Chapter 8

Reliability of Gadolinium-enhanced MRI findings and their correlation with clinical outcome in patients with sciatica

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

Submitted for publication

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 findings and clinical findings is debated and limited data is 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/SETTING

A randomized clinical trial with one year follow-up.

PATIENTS SAMPLE

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

OUTCOME MEASURES

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

METHODS

In total 204 patients underwent Gd-MRI at baseline and after one year. MRI findings were correlated to the outcome measures using The Mann-Whitney U test for continuous data and Fisher's exact tests for categorical data. 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.

RESULTS

Poor to moderate agreement was observed regarding gadolinium enhancement of the herniated disc and compressed nerve root (kappa <0.41) which was in contrast with excellent interobserver agreement about the disc level of the herniated disc and compressed nerve root (kappa >0.95). Of the 59 patients with an enhancing herniated disc at one year, 86% reported recovery compared to 100% of the 12 patients with non-enhancing herniated discs (P=0.34). Of the 12 patients with enhancement of the most affected nerve root at one year 83% reported

recovery compared to 85% of the 192 patients with no enhancement (P=0.69). Patients with and without enhancing herniated discs or nerve roots at one year reported comparable outcomes on RDQ and VAS-leg pain.

CONCLUSION

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

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] Since 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 following 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 while 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, as it has been proposed to differentiate between postoperative epidural scar tissue and recurrent disc herniation: scar tissue has a homogenous enhancement pattern while 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 one year follow-up a considerable proportion of patients still had a visible disc herniation on MRI (21% of surgically compared to 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 specific objectives of the present study were to evaluate interobserver agreement among experienced readers regarding MRI enhancement findings, and how well enhancement of nerve root and disc herniation are correlated with clinical outcome and neurological findings at baseline and after one year.

METHODS

STUDY POPULATION AND RANDOMIZATION

Patients for this study were participants in a multicentre randomized controlled trial among patients with 6-12 weeks 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 to 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.[22] 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). One 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.[28]

MRI PROTOCOL AND IMAGE EVALUATION

Patients underwent MRI at baseline and after one year follow-up. The 12 month evaluation period was selected since postoperative fibrosis stabilizes by 6 months, with no further changes at 12 months.[29]

MRI scans were performed in all 9 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 (Gadolinium-DTPA at a standard dose of 0.1 mmol/kg body weight) T1 fat suppressed sagittal and axial images were obtained.

Two neuroradiologists (BK and GL) and one neurosurgeon (CV) independently evaluated all MR images. 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 post-residency for the neuroradiologists and 4 years post-residency for the neurosurgeon.

First, the blinded readers had to decide on the baseline MRI which disc level 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 one-year MRI. On the one-year MRI the readers had also to assess whether scar tissue was present (no, moderate or severe). The readers evaluated the enhancement on the baseline and one year MRI of the following structures using different categories (Appendix Table S1): 1. Disc herniation (if present): no, any edge, complete circumferential or diffuse enhancement, 2. Most affected nerve root: no, mild or strong enhancement and 3. Scar tissue (if present at one year): yes vs no enhancement. Structures were considered enhanced when brighter compared with the precontrast image.

NEUROLOGICAL EXAMINATION

Patients underwent a standardized neurological 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 MRC 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),[30] the 100-mm visual-analogue scale (VAS) for leg pain (with 0 representing no pain and 100 the worst pain ever experienced),[31] and a 7-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 7-point Likert scale was used in dichotomized form: "Complete" or "nearly complete disappearance of symptoms" was defined as "perceived recovery", while a score in the remaining five categories (varying from "minimally improved" to "very much worse") was marked as "no recovery".[22] 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

Interobserver agreement regarding the MRI findings was determined by use of absolute percentages of agreement and kappa values (weighted in case of ordered data). Since the kappa

statistic is affected by the prevalence of the events,[32, 33] kappa values were only calculated for findings reported in more than 10% and less than 90% of all reports.[34] 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 (e.g. 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). Values of less than 0.00 indicated poor; 0.00-0.20 slight; 0.21-0.40 fair; 0.41-0.60 moderate; 0.61-0.80 substantial; and 0.81-1.00 excellent or almost perfect reliability.[35]

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 2 out of 3 readers). In analyses comparing enhancement/no enhancement of disc herniation, ratings were categorized as 1, 2, 3 (any edge, complete circumferential, or diffuse enhancement) vs. 4 (no enhancement). In analyses comparing enhancement/no enhancement of the affected nerve, ratings were categorized as mild or strong enhancement vs. no enhancement. Differences between MRI findings were assessed by using The Mann-Whitney U test for continuous data and Fisher's exact tests for categorical data. Statistical significance was defined as $P< 0.05$.

RESULTS

Of 599 patients screened for the Trial, 283 patients were randomized.[22] One year after randomization a second MRI was available for 267 (94.3%) patients. However, at baseline 230 (81%) underwent MRI with gadolinium and at one 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 one 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).

Table 1 Baseline characteristics of the intention-to-treat groups and the 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	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 in weeks	9.5±2.3	9.5±2.2	9.5±2.3	9.6±2.2
Suspected disc level				
L3L4	5 (5)	2 (2)	6 (5)	1 (1)
L4L5	48 (46)	35 (35)	54 (42)	29 (39)
L5S1	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)
Weeks between baseline and follow-up MRI	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 in mm §	66.1±20.0	62.0±21.1	65.6±20.5	61.7±20.8
VAS back pain in mm §	33.4±29.0	28.5±25.9	32.7±29.7	28.0±23.2

Values are n (%) or means ± SD.
No significant baseline differences were observed in the intention-to-treat group and the 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 visual analogue scale (VAS) with 0 representing no pain and 100 the worst pain ever experienced.

INTERAGREEMENT ANALYSIS ONE YEAR

After one 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 one 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 one year (kappa=0.10). For the presence of scar tissue at one 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%.

Table 2 Interobserver agreement regarding MRI findings at baseline.

	A vs. B		A vs. C		B vs. C		All observers	
	% agree-ment	kappa	% agree-ment	kappa	% agree-ment	kappa	% agree-ment	kappa
Disc level with most severe nerve root compression	97.4	0.95	99.1	0.98	97.4	0.95	97.0	0.96
Probability of disc herniation	96.5	*	99.6	*	96.1	*	96.1	*
Enhancement disc herniation (4 categories) ò	55.0	0.42	50.0	0.34	64.3	0.48	47.8	0.41
Enhancement disc herniation (2 categories) ├	78.2	0.38	77.5	0.35	78.1	0.50	66.5	0.40
Probability nerve root compression	100.0	1.00	100.0	1.00	100.0	1.00	100.0	1.00
Most affected nerve root	97.8	0.97	97.0	0.96	96.5	0.96	95.7	0.96
Enhancement most affected nerve root	58.2	0.27	53.2	0.23	84.8	0.60	48.4	0.28

A en B represent the two neuroradiologists, while 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 (e.g. 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).

ò 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"

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

MRI FINDINGS

Baseline

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

One year

Of the 129 surgically treated patients, 26 still had a herniated disc at one 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 one year and 80% of these herniations enhanced. Of these 45 disc herniations, 32 (71%) were small.

Five percent of surgically treated patients showed one-year enhancement of the affected nerve root as compared to 7% of conservatively treated patients (P=0.76) (Table 4).

Of the 115 patients diagnosed with scar tissue at one 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 baseline showed comparable baseline scores on the RDQ and VAS for leg and back pain (Table 5). At baseline, 80% of the patients with enhancing disc herniation had muscle weakness compared to 62% of the patients with

Table 3 Interobserver agreement regarding MRI findings at one year.

	A vs. B		A vs. C		B vs. C		All observers	
	% agree-ment	kappa	% agree-ment	kappa	% agree-ment	kappa	% agree-ment	kappa
Probability of disc herniation	82.4	0.61	87.6	0.74	85.4	0.66	77.6	0.67
Enhancement disc herniation (4 categories) ò	48.2	0.32	57.5	0.35	55.4	0.32	36.4	0.32
Enhancement disc herniation (2 categories) ├	67.9	0.10	75.0	0.23	67.9	0.24	54.4	0.13
Probability nerve root compression	75.2	0.46	77.2	0.51	92.1	0.76	72.6	0.55
Enhancement of the nerve root that was most affected at baseline	78.8	0.24	73.5	0.03	92.7	*	72.3	0.10
Presence of scar tissue	87.8	0.75	74.2	0.51	77.0	0.55	69.5	0.59
Enhancement scar tissue ╞	99.2	**	97.9	**	97.7	**	97.6	**

A en B represent the two neuroradiologists, while 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 (e.g. when there was disagreement about whether at one year a herniated disc was still visible, this case did not contribute to the interagreement analysis regarding the enhancement of the herniated disc).

ò 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"

╞ Yes vs. no

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

** 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). Values are n (%). Total n=204

	Surgery (129)	No surgery (75)	P Value
Enhancement disc herniation at one year			
Enhanced	23 (18)	36 (48)	0.52
No enhancement	3 (2)	9 (12)	
Not applicable, no disc herniation at one year	103 (80)	30 (40)	
Enhancement at one year of the nerve root thought at baseline to cause symptoms			
Enhanced	7 (5)	5 (7)	0.76
No enhancement	122 (95)	70 (93)	

non-enhancing herniated discs (P=0.02). Patients with enhancing disc herniation had more frequently sensory loss compared to patients with non-enhancing herniated discs (74 vs. 54%, P=0.02). At baseline, 84 and 77% of the patients with enhancement of the affected nerve root had muscle weakness and sensory loss respectively compared to 72 and 68% of the patients with non-enhancing nerve roots (P=0.11 and P=0.24).

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 one year (Table 5). Patients with enhancing nerve roots reported lower VAS-back pain scores at one year compared to patients with no enhancing nerve roots at baseline (9.9

Table 5 Outcome measures at baseline and after one year stratified by enhancement of the herniated disc and affected nerve root at baseline. Of the 204 patients with both Gd-MRI at baseline and one year, 200 patients had a herniated disc at baseline. Values are n (%) or means ± SD.

	Enhancement disc herniation at baseline			**Enhancement of the affected nerve root at baseline**		
	Yes (n=161)	No (n=39)	P Value	Yes (n=61)	No (n=143)	P Value
Roland Disability ‡						
Baseline	16.3±4.0	14.9±4.7	0.10	16.5±3.4	15.9±4.4	0.58
One year	3.3±5.2	3.8±6.1	0.96	2.8±4.8	3.8±5.6	0.15
VAS-leg pain ¶						
Baseline	63.2±21.0	67.6±19.5	0.22	64.8±19.7	63.9±21.0	0.90
One year	10.5±18.9	12.2±21.7	0.72	9.1±16.2	11.5±20.4	0.56
VAS-back pain ¶						
Baseline	29.7±26.8	32.9±29.6	0.63	26.7±25.2	32.7±28.4	0.13
One year	13.5±20.1	18.1±25.8	0.87	9.9±17.2	16.2±22.6	0.02
Perceived recovery at one year	139 (86)	32 (82)	0.46	55 (90)	119 (83)	0.28
Muscle weakness						
Baseline	128 (80)	24 (62)	0.02	51 (84)	103 (72)	0.11
One year	34 (21)	10 (26)	0.53	14 (23)	30 (21)	0.85
Sensory loss						
Baseline	119 (74)	21 (54)	0.02	47 (77)	96 (68)	0.24
One year	49 (30)	15 (38)	0.34	20 (33)	45 (31)	0.87
Reflex loss						
Baseline	102 (64)	25 (64)	1.00	44 (72)	87 (61)	0.15
One year	70 (43)	19 (49)	0.59	26 (43)	64 (45)	0.88

‡ 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 visual analogue scale (VAS) with 0 representing no pain and 100 the worst pain ever experienced

Table 6 Clinical outcome measures at one year stratified by MRI findings at one year. Of the 204 patients with both Gd-MRI at baseline and one year, 71 still had a herniated disc at one year.

	Enhancement disc herniation at one year			One-year enhancement of the nerve root most affected at baseline		
	Yes (n=59)	No (n=12)	P Value	Yes (n=12)	No (n=192)	P Value
One year outcome						
Perceived recovery	51 (86)	12 (100)	0.34	10 (83)	164 (85)	0.69
Roland Disability ‡	3.4±4.9	2.2±3.7	0.34	2.8±3.9	3.5±5.5	0.83
VAS-leg pain ¶	11.1±20.7	4.0±6.1	0.43	5.6±7.5	11.1±19.7	1.00
VAS-back pain ¶	14.2±20.2	4.8±5.8	0.17	7.8±9.6	14.7±21.7	0.59
Muscle weakness	8 (14)	3 (25)	0.38	2 (17)	42 (22)	1.00
Sensory loss	17 (29)	6 (50)	0.18	2 (17)	63 (33)	0.35
Reflex loss	22 (37)	5 (42)	0.76	5 (42)	85 (44)	1.00

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 visual analogue scale (VAS) with 0 representing no pain and 100 the worst pain ever experienced

vs. 16.2mm, P=0.02). The same results were observed in both conservatively and surgically treated patients.

ONE-YEAR ENHANCEMENT IN RELATION TO ONE-YEAR CLINICAL DATA

Patients with and without enhancing herniated disc at one year did not significantly differ in perceived recovery (86% vs. 100% P=0.34) (Table 6). Of the few patients with one-year enhancement of the nerve root 83% reported perceived recovery compared to 85% of the patients with no enhancement (P=0.69). Patients with and without enhancing herniated discs or nerve roots showed comparable outcomes on RDQ, VAS-leg pain, VAS-back pain and neurological findings. Analyses stratified according to surgical status at one year yielded similar results (Table S2).

DISCUSSION

Within patients with symptomatic lumbar disc herniations at baseline who were followed for 1 year, this study presented poor to moderate agreement about gadolinium 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 one year. Furthermore, no relationship was observed between enhancement findings and clinical findings at one 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. found no association between nerve root enhancement and clinical outcome one year after surgery when patients with recurrent disc herniation were excluded.[26] Taneichi et al. did also not observe an association between nerve root-enhancement and radicular symptoms in the post-operative lumbar spine.[27]

Since the interobserver agreement regarding the enhancement findings was poor to moderate, one could question the added value of correlating enhancement findings with clinical findings. 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 radiological literature, values of agreement show a high variation depending on the variable investigated.[36] Even regarding the most involved disc level, important for making treatment decisions, in 3% of the cases disagreement arose in this study, which is in agreement with previous literature.[37] However, it is crucial that radiologists and clinicians strive to reduce variability in interpretations as inconsistency in interpretation may lead to alternative treatment options between clinicians and therefore may impact the outcome of patient treatment.[38, 39] 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, which causes the inputs in the prediction model to be faulty.[40]

The present study has several limitations. Firstly, the reported MRI findings and their relation with clinical outcome was timed only once, one 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. Secondly, 72% underwent Gd-MRI both at baseline and one year. We can not exclude the possibility of other findings if all patients underwent Gd-MRI. Finally, we did not use pixel values in the determination of nerve root enhancement and did also not measure the length of root enhancement,[18, 20] but presence or absence of enhancement was based on the readers visual intuitive impression as this is still the most common technique used in clinical practice.

In summary, reliability of MRI enhancement findings was poor to moderate and no relationship was observed between enhancement findings and clinical findings at one 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.

REFERENCES

1. Bejia I, Younes M, Zrour S, Touzi M, Bergaoui N. 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 (Phila Pa 1976) 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, Thomeer RT, Koes BW. 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 (Phila Pa 1976) 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 (Phila Pa 1976) 2008;33:2789-800.
12. Arts MP, Brand R, van den Akker ME, Koes BW, Bartels RH, Peul WC. 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 (Phila Pa 1976) 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. 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.
23. Vroomen PC, Van Hapert SJ, Van Acker RE, Beuls EA, Kessels AG, Wilmink JT. 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, Yamagata M, Murakami M, Moriya H. 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., Hunt S, Story JL. 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, Kloster R, Dullerud R. 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 23.
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 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.
29. 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 61-3.
30. Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB. Assessing health-related quality of life in patients with sciatica. Spine (Phila Pa 1976) 1995;20:1899-908; discussion 909.
31. Collins SL, Moore RA, McQuay HJ. The visual analogue pain intensity scale: what is moderate pain in millimetres? Pain 1997;72:95-7.
32. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43:543-9.
33. Gjorup T. The kappa coefficient and the prevalence of a diagnosis. Methods Inf Med 1988;27:184-6.
34. 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.
35. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
36. Pfirrmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N. Magnetic resonance classification of lumbar intervertebral disc degeneration. Spine (Phila Pa 1976) 2001;26:1873-8.
37. Lurie JD, Doman DM, Spratt KF, Tosteson AN, Weinstein JN. Magnetic resonance imaging interpretation in patients with symptomatic lumbar spine disc herniations: comparison of clinician and radiologist readings. Spine (Phila Pa 1976) 2009;34:701-5.
38. 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;6:177-84.
39. Ross JS. Babel 2.0. Radiology 2010;254:640-1.
40. Carrino JA, Lurie JD, Tosteson AN, et al. Lumbar spine: reliability of MR imaging findings. Radiology 2009;250:161-70.

Table S1 MRI study variables. For both the MRI at baseline and one year after randomization the three readers (2 neuroradiologists and one neurosurgeon) independently used the same case record forms, with the exception that the one-year case record forms also included questions regarding the presence of scar tissue and its enhancement.

MRI variable	Type	Categories
Disc level with the most severe nerve root compression	Disc level	1. L2L3 2. L3L4 3. L4L5 4. L5S1 5. Not applicable, all disc levels have a normal disc contour
	Disc contour at this disc level	1. Normal: no disc extension beyond the normal margins of the intervertebral disc space 2. Bulging: presence of disc tissue circumferentially (50-100%) beyond the edges of the ring apophyses 3. Consideration of a disc herniation: localized displacement of disc material beyond the normal margins of the intervertebral disc space
	Certainty about the presence of a disc herniation	1. Definite about the presence: no doubt about the presence 2. Probable about the presence: some doubt but probability > 50% 3. Possible about the presence: reason to consider but probability < 50% 4. Definite about the absence: no doubt about the absence of a disc herniation
If a herniation at the disc level is considered	Gadolineum enhancement of the intervertebral disc herniation	1. No enhancement 2. Any edge enhancement 3. Complete circumferential enhancement 4. Diffuse enhancement
	Size of this disc herniation in relation to spinal canal	1. Large stenosing: size >75% of the spinal canal 2. Large: size 75-50% of the spinal canal 3. Average: size 25-50% of the spinal canal 4. Small: size <25% of the spinal canal
Scar tissue	Presence	1. No: scar tissue absent 2. Yes, moderate scar tissue present 3. Yes, severe scar tissue present
	If present, place scar tissue	1. Scar tissue surrounds the nerve root 2. Scar tissue does not surround the nerve root
	Gadolineum enhancement	1. Yes 2. No
Nerve root compression	Probability of nerve root compression	1. Definite about the presence: no doubt about the presence 2. Probable about the presence: (probability > 50%) 3. Possible about the presence: reason to consider but probability < 50% 4. Definite about the absence: no doubt about the absence of a disc herniation

Table S1 (Continued)

MRI variable	Type	Categories
	If root compression present, which nerve root is affected	1. L3 2. L4 3. L5 4. S1 5. Not applicable, definitely no nerve root compression
	Side nerve root compression	1. Right 2. Left
	Gadolineum enhancement of the affected nerve root	1. No enhancement 2. Yes, mild enhancement 3. Yes, strong enhancement

Table S2A Clinical outcome measures at one year stratified by MRI findings at one year. Values are n (%) or means ± SD.
A) Group that underwent surgery during the first year. Of the 129 surgical patients with both Gd-MRI at baseline and one year, 26 still had a herniated disc at one year.
B) Group that underwent conservative care during the first year. Of the 75 conservative patients with both Gd-MRI at baseline and one year, 45 still had a herniated disc at one year.
S2A

	Enhancement disc herniation at one year			One-year enhancement of the nerve root most affected at baseline		
	Yes (n=23)	No (n=3)	P Value	Yes (n=7)	No (n=122)	P Value
One year outcome						
Perceived recovery	20 (87)	3 (100)	1.00	6 (86)	105 (86)	1.00
Roland Disability ‡	3.3±5.0	2.3±4.0	0.82	2.0±3.7	3.6±5.8	0.55
VAS-leg pain ¶	13.0±25.5	2.3±2.5	0.90	5.3±9.4	11.9±21.5	0.78
VAS-back pain ¶	13.7±22.4	3.7±3.2	0.93	5.9±10.8	14.8±22.9	0.29
Muscle weakness	2 (9)	1 (33)	0.32	1 (14)	29 (24)	1.00
Sensory loss	4 (17)	3 (100)	0.01	1 (14)	40 (33)	0.43
Reflex loss	10 (44)	3 (100)	0.22	5 (71)	54 (44)	0.25

‡ 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 visual analogue scale (VAS) with 0 representing no pain and 100 the worst pain ever experienced

S2B						
	Enhancement disc herniation at one year			One-year enhancement of the nerve root most affected at baseline		
	Yes (n=36)	No (n=9)	P Value	Yes (n=5)	No (n=70)	P Value
One year outcome						
Perceived recovery	31 (86)	9 (100)	0.57	4 (80)	59 (84)	1.00
Roland Disability ‡	3.4±5.0	2.1±3.8	0.32	4.0±4.3	3.4±4.9	0.69
VAS-leg pain ¶	9.8±17.2	4.6±6.9	0.33	6.0±4.6	9.6±16.1	0.80
VAS-back pain ¶	14.5±19.0	5.1±6.6	0.11	10.6±7.8	14.7±19.7	0.74
Muscle weakness	6 (17)	2 (22)	0.65	1 (20)	13 (19)	1.00
Sensory loss	13 (36)	3 (33)	1.00	1 (20)	23 (33)	1.00
Reflex loss	12 (33)	2 (22)	0.70	0 (0)	31 (44)	0.07

‡ 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 visual analogue scale (VAS) with 0 representing no pain and 100 the worst pain ever experienced