Chapter 6

Magnetic Resonance Imaging in follow-up assessment of sciatica

A. el Barzouhi, C.L.A.M. Vleggeert-Lankamp, G.J. Lycklama à Nijeholt, B.F. Van der Kallen, W.B. van den Hout, W.C.H. Jacobs, B.W. Koes, and W.C. Peul, for the Leiden–The Hague Spine Intervention Prognostic Study Group

N Engl J Med 2013 Mar 14;368(11):999-1007

ABSTRACT

BACKGROUND

Magnetic resonance imaging (MRI) is frequently performed during follow-up in patients with known lumbar-disk herniation and persistent symptoms of sciatica.
The association between findings on MRI and clinical outcome is controversial.

METHODS

We studied 283 patients in a randomized trial comparing surgery and prolonged conservative care for sciatica and lumbar-disk herniation. Patients underwent MRI at baseline and after 1 year. We used a 4-point scale to assess disk herniation on MRI, ranging from 1 for "definitely present" to 4 for "definitely absent." A favorable clinical outcome was defined as complete or nearly complete disappearance of symptoms at 1 year. We compared proportions of patients with a favorable outcome among those with a definite absence of disk herniation and those with a definite, probable, or possible presence of disk herniation at 1 year. The area under the receiver-operating characteristic (ROC) curve was used to assess the prognostic accuracy of the 4-point scores regarding a favorable or unfavorable outcome, with 1 indicating perfect discriminatory value and 0.5 or less indicating no discriminatory value.

RESULTS

At 1 year, 84% of the patients reported having a favorable outcome. Disk herniation was visible in 35% with a favorable outcome and in 33% with an unfavorable outcome (P = 0.70). A favorable outcome was reported in 85% of patients with disk herniation and 83% without disk herniation (P = 0.70). MRI assessment of disk herniation did not distinguish between patients with a favorable outcome and those with an unfavorable outcome (area under ROC curve, 0.48).

CONCLUSIONS

MRI performed at 1-year follow-up in patients who had been treated for sciatica and lumbar-disk herniation did not distinguish between those with a favorable outcome and those with an unfavorable outcome. (Funded by the Netherlands Organization for Health Research and Development and the Hoelen Foundation; Controlled Clinical Trials number, ISRCTN26872154.)

INTRODUCTION

Sciatica is a relatively common condition, with a lifetime incidence of 13 to 40%.[1] The most common cause of sciatica is a herniated disk. The natural history of sciatica is favorable, with spontaneous resolution of leg pain within 8 weeks in the majority of patients.[2] Surgery should be offered only if symptoms persist after a period of conservative treatment. However, contrary to what one might expect, given the advancements in diagnostic imaging and surgical techniques, the results after lumbar-disk surgery do not seem to have improved during recent decades. Both classical studies and randomized, controlled trials have shown that during longer follow-up at least 15 to 20% of patients report recurring or persistent symptoms after a first episode of sciatica, regardless of whether they underwent surgery.[3-6] Persistent or recurrent sciatica despite treatment leads to physical and emotional suffering for the patient and substantial costs in terms of treatment, sick leave, and pensions for society.[7, 8]

Magnetic resonance imaging (MRI), which is considered the imaging procedure of choice for patients in whom lumbar-disk herniation is suspected,[9, 10] is frequently performed in patients with persistent or recurrent symptoms of sciatica.[11] However, the association between findings on MRI and symptoms is controversial, with several studies showing a high prevalence of disk herniation, ranging from 20 to 76%, in persons without any symptoms.[9, 12] Even after disk surgery, MRI studies have shown disk herniation in up to 53% of asymptomatic persons.[13-15] Therefore, one could question the value of repeating MRI in clinical practice, given the high percentage of MRI abnormalities in persons with no clinical history of sciatica or physical findings of nerve root pain.[11, 16] Despite the scientific debate, physicians often order repeat MRI studies (usually with gadolinium) for patients with persistent or recurrent symptoms of sciatica.[11] Moreover, abnormal MRI findings frequently result in surgical treatment or other invasive procedures, such as epidural injections.[17, 18]

We previously reported the clinical outcome results of a randomized, controlled trial, which was designed to define the effect of timing of surgery for patients with sciatica.[4] The trial showed that recovery after early surgery was faster than a strategy of prolonged conservative care with surgery if needed, but there were no significant differences in clinical outcomes after 1 year. We now report on the radiologic findings at 1 year, changes in these findings over time, and their correlation with clinical outcome.

METHODS

STUDY POPULATION

Patients in this study were participants in the Sciatica Trial, a multicenter, randomized trial among patients with a history of 6 to 12 weeks of sciatica and disk herniation, as seen on MRI. Patients were included only 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 whose symptoms did not improve or who requested surgery earlier because of aggravating symptoms.[4, 19] The medical ethics committee at each of the nine participating hospitals approved the protocol, which is available with the full text of this article at NEJM.org. Written informed consent was obtained from all patients.

MRI PROTOCOL AND IMAGE EVALUATION

Patients underwent MRI at baseline and 1 year after randomization. The 1-year evaluation period was selected since postoperative fibrosis usually stabilizes by 6 months, with no further changes at 1 year.[20]

MRI scans were performed at each study center with the use of standardized protocols tailored to a 1.5-Tesla scanner. Sagittal T1-weighted images and axial T1-weighted spin–echo images of the lumbar spine were obtained, as well as T2-weighted sagittal and axial series and contrast enhanced (gadolinium) fat-suppressed T1-weighted images.

Two experienced neuroradiologists and one neurosurgeon independently evaluated all MRI scans. The readers were not provided any clinical information and had not been involved in the selection or care of the included patients.

Definitions of imaging characteristics were based on recommendations from the combined task forces of the North American Spine Society, the American Society of Spine Radiology, and the American Society of Neuroradiology for classification of lumbar-disk pathology.[21] Before the start of the study, the readers met in person to evaluate and refine the definitions. Standardized case-record forms with final definitions were used to evaluate the images (see Table S1 in the Supplementary Appendix, available at NEJM.org).

First, the readers had to decide which disk level showed the most severe nerve-root compression. At this level, the disk contour was categorized into one of three categories: disk herniation, bulging disk, and normal disk. Afterward, the readers used a 4-point scale to evaluate the scans for the presence of disk herniation and root compression as follows: 1 for definite presence, 2 for probable presence, 3 for possible presence, and 4 for definite absence.

Scans that were categorized as "definite absence" of disk herniation may have included those with either a normal or bulging disk. When a disk herniation was considered to be present (definite, probable, or possible), multiple characteristics of the disk herniation were additionally scored.

OUTCOMES

In the randomized trial, the original primary outcome measure that was used to define a favorable outcome at 1 year was the Roland Disability Questionnaire (RDQ) for Sciatica (with scores ranging from 0 to 23, with higher scores indicating worse functional status).[22] Original secondary outcome measures were the response of a 7-point Likert self-rating scale

of global perceived recovery (with a higher score indicating better recovery) and the 100-mm visual-analogue scale for leg pain (with 0 representing no pain and 100 the worst pain ever experienced).[23] Since the responsiveness of the RDQ score has been shown to depend on the external criteria used to assess pain or disability,[24] we decided to define a favorable outcome at 1 year as complete or nearly complete disappearance of symptoms on the patient-reported 7-point Likert scale for global perceived recovery.[4, 19] All outcome measures were assessed at baseline and at 2, 4, 8, 12, 26, 38, and 52 weeks.

Patients were not aware of results of earlier assessments and MRI findings. For the purposes of this study, the results at baseline and at 1 year were used in the analysis.

STATISTICAL ANALYSIS

The majority opinion of the three readers regarding the MRI characteristics (answered independently by at least two of the three) was used in the statistical analysis. Interobserver agreement regarding the MRI findings was determined with the use of absolute percentages of agreement and kappa values (weighted in cases of ordered data). In analyses comparing ratings for the presence or absence of disk herniation or root compression, the ratings were dichotomized (definitely, probably, or possibly present vs. definitely absent). Mean scores on the RDQ and visual-analogue scale for leg and back pain were stratified and compared according to MRI findings. In a subanalysis, MRI characteristics were also compared between patients without persistent leg or back pain and those with such pain, defined as a score on the visual-analogue scale of leg or back pain of a least 40 mm,[25, 26] or less than 30% of improvement in the score between baseline and 1 year.[27, 28] MRI characteristics were also compared between patients with a score on the RDQ of less than 14 and those with a score of 14 or more.[29]

Analysis of the receiver-operating-characteristic (ROC) curve was used to assess the diagnostic accuracy of ordinal 1-year MRI findings (4-point scale for assessing disk herniation and root compression) for a favorable outcome at 1 year. The area under the ROC curve (AUC) ranges from 0 to 1 and provides a measure of a test's ability to discriminate between participants who have the outcome of interest and those who do not.[30] A test that correctly classifies all participants has an AUC of 1.0, and a test with no discriminatory value has an AUC of 0.5 or less.[30]

We also used basic measures of diagnostic test accuracy: sensitivity (proportion of patients with an unfavorable outcome who had an abnormal test finding), specificity (proportion of patients with a favorable outcome with no abnormal test finding), positive predictive value, and negative predictive value. For binary variables, these measures were derived from two-by-two tables. For ordinal variables (e.g., presence of disk herniation and root compression), these measures were derived by varying the cutoff point used to define a positive test. Differences between groups for continuous data were assessed by means of Student's t-test. In logistic-regression models, the association between MRI findings and clinical outcome was adjusted for randomized treatment and treatment received. Model-based multiple imputation was used

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

	Intention to treat Analysis¶		As- treated Analysis¶	
	Randomized to early surgery (N=131)	Randomized to prolonged conservative care (N=136)	Received surgery (n=170)	Received no surgery (n=97)
Age-yr	41.7±9.9	43.2±9.2	41.8±9.8	43.5±9.2
Male sex	84 (64.1)	96 (70.6)	111 (65.3)	69 (71.1)
Body-mass indexò *	26.0±4.1	25.6±3.3	26.2±3.9	25.1±3.4
Duration of sciatica in weeks	9.5±2.4	9.6±2.2	9.5±2.4	9.6±2.1
Receipt of pain medication	121 (92)	120 (88)	87 (91)	154 (91)
Suspected disk level and type of displacement on MRI				
L3L4 Herniation	5 (3.8)	4 (2.9)	7 (4.1)	2 (2.1)
L4L5 Herniation	59 (45.0)	50 (36.8)	71 (41.8)	38 (39.2)
L4L5 Bulging	2 (1.5)	1 (0.7)	3 (1.8)	0 (0)
L5S1 Herniation	64 (48.9)	79 (58.1)	88 (51.8)	55 (56.7)
L5S1 Bulging	1 (0.8)	2 (1.5)	1 (0.6)	2 (2.1)
Nerve root compression on MRI				
Definite	82 (62.6)	96 (70.6)	112 (65.9)	66 (68.0)
Probable	35 (26.7)	29 (21.3)	42 (24.7)	22 (22.7)
Possible	11 (8.4)	10 (7.4)	13 (7.6)	8 (8.2)
Definitely not	3 (2.3)	1 (0.7)	3 (1.8)	1 (1.0)
Time between baseline and follow-up MRI-wk	53.4±3.1	52.7±3.8	53.0±3.7	53.0±3.2
Sensory loss	84 (64.1)	102 (75.0)	118 (69.4)	68 (70.1)
Abnormal reflexes ╪	82 (62.6)	97 (71.3)	111 (65.3)	68 (70.1)
Muscle weakness ‖	94 (71.8)	109 (80.1)	130 (76.5)	73 (75.3)
Abnormal result on neurological test ╞ **	122 (93.1)	124 (91.2)	162 (95.3)	84 (86.6)
Roland Disability score ‡ **	16.4±4.5	16.1±4.0	16.7±4.2	15.4±4.1
VAS leg pain in mm § *	66.9±20.0	63.5±21.2	67.2±19.9	61.5±21.5
VAS back pain in mm §	33.8±29.3	30.7±27.0	34.1±30.2	28.9±23.8

Values are n (%) or means ± SD.

No significant baseline differences were observed in the intention-to-treat group

* P<0.05 for the difference in the as-treated group

** P<0.01 for the difference in the as-treated group

¶ Based on n=267 as one year after randomization a second MRI was available for 267 of the 283 randomized patients

ò Body-mass index is the weight in kilograms divided by the square of the height in meters

╪ Reflexes were rated as abnormal if absent, less than the other side, or in case of an extensor plantar response (Babinski sign).

‖ Muscle strength was considered normal in case of MRC Grade 5 whereas Grade 4 or less was rated abnormal.

╞ Six neurological tests were performed (Lasègue's sign, Crossed straight-leg raising, Kemp's sign, Bragard's Sign, walking on heels and walking on toes). One or more abnormal tests was considered to be an abnormal result.
‡ 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.

Table 2 Differences in 1-year MRI findings between patients who actually received surgery and those who did not receive surgery during the first year (as-treated).

	Surgery (170)	No surgery (97)	P Value
Disk herniation¶			**<0.001**
Definite	15 (8.8)	25 (25.8)	
Probable (some doubt but probability > 50%)	18 (10.6)	26 (26.8)	
Possible (reason to consider, but probability < 50%)	2 (1.2)	7 (7.2)	
Definitely not			
Normal disk	106 (62.4)	23 (23.7)	
Bulging disk	29 (17.1)	16 (16.5)	
As compared with baseline			**<0.001**
Disappeared	134 (78.8)	37 (38.1)	
Reduced in size	24 (14.1)	51 (52.6)	
Unchanged or enlarged in size	8 (4.7)	7 (7.2)	
Not applicable, no disk herniation at baseline	4 (2.4)	2 (2.1)	
Nerve-root compression‡			**<0.001**
Definite	5 (2.9)	6 (6.2)	
Probable (some doubt but probability > 50%)	6 (3.5)	6 (6.2)	
Possible (reason to consider, but probability < 50%)	16 (9.4)	26 (26.8)	
Definitely not	143 (84.1)	59 (60.8)	
As compared with baseline			**<0.001**
Disappeared	140 (82.4)	58 (59.8)	
Reduced	16 (9.4)	29 (29.9)	
Unchanged or increased	14 (8.2)	10 (10.3)	

Values are n (%).
* Shown are values at 1 year for 267 of 283 patients for whom data were available on a second magnetic resonance imaging (MRI) scan.
¶ A four point scale was used for the presence of disk herniation ranging from 1 (definitely present) to 4 (definitely absent). When a disk herniation was definitely absent the disk contour could be either normal or a bulging disk.
‡ A four point scale was used for the presence of nerve root compression ranging from 1 (definitely present) to 4 (definitely absent).

to account for missing data with respect to clinical outcome at 1 year (with the use of variables mentioned in Tables 1 and 2). As sensitivity analyses, we performed analysis as observed (e.g., no imputation), analysis using the last-observation-carried-forward method, and analysis in which all three readers agreed about the MRI findings.

RESULTS

PATIENTS

Of the 599 patients who were screened for the Sciatica Trial, 283 underwent randomization in our study.[4] One year after randomization, results on a second MRI were available for 267 patients (94.3%) (Table S2 in the Supplementary Appendix). Baseline characteristics were similar among patients for whom a second MRI was available, as compared with those for whom a second scan was not available.

Of the 267 patients who were eligible for analysis, 131 had been randomly assigned to undergo early surgery and 136 to receive prolonged conservative care. Of the 131 patients in the surgery group, 15 recovered before surgery could be performed. Of the 136 patients in the conservative care group, 54 eventually underwent surgery within the first year. Baseline characteristics of the intention-to-treat and the as-treated groups are shown in Table 1.

One year after randomization, 84% of the patients reported having a favorable outcome on the basis of the global perceived recovery scale. Clinical outcomes at 1 year were missing for 2 to 3% of the patients (Table S3 in the Supplementary Appendix).

Moderate-to-substantial agreement was found for the MRI assessment of the presence of a herniated disk (kappa range, 0.57 to 0.67), nerveroot compression (kappa range, 0.46 to 0.74), and scar tissue (kappa range, 0.50 to 0.77) (Table S4 in the Supplementary Appendix).

MRI FINDINGS AT 1 YEAR

At 1-year follow-up in the as-treated analysis, a herniated disk was considered to be present in 21% of patients who had undergone surgery and in 60% of those who had received conservative treatment (P<0.001) (Table 2). Nerve-root compression was observed significantly more frequently in patients who had received conservative treatment than in those who had undergone surgery (39% vs. 16%, P<0.001). As compared with baseline, root compression had disappeared in 82% of patients who had undergone surgery and in 60% of those who had received conservative treatment (P<0.001).

In the intention-to-treat analysis, results according to randomized group are shown in Table S5 in the Supplementary Appendix. At 1-year followup, a herniated disk was considered to be present in 22% of patients in the surgery group and in 47% of patients in the conservative-care group (P<0.001).

Table 3 MRI findings according to favorable outcome at one year.

MRI findings	Unfavorable outcome (n=43)	Favorable outcome (n=224)	Difference in proportion (95% CI)	P Value
Disk herniation				
Presence at 1 year¶	14 (33)	79 (35)	-2.7 (-18.8 to 12.6)	0.70
Size at 1 yr, as compared with baseline size				
Disappeared	28 (65)	143 (64)	1.3 (-14.2 to 17.5)	0.84
Reduced	9 (21)	66 (29)	-8.5 (-22.5 to 7.2)	0.31
Unchanged	3 (7)	7 (3)	3.9 (-0.0 to 10.0)	0.23
Enlarged	2 (5)	3 (1)	3.3 (-0.0 to 0.07)	0.43
Not applicable, no disk herniation at baseline	1 (2)	5 (2)	0.1 (-4.8 to 4.9)	0.98
Nerve-root compression				
Presence at one year‡	11 (26)	54 (24)	1.5 (-13.1 to 15.4)	0.87
Visibility on MRI at 1 yr, as compared with baseline				
Disappeared	29 (67)	169 (75)	-8.0 (-22.1 to 6.8)	0.30
Reduced	6 (14)	39 (17)	-3.5 (-15.0 to 10.0)	0.69
Unchanged	6 (14)	13 (6)	8.1 (-0.00 to 16.4)	0.06
Increased	2 (5)	3 (1)	3.3 (-0.03 to 0.07)	0.43

Favorable outcome was defined as complete or nearly complete disappearance of symptoms according to the 7-point Likert scale for global perceived recovery. CI denotes confidence interval. Values are n (%). Total n=267

¶ A four point scale was used for the presence of disk herniation ranging from 1 (definitely present) to 4 (definitely absent). Cases with definite, probable, or possible disk herniation are presented.

‡ A four point scale was used for the presence of nerve root compression ranging from 1 (definitely present) to 4 (definitely absent). Cases with nerve root compression (definite, probable, or possible nerve root compression are presented.

ASSOCIATION BETWEEN MRI FINDINGS AND CLINICAL OUTCOME

At 1 year, disk herniation was visible in 35% of the patients with a favorable outcome and in 33% of those with an unfavorable outcome (95% confidence interval [CI] for difference in proportion, –18.8 to 12.6; P = 0.70) (Table 3). Nerve-root compression was considered to be present in 24% of the patients with a favorable outcome and in 26% of the patients with an unfavorable outcome. Similar results were observed in patients with persistent leg and back pain at 1 year and in those without such pain and in those with an RDQ score of at least 14 and those with a score of less than 14 (Table S6 in the Supplementary Appendix).

Readers' ratings on the 4-point scale assessing the presence of disk herniation on MRI did not distinguish between patients with a favorable outcome versus those with an unfavorable outcome (AUC, 0.48; 95% CI, 0.39 to 0.58) (Fig. S1A in the Supplementary Appendix). Depending on the cutoff point on the 4-point scale that was used to determine a positive test, sensitivity ranged from 0.14 to 0.32 and specificity from 0.65 to 0.85 (Table S7 in the

Supplementary Appendix). The AUC for MRI-assessed nerve-root compression was 0.52 (95% CI, 0.42 to 0.61) (Fig. S1B in the Supplementary Appendix).

Table 4 Clinical Outcomes at 1 year, According to MRI findings.

	Presence of a herniated disk		P Value	Presence of nerve root compression		P Value
	Yes (n=93)	No (n=174)		Yes (n=65)	No (n=202)	
Outcome						
Favorable clinical outcomeò	79 (85)	145 (83)	0.70	54 (83)	170 (84)	0.87
Roland Disability Questionnaire ‡	3.4±5.3	3.4±5.5	0.98	3.8±5.4	3.3±5.5	0.57
VAS-Leg pain¶	11.7±21.9	10.5±18.4	0.66	11.4±21.7	10.8±19.1	0.85
VAS-back pain¶	15.8±23.7	15.0±21.5	0.79	13.3±19.9	15.8±23.0	0.52

Values are n (%) or means ± SD. Total n=267

ò Favorable clinical outcome was defined as complete or nearly complete disappearance of symptoms according to the Likert-7 point scale.

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

Table 5 Unadjusted and multivariable analyses of association between one-year MRI findings and favorable outcome at one year.

Outcome	Univariate analysis			Adjusted for assigned treatment ¶			Adjusted for received treatment ‡			Multivariate analysis ╞		
	OR	95% CI	P-value	OR	95%CI	P-value	OR	95%CI	P-value	OR	95%CI	P-value
No Disk herniation§	0.87	0.43-1.76	0.70	0.82	0.40-1.71	0.60	0.76	0.35-1.65	0.49	0.97	0.39-2.52	0.95
No nerve-root compression ‖	1.07	0.50-2.29	0.87	1.03	0.48-2.25	0.93	1.00	0.45-2.21	0.99	1.84	0.66-5.12	0.24

Favorable outcome was defined as complete or nearly complete disappearance of symptoms according to the 7-point Likert scale for global perceived recovery. OR denotes odds ratio. CI denotes confidence interval. Total n=267

§ A four point scale was used for the presence of disk herniation ranging from 1 (definitely present) to 4 (definitely absent).

‖ A four point scale was used for the presence of nerve root compression ranging from 1 (definitely present) to 4 (definitely absent).

¶ An early surgery strategy vs. prolonged conservative care for an additional 6 months followed by surgery if needed.

‡ Analysis adjusted for actual received treatment (surgery vs. no surgery during the first year).

╞ Analysis adjusted for randomized treatment, age, gender, body-mass index, smoking, Roland Disability Questionnaire score at baseline, Visual Analogue scale for leg and back pain at baseline and presence of one or more abnormal neurological tests (Lasègue's sign, Crossed straight-leg raising, Kemp's sign, Bragard's Sign, walking on heels and walking on toes).

Of the patients with disk herniation at 1 year, 85% reported a favorable outcome, as compared with 83% with no disk herniation at 1 year (P = 0.70) (Table 4). Of the 93 herniated disks, 70% were classified as protrusion and 30% as extrusion. Of the patients with a protrusion, 16% reported having an unfavorable outcome, as compared with 14% of the patients with an extrusion (P = 0.87).

Of the 170 patients who underwent surgery during the first year, 150 (88%) had visible scar tissue on MRI. Of the patients with visible scar tissue, 86% reported a favorable outcome, as compared with 75% with no visible scar tissue (P = 0.19). Of the 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.

After adjustment for randomized treatment, the presence of disk herniation on MRI was not associated with a favorable outcome at 1 year (odds ratio, 0.82; 95% CI, 0.40 to 1.71; P = 0.60), nor was MRI-assessed nerve-root compression (odds ratio, 1.03; 95% CI, 0.48 to 2.25; P = 0.93), the size of the disk herniation (odds ratio, 1.48; 95% CI, 0.43 to 5.01; P = 0.53), or the herniation form (protrusion vs. extrusion) (odds ratio, 0.88; 95% CI, 0.25 to 3.16; P = 0.85) (Table 5, and Table S8 in the Supplementary Appendix). Sensitivity analyses that were performed to account for missing data and interobserver agreement yielded similar results (see the Supplementary Appendix).

DISCUSSION

In this study of patients with symptomatic lumbar disk herniation at baseline who were treated with either surgery or conservative treatment and followed for 1 year, the presence of disk herniation on MRI at 1-year follow-up did not distinguish patients with a favorable clinical outcome from those with an unfavorable outcome. Therefore, patients asking for reimaging because of persistent or recurrent symptoms should be informed about the difficulty in MRI interpretation after a first episode of acute sciatica. A recent systematic review concluded that even in the acute setting of sciatica, evidence for the diagnostic accuracy of MRI is not conclusive.[10]

Other studies have reported results similar to our findings.[7, 13] In a report on 154 conservatively treated patients, Jensen et al.[7] did not observe any correlation between improvement in symptoms and improvement of disk herniation and nerve-root compression on MRI at 14 months. Bath et al. [13] observed a high incidence (approximately 67%) of extrusions and protrusions 2 years postoperatively, although these findings did not correlate with clinical outcome. In a retrospective evaluation of morphologic changes on MRI in 77 patients who had received conservative treatment for sciatica, Komori et al.[31] found that such changes did correspond with clinical results. However, the investigators found that morphologic changes tended to lag behind actual improvement in leg pain.

In a landmark study, Jensen et al.[12] suggested that by considering protrusions and extrusions as two different types of herniation, MRI interpretations could gain specificity for clinically important disk lesions. The authors reached this hypothesis because of the high prevalence (approximately 30%) of disk protrusions among their asymptomatic volunteers, whereas only 1% had an extrusion. However, in our study, distinguishing between protrusions and extrusions did not have diagnostic value. A limitation of the study by Jensen et al. was that it involved only asymptomatic volunteers.

The postoperative formation of epidural scars is a common phenomenon[32] and is hypothesized to cause mechanical traction on the dura or nerve roots, resulting in persistent back and leg pain after spinal surgery. Some studies have supported this hypothesis,[20, 33] whereas other studies have not shown a correlation between epidural-scar formation and clinical outcome.[34, 35] We did not find a positive correlation between the presence of scar tissue and symptoms. One of the strengths of our study is that the presence of scar tissue was examined by three observers. Our results show that clinicians should not automatically ascribe recurrent or persistent symptoms to visible scar formation on MRI.

An important limitation of our study is that the reported MRI findings and their relation with clinical outcome was only once, at 1 year after randomization. It is uncertain whether we would have found similar results at other time points. In addition, some observers might view the agreement among MRI readers as suboptimal. However, the kappa values are similar to those in previous studies,[12, 36, 37] and therefore one might consider them to reflect existing agreement among expert readers in clinical practice.

In summary, in patients who had undergone repeated MRI 1 year after treatment for symptomatic lumbar-disk herniation, anatomical abnormalities that were visible on MRI did not distinguish patients with persistent or recurrent symptoms of sciatica from asymptomatic patients. Further research is needed to assess the value of MRI in clinical decision making for patients with persistent or recurrent sciatica.

Supported by a grant from the Netherlands Organization for Health Research and Development and the Hoelen Foundation. Dr. Vleggeert-Lankamp reports receiving grant support from B. Braun Medical, Medtronic, and Paradigm Spine and lecture fees from B. Braun Medical. Dr. Peul reports receiving grant support from the Netherlands Organization for Health Research and Development, College voor zorgverzekeringen (CVZ, Dutch health care insurance board), Medtronic, B. Braun Aesculap, and Paradigm Spine, receiving lecture fees from Zorgverzekeraars Nederland (Dutch association of insurance companies) and CVZ, and providing legal expert testimony for medicolegal cases. No other potential conflict of interest relevant to this article was reported. Disclosure forms provided by the authors are available with the full text of this article at NEJM.org. We thank Ronald Brand for reviewing an earlier draft of the manuscript, and the patients for participating in the study.

REFERENCES

1. Stafford MA, Peng P, Hill DA. Sciatica: a review of history, epidemiology, pathogenesis, and the role of epidural steroid injection in management. Br J Anaesth 2007;99:461-73.
2. Vroomen PC, de Krom MC, Slofstra PD, Knottnerus JA. Conservative treatment of sciatica: a systematic review. J Spinal Disord 2000;13:463-9.
3. 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.
4. 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.
5. Vucetic N, Astrand P, Guntner P, Svensson O. Diagnosis and prognosis in lumbar disc herniation. Clin Orthop Relat Res 1999:116-22.
6. Weber H. Lumbar disc herniation. A controlled, prospective study with ten years of observation. Spine (Phila Pa 1976) 1983;8:131-40.
7. Jensen TS, Albert HB, Sorensen JS, Manniche C, Leboeuf-Yde C. Magnetic resonance imaging findings as predictors of clinical outcome in patients with sciatica receiving active conservative treatment. J Manipulative Physiol Ther 2007;30:98-108.
8. Legrand E, Bouvard B, Audran M, Fournier D, Valat JP. Sciatica from disk herniation: Medical treatment or surgery? Joint Bone Spine 2007;74:530-5.
9. Boos N, Rieder R, Schade V, Spratt KF, Semmer N, Aebi M. 1995 Volvo Award in clinical sciences. The diagnostic accuracy of magnetic resonance imaging, work perception, and psychosocial factors in identifying symptomatic disc herniations. Spine (Phila Pa 1976) 1995;20:2613-25.
10. Wassenaar M, van Rijn RM, van Tulder MW, et al. Magnetic resonance imaging for diagnosing lumbar spinal pathology in adult patients with low back pain or sciatica: a diagnostic systematic review. Eur Spine J 2012;21:220-7.
11. 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.
12. Jensen MC, Brant-Zawadzki MN, Obuchowski N, Modic MT, Malkasian D, Ross JS. Magnetic resonance imaging of the lumbar spine in people without back pain. N Engl J Med 1994;331:69-73.
13. Barth M, Diepers M, Weiss C, Thome C. Two-year outcome after lumbar microdiscectomy versus microscopic sequestrectomy: part 2: radiographic evaluation and correlation with clinical outcome. Spine (Phila Pa 1976) 2008;33:273-9.
14. 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.
15. Van Goethem JW, Van de Kelft E, Biltjes IG, et al. MRI after successful lumbar discectomy. Neuroradiology 1996;38 Suppl 1:S90-6.
16. 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.
17. Cheng F, You J, Rampersaud YR. Relationship between spinal magnetic resonance imaging findings and candidacy for spinal surgery. Can Fam Physician 2010;56:e323-30.
18. Lurie JD, Birkmeyer NJ, Weinstein JN. Rates of advanced spinal imaging and spine surgery. Spine (Phila Pa 1976) 2003;28:616-20.
19. 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.

20. 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.
21. Fardon DF, Milette PC. Nomenclature and classification of lumbar disc pathology. Recommendations of the Combined task Forces of the North American Spine Society, American Society of Spine Radiology, and American Society of Neuroradiology. Spine (Phila Pa 1976) 2001;26:E93-E113.
22. 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.
23. Collins SL, Moore RA, McQuay HJ. The visual analogue pain intensity scale: what is moderate pain in millimetres? Pain 1997;72:95-7.
24. Kuijer W, Brouwer S, Dijkstra PU, Jorritsma W, Groothoff JW, Geertzen JH. Responsiveness of the Roland-Morris Disability Questionnaire: consequences of using different external criteria. Clin Rehabil 2005;19:488-95.
25. 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.
26. 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 (Phila Pa 1976) 2006;31:2602-8.
27. Farrar JT, Young JP, Jr., LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001;94:149-58.
28. Ostelo RW, Deyo RA, Stratford P, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine (Phila Pa 1976) 2008;33:90-4.
29. Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. Spine (Phila Pa 1976) 1983;8:141-4.
30. Obuchowski NA. Receiver operating characteristic curves and their use in radiology. Radiology 2003;229:3-8.
31. Komori H, Shinomiya K, Nakai O, Yamaura I, Takeda S, Furuya K. The natural history of herniated nucleus pulposus with radiculopathy. Spine (Phila Pa 1976) 1996;21:225-9.
32. Ronnberg K, Lind B, Zoega B, et al. Peridural scar and its relation to clinical outcome: a randomised study on surgically treated lumbar disc herniation patients. Eur Spine J 2008;17:1714-20.
33. Maroon JC, Abla A, Bost J. Association between peridural scar and persistent low back pain after lumbar discectomy. Neurol Res 1999;21 Suppl 1:S43-6.
34. Nygaard OP, Kloster R, Dullerud R, Jacobsen EA, Mellgren SI. No association between peridural scar and outcome after lumbar microdiscectomy. Acta Neurochir (Wien) 1997;139:1095-100.
35. Vogelsang JP, Finkenstaedt M, Vogelsang M, Markakis E. Recurrent pain after lumbar discectomy: the diagnostic value of peridural scar on MRI. Eur Spine J 1999;8:475-9.
36. Brant-Zawadzki MN, Jensen MC, Obuchowski N, Ross JS, Modic MT. Interobserver and intraobserver variability in interpretation of lumbar disc abnormalities. A comparison of two nomenclatures. Spine (Phila Pa 1976) 1995;20:1257-63; discussion 64.
37. Jarvik JG, Haynor DR, Koepsell TD, Bronstein A, Ashley D, Deyo RA. Interreader reliability for a new classification of lumbar disk disease. Acad Radiol 1996;3:537-44.

Supplementary Appendix belonging to the article "Magnetic Resonance Imaging in follow-up assessment of sciatica"

Authors and participants in the Leiden–The Hague Spine Intervention Prognostic Study Group and their role in the currently presented study were as follows:

Conception and design of the clinical trial: B.W. Koes, W.B. van den Hout, R.T.W.M. Thomeer, and W.C. Peul
Conception and design of the diagnostic part (with regard to the value of MRI) of the clinical trial: A. el Barzouhi, W.B. van den Hout, and W.C. Peul
Data Collection and Management was done by the following research nurses: M. Nuyten, P. Bergman, G. Holtkamp, S. Dukker, A. Mast, L. Smakman, C. Waanders, L. Polak, and A. Nieborg
MRI blinded reading: C.L.A.M. Vleggeert-Lankamp, G.J. Lycklama à Nijeholt, B.F. Van der Kallen
Statistical Analysis of the data: A. el Barzouhi and W.B. van den Hout
Statistical assistance and review of the statistical analysis: R. Brand
Interpretation of the data: A. el Barzouhi, C.L.A.M. Vleggeert-Lankamp, G.J. Lycklama à Nijeholt, B.F. Van der Kallen, WB van den Hout, W.C.H Jacobs, B.W. Koes, and W.C. Peul *Vouch for the data and analysis*: A. el Barzouhi, C.L.A.M. Vleggeert-Lankamp, G.J. Lycklama à Nijeholt, B.F. Van der Kallen, WB van den Hout, W.C.H Jacobs, B.W. Koes, and W.C. Peul
Manuscript Preparation (first draft): A. el Barzouhi
Critical revision for important intellectual content: C.L.A.M. Vleggeert-Lankamp, G.J. Lycklama à Nijeholt, B.F. Van der Kallen, WB van den Hout, W.C.H. Jacobs, B.W. Koes, and W.C. Peul
Decision to publish the paper: A. el Barzouhi, C.L.A.M. Vleggeert-Lankamp, G.J. Lycklama à Nijeholt, B.F. Van der Kallen, WB van den Hout, W.C.H Jacobs, B.W. Koes, and W.C. Peul
Sponsors of the study: a grant from the Netherlands Organisation for Health Research and Development (ZonMW) and the Hoelen Foundation, The Hague.
Role of the sponsor: there were no agreements concerning confidentiality of the data between the sponsor and the authors or the institutions named in the article or in this supplement. The sponsors did also not have any role in the writing or analysis of this study.
Protocol Committee: W.C. Peul, B.W. Koes, and R.T.W.M. Thomeer
Steering Committee: B.W. Koes, R.T.W.M. Thomeer, J.A.H. Eekhof, J.T.J. Tans, W.B. van den Hout, W.C. Peul, R. Brand, and H.C. van Houwelingen
Participating Hospitals and Coordinating Physicians: Medical Center Haaglanden, The Hague — J.T.J. Tans and R. Walchenbach; Diaconessen Hospital, Leiden — J. van Rossum, P. Schutte, and R.T.W.M. Thomeer; Groene Hart Hospital, Gouda — G.A.M. Verheul, J.E. Dalman, and J.A.L. Wurzer; Reinier de Graaf Hospital, Delft/Voorburg — J.W.A. Sven and A. Kloet; Spaarne Hospital, Heemstede/Haarlem — I.S.J. Merkies and H. van Dulken; Bronovo Hospital, The Hague — P.C.L.A. Lambrechts and J.A.L. Wurzer; Haga Hospital, The Hague

— R.W.M. Keunen and C.F.E. Hoffmann; Rijnland Hospital, Leiderdorp/Alphen ad Rijn — J. Haan and H. van Dulken; Lange Land Hospital, Zoetermeer — R. Groen and R.R.F. Kuiters; Leiden University Medical Center, Leiden — R.A.C. Roos and J.H.C. Voormolen; Public Health and Primary Care, Leiden University, Leiden — J.A.H. Eekhof.

No writing assistance was provided.

Methods sensitivity analysis

1. ANALYSES TO ACCOUNT FOR MISSING CLINICAL DATA

Depending on the clinical outcome, data at one year was missing in 2 to 3% of the included cases (see Table S3 of this Appendix). In the main analysis we used model-based multiple imputation to account for missing clinical outcome data at one year (using the variables mentioned in Table 1 and 2 of the manuscript to predict the missing values).

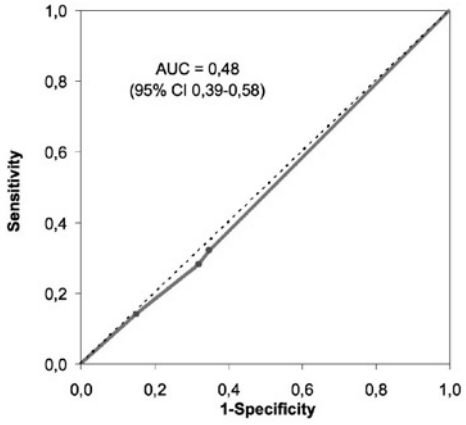
As sensitivity analyses to account for these missing data, we performed analysis as observed (e.g., no imputation, thus depending on the clinical outcome 6 or 7 patients with missing data were excluded from the analysis) and analysis using the last-observation-carried-forward method (depending on the clinical outcome the last observation was carried forward for 6 or 7 patients. These last observations were derived from the period 8-52 weeks after randomization).

All sensitivity analyses performed to account for missing data yielded similar results as the analyses presented in the manuscript. In this Appendix we include some examples of the sensitivity analyses by presenting the ROC curves. Figure S2A and S2B of this Appendix show the ROC curves of one-year MRI findings when the last-observation-carried-forward method was used. Figure S3A and S3B of this Appendix show the ROC curves of one-year MRI findings when the cases with no reported clinical data at one year were excluded.

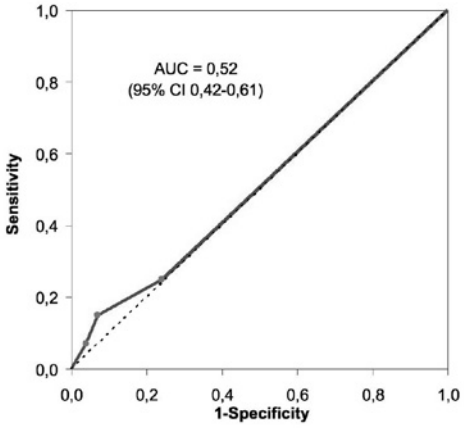
2. Analyses to account for interobserver agreement

In the main statistical analysis, as presented in the manuscript, we used the majority opinion of the three readers regarding the MRI characteristics (answer independently given by minimum 2 out of 3 readers). As sensitivity analyses we reproduced all analyses using only the cases in whom all 3 readers independently agreed regarding the presence of an MRI characteristic. All analyses yielded similar results. In this Appendix we include some examples of the sensitivity analyses by presenting the area under the ROC curve for MRI assessed disc herniation and nerve root compression and the clinical outcomes stratified by the MRI findings at one year.

Figure S1 Receiver operating characteristic (ROC) curve of one-year MRI findings. The curves show the ability of MRI variables to differentiate between patients with favorable outcome (defined as complete or nearly complete disappearance of symptoms on the 7-point Likert scale, n=43) and patients with unfavorable outcome at one year (n=224). The dotted line is a reference line with an area under the curve of 0.5, indicating no discriminatory value.
A) ROC curve of the MRI assessed presence of a herniated disc at one year.
B) ROC curve of the MRI assessed nerve root compression at one year.
The points in the curves indicate the actual results (sensitivity and 1-specificity) associated with different MRI interpretations. For both the presence of a herniated disc and root compression an ordinal four point scale was used, ranging from 1 (definitely present) to 4 (definitely absent). AUC denotes area under the curve. CI denotes confidence interval.

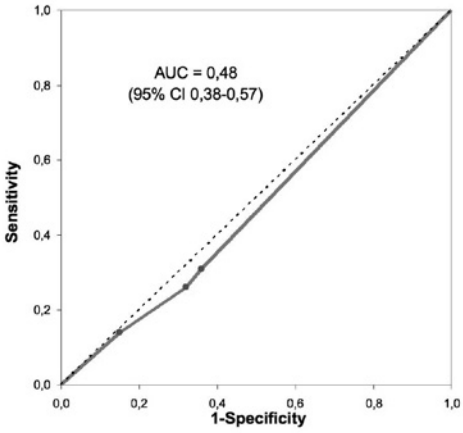


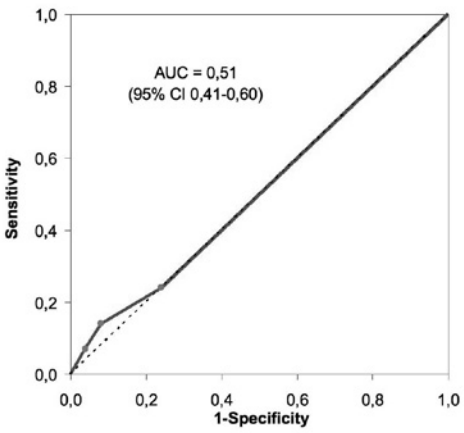
S1A



S1B

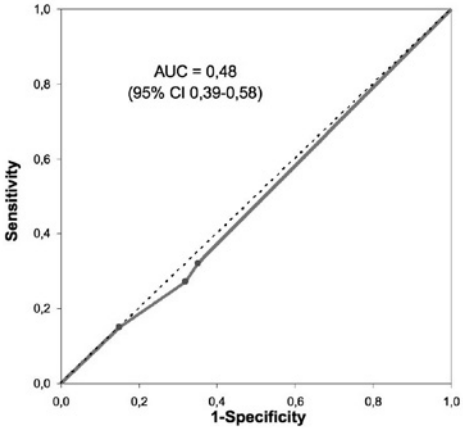
Figure S2 Receiver operating characteristic (ROC) curve of one-year MRI findings. The curves show the ability of MRI variables to differentiate between patients with favorable outcome (defined as complete or nearly complete disappearance of symptoms on the 7-point Likert scale, n=42) and patients with unfavorable outcome at one year (n=225). The dotted line is a reference line with an area under the curve of 0.5, indicating no discriminatory value. Seven (3%) patients had missing clinical outcome data at one year. *The last-observation-carried-forward method was used to account for this missing data.*
A) ROC curve of the MRI assessed presence of a herniated disc at one year.
B) ROC curve of the MRI assessed nerve root compression at one year.
The points in the curves indicate the actual results (sensitivity and 1-specificity) associated with different MRI interpretations. For both the presence of a herniated disc and root compression an ordinal four point scale was used, ranging from 1 (definitely present) to 4 (definitely absent). AUC denotes area under the curve. CI denotes confidence interval.

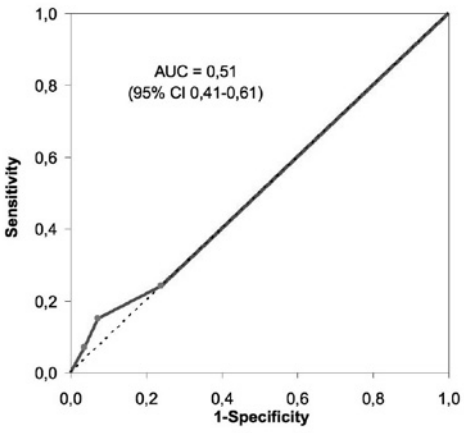




S2A **S2B**

Figure S3 Receiver operating characteristic (ROC) curve of one-year MRI findings. The curves show the ability of MRI variables to differentiate between patients with favorable outcome (defined as complete or nearly complete disappearance of symptoms on the 7-point Likert scale, n=41) and patients with unfavorable outcome at one year (n=219). *Seven (3%) patients had missing clinical outcome data at one year. These seven patients were excluded from the analysis (so n=260 instead of n=267)*
A) ROC curve of the MRI assessed presence of a herniated disc at one year.
B) ROC curve of the MRI assessed nerve root compression at one year.
The points in the curves indicate the actual results (sensitivity and 1-specificity) associated with different MRI interpretations. For both the presence of a herniated disc and root compression an ordinal four point scale was used, ranging from 1 (definitely present) to 4 (definitely absent). AUC denotes area under the curve. CI denotes confidence interval.





S3A **S3B**

EXAMPLE SENSITIVITY ANALYSES TO ACCOUNT FOR INTEROBSERVER AGREEMENT

- Using MRI assessed presence of disc herniation to discriminate between subjects with favorable outcome versus subjects with unfavorable outcome revealed an AUC of 0.48 (95% CI 0.39-0.58) when all participants (n=267) were included.
- *Using MRI assessed presence of disc herniation to discriminate between subjects with favorable outcome versus subjects with unfavorable outcome revealed an AUC of 0.48 (95% CI 0.38-0.59) when only participants with full agreement [n=200]) were included.*
- Using MRI assessed nerve root compression to discriminate between subjects with favorable outcome versus subjects with unfavorable outcome revealed an AUC of 0.52 (95% CI 0.42-0.61) when all participant (n=267) were included.
- *Using MRI assessed nerve root compression to discriminate between subjects with favorable outcome versus subjects with unfavorable outcome revealed an AUC of 0.49 (95% CI 0.37-0.61) when only participants with full agreement [n=200]) were included.*

In this Table only cases were included in whom all three readers independently agreed regarding the presence of the MRI characteristic

	Presence of a herniated disc on MRI at one year (total n=200)		P Value§	Presence of root compression on MRI at one year (total n=189)		P Value§
	Yes (n=62)	No (n=138)		Yes (n=46)	No (n=143)	
Clinical outcome at one year						
Favorable clinical outcome	53 (86)	115 (83)	0.84	40 (87)	121 (85)	0.81
Roland Disability Questionnaire	3.7±5.1	3.7±5.8	0.59	4.0±5.6	3.4±5.5	0.32
VAS-Leg pain	9.5±18.4	11.6±19.5	0.93	11.8±23.3	11.5±19.7	0.84
VAS-back pain	12.4±16.9	15.2±20.9	0.72	13.7±20.1	15.6±22.5	0.70

ò Favorable clinical outcome was defined as complete or nearly complete disappearance of symptoms according to the Likert-7 point scale.
‡ 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 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 how the size of the disc herniation was related to the baseline size.

MRI variable	Type	Categories
Disc level that most likely caused the lumbosacral radicular syndrome at baseline	Disc level	1. L2L3 2. L3L4 3. L4L5 4. L5S1
	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.
	Size disc herniation in relation to baseline size	1. Not applicable, herniation completely disappeared 2. Disc herniation reduced in size 3. No size reduction of disc herniation 4. Herniation increased in size
If a herniation at the disc level is considered	Side of this disc herniation	1. Right 2. Left 3. Right and left
	Location on axial view of this disc herniation	1. Central zone: zone within the vertebral canal between sagittal planes through the medial edges of each facet 2. Sub-articular zone: zone, within the vertebral canal, sagittally between the plane of the medial edges of the pedicles and the plane of the medial edges of the facets, and coronally between the planes of the posterior surfaces of the vertebral bodies and the under anterior surfaces of the superior facets. 3. Foraminal zone: zone between planes passing through the medial and lateral edges of the pedicles 4. Extra-foraminal zone: the zone beyond the sagittal plane of the lateral edges of the pedicles, having no well-defined lateral border
	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

Table S1 (Continued)

MRI variable	Type	Categories
	Form disc herniation	1. Protrusion: localized displacement of disc material beyond the intervertebral disc space, with the base against the disc of origin broader than any other dimension of the protrusion. 2. Extrusion: localized displacement of disc material beyond the intervertebral disc space, with the base against the disc of origin narrower than any one distance between the edges of the disc material beyond the disc space measured in the same plane, or when no continuity exists between the disc material beyond the disc space and that within the disc space.
Nerve root compression	Probability of nerve root compression	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 no clinical relevant nerve root compression
	If nerve root compression present, which nerve root is affected	1. L3 2. L4 3. L5 4. S1
	Side nerve root compression	1. Right 2. Left
Scar tissue	Presence	1. Yes: scar tissue present 2. No: scar tissue absent
	If present, place scar tissue	1. Scar tissue surrounds the nerve root 2. Scar tissue does not surround the nerve root

Table S2 One year after randomization a second MRI was available for 267 (94.3%) out of 283 participants. Reasons for why no second MRI at one year was available for the remaining 16 patients are listed in the Table.

Number of patients (total n=16)	Reason why no second MRI was available one year after randomization
3	Stopped participating in the study after 8 weeks
1	Stopped participating in the study after 12 weeks
1	Stopped participating in the study after 16 weeks
1	Stopped participating in the study after 26 weeks
1	Did not show up on the scheduled appointment
1	Pregnancy
5	A second MRI was actually performed at 52 weeks, but we were not able to retrieve these MRIs. These 5 MRI's might have been lost during the storage process at the centers were the MRI's were performed or during the collection of the MRI's
3	Reason unknown

Table S3 Outcome measurements available at 52 weeks after randomization. The mentioned outcome measures were assessed at baseline, 2, 4, 8, 12, 26, 38, and 52 weeks. Values are n (%).

	Number of patients with available clinical outcome at 52 weeks Total n=267
Outcome	
Global perceived recovery on a 7-point Likert scale at 52 weeks	260 (97)
Roland disability questionnaire at 52 weeks‡	261 (98)
Visual Analogue scale for leg pain at 52 weeks¶	261 (98)
Visual Analogue scale for back pain at 52 weeks¶	260 (97)

‡ 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 S4 Interobserver agreement regarding characteristics of the lumbar vertebral disc level at one year. Reader A en B represent the two neuroradiologists, while reader C represents the neurosurgeon. Significant kappa values are in bold, which means that the kappa values significantly differed from value zero. Guidelines proposed by Landis and Koch were used for interpretation. Values of less than 0.00 indicated poor reliability; 0.00 to 0.20, slight reliability; 0.21 to 0.40, fair reliability; 0.41 to 0.60, moderate reliability; 0.61 to 0.80, substantial agreement; and 0.81 to 1.00, excellent or almost perfect agreement. Kappa values and percentages of agreement for the *place of scar tissue and characteristics of disc herniation* were only calculated if the observers agreed about their presence (e.g. when a reading pair showed disagreement about the presence of disc herniation, this patient did not contribute to the interagreement analysis regarding the characteristics of the herniated disc).

	A vs B		A vs C		B vs C		All observers	
MRI characteristics at one year	% agree-ment	kappa	% agree-ment	kappa	% agree-ment	kappa	% agree-ment	multi-rater kappa
Probability presence of disc herniation (4 categories)‡	77.6	**0.64**	74.5	**0.67**	79.3	**0.67**	69.0	**0.57**
Probability presence of disc herniation (2 categories) ¶	81.9	**0.61**	87.5	**0.74**	85.4	**0.67**	77.5	**0.67**
Probability presence of nerve root compression (4 categories)‡	68.7	**0.52**	68.8	**0.53**	88.1	**0.74**	64.7	**0.46**
Probability presence of nerve root compression (2 categories) ¶	76.1	**0.48**	88.0	**0.53**	92.0	**0.76**	73.3	**0.57**
Presence of scar tissueò	88.7	**0.77**	73.6	**0.50**	76.1	**0.53**	69.1	**0.59**
Place of scar tissue*	97.8	**0.66**	95.1	**0.43**	100.0	**1.00**	96.8	**0.49**
Characteristics of the disc herniation								
Size in relation to baseline size ‖	79.8	**0.67**	84.9	**0.74**	84.2	**0.69**	71.7	**0.65**
Location ⊨	87.3	**0.79**	81.8	**0.72**	91.9	**0.86**	85.5	**0.82**

Table S4 (Continued)

	A vs B		A vs C		B vs C		All observers	
MRI characteristics at one year	% agree-ment	kappa	% agree-ment	kappa	% agree-ment	kappa	% agree-ment	multi-rater kappa
Size disc herniation in relation to spinal canal§	68.3	**0.56**	72.7	**0.55**	82.3	**0.70**	61.3	**0.51**
Herniation form**	75.8	**0.51**	86.2	**0.65**	83.6	**0.67**	73.8	**0.62**

‡ The four categories were: 1) "Definite about the presence" if there was no doubt about the presence 2) "Probable about the presence" if there was some doubt but the probability was greater than 50% 3) "Possible about the presence" if there was reason to consider but the probability was less than 50%, and 4) "Definite about the absence" if there was no doubt about the absence (Table 1 Supplementary appendix).

¶ 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).

ò The categories were: 1) "Yes" or 2) "No" (Table 1 Supplementary appendix).

* The categories were: 1) "Scar tissue surrounds the nerve root" or 2) "Scar tissue does not surround the nerve root" (Table 1 Supplementary appendix).

‖ The categories were: 1) "Disc herniation completely disappeared" 2) "Disc herniation reduced in size" 3) "No size reduction of disc herniation" and 4) "Herniation increased in size" (Table 1 Supplementary appendix).

╞ The categories were: 1) "Central zone" 2) "Sub-articular zone" 3) "Foraminal zone" and 4) "Extra-foraminal zone" (Table 1 Supplementary appendix).

§ The categories were: 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" and 4) "Small: size <25% of the spinal canal" (Table 1 Supplementary appendix).

** The categories were: 1) "Protrusion" and 2) "Extrusion" (Table 1 Supplementary appendix).

Table S5 Differences in 1-year MRI findings and clinical outcome between patients who were randomized to early surgery and those who were randomized to prolonged conservative care (intention-to-treat). Values are n (%). Total n=267

	Early surgery (n=131)	Prolonged conservative care (n=136)	P Value
Clinial outcome at one year			
Favorable clinical outcomeò	111 (85)	113 (83)	0.65
Roland Disability‡	3.4±5.8	3.5±5.1	0.84
VAS-Leg pain¶	11.3±20.8	10.6±18.6	0.77
VAS-back pain¶	14.9±22.5	15.6±22.1	0.82
MRI findings			
Disc contour one year after randomization			
Normal	79 (60)	50 (37)	**<0.001**
Bulging	23 (18)	22 (16)	
Definite (100%) herniation	12 (9)	28 (21)	
Probable (some doubt but probability > 50%) herniation	16 (12)	28 (21)	

Table S5 (Continued)

	Early surgery (n=131)	Prolonged conservative care (n=136)	P Value
Possible (reason to consider but probability < 50%) herniation	1 (1)	8 (6)	
Disc herniation one year after randomization compared to baseline			
Disappeared	100 (76)	71 (52)	**<0.001**
Reduced in size	22 (17)	53 (39)	
Unchanged or enlarged in size	6 (5)	9 (7)	
Not applicable, no disc herniation at baseline	3 (2)	3 (2)	
Nerve root compression one year after randomization			
Definitely no root compression	109 (83)	93 (68)	**0.021**
Possible: reason to consider but probability < 50%	16 (12)	26 (19)	
Probable: some doubt but probability > 50%	2 (2)	10 (7)	
Definite: no doubt about the presence	4 (3)	7 (5)	
Nerve root compression one year after randomization compared to baseline			
Disappeared	106 (81)	92 (68)	**0.038**
Reduced	15 (11)	30 (22)	
Unchanged or increased	10 (8)	14 (10)	
Clinical outcome			

ò Favorable clinical outcome was defined as complete or nearly complete disappearance of symptoms according to the Likert-7 point scale.

‡ 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 S6 MRI differences stratified according to clinical outcome at one year.

A) MRI differences between patients with a Visual Analogue Scale (VAS) for *leg pain* of at least 40mm and patients with VAS for leg pain less than 40mm. This cutt-off value is often used when an absolute VAS score (with 0 representing no pain and 100 the worst pain ever experienced) is categorized into favorable and unfavorable outcome.[1, 2] Values are n (%).

B) MRI differences between patients with less than 30% improvement and patients with at least 30% improvement in *Vas-leg pain* between baseline and one year, since a 30% improvement has been proposed to be a clinically meaningful improvement when comparing before and after measures of pain and functional status for individual patients.[3-5] Total N=266 instead of 267 as one patients had at baseline a VAS-leg of 0. Values are n (%).

C) MRI differences between patients with a VAS for *back pain* of at least 40mm and patients with VAS for back pain less than 40mm.[1, 2] Values are n (%).

D) MRI differences between patients with less than 30% improvement and patients with at least 30% improvement in *Vas-back pain* between baseline and one year.[3-5] Total N=232 as 35 patients had at baseline a VAS-back of 0. Values are n (%).
E) MRI differences between patients with a *Roland disability questionnaire (RDQ)* score of least 14 and patients with an RDQ less than 14. This cut-off value is often used when the RDQ is dichotomized into favorable and unfavorable outcome.[6,7] Values are n (%).
F) MRI differences between patients with less than 30% improvement and patients with at least 30% improvement in *RDQ* between baseline and one year.[3-5] Values are n (%).

S6A

	VAS-leg pain ≥40 at one year (n=24)	VAS-leg pain <40 at one year (n=243)	P Value
MRI findings at one year			
Presence of disc herniation at one year	10 (42)	83 (34)	0.43
MRI assessed presence of nerve root compression	6 (25)	59 (24)	0.87

S6B

	<30% improvement in VAS-leg pain between baseline and one year (n=23)	≥30% improvement in VAS-leg pain between baseline and one year (n=243)	P Value
MRI findings at one year			
Presence of disc herniation at one year	9 (39)	84 (35)	0.63
MRI assessed presence of nerve root compression	5 (22)	60 (25)	0.84

S6C

	VAS-back pain ≥40 at one year (n=34)	VAS-back pain <40 at one year (n=233)	P Value
MRI findings at one year			
Presence of disc herniation at one year	11 (32)	82 (35)	0.92
MRI assessed presence of nerve root compression	5 (15)	60 (26)	0.25

S6D

	<30% improvement in VAS-back pain between baseline and one year (n=66)	≥30% improvement in VAS-back pain between baseline and one year (n=166)	P Value
MRI findings at one year			
Presence of disc herniation at one year	23 (35)	58 (35)	0.93
MRI assessed presence of nerve root compression	15 (23)	41 (25)	0.85

S6E

	RDQ≥14 at one year (n=22)	RDQ< 14 at one year (n=245)	P Value
MRI findings at one year			
Presence of disc herniation at one year	6 (27)	87 (36)	0.49
MRI assessed presence of nerve root compression	5 (23)	60 (24)	0.92

S6F

	<30% improvement in RDQ between baseline and one year (n=29)	≥30% improvement in RDQ between baseline and one year (n=238)	P Value
MRI findings at one year			
Presence of disc herniation at one year	9 (31)	84 (35)	0.69
MRI assessed presence of nerve root compression	7 (24)	58 (24)	0.98

Table S7 Accuracy measures of one-year MRI findings for favorable outcome at one year.
Favorable outcome was defined as complete or nearly complete disappearance of symptoms according to the Likert-7 point scale. Total n=267

	Sensitivity¶ (95% CI)	Specificity§ (95% CI)	Positive predictive value ╞ (95% CI)	Negative predictive value‡ (95% CI)
MRI assessed presence of disc herniation at one year				
Definite (no doubt about the presence)	0.14 (0.04-0.24)	0.85 (0.80-0.90)	0.15 (0.04-0.26)	0.84 (0.79-0.88)
Definite or probable (Probability >50%)	0.28 (0.14-0.41)	0.68 (0.62-0.74)	0.14 (0.07-0.22)	0.83 (0.77-0.88)
Definite, probable or possible (Probability >0%)	0.32 (0.18-0.46)	0.65 (0.58-0.71)	0.15 (0.08-0.22)	0.83 (0.78-0.89)
Characteristic of the herniated disc				
Size >25% in relation to spinal canal	0.40 (0.14-0.66)	0.70 (0.60-0.81)	0.19 (0.04-0.34)	0.87 (0.78-0.95)
Extrusion instead of protrusion	0.29 (0.05-0.53)	0.69 (0.59-0.79)	0.14 (0.01-0.27)	0.84 (0.75-0.94)
MRI assessed presence of nerve root compression at one year				
Definite (no doubt about the presence)	0.07 (0.00-0.14)	0.96 (0.94-0.99)	0.27 (0.01-0.54)	0.84 (0.80-0.89)
Definite or probable (Probability >50%)	0.15 (0.04-0.26)	0.93 (0.89-0.96)	0.29 (0.10-0.48)	0.85 (0.80-0.89)

Table S7 (Continued)

	Sensitivity¶ (95% CI)	Specificity§ (95% CI)	Positive predictive value ╞ (95% CI)	Negative predictive value‡ (95% CI)
Definite, probable or possible (Probability >0%)	0.25 (0.12-0.38)	0.76 (0.70-0.81)	0.17 (0.08-0.26)	0.84 (0.79-0.89)

¶ Sensitivity indicates the proportion of patients with unfavorable outcome who had an abnormal test finding.
§ Specificity indicates the proportion of patients with favorable outcome with no abnormal test finding.
╞ Positive predictive value indicates the proportion of patients with an abnormal test finding who did report unfavorable outcome.
‡ Negative predictive value indicates the proportion of patients with no abnormal test finding who did report favorable outcome.

Table S8 Uni- and multivariate analysis of the characteristics of the disc herniation at one year to determine predictive value on favorable outcome at one year.

Comparison (%)	Univariate analysis			Adjusted for randomized treatment ¶			Adjusted for received treatment ‡			Multivariate adjustment ╞		
	OR	95% CI	P-value	OR	95% CI	P-value	OR	95% CI	P-value	OR	95% CI	P-value
Size disc herniation <25% in relation to spinal canal (69) vs. size 25-75% in relation to spinal canal (31)	1.58	0.5-5.3	0.46	1.48	0.4-5.0	0.53	1.54	0.5-5.2	0.49	0.74	0.1-5.8	0.77
Protrusion (70) vs. extrusion (30)	0.90	0.3-3.2	0.87	0.88	0.2-3.2	0.85	0.88	0.2-3.2	0.85	1.96	0.3-13.6	0.50

Favorable outcome was defined as complete or nearly complete disappearance of symptoms according to the 7-point Likert scale for global perceived recovery. OR denotes odds ratio. CI denotes confidence interval. Total n=93
¶ An early surgery strategy vs. 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.
‡ Analysis adjusted for actual received treatment (surgery vs. no surgery during the first year).
╞ Analysis adjusted for randomized treatment, age, gender, body-mass index, smoking, Roland Disability Questionnaire score at baseline, Visual Analogue scale for leg and back pain at baseline and presence of disturbed neurological tests (six neurological tests were performed [Lasègue's sign, Crossed straight-leg raising, Kemp's sign, Bragard's Sign, walking on heels and walking on toes]. One or more disturbed tests was considered to be an abnormal result).

REFERENCES

1. 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.
2. 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 (Phila Pa 1976) 2006;31:2602-8.
3. Farrar JT, Young JP, Jr., LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001;94:149-58.
4. Jordan K, Dunn KM, Lewis M, Croft P. A minimal clinically important difference was derived for the Roland-Morris Disability Questionnaire for low back pain. J Clin Epidemiol 2006;59:45-52.
5. Ostelo RW, Deyo RA, Stratford P, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine (Phila Pa 1976) 2008;33:90-4.
6. Ostelo RW, Vlaeyen JW, van den Brandt PA, de Vet HC. Residual complaints following lumbar disc surgery: prognostic indicators of outcome. Pain 2005;114:177-85.
7. Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. Spine (Phila Pa 1976) 1983;8:141-4.