ELSEVIER

CrossMark

The Spine Journal 14 (2014) 2598–2607

THE SPINE JOURNAL

Clinical Study

Reliability of gadolinium-enhanced magnetic resonance imaging findings and their correlation with clinical outcome in patients with sciatica

Abdelilah el Barzouhi, MD, MSc[a,*], Carmen L.A.M. Vleggeert-Lankamp, MD, PhD[a], Geert J. Lycklama à Nijeholt, MD, PhD[b], Bas F. Van der Kallen, MD[b], Wilbert B. van den Hout, PhD[c], Bart W. Koes, PhD[d], Wilco C. Peul, MD, PhD[a,e] for the Leiden–The Hague Spine Intervention Prognostic Study Group

[a]*Department of Neurosurgery, Leiden University Medical Center, Postbus 9600, 2300 RC, Leiden, The Netherlands*
[b]*Department of Radiology, Medical Center Haaglanden, Lijnbaan 32, 2512 VA, The Hague, The Netherlands*
[c]*Department of Medical Decision Making, Leiden University Medical Center, Postbus 9600, 2300 RC, Leiden, The Netherlands*
[d]*Department of General Practice, Erasmus Medical Center, University Medical Center, Postbus 2040, 3000 CA, Rotterdam, The Netherlands*
[e]*Department of Neurosurgery, Medical Center Haaglanden, Lijnbaan 32, 2512 VA, The Hague, The Netherlands*

Received 29 July 2013; revised 28 January 2014; accepted 15 February 2014

Abstract

BACKGROUND CONTEXT: Gadolinium-enhanced magnetic resonance imaging (Gd-MRI) is often performed in the evaluation of patients with persistent sciatica after lumbar disc surgery. However, correlation between enhancement and clinical findings is debated, and limited data are available regarding the reliability of enhancement findings.

PURPOSE: To evaluate the reliability of Gd-MRI findings and their correlation with clinical findings in patients with sciatica.

STUDY DESIGN: Prospective observational evaluation of patients who were enrolled in a randomized trial with 1-year follow-up.

PATIENTS SAMPLE: Patients with 6- to 12-week sciatica, who participated in a multicentre randomized clinical trial comparing an early surgery strategy with prolonged conservative care with surgery if needed. In total 204 patients underwent Gd-MRI at baseline and after 1 year.

OUTCOME MEASURES: Patients were assessed by means of the Roland Disability Questionnaire (RDQ) for sciatica, visual analog scale (VAS) for leg pain, and patient-reported perceived recovery at 1 year. Kappa coefficients were used to assess interobserver reliability.

METHODS: In total, 204 patients underwent Gd-MRI at baseline and after 1 year. Magnetic resonance imaging findings were correlated to the outcome measures using the Mann-Whitney *U* test for continuous data and Fisher exact tests for categorical data.

RESULTS: Poor-to-moderate agreement was observed regarding Gd enhancement of the herniated disc and compressed nerve root (kappa<0.41), which was in contrast with excellent interobserver agreement of the disc level of the herniated disc and compressed nerve root (kappa>0.95). Of the 59 patients with an enhancing herniated disc at 1 year, 86% reported recovery compared with 100% of the 12 patients with nonenhancing herniated discs (p=.34). Of the 12 patients with enhancement of the most affected nerve root at 1 year, 83% reported recovery compared with 85% of the 192 patients with no enhancement (p=.69). Patients with and without enhancing herniated discs or nerve roots at 1 year reported comparable outcomes on RDQ and VAS-leg pain.

FDA device/drug status: Not applicable.

Author disclosures: ***AeB:*** Nothing to disclose. ***CLAMV-L:*** Nothing to disclose. ***GJLàN:*** Nothing to disclose. ***BFVdK:*** Nothing to disclose. ***WBvdH:*** Nothing to disclose. ***BWK:*** Nothing to disclose. ***WCP:*** Nothing to disclose.

This study was supported by a grant from the Netherlands Organisation for Health Research and Development (ZonMW) and the Hoelen Foundation, The Hague. None of the authors of this study has any conflict of interest.

* Corresponding author. Department of Neurosurgery, Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, The Netherlands. Tel.: (31) 646570264.

E-mail address: A.el_barzouhi@lumc.nl (A. el Barzouhi)

http://dx.doi.org/10.1016/j.spinee.2014.02.028

CONCLUSIONS: Reliability of Gd-MRI findings was poor-to-moderate and no correlation was observed between enhancement and clinical findings at 1-year follow-up.

Keywords: Sciatica; Disc herniation; Gadolinium-enhanced MRI findings; Surgery; Conservative treatment; Follow-up; Clinical outcome

Introduction

Sciatica is one of the most common lumbar spine disorders and a major source of lost productivity [1,2]. The most common cause of sciatica is a disc herniation [3]. Because the natural history of sciatica is favorable, surgery should be offered only if symptoms persist after a period of conservative treatment [4,5]. The reported prevalence of satisfactory results after initial surgery varies between 80% and 95% [6–12]. However, repeated surgery is less successful: only 60% to 82% of patients with recurrent disc herniation improve after surgery [13–16]. In patients who have only epidural scar tissue and no other abnormalities, the success rate of repeat surgery is even lower: 17% to 38% [14,16,17]. Therefore, evidence of scar tissue alone is often regarded as a contraindication for repeat surgery, whereas evidence of (recurrent) disc herniation may be an indication for a repeated surgical procedure [18]. Contrast-enhanced magnetic resonance imaging (MRI) is frequently performed in patients with persistent or recurrent symptoms of sciatica after surgical treatment because it has been proposed to differentiate between postoperative epidural scar tissue and recurrent disc herniation: scar tissue has a homogenous enhancement pattern, whereas disc herniation usually lacks central enhancement [16,18–21]

The investigators previously reported the 1-year MRI results of patients with symptomatic lumbar disc herniations at baseline, who were treated with either surgery or conservative treatment [22]. At 1-year follow-up, a considerable proportion of patients still had a visible disc herniation on MRI (21% of surgically compared with 60% of conservatively treated patients). However, presence of disc herniation on MRI did not correlate to the clinical status and could not distinguish patients with persistent or recurrent symptoms of sciatica from asymptomatic patients. In the search for causes for persistent sciatica, previous studies have observed an association between enhancement of the nerve root and clinical findings in sciatica [18,20,23–25]. However, other studies have not shown an association between nerve root enhancement and clinical outcome [26,27]. Moreover, as with any diagnostic radiographic study, interpretation of the results regarding the assessment of contrast enhancement may become inconsistent between examiners. The reliability of enhancement findings has been poorly investigated in previous literature.

The primary objective of the present study was to understand the role of gadolinium (Gd) enhancement in MRI, by correlating MRI enhancement findings to clinical outcomes and neurologic findings. A secondary objective was to evaluate interobserver agreement regarding MRI enhancement findings.

Methods

Study population and randomization

Patients for this study were participants in a multicentre randomized controlled trial among patients with 6- to 12-week sciatica with a disc herniation on MRI. Patients were only included if they had a dermatomal pattern of pain distribution, with concomitant neurologic disturbances that correlated with the same nerve root being affected on MRI. An early surgery strategy was compared with prolonged conservative care for an additional 6 months, followed by surgery for patients who did not improve or who did request it earlier because of aggravating symptoms [28,29]. Patients were excluded if they were presenting with cauda equina syndrome, insufficient strength to move against gravity, identical complaints in the previous 12 months, previous spine surgery, pregnancy, spinal stenosis, spondylolisthesis, or severe coexisting disease.

A computer-generated permuted-block scheme was used for randomization, with patients stratified according to center (n=9). An hour before randomization, the patients were evaluated again and patients who had recovered from their symptoms were excluded from the trial. For patients who were included, the next numbered opaque envelope containing the assigned treatment was opened and the patient was assigned to a treatment group.

Surgery was performed in the conventional manner with microscope or loupe magnification. During a consensus meeting before the trial, the surgical method was discussed and no alternative methods of surgery were allowed. The goal of surgery was to decompress the nerve root and reduce the risk of recurrent disc herniation by performing an annular fenestration, curettage, and removal of loose degenerated disc material from the disc space.

The medical ethics committees at the nine participating hospitals approved the protocol. Written informed consent was obtained from all patients. Details of the design and study protocol were published previously [29]. For the objectives of this study, it can be regarded as a prospective observational study of patients who were enrolled in a randomized controlled trial.

EVIDENCE & METHODS

Context
Gadolinium (Gd) enhanced MRI is frequently used in the setting of recurrent disc herniation as a means to differentiate between extruded disc material and scar tissue. Inter-rater reliability of Gd enhanced MRI in this setting has not been well characterized, nor have correlations been made between imaging findings and clinical results.

Contribution
The authors report poor to moderate inter-rater reliability for the interpretation of Gd enhanced MRI. In addition, no correlation was appreciated between findings on MRI with Gd and clinical results one year after surgery.

Implications
Even in the idealized study setting created by the authors, Gd enhanced MRI demonstrated only poor to moderate inter-rater reliability. Disc and nerve root enhancement were also insufficient for the purposes of prognosticating outcomes at one year following surgery. The study design, which utilized a sample of patients collected for other purposes (in this case a randomized trial intended to evaluate the benefit of early surgery for disc herniation) raises the possibility that this investigation was actually underpowered to detect differences in outcome based on the results of Gd enhanced MRI. Nonetheless, the authors' findings regarding inter-rater reliability should be noted by other researchers who intend to employ this modality in similar efforts.

—The Editors

MRI protocol and image evaluation

Patients underwent MRI at the baseline and after 1-year follow-up. The 12-month evaluation period was selected since postoperative fibrosis stabilizes by 6 months, with no further changes at 12 months [30]

Magnetic resonance imaging scans were performed in all nine participating hospitals, using standardized protocols tailored to a 1.5 Tesla scanner. Sagittal T1 and axial T1 spin echo images of the lumbar spine were acquired. In addition, T2-weighted sagittal and axial images were obtained. For research purposes also contrast-enhanced (Gd-diethylenetriamine pentaacetic acid at a standard dose of 0.1 mmol/Kg body weight) T1 fat suppressed sagittal and axial images were obtained.

Two neuroradiologists (BFK and GLaN) and one neurosurgeon (CLAMV-L) independently evaluated all MRIs. The readers were not provided any clinical information and had not been involved in the selection or care of the included patients. Before the start of the study, the readers met in person to evaluate and refine standardized definitions of imaging characteristics. After reaching final consensus, standardized case record forms with these final definitions were used (Appendix Table S1). Observer experience in reading spine MRIs was 7 and 6 years postresidency for the respective neuroradiologists and 4 years postresidency for the neurosurgeon.

First, the blinded readers had to decide on the baseline MRI, the disc level that showed the most severe nerve root compression. For both the presence of disc herniation and nerve root compression a four point scale was used, ranging from 1 (definitely present) to 4 (definitely absent). The size of the disc herniation was also evaluated. The same disc level thought to cause symptoms at baseline was evaluated on the 1-year MRI. On the 1-year MRI, the readers had to also assess whether scar tissue was present (no, moderate, or severe). The readers evaluated the enhancement on the baseline and 1-year MRI of the following structures using different categories (Appendix Table S1): disc herniation (if present): no, any edge, complete circumferential, or diffuse enhancement; most affected nerve root: no, mild, or strong enhancement; and scar tissue (if present at 1 year): yes versus no enhancement. Structures were considered enhanced when brighter compared with the precontrast image.

Neurologic examination

Patients underwent a standardized neurologic examination by trained research nurses. The examination was performed blind to the MRI results. Sensation was dichotomized as normal or abnormal for each dermatome. Muscle strength Medical Research Council Grade 5 was considered normal, whereas Grade 4 or less was rated abnormal. Reflexes were rated as abnormal if absent, less than the contralateral side, or in case of an extensor plantar response.

Outcomes

The outcome measures of the trial were the Roland Disability Questionnaire (RDQ) for sciatica (scores range from 0 to 23, with higher scores indicating worse functional status) [31], the 100-mm visual analogue scale (VAS) for leg pain (with 0 representing no pain and 100 the worst pain ever experienced) [32], and a seven-point Likert self-rating scale of global perceived recovery given by the question whether the patient experienced recovery, with answers ranging from completely recovered to much worse. Perceived recovery on the seven-point Likert scale was used in dichotomized form: ''complete'' or ''nearly complete disappearance of symptoms'' was defined as ''perceived recovery,'' whereas a score in the remaining five categories (varying from ''minimally improved'' to ''very much worse'') was marked as ''no recovery'' [28]. These outcome measures were assessed at baseline, 2, 4, 8, 12, 26, 38, and 52 weeks.

Patients were blinded to results of earlier assessments and MRI findings. For the purpose of the present study, the results at baseline and 52 weeks were used in the analysis.

Statistical analysis

We used STATA (version 11.2, StataCorp, College Station, TX, USA)) for our statistical analysis. Interobserver agreement regarding the MRI findings was determined by use of absolute percentages of agreement and kappa values. Percentage of absolute agreement equals the number of cases for which the observers fully agree, proportional to the total number of cases [33]. However, the absolute percentage of agreement is inadequate because it does not discriminate between actual agreement and agreement that arises due to chance [34]. A measure which attempts to correct for this is the kappa statistic [35]. In case of ordered data, we calculated weighted kappa scores that are based on the idea that in any ordered scale, some possible disagreements are more serious than others.

The kappa statistic is affected by the prevalence of the events [36,37], so that findings with very high or low prevalence lead to very low kappa values, even if the observer agreement is high [38]. Therefore, kappa values were only calculated for findings reported in more than 10% and less than 90% of all reports [39]. Kappa values and percentages of agreement for the enhancement of the structures were also only calculated if the observers marked the same structure as affected (eg, when there was disagreement about the most affected nerve root in a patient, this patient did not contribute to the interagreement analysis regarding the enhancement of the most affected nerve root).

Both weighted and unweighted kappa statistics were computed for all possible pairings of observers. In addition, we computed overall unweighted kappa coefficients for multiple raters. STATA (like other reliable statistical packages such as SPSS and SAS) is currently not able to calculate weighted kappa coefficients for multiple rates. Although no absolute definitions have been accepted for the interpretation of kappa values, we used guidelines proposed by Landis and Koch [40] for interpretation. Values of less than 0.00 indicated poor; 0.00 to 0.20 slight; 0.21 to 0.40 fair; 0.41 to 0.60 moderate; 0.61 to 0.80 substantial; and 0.81 to 1.00 excellent or almost perfect agreement.

When the MRI findings were correlated with clinical outcome, the majority opinion of the three readers regarding the MRI findings was used (answer independently given by minimum two out of three readers). In analyses comparing enhancement/no enhancement of disc herniation, ratings were categorized as 1, 2, 3 (any edge, complete circumferential, or diffuse enhancement) versus 4 (no enhancement). In analyses comparing enhancement/no enhancement of the affected nerve, ratings were categorized as mild or strong enhancement versus no enhancement. Differences between MRI findings were assessed by using the Mann-Whitney *U* test for continuous data and Fisher exact tests for categorical data. In a subanalysis, we stratified according to the presence of leg pain at baseline, defined as a VAS of leg of at least 40 mm, as this cutoff value is regularly used when the VAS is categorized into favorable and unfavorable outcomes [41,42]. In both subgroups, we compared the outcome measures between patients with and without

Table 1
Baseline characteristics of the intention-to-treat and as-treated groups

	Intention to treat		As treated	
	Randomized to early surgery (N=105)	Randomized to prolonged conservative care (N=99)	Received surgery (n=129)	Received no surgery (n=75)
Age (y)	42.4±10.4	43.0±9.5	42.1±10.2	43.6±9.4
Male gender	66 (63)	71 (72)	80 (62)	57 (76)
Duration of sciatica (wk)	9.5±2.3	9.5±2.2	9.5±2.3	9.6±2.2
Suspected disc level				
L3–L4	5 (5)	2 (2)	6 (5)	1 (1)
L4–L5	48 (46)	35 (35)	54 (42)	29 (39)
L5–S1	52 (50)	62 (63)	69 (53)	45 (60)
MRI assessed nerve root compression				
Definite	66 (63)	70 (71)	87 (67)	49 (65)
Probable	30 (29)	22 (22)	32 (25)	20 (27)
Possible	8 (8)	6 (6)	9 (7)	5 (7)
Definitely no root compression	1 (1)	1 (1)	1 (1)	1 (1)
Duration between baseline and follow-up MRIs (wk)	53.3±2.9	52.7±3.9	52.9±3.6	53.2±3.2
Roland Disability score*	16.2±4.3	15.9±3.9	16.2±4.3	15.7±3.8
VAS-leg pain (mm)†	66.1±20.0	62.0±21.1	65.6±20.5	61.7±20.8
VAS-back pain (mm)†	33.4±29.0	28.5±25.9	32.7±29.7	28.0±23.2

MRI, magnetic resonance imaging; VAS, visual analog scale; SD, standard deviation.

Note: Values are n (%) or means±SD.

No significant baseline differences were observed in the intention-to-treat and as-treated groups.

* The Roland Disability Questionnaire for sciatica measures the functional status of patients with pain in the leg or back. Scores range from 0 to 23, with higher scores indicating worse functional status.

† The intensity of pain is indicated on a horizontal 100-mm VAS, with 0 representing no pain and 100 the worst pain ever experienced.

Table 2
Interobserver agreement regarding MRI findings at baseline

	A vs. B		A vs. C		B vs. C		All observers	
	% Agreement	Kappa (95% CI)	% Agreement	Kappa (95% CI)	% Agreement	Kappa (95% CI)	% Agreement	Kappa (95% CI)
Disc level with most severe nerve root compression	97.4	0.95 (0.90–0.98)	99.1	0.98 (0.96–1.00)	97.4	0.95 (0.90–0.98)	97.0	0.96 (0.93–0.98)
Probability of disc herniation*	96.5	†	99.6	†	96.1	†	96.1	†
Enhancement of disc herniation (four categories)‡	55.0	**0.42** (0.35–0.51)	50.0	**0.34** (0.26–0.44)	64.3	**0.48** (0.38–0.57)	47.8	0.41 (0.33–0.49)
Enhancement of disc herniation (two categories)§	78.2	0.38 (0.26–0.51)	77.5	0.35 (0.23–0.47)	78.1	0.50 (0.37–0.62)	66.5	0.40 (0.29–0.51)
Probability of nerve root compression*	100.0	1.00 (1.00–1.00)	100.0	1.00 (1.00–1.00)	100.0	1.00 (1.00–1.00)	100.0	1.00 (1.00–1.00)
Most affected nerve root	97.8	0.97 (0.94–0.99)	97.0	0.96 (0.93–0.99)	96.5	0.96 (0.92–0.98)	95.7	0.96 (0.94–0.98)
Enhancement of most affected nerve root¶	58.2	0.27 (0.19–0.36)	53.2	0.23 (0.16–0.30)	84.8	0.60 (0.49–0.72)	48.4	0.28 (0.19–0.38)

MRI, magnetic resonance imaging; CI, confidence interval.

Note: In bold are weighted kappa coefficients.

A and B represent the two neuroradiologists, whereas C represents the neurosurgeon. Kappa values and percentages of agreement for the enhancement of the structures were only calculated if the observers marked the same structure as affected (eg, when there was disagreement about the most affected nerve root in a patient, this patient did not contribute to the interagreement analysis regarding the enhancement of the most affected nerve root).

* The categories "definite, probable, and possible about the presence" were combined to one category. The other category was "definite about the absence".

† Prevalence of one category too low (<10% of the reports) to calculate kappa values.

‡ The categories were: 1) No, 2) Any edge, 3) Complete circumferential, and 4) Diffuse enhancement.

§ The categories "any edge, complete circumferential, and diffuse enhancement" were combined to one category. The other category was "no enhancement."

¶ The categories "mild or strong enhancement" were combined to one category. The other category was "no enhancement."

MRI enhancement findings. Statistical significance was defined as $p<.05$.

Results

Of 599 patients screened for the trial, 283 patients were randomized [28]. A year after randomization, a second MRI was available for 267 (94.3%) patients (Table S2). However, at baseline, 230 (81%) underwent MRI with Gd and at 1 year, 245 (87%) patients. No significant differences in patient characteristics existed between patients who underwent Gd-MRI and conventional MRI. In total, 204 patients (72%) underwent Gd-MRI, both at baseline and 1 year. Of the 204 patients who were eligible to be analyzed for the present study, 105 patients were randomized to early surgery and 99 to prolonged conservative care. Of the 105 patients randomized to early surgery, 12 patients recovered before surgery could be performed. Of the 99 patients randomized to prolonged conservative care, 36 eventually received surgery within the first year. Thus, during the first year after randomization, 129 patients underwent surgery and 75 patients conservative care. Baseline characteristics of the intention-to-treat and the as-treated groups are demonstrated in Table 1.

Interagreement analysis at baseline

At baseline, interobserver agreement was excellent regarding the disc level with the most severe nerve root compression (kappa=0.96), most affected nerve root (kappa=0.96), and probability of nerve root compression (kappa=1.0) (Table 2). However, interobserver agreement was only fair-to-moderate regarding enhancement of the herniated disc (kappa=0.40–0.41) and the most affected nerve root (kappa=0.28).

Interagreement analysis 1 year

After 1 year, substantial interobserver agreement was found regarding the question whether the disc herniation was still present (kappa=0.67) (Table 3). However, when disc herniation was still considered present at 1 year, the MRI assessors reached only slight-to-fair agreement regarding its enhancement (kappa=0.13–0.32). Interobserver agreement was only slight regarding the question whether the affected nerve root was enhanced at 1 year (kappa=0.10). For the presence of scar tissue at 1 year, interobserver agreement was moderate-to-substantial (kappa=0.59). All readers marked scar tissue as enhanced in at least 97% when they considered it present, which led to a multirater agreement regarding the enhancement of scar tissue of 97.6%.

MRI findings

Baseline

When using the majority opinion of the three readers regarding the MRI findings, of the 204 patients, 81% showed enhancement of the herniated disc and 30% showed enhancement of the affected nerve root.

Table 3
Interobserver agreement regarding MRI findings at 1 year

	A vs. B		A vs. C		B vs. C		All observers	
	% Agreement	Kappa (95% CI)	% Agreement	Kappa (95% CI)	% Agreement	Kappa (95% CI)	% Agreement	Kappa (95% CI)
Probability of disc herniation*	82.4	0.61 (0.51–0.71)	87.6	0.74 (0.65–0.83)	85.4	0.66 (0.57–0.76)	77.6	0.67 (0.59–075)
Enhancement of disc herniation (four categories)†	48.2	**0.32** (0.14–0.49)	57.5	**0.35** (0.19–0.53)	55.4	**0.32** (0.12–0.53)	36.4	0.32 (0.18–0.49)
Enhancement of disc herniation (two categories)‡	67.9	0.10 (−0.09 to 0.28)	75.0	0.23 (0.02–0.44)	67.9	0.24 (−0.02 to 0.51)	54.4	0.13 (−0.05 to 0.33)
Probability of nerve root compression*	75.2	0.46 (0.36–0.56)	77.2	0.51 (0.41–0.61)	92.1	0.76 (0.66–0.86)	72.6	0.55 (0.46–0.65)
Enhancement of the nerve root that was most affected at baseline§	78.8	0.24 (0.11–0.36)	73.5	0.03 (−0.05 to 0.11)	92.7	¶	72.3	0.10 (0.00–0.21)
Presence of scar tissue	87.8	0.75 (0.67–0.83)	74.2	0.51 (0.41–0.60)	77.0	0.55 (0.45–0.65)	69.5	0.59 (0.52–0.66)
Enhancement of scar tissue‖	99.2	#	97.9	#	97.7	#	97.6	#

MRI, magnetic resonance imaging; CI, confidence interval.

Note: In bold are weighted kappa coefficients. A and B represent the two neuroradiologists, whereas C represents the neurosurgeon. Kappa values and percentages of agreement for the enhancement of the structures were only calculated if the observers marked the same structure as affected (eg, when there was disagreement about whether at 1 year a herniated disc was still visible, this case did not contribute to the interagreement analysis regarding the enhancement of the herniated disc).

* The categories "definite, probable, and possible about the presence" were combined to one category. The other category was "definite about the absence" (Table 1 Supplementary appendix).

† The categories were: 1) No, 2) Any edge, 3) Complete circumferential, and 4) Diffuse enhancement.

‡ The categories "any edge, complete circumferential, and diffuse enhancement" were combined to one category. The other category was "no enhancement."

§ The categories "mild or strong enhancement" were combined to one category. The other category was "no enhancement."

¶ Prevalence of "mild and strong enhancement" too low (<10%) to calculate kappa values.

‖ Yes versus no.

\# Prevalence of "no enhanced scar tissue" too low (<10% of the reports) to calculate kappa values.

Table 4
Differences in 1-year MRI findings between patients with and without surgery during the first year (as-treated)

	Surgery (129)	No surgery (75)	p
Enhancement disc herniation at 1 yr			
Enhanced	23 (18)	36 (48)	.52
No enhancement	3 (2)	9 (12)	
Not applicable, no disc herniation at 1 yr	103 (80)	30 (40)	
Enhancement at 1 yr of the nerve root thought at baseline to cause symptoms			
Enhanced	7 (5)	5 (7)	.76
No enhancement	122 (95)	70 (93)	

MRI, magnetic resonance imaging.
Note: Values are n (%). Total n=204.

1 year

Of the 129 surgically treated patients, 26 still had a herniated disc at 1 year and 88% of these herniations enhanced. Of these 26 disc herniation, 17 (65%) were small (size <25% of spinal canal). Of the 75 conservatively treated patients, 45 still had a herniated disc at 1 year and 80% of these herniations enhanced. Of these 45 disc herniations, 32 (71%) were small.

Five percentage of surgically treated patients showed 1-year enhancement of the affected nerve root as compared with 7% of conservatively treated patients (p=.76) (Table 4).

Of the 115 patients diagnosed with scar tissue at 1 year (108 had moderate scar tissue and 7 severe), 113 (98%) had undergone surgery. Of the 115 patients with visible scar tissue, 96% had scar tissue that surrounded the nerve root and 4% had scar tissue that did not surround the nerve root.

Baseline enhancement findings in relation to clinical data

Patients with and without an enhancing herniated disc at the baseline showed comparable baseline scores on the RDQ and VAS for leg and back pain (Table 5). At baseline, 80% of patients with enhancing disc herniation had muscle weakness compared with 62% with nonenhancing herniated discs (p=.02). Patients with enhancing disc herniation had more frequent sensory loss compared with patients with nonenhancing herniated discs (74% vs. 54%, p=.02). At baseline, 84% and 77% of the patients with enhancement of the affected nerve root had muscle weakness and sensory loss, respectively, compared with 72% and 68% with nonenhancing nerve roots (p=.11 and p=.24, respectively).

Patients with and without enhancement of the herniated disc or affected nerve root at baseline showed comparable scores on RDQ, VAS-leg, and Likert scale of global perceived recovery after 1 year (Table 5). Patients with enhancing nerve roots reported lower VAS-back pain scores at 1 year compared with patients with no enhancing nerve roots at baseline (9.9 vs. 16.2 mm, p=.02). The same results were observed in both conservatively and surgically treated patients.

Table 5
Outcome measures at baseline and after 1 year, stratified by enhancement of the herniated disc and affected nerve root at baseline

	Enhancement of disc herniation at baseline				Enhancement of the affected nerve root at baseline			
	Yes (n=161)	No (n=39)	Difference (95% CI)	p	Yes (n=61)	No (n=143)	Difference (95% CI)	p
Roland Disability*								
Baseline	16.3±4.0	14.9±4.7	1.4 (0.0–2.9)	.10	16.5±3.4	15.9±4.4	0.6 (−0.6 to 1.9)	.58
1 yr	3.3±5.2	3.8±6.1	0.5 (−1.4 to 2.4)	.96	2.8±4.8	3.8±5.6	1.0 (−0.6 to 2.6)	.15
VAS-leg pain†								
Baseline	63.2±21.0	67.6±19.5	4.5 (−2.8 to 11.7)	.22	64.8±19.7	63.9±21.0	1.0 (−5.3 to 7.2)	.90
1 yr	10.5±18.9	12.2±21.7	1.6 (−5.2 to 8.5)	.72	9.1±16.2	11.5±20.4	2.3 (−3.5 to 8.1)	.56
VAS-back pain†								
Baseline	29.7±26.8	32.9±29.6	3.2 (−6.4 to 12.8)	.63	26.7±25.2	32.7±28.4	6.0 (−2.3 to 14.4)	.13
1 yr	13.5±20.1	18.1±25.8	4.6 (−2.9 to 12.1)	.87	9.9±17.2	16.2±22.6	6.3 (0.0–12.7)	.02
Perceived recovery at 1 yr	139 (86)	32 (82)	4.3 (−8.2 to 16.7)‡	.46	55 (90)	119 (83)	6.9 (−3.7 to 17.6)‡	.28
Muscle weakness								
Baseline	128 (80)	24 (62)	18.0 (3.7–33.3)‡	.02	51 (84)	103 (72)	11.6 (−1.8 to 24.0)‡	.11
1 yr	34 (21)	10 (26)	4.5 (−10.1 to 19.2)‡	.53	14 (23)	30 (21)	2.0 (−10.5 to 14.4)‡	.85
Sensory loss								
Baseline	119 (74)	21 (54)	20.1 (4.6–36.4)‡	.02	47 (77)	96 (68)	9.9 (−4.3 to 23.2)‡	.24
1 yr	49 (30)	15 (38)	8.0 (−8.4 to 24.5)‡	.34	20 (33)	45 (31)	1.3 (−12.8 to 15.4)‡	.87
Reflex loss								
Baseline	102 (64)	25 (64)	0.7 (−17.4 to 16.7)‡	1.00	44 (72)	87 (61)	11.3 (−3.6 to 25.3)‡	.15
1 yr	70 (43)	19 (49)	5.2 (−12.3 to 22.8)‡	.59	26 (43)	64 (45)	2.1 (−12.9 to 17.2)‡	.88

VAS, visual analog scale; CI, confidence interval; MRI, magnetic resonance imaging; SD, standard deviation.
Note: Of the 204 patients with both Gd-MRIs at baseline and 1 year, 200 patients had a herniated disc at baseline. Values are n (%) or means±SD.
* The Roland Disability Questionnaire for sciatica is a disease-specific disability scale that measures the functional status of patients with pain in the leg or back. Scores range from 0 to 23, with higher scores indicating worse functional status.
† The intensity of pain is indicated on a horizontal 100-mm VAS, with 0 representing no pain and 100 the worst pain ever experienced.
‡ The value is the percentage-point difference between the two, with a 95% CI.

Table 6
Clinical outcome measures at 1 year, stratified by MRI findings at 1 year

	Enhancement disc herniation at 1 yr				1-yr enhancement of the nerve root most affected at baseline			
	Yes (n=59)	No (n=12)	Difference (95% CI)	p	Yes (n=12)	No (n=192)	Difference (95% CI)	p
1 yr outcome								
Roland Disability*	3.4±4.9	2.2±3.7	1.2 (−1.8 to 4.2)	.34	2.8±3.9	3.5±5.5	0.7 (−2.5 to 3.9)	.83
VAS-leg pain†	11.1±20.7	4.0±6.1	7.1 (−5.0 to 19.2)	.43	5.6±7.5	11.1±19.7	5.5 (−5.8 to 16.8)	1.00
VAS-back pain†	14.2±20.2	4.8±5.8	9.4 (−2.4 to 21.2)	.17	7.8±9.6	14.7±21.7	6.9 (−5.6 to 19.4)	.59
Perceived recovery	51 (86)	12 (100)	13.6 (−6.4 to 33.6)‡	.34	10 (83)	164 (85)	2.1 (−18.8 to 23.0)‡	.69
Muscle weakness	8 (14)	3 (25)	11.4 (−11.6 to 34.5)‡	.38	2 (17)	42 (22)	5.2 (−19.0 to 29.4)‡	1.00
Sensory loss	17 (29)	6 (50)	21.2 (−8.4 to 50.7)‡	.18	2 (17)	63 (33)	16.1 (−11.2 to 43.5)‡	.35
Reflex loss	22 (37)	5 (42)	4.4 (−26.7 to 35.5)‡	.76	5 (42)	85 (44)	2.6 (−26.7 to 31.9)‡	1.00

MRI, magnetic resonance imaging; CI, confidence interval; VAS, visual analog scale; SD, standard deviation.

Note: Of the 204 patients with both Gd-MRIs at baseline and 1 year, 71 still had a herniated disc at 1 year. Values are n (%) or means±SD.

* The Roland Disability Questionnaire for sciatica is a disease-specific disability scale that measures the functional status of patients with pain in the leg or back. Scores range from 0 to 23, with higher scores indicating worse functional status.

† The intensity of pain is indicated on a horizontal 100-mm VAS, with 0 representing no pain and 100 the worst pain ever experienced.

‡ The value is the percentage-point difference between the two, with a 95% CI.

1-year enhancement in relation to 1-year clinical data

Patients with and without enhancing herniated disc at 1 year did not significantly differ in perceived recovery (86% vs. 100%; p=.34) (Table 6). Of the few patients with 1-year enhancement of the nerve root, 83% reported perceived recovery compared with 85% with no enhancement (p=.69). Patients with and without enhancing herniated discs or nerve roots showed comparable outcomes on RDQ, VAS-leg pain, VAS-back pain, and neurologic findings. Analyses stratified according to surgical status at 1 year yielded similar results (Table S3). The stratified analysis according to the presence of leg pain at baseline also yielded similar results (Table S4).

Discussion

Within patients with symptomatic lumbar disc herniations at baseline who were followed for 1 year, this study presented poor-to-moderate agreement about Gd enhancement in lumbar spine MRIs between observers, which is in firm contrast with their excellent agreement about the disc level of the herniated disc and compressed nerve root. This study also showed that even with Gd-MRI only, moderate agreement was reached regarding the presence of scar tissue at 1 year. Furthermore, no relationship was observed between enhancement and clinical findings at 1 year.

Previous studies reported contradictory results regarding the clinical value of nerve root enhancement in patients with sciatica [18,20,23–26]. Two studies reported a correlation between nerve root enhancement on MRI and clinical symptoms in patients who had undergone lumbar disc surgery [18,20]. Unfortunately, these two studies included only patients with residual or recurrent sciatica after surgery and thus, lacked comparisons with asymptomatic patients (as control subjects). In a prospective cohort study, in which symptomatic and asymptomatic persons were evaluated, Nygaard et al. [26] found no association between nerve root enhancement and clinical outcome 1 year after surgery, when patients with recurrent disc herniation were excluded. Taneichi et al. [27] also did not observe an association between nerve root enhancement and radicular symptoms in the postoperative lumbar spine.

Because the interobserver agreement regarding the enhancement findings was poor–to-moderate, one could question the added value of correlating enhancement with clinical findings. The lack of an association between MRI enhancement and clinical outcomes could be due in part to misclassification of MRI findings, a true lack of association, or both. With the exception of one study (kappa=0.66 for nerve root enhancement between two radiologists) [26], no prevailing studies reported on the interobserver agreement with regard to the enhancement findings. Within the radiologic literature, values of agreement show a high variation depending on the variable investigated [43]. Even regarding the most involved disc level, important for making treatment decisions, disagreement arose in 3% of the cases in this study that is in agreement with previous literature. [44]. However, it is crucial that radiologists and clinicians strive to reduce variability in interpretations because inconsistency in interpretation may lead to alternative treatment options between clinicians and therefore, may impact the outcome of patient treatment [45,46]. Moreover, to gain more insight in the relationship between specific imaging characteristics and patient outcomes, those interpreting the images must reliably assess the finding. One reason that a prediction model might lose its predictive power is the incorrect assessment of MRI findings that causes the inputs in the prediction model to be faulty [47]. As a first step in the attempt to achieve better agreements between observers, the language for image interpretation for degenerative disc disease has to be defined [48]. Radiologists and clinicians should strive to define a nomenclature that has the best support among clinicians and radiologists. In addition to defining the language for image interpretation for degenerative disc disease, reading training might be an important next

step [47,49]. In support are the results of two reliability studies of The Spine Patient Outcomes Research Trial [44,50]. In one of the two studies, the reported agreement on disc morphology was only fair (kappa=0.24) between the clinicians and radiologists [44]. In another study, interreader reliability for disc morphology was excellent (kappa=0.81) between three radiologists and one orthopedic surgeon [50]. The observation of a much better agreement in the second study might be explained by better training of the MRI assessors, as in that study the MRI assessors, before beginning the study, first evaluated a sample set of images with use of definitions and afterward they met in person to review each image, enabling them to better streamline the way of interpreting the images.

In this study, we observed that nerve root enhancement decreased during 1-year follow-up (30% showed enhancement of nerve root at baseline, whereas at 1 year, 5% of surgically and 7% of conservatively treated patients showed nerve root enhancement). It is generally believed that enhancement with Gd in an MRI is caused by hypervascularity of the periradicular vessels by epidural fibrosis in response to the herniated disc or disruption of the blood-nerve barrier [20,24]. In the preoperative MRI study of Vroomen et al. [23], 25% showed nerve root enhancement and 76% showed enhancement of the disc herniation, results comparable to our study. Unfortunately, only limited studies are available that studied changes in enhancement over time. Boden et al. [51] examined 15 patients (16 levels) with MRI at 3 weeks, 3 months, and 6 months after successful surgery for lumbar disc herniation. They observed intradural contrast enhancement in the nerve root in 10 patients (62%) after 3 weeks, in 2 (12%) after 3 months, and in none after 6 months. In two studies, the total disappearance of nerve root enhancement was seen within 3 months after surgery [24] and within 6 months of conservative treatment [52]. The explanation for a decrease in enhancement over time in these studies and our study could be that in most of the patients, the enhancement disappeared owing to repair of the blood-nerve barrier after the acute stage of sciatica.

The previous publication reported on the correlation between 1-year MRI findings (presence of disc herniation, nerve root compression, and scar tissue) and 1-year clinical outcome [22]. It did not investigate interobserver agreement regarding enhancement findings and did not correlate baseline and 1-year MRI enhancement findings with clinical outcome. Furthermore, no correlation was established between MRI and neurologic examination findings.

The present study has several limitations. The reported MRI findings and their relation with clinical outcome was timed only once, 1 year after randomization. Although seemingly generalizable to other time points during the first year, it is scientifically uncertain if we would have found comparable results at other moments. Another potential limitation is that about one-quarter of patients enrolled in the randomized trial did not undergo Gd-enhanced MRI (Gd-MRI) at baseline and at 1 year. Although we did not observe any differences in characteristics between patients with and without Gd-MRIs, the results might have been different if all patients had undergone Gd-MRI. Furthermore, we did not use pixel values in the determination of nerve root enhancement and did not measure the length of root enhancement [18,20], but presence or absence of enhancement was based on the readers' visual qualitative impression, as this is still the most common technique used in clinical practice. Finally, usual reliable statistical packages (STATA, SAS, SPSS) are only able to calculate unweighted kappa coefficients for multiple raters. However, unweighted kappa coefficients are inappropriate for ordinal scales because they treat all disagreements equally [53]. We encourage the development of statistical software that will solve this problem.

In summary, reliability of MRI enhancement findings was poor-to-moderate and no relationship was observed between enhancement and clinical findings at 1 year. Further research is needed to assess the value of Gd-MRI in clinical decision making of patients with acute and persistent or recurrent sciatica.

Appendix

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.spinee.2014.02.028.

References

[1] Bejia I, Younes M, Zrour S, et al. Factors predicting outcomes of mechanical sciatica: a review of 1092 cases. Joint Bone Spine 2004;71:567–71.
[2] Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. Spine J 2008;8:8–20.
[3] Koes BW, van Tulder MW, Peul WC. Diagnosis and treatment of sciatica. BMJ 2007;334:1313–7.
[4] Vroomen PC, de Krom MC, Slofstra PD, Knottnerus JA. Conservative treatment of sciatica: a systematic review. J Spinal Disord 2000;13:463–9.
[5] Weber H, Holme I, Amlie E. The natural course of acute sciatica with nerve root symptoms in a double-blind placebo-controlled trial evaluating the effect of piroxicam. Spine 1993;18:1433–8.
[6] Cihangiroglu M, Yildirim H, Bozgeyik Z, et al. Observer variability based on the strength of MR scanners in the assessment of lumbar degenerative disc disease. Eur J Radiol 2004;51:202–8.
[7] Vucetic N, Astrand P, Guntner P, Svensson O. Diagnosis and prognosis in lumbar disc herniation. Clin Orthop Relat Res 1999; 116–22.
[8] Peul WC, Brand R, Thomeer RT, Koes BW. Influence of gender and other prognostic factors on outcome of sciatica. Pain 2008;138: 180–91.
[9] Peul WC, van den Hout WB, Brand R, et al. Leiden-The Hague Spine Intervention Prognostic Study G. Prolonged conservative care versus early surgery in patients with sciatica caused by lumbar disc

herniation: two year results of a randomised controlled trial. BMJ 2008;336:1355–8.
[10] Osterman H, Seitsalo S, Karppinen J, Malmivaara A. Effectiveness of microdiscectomy for lumbar disc herniation: a randomized controlled trial with 2 years of follow-up. Spine 2006;31:2409–14.
[11] Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonoperative treatment for lumbar disc herniation: four-year results for the Spine Patient Outcomes Research Trial (SPORT). Spine 2008;33:2789–800.
[12] Arts MP, Brand R, van den Akker ME, et al. Tubular diskectomy vs conventional microdiskectomy for sciatica: a randomized controlled trial. JAMA 2009;302:149–58.
[13] Bernard TN Jr. Repeat lumbar spine surgery. Factors influencing outcome. Spine 1993;18:2196–200.
[14] Fandino J, Botana C, Viladrich A, Gomez-Bueno J. Reoperation after lumbar disc surgery: results in 130 cases. Acta Neurochir (Wien) 1993;122:102–4.
[15] Herron L. Recurrent lumbar disc herniation: results of repeat laminectomy and discectomy. J Spinal Disord 1994;7:161–6.
[16] Van Goethem JW, Parizel PM, Jinkins JR. Review article: MRI of the postoperative lumbar spine. Neuroradiology 2002;44:723–39.
[17] Jonsson B, Stromqvist B. Repeat decompression of lumbar nerve roots. A prospective two-year evaluation. J Bone Joint Surg Br 1993;75:894–7.
[18] Grane P, Lindqvist M. Evaluation of the post-operative lumbar spine with MR imaging. The role of contrast enhancement and thickening in nerve roots. Acta Radiol 1997;38:1035–42.
[19] Babar S, Saifuddin A. MRI of the post-discectomy lumbar spine. Clin Radiol 2002;57:969–81.
[20] Lee YS, Choi ES, Song CJ. Symptomatic nerve root changes on contrast-enhanced MR imaging after surgery for lumbar disk herniation. AJNR Am J Neuroradiol 2009;30:1062–7.
[21] Grane P. The postoperative lumbar spine. A radiological investigation of the lumbar spine after discectomy using MR imaging and CT. Acta Radiol Suppl 1998;414:1–23.
[22] el Barzouhi A, Vleggeert-Lankamp CL, Lycklama a Nijeholt GJ, et al. Magnetic resonance imaging in follow-up assessment of sciatica. N Engl J Med 2013;368:999–1007.
[23] Vroomen PC, Van Hapert SJ, Van Acker RE, et al. The clinical significance of gadolinium enhancement of lumbar disc herniations and nerve roots on preoperative MRI. Neuroradiology 1998;40:800–6.
[24] Toyone T, Takahashi K, Kitahara H, et al. Visualisation of symptomatic nerve roots. Prospective study of contrast-enhanced MRI in patients with lumbar disc herniation. J Bone Joint Surg Br 1993;75:529–33.
[25] Jinkins JR, Osborn AG, Garrett D Jr, et al. Spinal nerve enhancement with Gd-DTPA: MR correlation with the postoperative lumbosacral spine. AJNR Am J Neuroradiol 1993;14:383–94.
[26] Nygaard OP, Jacobsen EA, Solberg T, et al. Nerve root signs on postoperative lumbar MR imaging. A prospective cohort study with contrast enhanced MRI in symptomatic and asymptomatic patients one year after microdiscectomy. Acta Neurochir (Wien) 1999;141:619–22; discussion 623.
[27] Taneichi H, Abumi K, Kaneda K, Terae S. Significance of Gd-DTPA-enhanced magnetic resonance imaging for lumbar disc herniation: the relationship between nerve root enhancement and clinical manifestations. J Spinal Disord 1994;7:153–60.
[28] Peul WC, van Houwelingen HC, van den Hout WB, et al. Surgery versus prolonged conservative treatment for sciatica. N Engl J Med 2007;356:2245–56.
[29] Peul WC, van Houwelingen HC, van der Hout WB, et al. Prolonged conservative treatment or "early" surgery in sciatica caused by a lumbar disc herniation: rationale and design of a randomized trial [ISRCT 26872154]. BMC Musculoskelet Disord 2005;6:8.
[30] Ross JS, Robertson JT, Frederickson RC, et al. Association between peridural scar and recurrent radicular pain after lumbar discectomy: magnetic resonance evaluation. ADCON-L European Study Group. Neurosurgery 1996;38:855–61; discussion 861–3.
[31] Patrick DL, Deyo RA, Atlas SJ, et al. Assessing health-related quality of life in patients with sciatica. Spine 1995;20:1899–908; discussion 1909.
[32] Collins SL, Moore RA, McQuay HJ. The visual analogue pain intensity scale: what is moderate pain in millimetres? Pain 1997;72:95–7.
[33] Lynn MR. Determination and quantification of content validity. Nurs Res 1986;35:382–5.
[34] Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. BMJ 1992;304:1491–4.
[35] Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:37–46.
[36] Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43:543–9.
[37] Gjorup T. The kappa coefficient and the prevalence of a diagnosis. Methods Inf Med 1988;27:184–6.
[38] Kovacs FM, Royuela A, Jensen TS, et al. Agreement in the interpretation of magnetic resonance images of the lumbar spine. Acta Radiol 2009;50:497–506.
[39] Arana E, Royuela A, Kovacs FM, et al. Lumbar spine: agreement in the interpretation of 1.5-T MR images by using the Nordic Modic Consensus Group classification form. Radiology 2010;254:809–17.
[40] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
[41] Peters ML, Sommer M, de Rijke JM, et al. Somatic and psychologic predictors of long-term unfavorable outcome after surgical intervention. Ann Surg 2007;245:487–94.
[42] Yamashita K, Ohzono K, Hiroshima K. Patient satisfaction as an outcome measure after surgical treatment for lumbar spinal stenosis: testing the validity and discriminative ability in terms of symptoms and functional status. Spine 2006;31:2602–8.
[43] Pfirrmann CW, Metzdorf A, Zanetti M, et al. Magnetic resonance classification of lumbar intervertebral disc degeneration. Spine 2001;26:1873–8.
[44] Lurie JD, Doman DM, Spratt KF, et al. Magnetic resonance imaging interpretation in patients with symptomatic lumbar spine disc herniations: comparison of clinician and radiologist readings. Spine 2009;34:701–5.
[45] Mulconrey DS, Knight RQ, Bramble JD, et al. Interobserver reliability in the interpretation of diagnostic lumbar MRI and nuclear imaging. Spine J 2006;6:177–84.
[46] Ross JS. Babel 2.0. Radiology 2010;254:640–1.
[47] Carrino JA, Lurie JD, Tosteson AN, et al. Lumbar spine: reliability of MR imaging findings. Radiology 2009;250:161–70.
[48] Milette PC. Reporting lumbar disk abnormalities: at last, consensus! AJNR Am J Neuroradiol 2001;22:428–9.
[49] Jarvik JG, Deyo RA. Moderate versus mediocre: the reliability of spine MR data interpretations. Radiology 2009;250:15–7.
[50] 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.
[51] Boden SD, Davis DO, Dina TS, et al. Contrast-enhanced MR imaging performed after successful lumbar disk surgery: prospective study. Radiology 1992;182:59–64.
[52] Modic MT, Ross JS, Obuchowski NA, et al. Contrast-enhanced MR imaging in acute lumbar radiculopathy: a pilot study of the natural history. Radiology 1995;195:429–35.
[53] Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 2005;85:257–68.