_\_ Evening \_\_\_\_\_  
E-mail Address \_\_\_\_\_

\* Add \$32 for each additional shipping address

### Payment and Credit Card Details

**Enclosed:** Personal Check \_\_\_\_\_  
Credit Card Payment Details \_\_\_\_\_  
Checks must be paid in U.S. dollars and drawn on a U.S. Bank.  
Credit Card:  VISA  Am. Exp.  MasterCard  
Card Number \_\_\_\_\_  
Expiration Date \_\_\_\_\_  
Signature: \_\_\_\_\_

Please send your order form and prepayment made payable to:

**Cadmus Reprints**  
**P.O. Box 751903**  
**Charlotte, NC 28275-1903**

*Note: Do not send express packages to this location, PO Box.*  
FEIN #:541274108

Signature \_\_\_\_\_ Date \_\_\_\_\_

Signature is required. By signing this form, the author agrees to accept the responsibility for the payment of reprints and/or all charges described in this document.

### Invoice or Credit Card Information

#### Invoice Address Please Print Clearly

Please complete Invoice address as it appears on credit card statement

Name \_\_\_\_\_  
Institution \_\_\_\_\_  
Department \_\_\_\_\_  
Street \_\_\_\_\_  
City \_\_\_\_\_ State \_\_\_\_\_ Zip \_\_\_\_\_  
Country \_\_\_\_\_  
Phone \_\_\_\_\_ Fax \_\_\_\_\_  
E-mail Address \_\_\_\_\_

**Cadmus will process credit cards and Cadmus Journal  
Services will appear on the credit card statement.**

*If you don't mail your order form, you may fax it to 410-820-9765 with  
your credit card information.*

# Radiology 2008

## Black and White Reprint Prices

Domestic (USA only)						
# of Pages	50	100	200	300	400	500
1-4	\$221	\$233	\$268	\$285	\$303	\$323
5-8	\$355	\$382	\$432	\$466	\$510	\$544
9-12	\$466	\$513	\$595	\$652	\$714	\$775
13-16	\$576	\$640	\$749	\$830	\$912	\$995
17-20	\$694	\$775	\$906	\$1,017	\$1,117	\$1,220
21-24	\$809	\$906	\$1,071	\$1,200	\$1,321	\$1,471
25-28	\$928	\$1,041	\$1,242	\$1,390	\$1,544	\$1,688
29-32	\$1,042	\$1,178	\$1,403	\$1,568	\$1,751	\$1,924
Covers	\$97	\$118	\$215	\$323	\$442	\$555

## Color Reprint Prices

Domestic (USA only)						
# of Pages	50	100	200	300	400	500
1-4	\$223	\$239	\$352	\$473	\$597	\$719
5-8	\$349	\$401	\$601	\$849	\$1,099	\$1,349
9-12	\$486	\$517	\$852	\$1,232	\$1,609	\$1,992
13-16	\$615	\$651	\$1,105	\$1,609	\$2,117	\$2,624
17-20	\$759	\$787	\$1,357	\$1,997	\$2,626	\$3,260
21-24	\$897	\$924	\$1,611	\$2,376	\$3,135	\$3,905
25-28	\$1,033	\$1,071	\$1,873	\$2,757	\$3,650	\$4,536
29-32	\$1,175	\$1,208	\$2,122	\$3,138	\$4,162	\$5,180
Covers	\$97	\$118	\$215	\$323	\$442	\$555

International (includes Canada and Mexico)						
# of Pages	50	100	200	300	400	500
1-4	\$272	\$283	\$340	\$397	\$446	\$506
5-8	\$428	\$455	\$576	\$675	\$784	\$884
9-12	\$580	\$626	\$805	\$964	\$1,115	\$1,278
13-16	\$724	\$786	\$1,023	\$1,232	\$1,445	\$1,652
17-20	\$878	\$958	\$1,246	\$1,520	\$1,774	\$2,030
21-24	\$1,022	\$1,119	\$1,474	\$1,795	\$2,108	\$2,426
25-28	\$1,176	\$1,291	\$1,700	\$2,070	\$2,450	\$2,813
29-32	\$1,316	\$1,452	\$1,936	\$2,355	\$2,784	\$3,209
Covers	\$156	\$176	\$335	\$525	\$716	\$905

International (includes Canada and Mexico))						
# of Pages	50	100	200	300	400	500
1-4	\$278	\$290	\$424	\$586	\$741	\$904
5-8	\$429	\$472	\$746	\$1,058	\$1,374	\$1,690
9-12	\$604	\$629	\$1,061	\$1,545	\$2,011	\$2,494
13-16	\$766	\$797	\$1,378	\$2,013	\$2,647	\$3,280
17-20	\$945	\$972	\$1,698	\$2,499	\$3,282	\$4,069
21-24	\$1,110	\$1,139	\$2,015	\$2,970	\$3,921	\$4,873
25-28	\$1,290	\$1,321	\$2,333	\$3,437	\$4,556	\$5,661
29-32	\$1,455	\$1,482	\$2,652	\$3,924	\$5,193	\$6,462
Covers	\$156	\$176	\$335	\$525	\$716	\$905

Minimum order is 50 copies. For orders larger than 500 copies, please consult Cadmus Reprints at 800-407-9190.

### Reprint Cover

Cover prices are listed above. The cover will include the publication title, article title, and author name in black.

### Shipping

Shipping costs are included in the reprint prices. Domestic orders are shipped via UPS Ground service. Foreign orders are shipped via a proof of delivery air service.

### Multiple Shipments

Orders can be shipped to more than one location. Please be aware that it will cost \$32 for each additional location.

### Delivery

Your order will be shipped within 2 weeks of the journal print date. Allow extra time for delivery.

### Tax Due

Residents of Virginia, Maryland, Pennsylvania, and the District of Columbia are required to add the appropriate sales tax to each reprint order. For orders shipped to Canada, please add 7% Canadian GST unless exemption is claimed.

### Ordering

Reprint order forms and purchase order or prepayment is required to process your order. Please reference journal name and reprint number or manuscript number on any correspondence. You may use the reverse side of this form as a proforma invoice. Please return your order form and prepayment to:

**Cadmus Reprints**  
P.O. Box 751903  
Charlotte, NC 28275-1903

*Note: Do not send express packages to this location, PO Box. FEIN #:541274108*

Please direct all inquiries to:

**Rose A. Baynard**  
800-407-9190 (toll free number)  
410-819-3966 (direct number)  
410-820-9765 (FAX number)  
[baynardr@cadmus.com](mailto:baynardr@cadmus.com) (e-mail)

**Reprint Order Forms and purchase order or prepayments must be received 72 hours after receipt of form.**