Published in final edited form as:
Spine (Phila Pa 1976). 2009 April 1; 34(7): 701–705. doi:10.1097/BRS.0b013e31819b390e.

Magnetic Resonance Imaging Interpretation in Patients With Symptomatic Lumbar Spine Disc Herniations:

Comparison of Clinician and Radiologist Readings

Jon D. Lurie, MD, MS[*], **David M. Doman, MD**[†], **Kevin F. Spratt, PhD**[‡], **Anna N. A. Tosteson, ScD**[*], and **James N. Weinstein, DO, MS**[‡]

[*]Department of Medicine, Hanover, NH

[†]Dartmouth Medical School, Hanover, NH

[‡]Department of Orthopaedics; Hanover, NH

Abstract

Study Design—Retrospective review of imaging data from a clinical trial.

Objective—To compare the interpretation of lumbar spine magnetic resonance imaging (MRIs) by clinical spine specialists and radiologists in patients with lumbar disc herniation.

Summary of Background Data—MRI is the imaging modality of choice for evaluation of the lumbar spine in patients with suspected lumbar disc herniation. Guidelines provide standardization of terms to more consistently describe disc herniation. The extent to which these guidelines are being followed in clinical practice is unknown.

Methods—We abstracted data from radiology reports from patients with lumbar intervertebral disc herniation enrolled in the Spine Patient Outcomes Research Trial. We evaluated the frequency with which morphology (*e.g.*, protrusions, extrusions, or sequestrations) was reported as per guidelines and when present we compared the morphology ratings to those of clinicians who completed a structured data form as part of the trial. We assessed agreement using percent agreement and the κ statistic.

Results—There were 396 patients with sufficient data to analyze. Excellent agreement was observed between clinician and radiologist on the presence and level of herniation (93.4%), with 3.3% showing disagreement regarding level, of which a third could be explained by the presence of a transitional vertebra. In 3.3% of the cases in which the clinician reported a herniation (protrusion, extrusion, or sequestration), the radiologist reported no herniation on the MRI.

The radiology reports did not clearly describe morphology in 42.2% of cases. In the 214 cases with clear morphologic descriptions, agreement was fair ($\kappa = 0.24$) and the disagreement was asymmetric (Bowker's test of symmetry $P < 0.0001$) with clinicians more often rating more abnormal morphologic categories. Agreement on axial location of the herniation was excellent ($\kappa = 0.81$). There was disagreement between left or right side in only 3.3% of cases ($\kappa = 0.93$).

Conclusion—Radiology reports frequently fail to provide sufficient detail to describe disc herniation morphology. Agreement between MRI readings by clinical spine specialists and

Address correspondence and reprint requests to Jon D. Lurie, MD, MS, Dartmouth Hitchcock Medical Center, 1 Medical Center Drive, HB 7999, Lebanon, NH 03756..

radiologists was excellent when comparing herniation vertebral level and location within level, but only fair comparing herniation morphology.

Keywords

herniated disc; MRI; SPORT; reliability; imaging

MRI is the imaging modality of choice for evaluation of suspected lumbar disc herniation. Although MRI provides excellent anatomic detail, the relationship between pathoanatomy and clinical symptoms is controversial. It can be difficult to know which details of the anatomic picture are important, and what findings are more likely than others to manifest clinically.

Several studies[1-4] have shown that MRIs of asymptomatic patients have a high prevalence of bulges and protrusions in the lumbar spine, but extrusions were rarely observed. The Combined Task Force of the North American Spine Society, American Society of Spine Radiology, and American Society of Neuroradiology has issued guidelines that provide standardization of terms to characterize disc herniation, as well as other disc pathology.[5] To what extent these guidelines are being followed in clinical practice is unknown.

The Spine Patient Outcomes Research Trial (SPORT) is a clinical trial with both randomized and observational cohorts conducted at 13 sites with multidisciplinary spine practices across 11 states.[6-8] Using patients with disc herniation from the randomized cohort of this study, we compared the interpretation of a radiologist and a clinician reading the same image. Based on our review of the literature, this has not been reported previously. We compared radiologists' reports from free text dictation to the data from an imaging form completed by the enrolling physician.

Materials and Methods

Patient Population

Patients for this study were participants in the randomized cohort of SPORT with a diagnosis of intervertebral disc herniation.[7,8] These patients all had radicular pain and evidence of nerve root compression by physical examination and evidence of disc herniation at the level and side of symptoms by MRI or computed tomography (CT). Exclusion criteria included *caudaequina* syndrome, progressive neurologic deficit, malignancy, significant deformity, prior back surgery, and other established contraindications to elective surgery. Overall, 501 IDH patients were randomized in SPORT; the study population had a mean age of 42, with majorities being men, white, having attended at least some college, and working; 16% were receiving disability compensation. All patients had symptoms for at least 6 weeks; about 20% had symptoms for greater than 6 months. Most of the herniations were at L5-S1, posterolateral, and were extrusions by imaging criteria.[7,8] Although all had advanced imaging (97% MRI, 3% CT), 105 patients did not have MRI radiology reports available, leaving a total patient population of 396 for this study.

Imaging

The patients' baseline MRIs provided the initial diagnosis of disc herniation. There was no specified protocol for imaging; scanner and imaging protocols in routine clinical use at the study sites were used. These images were read by both the clinician caring for the patient as part of the inclusion criteria for SPORT, as well as a radiologist as part of routine clinical practice at each site. Each study was thus read by 1 clinician and 1 radiologist whose interpretations were compared. Clinicians were specialists in the areas of orthopedic spine surgery, neurosurgery or nonoperative spine care. The data form included level, morphology,

and axial location of the disc herniation (Table 1). The radiology reports were clinical reports copied from the medical record with written patient informed consent.

Data Abstraction

Using standardized rules, the radiologists' free text reports were abstracted into categories equivalent or nearly equivalent to those on the clinician imaging forms. These categories are shown in Table 1. Care was taken to avoid making assumptions about the radiologists' dictations; protrusions, extrusions, and sequestered fragments were selected only if the radiologist used these terms or some clear derivative. When the report was inconclusive between protrusion/extrusion or extrusion/sequestration, such as if 1 term was used in the body of the report but a different term in the impression, the more abnormal category (*e.g.*, extrusion or sequestration) was selected.

Data Analysis

For the analysis of agreement on vertebral level, the radiologist had the option of choosing several levels, but the clinician could choose only 1 level—the level of the symptomatic herniation for which the patient was enrolled in SPORT. For the purposes of analysis, the data were reformatted with 1 option for the clinician (level of herniation) and 2 options for the radiologist (agree or disagree with clinician level). Therefore the κ statistic could not be used to assess agreement, as a 1×2 table resulted. Instead of κ, percent agreement and adjusted percent agreement were calculated for this parameter.[9]

For comparison of morphology where categories were considered nominal, an unweighted κ statistic was calculated from the 3×3 table after the reports that did not indicate a specific morphology were excluded. Asymmetry was noted by constructing a histogram of the difference between the clinician and radiologist morphology scores. Bowker's test of symmetry[9] was used to determine the significance of asymmetry observed in the morphologic data.

For comparison of axial location data, where were considered ordinal, a weighted κ statistic was calculated using FleissCohen quadratic weights in a 7×10 table, excluding the radiologists' reports that did not indicate a specific location.[10] The radiologists' left/right unclear categories (Table 1) were given 0.89 weight if the clinician agreed on the side, and 0.56 weight if center. If the radiologist chose 2 locations, weight of 1.0 was given if the clinician chose either one of these locations. If 3 or 5 locations were chosen and the clinician chose the center, or if 4 was chosen and the clinician chose either of the two center locations, full weight of 1.0 was given. In the cases of multiple locations that did not meet these criteria, but were 1 category away, a weight of 0.56 was assigned.

The location data were also analyzed for left-right agreement. In the cases of multiple locations listed by the radiologist, the same method described above was used to determine left, right or center. Those patients with central herniations, as described by either the clinician or the radiologist, were excluded from this analysis. An unweighted κ statistic was calculated using a 2×2 table. Reliability of the data abstraction was tested on a random sample of 10% of reports using an unweighted κ.

Calculations were performed with Intercooled Stata 6.0 and Microsoft Excel 2003. Although the interpretation of strength of agreement based on κ values is controversial, in this article we followed the schema of Landis and Koch:[11] <0 = poor; 0 to 0.20 = slight; 0.21 to 0.40 = fair; 0.41 to 0.60 = moderate; 0.61 to 0.80 = substantial; 0.81 to 1.00 = almost perfect.

Results

Reliability of Abstraction Process

Reliability of the data showed excellent reproducibility. κs for the various parameters ranged from 0.89 to 1.00.

Herniation Level

Of the 396 cases included in the study, the clinician and radiologist agreed on the vertebral level of the herniation for 370 (93.4%), which corresponds to an adjusted percent agreement of 87%. Of the 26 (6.6%) patients where there was disagreement about the level, the radiologists reported a herniation at a different level in 13 patients, and reported no herniation at any level in another 13 patients (Table 2).

A closer look at these 13 cases where the clinician and radiologist agreed that a herniation was present but disagreed about level revealed the presence of a transitional vertebra in 4 patients; this may have lead to a difference in numbering the vertebrae. Three of the remaining 9 patients had available operative reports which showed that 2 patients had surgery on the herniation level reported by the clinician and 1 had surgery on the herniation level reported by the radiologist.

In the 13 cases where the radiologist reported no herniation, all available MRI reports showed a bulge at that same level with the exception of 1 case that instead described a disc/osteophyte complex. Five bulges were reported to be asymmetric and resulted in narrowing of the vertebral foramen. In 1 of these cases, the asymmetric bulge was reported to directly contact a nerve root.

Herniation Location

The 370 cases in which the radiologist and clinician agreed on the vertebral level showed excellent agreement on both the axial ($\kappa = 0.81$) and left/right ($\kappa = 0.93$) location of the herniation. In 10 patients (3.3%), the clinician and the radiologist disagreed on the side of the herniation. Eight patients (2.2%) were not used in this analysis because the radiologist noted a herniation but did not specify its location within the vertebral level. Another 58 patients (15.8%) had central herniations and were excluded from the left/right analysis (Table 2).

Herniation Morphology

Although clinician and radiologist agreed on the vertebral level of the herniation 93% of the time, agreement on the morphology of the herniation was only fair. Of 370 cases where there was agreement of the vertebral level of the herniation, the specific morphology of the herniation was not reported by the radiologist in 42.2% of cases (Table 2). In the 214 cases where a specific morphology was identified, the κ statistic ($\kappa = 0.24$) showed fair agreement (Table 3). Analysis of this disagreement showed an asymmetric distribution of morphology ($S = 35.75$, $P < 0.0001$), with the clinician more often reporting the more abnormal morphologic category (Figure 1). In other words, clinicians reported an extrusion in cases where the radiologist reported a protrusion significantly more often than clinicians reported a protrusion when the radiologist called it an extrusion.

Discussion

Herniation Level

The agreement between the radiologists with the clinicians on the presence and vertebral level of the disc herniation was excellent, with 93.4% agreement. Of the 26 (6.6%) patients with

disagreements, 13 patients had raters noting herniation at different levels. A third of these patients were also noted to have transitional vertebrae, which may explain this disagreement between levels as a difference in vertebral numbering. We believe that these 2 raters likely saw the same herniation, but described it differently. Of the remaining 9 patients, operative reports were available for 3, with surgical findings supporting the clinician's interpretation twice and radiologist's once.

When the radiologist failed to report any herniation, they typically reported asymmetric bulges that were interpreted as herniations by the clinicians. Interestingly, of these 13 (3.3%) cases, 7 (1.7%) were graded as an extrusion by the clinician, suggesting a substantial difference in morphologic interpretation between the clinicians and radiologists in these cases.

Herniation Location

The agreement between radiologist and clinician on the herniation location within the vertebral level was excellent. The data were examined using the discrete locations provided by the raters (Table 1), and also using the data after it was reformatted into left, right, and center locations.

Despite the excellent agreement in the left/right data, there was disagreement on the lateralization of the herniation for 6 patients (2.0%). This may be important since clinical symptoms usually lateralize, and disagreement about the side of the disc herniation relative to the patient's symptoms could alter treatment recommendations. Unfortunately, 5 of the 6 patients with left-right disagreements were randomized to conservative management, resulting in the actual location of the herniation being confirmed in only 1 patient. In this patient, the clinician reported herniation on the left, which was confirmed in the operation, and the radiologist reported a herniation on the right. We speculate that these 6 cases of left/right confusion are because of human error rather than real ambiguity in the images. Therefore, it is important for clinicians who treat patients with back pain but do not regularly read MRIs to understand that left/right confusion does occur, even among experienced readers. Although 2.0% is a small minority of patients, it is large enough to be observed on a regular basis in clinical practice.

Herniation Morphology

The data regarding herniation morphology is a good barometer of the standard practice of the radiologists describing disc herniation morphology compared to the task force guidelines. In 42.2% of the cases, the radiology dictation did not provide enough detail to classify the herniation as a protrusion, extrusion, or sequestered fragment (most said just "herniation"). Although the true clinical significance of these terms has not yet been determined, the difference in prevalence of protrusions and extrusions among asymptomatic subjects in prior studies suggests that this morphologic distinction may be important. An increased effort by lumber spine MRI readers to use these terms in accordance with published guidelines can only help in future studies.[5]

In the 214 patients with clearly defined morphology, the data shows fair agreement ($\kappa = 0.32$), in contrast to formal reliability studies that have reported moderate to substantial agreement. Using the same classification scheme, Brant-Zawadzki *et al* found moderate inter-reader agreement (unweighted $\kappa = 0.59$) and Jarvik *et al* found moderate to substantial inter-reader agreement with weighted κs of 0.50 to 0.75 across reader pairs.[12,13] Similarly, Solgaard *et al* and Weishaupt *et al* found substantial agreement for classifying disc morphology, with inter-reader κs of 0.79 and 0.68, respectively.[14,15] A formal reliability study using many of the images included in this study, read by 4 independent blinded readers using structured data forms, showed overall substantial to almost perfect agreement on disc morphology (summary $\kappa = 0.81$ (95% CI: 0.78, 0.85).[16] The lower reliability in the current study may relate to the

lack of a structured data form for the radiologist, the difference in context between a formal study and clinical practice, the level of expertise of radiologists participating in formal reliability research, the additional chance for error introduced by the need to abstract data from free text reports, or a combination of factors.

The disagreements in the current study were also significantly asymmetric, with morphology scores higher for clinicians compared with radiologists. This difference occurred despite the methodologic design of choosing the more abnormal morphologic category when the radiology report was equivocal between 2 categories. The reason why clinicians' ratings of morphology were more abnormal than the radiologists' is unclear. It is possible that clinicians were influenced by the patient's symptoms. In a study by van Rijn *et al*, comparing MRI interpretation with and without clinical information, the addition of clinical information lowered the threshold for reporting bulges, though no difference in reported herniation was seen.[17] The asymmetry may also result from differences in training; however,in a formal reliability study using some of the same cases studied here, we found no difference in agreement between radiologists and between radiologists and an orthopedic surgeon.[16]

Limitations

This study relies on the abstraction of data from free text dictations, a process that is not optimal nor completely comparable with the clinician's data. However, it is the only option to compare a large number of standard practice radiologist outcomes. Efforts were made to minimize the inaccuracies inherent in the abstraction as described in the methods. In addition, the clinicians in this study were a specialized group whose practices are dominated by spine problems and who have extensive experience interpreting spine MRIs. As a result, our results may, in fact, underestimate the true occurrence of disagreements between clinicians and radiologists in routine clinical practice.

Conclusion

For MRI readings on patients with lumbar disc herniation, agreement between clinical spine specialists completing a data form and radiologists using free-text dictation is excellent when comparing herniation level and location within the level, but agreement is only fair when comparing disc morphology. Care should be taken to avoid common pitfalls. Transitional vertebrae may lead to confusion between vertebral levels. Disc morphology should be described as per the guidelines, because the prevalence of different morphologies varies substantially in asymptomatic populations. Also, left/right confusion must be considered a potential reason for discrepancy between lateralization of clinical symptoms and lateralization of a herniation on a radiology report. Potential consequences related to surgical treatments must not be underestimated, as wrong site, wrong level, and wrong side surgeries can occur and, in fact, may be underreported, given the potential difficulties in identifying side, level, and location, not only before surgery but intraoperatively as well.

Key Points

- Radiation reports frequently fail to provide sufficient detail to describe herniation morphology.
- When comparing MRI interpretations, radiologists and clinicians were found to agree on presence and level of herniation 93.4% of the time.
- Radiologists and clinicians had only fair agreement ($\kappa = 0.24$) when interpreting herniation morphology.

Acknowledgments

Acknowledgment date: June 21, 2008. Revision date: October 8, 2008. Acceptance date: October 16, 2008

The manuscript submitted does not contain information about medical device(s)/drug(s).

Federal funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.

Supported by The National Institute of Arthritis and Musculoskeletal and Skin Diseases (U01-AR45444-01A1) and the Office of Research on Women's Health, the National Institutes of Health, and the National Institute of Occupational Safety and Health, the Centers for Disease Control and Prevention; and a Research Career Award from NIAMS (1 K23 AR 048138-01) (to J.D.L.). The Multidisciplinary Clinical Research Center in Musculoskeletal Diseases is funded by National Institute of Arthritis and Musculoskeletal and Skin Diseases (P60-AR048094-01A1).

References

1. Dora C, Walchli B, Elfering A, et al. The significance of spinal canal dimensions in discriminating symptomatic from asymptomatic disc herniations. Eur Spine J 2002;11:575–81. [PubMed: 12522716]
2. Jarvik JJ, Hollingworth W, Heagerty P, et al. The longitudinal assessment of imaging and disability of the back (LAIDBack) study: baseline data. Spine 2001;26:1158–66. [PubMed: 11413431]
3. Jensen M, Brant-Zawadzki M, Obuchowski N. Magnetic resonance imaging of the lumbar spine in people without back pain. N Engl J Med 1994;331:69–73. [PubMed: 8208267]
4. Stadnik TW, Lee RR, Coen HL, et al. Annular tears and disk herniation: prevalence and contrast enhancement on MR images in the absence of low back pain or sciatica. Radiology 1998;206:49–55. [PubMed: 9423651]
5. Fardon DF, Milette PC. Nomenclature and classification of lumbar disc 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:E93–113. [PubMed: 11242399]
6. Birkmeyer NJ, Weinstein JN, Tosteson AN, et al. Design of the Spine Patient outcomes Research Trial (SPORT). Spine 2002;27:1361–72. [PubMed: 12065987]
7. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical vs. nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT) observational cohort. JAMA 2006;296:2451–9. [PubMed: 17119141]
8. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical vs. nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT): a randomized trial. JAMA 2006;296:2441–50. [PubMed: 17119140]
9. Brennan RL, Prediger DJ. Coefficient kappa: some uses and alternatives. Educ Psychol Meas 1981;41:687–99.
10. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educ Psychol Meas 1973;33:613–9.
11. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74. [PubMed: 843571]
12. 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. [PubMed: 7660234]discussion 64
13. Jarvik JG, Haynor DR, Koepsell TD, et al. Interreader reliability for a new classification of lumbar disk disease. Acad Radiol 1996;3:537–44. [PubMed: 8796714]
14. Solgaard Sorensen J, Kjaer P, Jensen ST, et al. Low-field magnetic resonance imaging of the lumbar spine: reliability of qualitative evaluation of disc and muscle parameters. Acta Radiol 2006;47:947–53. [PubMed: 17077047]
15. Weishaupt D, Zanetti M, Hodler J, et al. 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;209:661–6. [PubMed: 9844656]

16. Lurie JD, Tosteson AN, Tosteson TD, et al. Reliability of magnetic resonance imaging readings for lumbar disc herniation in the Spine Patient Outcomes Research Trial (SPORT). Spine 2008;33:991–8. [PubMed: 18427321]

17. van Rijn JC, Klemetsö N, Reitsma JB, et al. Observer variation in MRI evaluation of patients suspected of lumbar disk herniation. AJR 2005;184:299–303. [PubMed: 15615992]

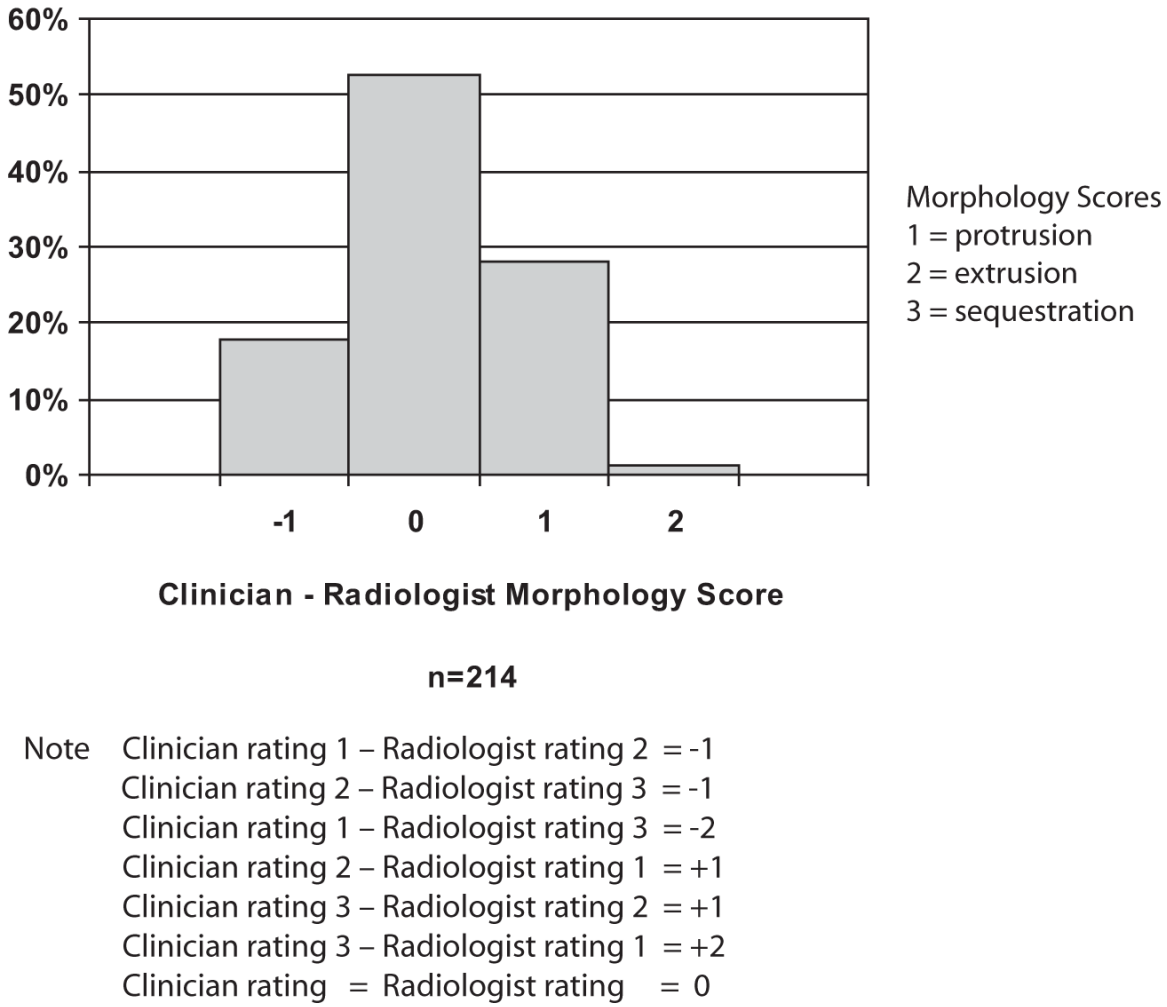

Figure 1.
Histogram of morphology score difference.

Table 1

Clinician and Radiologist Category Options

	Clinician	Radiologist
Level	L2-L3	L2-L3
	L3-L4	L3-L4
	L4-L5	L4-L5
	L5-S1	L5-S1
Morphology		Bulge
	Protrusion	Protrusion
	Extrusion	Extrusion
	Sequestered fragment	Sequestered fragment
		Unclear
Location		Left side, unclear
	Left far lateral	Left far lateral
	Left foraminal	Left foraminal
	Left posterolateral	Left posterolateral
	Center	Center
	Right posterolateral	Right posterolateral
	Right foraminal	Right foraminal
	Right far lateral	Right far lateral
		Right side, unclear
		Unclear

Table 2

Results Summary

	Total	Unclear	n	Agreement	Comments
Level	396	0	396	370 (93.4%)	13 (3.3%) different level 13 (3.3%) no herniation
Morphology	370	156 (42.2%)	214	$\kappa = 0.24$	Fair agreement, asymmetric Clinician scores higher
Axial location	370	8 (2.2%)	362	$\kappa = 0.81$	Excellent agreement
Left/right location Excluded 58 central herniations	312	8	304	$\kappa = 0.93$	Excellent agreement 10 (3.3%) disagree on left/right

Table 3

Morphology

	Radiologist		
Clinician	**Protrusion**	**Extrusion**	**Sequestration**
Protrusion	38	11	0
Extrusion	50	63	27
Sequestration	3	10	12

n = 214.

κ = 0.24 (unweighted).