Chapter 3

Prognostic value of interpretation consistency of MR Imaging in sciatica

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

Submitted for publication

ABSTRACT

PURPOSE

To evaluate the clinical outcome results of patients with sciatica according to consistency in Magnetic resonance imaging (MRI) interpretation among spine specialists.

METHODS

Patients for this study were participants who underwent a baseline MRI to assess the eligibility for a randomized trial designed to compare the efficacy of early surgery with prolonged conservative care for patients with sciatica. Two neuroradiologists and one neurosurgeon independently evaluated all MRIs. (In)consistent MRI interpretation was correlated with patient's report of perceived recovery, Roland Disability Questionnaire (RDQ) and visual-analogue scale (VAS) for leg and back pain at one year.

RESULTS

Of the 389 patients the three MRI observers agreed in 296 (76%) patients about the presence of a disc herniation, disagreed in 48 (12%) patients about its presence and agreed in 45 (12%) patients about its absence. Of the patients with a (consistent) disc herniation on MRI 84% reported perceived recovery after one year compared to 75% of the patients with inconsistent interpretation and 58% of the patients in whom all three readers agreed about the absence of a herniated disc ($P<0.001$). The same pattern was observed with the RDQ score ($P=0.007$), VAS leg pain ($P=0.06$) and VAS back pain ($P=0.001$). Patients with a (consistent) disc herniation had the highest speed of perceived recovery, followed by patients with inconsistent interpretation and those with absence of disc herniation ($P=0.006$).

CONCLUSION

At one year follow-up the most favorable clinical outcome results were reported by those patients in whom all three MRI observers independently agreed about the presence of disc herniation, followed by those with inconsistent interpretation and by those with absence of those findings at baseline.

INTRODUCTION

Sciatica is one of the most common lumbar-spine disorders. It is characterized by radiating pain in an area of the leg typically served by one nerve root in the lumbar or sacral spine and is occasionally acompanied by neurological deficit.[1] The most common cause of sciatica is a herniated disc.[1] The prevalence of sciatic symptoms varies considerably ranging from 1.6% in the general population to 43% in a selected working population.[2] Sciatica results in severe pain and disability for the individual patient and significant costs in terms of treatment, sick leave, and pensions for society.[3,4]

Magnetic resonance imaging (MRI) is considered the imaging procedure of choice for patients suspected of lumbar herniated discs [5,6] and is indicated in patients with severe symptoms who fail to respond to conservative care for at least 6 to 8 weeks.[1] If a herniated disc with nerve root compression is indeed present surgery as a treatment modality might be considered. About 20 to 30% of the patients with sciatica finally receives surgery.[7] Depending upon the used outcome measure, the results of lumbar disc surgery are unsatisfactory in 15 to 40% of the patients.[8-10] It has been suggested that the poor outcomes following lumbar disc surgery may be more often due to the errors in diagnosis than failure of the surgical intervention or its complications.[11,12] For example, a false-positive diagnosis of a herniated disc with nerve root compression on MRI may lead to unwarranted surgery and subsequently a poor clinical outcome. Therefore, we hypothesized that patients in whom spine specialists independently agree about the presence of a disc herniation might fare better than those with inconsistent interpretation or those in whom spine specialists independently agree about the absence of a disc herniation.

The researchers previously reported the results of a randomized controlled trial comparing early surgery with prolonged conservative care for patients with sciatica.[13] The trial showed faster recovery after early surgery compared to a strategy of prolonged conservative care with surgery if needed, but without any differences in the clinical outcomes after one year. The randomized patients were part of a larger group of patients with sciatica who underwent MRI and were followed up for one year. We now report on the clinical outcome of patients with sciatica in whom spine specialists independently agreed about the presence of a disc herniation or nerve root compression, those with inconsistent MRI interpretation, and those in whom spine specialists independently agreed about the absence of those findings.

METHODS

STUDY POPULATION

Patients for this study were participants who underwent an MRI to assess the eligibility for a multicenter, prospective, randomized controlled trial designed to determine whether early

surgery results in a more effective outcome compared to a strategy of prolonged conservative treatment with surgery if needed among patients with 6 to 12 weeks sciatica.[13] Patients who had symptoms being so severe that they were eligible for surgery according to their general practitioners were referred to the neurologist who subsequently evaluated whether these patients were eligible to participate in the trial. 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 or severe coexisting disease. Patients who were not between 18 to 65 years of age were also excluded. All participants who were not meeting one or more of the aforementioned exclusion criteria and had a lumbosacral radicular syndrome lasting between 6 to 12 weeks underwent MRI. MRIs of all patients, regardless of participation in the randomized clinical trial, were again evaluated by independent observers. 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 of the randomized controlled trial have been published previously.[14] In the present study, however, the data were analyzed as a cohort study.

MRI PROTOCOL AND EVALUATION

MRI scans were performed in all 9 participating hospitals 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 contrastenhanced (gadolinium) fat-suppressed T1-weighted images.

Two neuroradiologists (BK and GL) and one neurosurgeon (CV) independently evaluated all MR images, blinded to clinical information. None of the readers had been involved in either the selection or care of the included patients. Observer experience in reading spine MRI's was 7 and 6 years post-residency for the neuroradiologists and 4 years post-residency for the neurosurgeon. The observers hold senior positions in busy spinal clinics with a focus on advanced spine surgery, and are confronted with spinal MRIs on a daily basis.

For both the presence of disc herniation and nerve root compression a four point scale was used: 1 for definite presence, 2 for probable presence if there was some doubt but the probability was greater than 50%, 3 for possible presence if there was reason to consider but the probability was less than 50%, and 4 for definite absence. For each MRI observer the evaluations on the 4 point scale were dichotomized: the first two categories were combined and marked as herniated disc or nerve root compression present, the last two categories were combined and marked as absence of the abnormalities. Readings between the MRI observers were considered inconsistent when one of the three MRI observers had a different evaluation based on the dichotomized (made) scale.

OUTCOMES

The patients were assessed by means of the Roland Disability Questionnaire for Sciatica (RDQ, scores range from 0 to 23, with higher scores indicating worse functional status),[15] the 100-mm visual-analogue scale (VAS) for leg and back pain (with 0 representing no pain and 100 the worst pain ever experienced),[16] 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 was defined as "complete" or "nearly complete disappearance of symptoms" on the patient-reported 7-point Likert scale for global perceived recovery, while a score in the remaining five categories was marked as "no recovery".[13] Outcome measures were assessed at baseline, 2, 4, 8, 12, 26, 38 and 52 weeks. Patients were blinded to results of earlier assessments.

STATISTICAL ANALYSIS

The total study population was divided into three groups: a group with consistent MRI interpretation regarding the presence (i.e. all three readers independently agreed about the presence of a disc herniation or nerve root compression), a group with inconsistent MRI interpretation (i.e. one reader disagreed with the other two), and a group with consistent MRI interpretation regarding the absence. Differences between the three groups in clinical outcome at one year were assessed by using one-way analysis of variance (ANOVA) for continuous data and Chi-square test for categorical data. Time from baseline until perceived recovery (as determined by the prescheduled moments of outcome registration during the first year) was compared between the three groups by use of Kaplan-Meier curves and analyzed by Cox proportional-hazards models.

We assumed clinical outcome data to be missing at random and used model-based multiple imputation to impute the outcome values, a method in which the distribution of the observed data is used to construct sets of plausible values for the missing observations (10 imputed datasets). Variables included in the model were age, gender, body-mass index, duration of symptoms, smoking, treatment group (randomized to surgery, randomized to prolonged conservative care or non-randomized), MRI variables (presence of disc herniation, presence of nerve root compression and corresponding disc level), and baseline and other follow-up measurements of the outcomes being predicted. Complete case analysis (i.e. no imputation) was performed as a sensitivity analysis. Statistical significance was defined as $P < 0.05$.

RESULTS

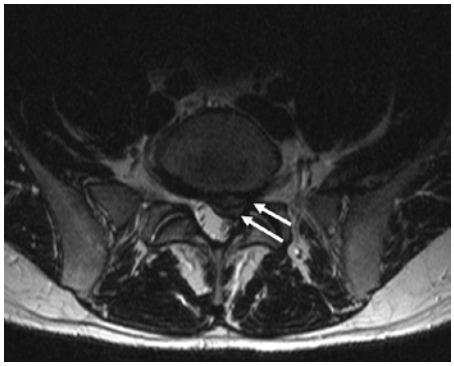
Of the 599 patients screened for the study, 395 patients considered eligible for inclusion underwent MRI of whom 283 patients were randomized.[13] In total, 106 baseline MRIs of the 112 non-randomized patients and 283 MRIs of the 283 randomized patients could be retrieved,

bringing the total to 389 MRIs. The study population had a mean age of 43.2 years with the majority being men (63%). The mean duration of sciatica was 9.3 weeks. At baseline, the study population reported a mean RDQ of 16.0, VAS-leg pain of 63.2 mm and VAS-back pain of 33.5 mm. Clinical outcome at 52 weeks was missing in 13-14% of patients (Appendix Table

Table 1 Agreement among the three observers regarding the presence of disc herniation and nerve root compression on MRI. Values are n (%).

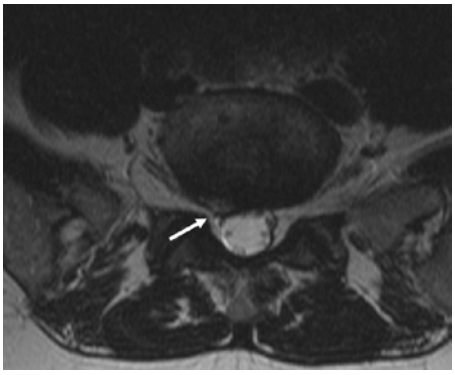
	Presence of disc herniation	Presence of nerve root compression
All 3 observers independently agreed about the presence	296 (76)	262 (67)
All 3 observers independently agreed about the absence	45 (12)	57 (15)
Disagreement		
2 of the 3 observers independently considered it present and 1 observer considered it absent	34 (9)	47 (12)
1 of the 3 observer considered it present and 2 observers considered it absent	14 (4)	23 (6)

1A T2-weighted sagittal image



1B T2-weighted axial image

1C T2-weighted sagittal image



1D T2-weighted axial image

Figure 1 Axial and sagittal T2 images of 1 patient in whom all 3 MRI observers agreed about the presence of disc herniation at disc level L5-S1 (A and B, arrows), and of 1 patient in whom the 3 MRI observers disagreed whether a disc herniation is visible at disc level L5-S1 (C and D, arrows)

S1). RDQ, VAS-leg and VAS-back pain were comparable among patients for whom clinical outcome at 52 weeks was available and those for whom not (P-value range 0.20-0.39).

Of the 389 patients, the three observers independently agreed in 296 (76%) patients about the presence of a disc herniation, disagreed in 48 (12%) patients about its presence and agreed in 45 (12%) patients about its absence (Table 1). An example of complete agreement among the three observers about the presence of a disc herniation and an example of disagreement is shown in Figure 1.

Of the patients with a (consistent interpretation of) disc herniation on MRI 84% reported perceived recovery after one year compared to 75% of the patients with inconsistent interpretation and 58% of the patients in whom all three readers agreed about the absence of a herniated disc (P<0.001). The same pattern was observed with the RDQ score (P=0.007), VAS leg pain (P=0.06) and VAS back pain (P=0.001) (Table 2). When comparing the three groups pairwise no statistical significant difference was observed between patients with a disc herniation and those with inconsistent interpretation in any of the four outcome measures, but statistical significant differences were observed between the group with absence of a disc herniation compared with the group with a disc herniation or the group with inconsistent interpretation.

Of the 389 patients, the three observers agreed in 262 (67%) patients about the presence of nerve root compression, disagreed in 70 (18%) patients about its presence and agreed in 57 (15%) patients about its absence (Table 1). Of the patients with (consistent interpretation of) nerve root compression 86% reported perceived recovery after one year compared to 83% of the patients with inconsistent MRI interpretation and 49% of the patients in whom all three readers agreed about the absence of nerve root compression (P<0.001) (Table 2). The same pattern was observed with the RDQ score (P<0.001), VAS leg pain (P=0.001) and VAS back pain (P<0.001). Again, when comparing the three groups pairwise no statistical significant difference was observed between the patients with nerve root compression and those with inconsistent interpretation in any of the four outcome measures, but statistical significant differences were observed between the group with absence of nerve root compression compared with the group with nerve root compression or the group with inconsistent interpretation.

Results stratified by treatment group are shown in Table 3. Only 4 patients had absence of disc herniation in the group randomized to surgery or prolonged conservative care In all three treatment groups the same pattern was observed regarding nerve root compression: patients with absence of nerve root compression had the worse clinical results compared to those with nerve root compression or inconsistent interpretation (Table 3B).

The Kaplan-meier curves show that patients with a (consistent interpretation of) disc herniation had the highest speed of perceived recovery, followed by patients with inconsistent interpretation and those with absence of disc herniation (P=0.006) (Figure 2A). Patients with absence of nerve root compression also had a lower speed compared to those with nerve root compression or inconsistent interpretation (P=0.006) (Figure 2B and Table 4).

Sensitivity analyses performed to account for missing clinical data yielded similar results (Table S2 and Table S3 Supplementary Appendix).

Table 2 Clinical outcome measures at one year stratified by consistency in MRI interpretation among three observers regarding the presence of disc herniation and nerve root compression at baseline. Values are n (%) or means ± SD.

	Presence of a herniated disc at baseline			P Value	Presence of nerve root compression at baseline			P Value
	Consistent Present (n=296)	Inconsistent interpretation (n=48)	Consistent Absent (n=45)		Consistent present (n=262)	Inconsistent interpretation (n=70)	Consistent Absent (n=57)	
1-year outcome								
Perceived recoveryò	250 (84)	36 (75)	26 (58)	<0.001	226 (86)	58 (83)	28 (49)	<0.001
Roland Disability‡	3.2±5.1	4.2±5.8	5.7±6.3	0.007	3.1±5.0	2.8±4.4	7.0±6.8	<0.001
VAS-Leg pain¶	10.2±18.5	13.1±21.6	16.1±24.2	0.06	9.5±17.6	9.9±18.0	20.9±26.7	0.001
VAS-back pain¶	14.4±20.7	19.6±25.3	26.7±28.8	0.001	13.5±20.0	15.1±20.6	31.7±29.9	<0.001

ò Perceived recovery 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 3 Clinical outcome measures at one year stratified by treatment group and consistency in MRI interpretation at baseline. Values are n (%) or means± SD.
3A Clinical outcome measures at one year stratified by treatment group and consistency in MRI interpretation regarding the presence of disc herniation at baseline.

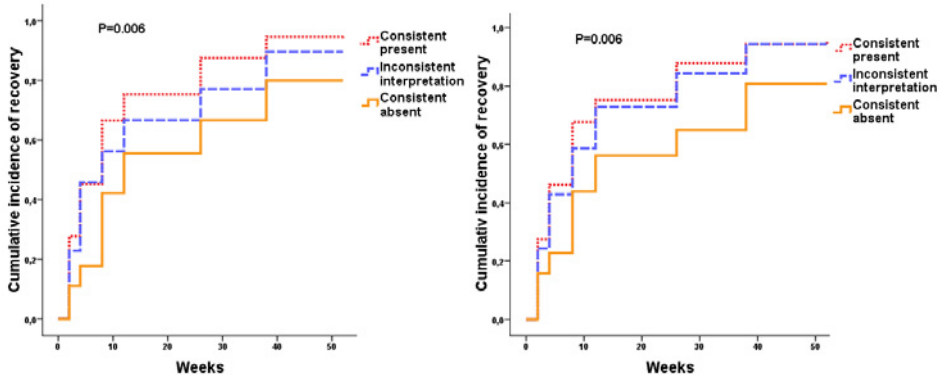
	Patients not randomized				Patients assigned to surgery				Patients assigned to conservative care			
	Consistent present (n=41)	Inconsistent interpretation (n=28)	Consistent absent (n=37)	P Value	Consistent present (n=124)	Inconsistent interpretation (n=13)	Consistent absent (n=4)	P Value	Consistent present (n=131)	Inconsistent interpretation (n=7)	Consistent absent (n=4)	P Value
Perceived recoveryò	34 (83)	22 (79)	19 (51)	0.009	108 (87)	9 (69)	4 (100)	0.55	108 (82)	6 (86)	3 (75)	0.73
Roland Disability‡	3.1±3.8	3.3±4.9	6.0±6.6	0.02	2.9±5.4	6.5±7.3	1.3±1.9	0.32	3.6±5.2	3.8±5.5	6.5±5.7	0.34
VAS-Leg pain¶	7.9±13.5	10.7±19.3	17.7±26.0	0.05	10.5±19.9	17.1±25.8	4.0±4.1	0.81	10.6±18.4	15.7±23.7	12.8±13.3	0.59
VAS-back pain¶	12.6±16.5	16.2±21.2	30.1±30.0	0.002	13.6±21.1	26.2±29.1	2.8±1.0	0.60	15.8±21.5	21.3±33.2	19.5±20.2	0.55

ò Perceived recovery 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.

3B Clinical outcome measures at one year stratified by treatment group and consistency in MRI interpretation regarding the presence of nerve root compression at baseline.

	Patients not randomized				Patients assigned to surgery				Patients assigned to conservative care			
	Consistent present (n=30)	Inconsistent interpretation (n=38)	Consistent absent (n=38)	P Value	Consistent present (n=112)	Inconsistent interpretation (n=18)	Consistent absent (n=11)	P Value	Consistent present (n=120)	Inconsistent interpretation (n=14)	Consistent absent (n=8)	P Value
Perceived recoveryò	25 (83)	30 (79)	19 (50)	0.010	101 (90)	16 (89)	5 (45)	0.001	101 (84)	12 (86)	4 (50)	0.05
Roland Disability‡	2.9±3.8	2.8±3.6	6.5±6.9	0.005	2.7±5.1	3.0±5.9	8.5±7.6	0.003	3.5±5.2	2.6±4.5	7.6±5.0	0.17
VAS-Leg pain¶	7.7±14.7	9.1±16.0	18.4±26.3	0.032	9.0±18.0	11.1±21.4	29.4±31.3	0.004	10.3±18.0	10.5±19.5	22.0±22.2	0.19
VAS-back pain¶	9.9±12.9	15.4±20.0	31.6±30.0	<0.001	12.3±19.8	15.7±23.2	34.3±31.2	0.002	15.6±21.5	13.4±20.0	29.0±30.8	0.25

ò Perceived recovery 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.



2A **2B**

Figure 2 Inverse Kaplan-Meier Curves estimating the cumulative incidence of perceived recovery within the first year after baseline. Recovery was defined as complete or nearly complete disappearance of symptoms on the patient-reported 7-point Likert scale for global perceived recovery.
2A Cumulative incidence of recovery for patients with consistent (all three readers agreed about the presence or absence) and inconsistent MRI interpretation (i.e. one reader disagreed with the other two) regarding the presence of disc herniation at baseline.
2B Cumulative incidence of recovery for patients with consistent (all three readers agreed about the presence or absence) and inconsistent MRI interpretation regarding the presence of nerve root compression at baseline.

Table 4 Time to perceived recovery within the first year according to consistency in baseline MRI interpretation among three observers. Perceived recovery was defined as "complete" or "nearly complete disappearance of symptoms" on the 7-point Likert scale. HR denotes hazard ratio. CI denotes confidence interval.

	HR	95% CI	P-value
Presence of disc herniation			0.006
Consistent present	1.67	1.16-2.41	0.006
Inconsistent interpretation	1.43	0.90-2.27	0.14
Consistent absent		Reference group	
Presence of nerve root compression			0.006
Consistent present	1.65	1.19-2.28	0.003
Inconsistent interpretation	1.60	1.08-2.38	0.020
Consistent absent		Reference group	

DISCUSSION

The present study analyzed the significance of MRI interobserver variability among three spine specialists for the one-year outcomes in patients with sciatica who were potential candidates for lumbar disc surgery based on clinical grounds. The most favorable clinical outcome results after one year follow-up were reported by those patients in whom all three MRI observers independently agreed about the presence of disc herniation or nerve root compression, followed

by those with inconsistent interpretation and finally by those in whom independent agreement was reached about the absence of those abnormalities.

The direct evaluation of herniated discs and nerve roots by MRI has been considered an important asset to facilitate decision making in patients with leg and/or back pain.[5,17,18] Uncertainty on the presence of a herniated disc with nerve root compression will in most cases result in conservative treatment, while a certain herniated disc with nerve root compression will in most cases result in surgery.[11] However, as with any diagnostic radiographic study, interpretation of the results regarding the presence of a herniated disc and nerve root compression may become inconsistent between examiners.[11,19-22] It has been suggested that inconsistency in interpretation may lead to alternative treatment options between clinicians and therefore may impact the outcome of patient treatment.[19] Variations in rates of spinal surgery may be related in part to substantial variability among physicians in interpreting the abnormalities identified with lumbar MRI.[23] The results of the present study suggest that based on the consistency in interpretation by the MRI assessors prognostic profiles can be made in sciatica, and that the mechanism behind these prognostic profiles is probably related to whether there is truly a disc herniation or nerve root compression present (if present a favorable prognosis compared with unfavorable when absent). The presence of nerve root compression in patients with sciatica has earlier been reported to be associated with favorable prognosis in primary care patients with sciatica.[24]

Principles of rational medicine suggest that outcomes can be improved by providing physicians better diagnostic data that clarify disease characteristics.[25,26] Clinical outcomes might be poorer when patient heterogeneity is not recognized, leading to a mismatch between patient subgroups and intervention type.[26,27] The current study show that indeed the best clinical treatment results after one year follow-up are reported by those patients in whom MRI observers agreed regarding the presence of a disc herniation or nerve root compression as compared to those with inconsistent interpretation or those with absence of those findings. Tremendous effort has been put in uncovering the relationship between specific imaging characteristics and patient outcomes in sciatica, which unfortunately remains controversial.[28,29] 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 (the predictors), which causes the inputs in the prediction model to be faulty.[30] Therefore, it is not only from a clinical but also from a research perspective crucial that radiologists and clinicians strive to reduce variability in interpretation.[31] In the current study the MRI observers disagreed in nearly one fifth regarding the presence of nerve root compression. Specific training and defining the language for image interpretation for degenerative disc disease have been proposed to reduce variability in interpretation.[30,31]

We deliberately did not organize a consensus meeting in which a sample of images was evaluated. Such a meeting could have caused the observers to adjust their diagnostic imaging

criteria and could have overestimated consistency compared to the situation as it existed before the study. However, our study has several shortcomings. The concordance found in this study may have been overestimated, since one reading pair consisted of two neuroradiologists who had nearly the same observer experience and also worked together which may have led to an informal agreement in their diagnostic criteria.[21] The concordance might also have been overestimated since a great part of our study sample consisted of a relatively homogeneous study sample with well-defined inclusion criteria and known sciatica due to previous confirmed disc herniation by another observer. The found concordance is likely to be higher compared to a study sample consisting of patients in whom diagnosis was not confirmed as well as those who are confirmed to have or not to have disc herniation.[32] And finally, the study population of the present study consisted of sciatica patients who had severe symptoms for at least 6 weeks and were referred to the neurologists. These patients were willing to undergo surgery, so patients with a clear preference for conservative treatment are underrepresented.[3]

In conclusion, at one year follow-up the most favorable clinical outcome results were reported by those patients in whom all three MRI observers independently agreed about the presence of disc herniation or nerve root compression, followed by those with inconsistent interpretation and finally by those with absence of those findings.

INFORMED CONSENT

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. Informed consent was obtained from all patients for being included in the study.

ACKNOWLEDGEMENT

This study was funded by grants from the Health Care Efficiency Research Program of Netherlands Organisation for Health Research and Development (ZonMw) and the Hoelen Foundation, The Hague. There were no agreements concerning confidentiality of the data between the funders and the authors or the participating institutions. The funders did also not have any role in the writing or analysis of this study, nor in the decision to publish the paper.

REFERENCES

1. Koes BW, van Tulder MW, Peul WC (2007) Diagnosis and treatment of sciatica. BMJ 334(7607):1313-1317
2. Konstantinou K, Dunn KM (2008) Sciatica: review of epidemiological studies and prevalence estimates. Spine (Phila Pa 1976) 33(22):2464-2472
3. Carragee EJ, Kim DH (1997) A prospective analysis of magnetic resonance imaging findings in patients with sciatica and lumbar disc herniation. Correlation of outcomes with disc fragment and canal morphology. Spine (Phila Pa 1976) 22(14):1650-1660
4. van Tulder MW, Koes BW, Bouter LM (1995) A cost-of-illness study of back pain in The Netherlands. Pain 62(2):233-240
5. Boos N, Rieder R, Schade V, Spratt KF, Semmer N, Aebi M (1995) 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) 20(24):2613-2625
6. Wassenaar M, van Rijn RM, van Tulder MW, et al. (2012) Magnetic resonance imaging for diagnosing lumbar spinal pathology in adult patients with low back pain or sciatica: a diagnostic systematic review. Eur Spine J 21(2):220-227
7. den Boer JJ, Oostendorp RA, Beems T, Munneke M, Oerlemans M, Evers AW (2006) A systematic review of bio-psychosocial risk factors for an unfavourable outcome after lumbar disc surgery. Eur Spine J 15(5):527-536
8. Findlay GF, Hall BI, Musa BS, Oliveira MD, Fear SC (1998) A 10-year follow-up of the outcome of lumbar microdiscectomy. Spine (Phila Pa 1976) 23(10):1168-1171
9. Korres DS, Loupassis G, Stamos K (1992) Results of lumbar discectomy: a study using 15 different evaluation methods. Eur Spine J 1(1):20-24
10. Peul WC, van den Hout WB, Brand R, Thomeer RT, Koes BW (2008) Prolonged conservative care versus early surgery in patients with sciatica caused by lumbar disc herniation: two year results of a randomised controlled trial. BMJ 336(7657):1355-1358
11. Cihangiroglu M, Yildirim H, Bozgeyik Z, et al. (2004) Observer variability based on the strength of MR scanners in the assessment of lumbar degenerative disc disease. Eur J Radiol 51(3):202-208
12. Vucetic N, Astrand P, Guntner P, Svensson O (1999) Diagnosis and prognosis in lumbar disc herniation. Clin Orthop Relat Res(361):116-122
13. Peul WC, van Houwelingen HC, van den Hout WB, et al. (2007) Surgery versus prolonged conservative treatment for sciatica. N Engl J Med 356(22):2245-2256
14. Peul WC, van Houwelingen HC, van der Hout WB, et al. (2005) 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 6:8
15. Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB (1995) Assessing health-related quality of life in patients with sciatica. Spine (Phila Pa 1976) 20(17):1899-1908; discussion 1909
16. Collins SL, Moore RA, McQuay HJ (1997) The visual analogue pain intensity scale: what is moderate pain in millimetres? Pain 72(1-2):95-97
17. Jarvik JJ, Hollingworth W, Heagerty P, Haynor DR, Deyo RA (2001) The Longitudinal Assessment of Imaging and Disability of the Back (LAIDBack) Study: baseline data. Spine (Phila Pa 1976) 26(10):1158-1166
18. van Rijn JC, Klemetso N, Reitsma JB, et al. (2006) Observer variation in the evaluation of lumbar herniated discs and root compression: spiral CT compared with MRI. Br J Radiol 79(941):372-377

19. Mulconrey DS, Knight RQ, Bramble JD, Paknikar S, Harty PA (2006) Interobserver reliability in the interpretation of diagnostic lumbar MRI and nuclear imaging. Spine J 6(2):177-184
20. Arana E, Royuela A, Kovacs FM, et al. (2010) Lumbar spine: agreement in the interpretation of 1.5-T MR images by using the Nordic Modic Consensus Group classification form. Radiology 254(3):809-817
21. Kovacs FM, Royuela A, Jensen TS, et al. (2009) Agreement in the interpretation of magnetic resonance images of the lumbar spine. Acta Radiol 50(5):497-506
22. van Rijn JC, Klemetso N, Reitsma JB, et al. (2005) Observer variation in MRI evaluation of patients suspected of lumbar disk herniation. AJR Am J Roentgenol 184(1):299-303
23. Weinstein JN, Lurie JD, Olson PR, Bronner KK, Fisher ES (2006) United States' trends and regional variations in lumbar spine surgery: 1992-2003. Spine (Phila Pa 1976) 31(23):2707-2714
24. Vroomen PC, Wilmink JT, de KM (2002) Prognostic value of MRI findings in sciatica. Neuroradiology 44(1):59-63
25. Bornstein BH, Emler AC (2001) Rationality in medical decision making: a review of the literature on doctors' decision-making biases. J Eval Clin Pract 7(2):97-107
26. Lotz JC, Haughton V, Boden SD, et al. (2012) New treatments and imaging strategies in degenerative disease of the intervertebral disks. Radiology 264(1):6-19
27. Wand BM, O'Connell NE (2008) Chronic non-specific low back pain - sub-groups or a single mechanism? BMC Musculoskelet Disord 9:11
28. Jensen MC, Brant-Zawadzki MN, Obuchowski N, Modic MT, Malkasian D, Ross JS (1994) Magnetic resonance imaging of the lumbar spine in people without back pain. N Engl J Med 331(2):69-73
29. Boden SD, Davis DO, Dina TS, Patronas NJ, Wiesel SW (1990) Abnormal magnetic-resonance scans of the lumbar spine in asymptomatic subjects. A prospective investigation. J Bone Joint Surg Am 72(3):403-408
30. Carrino JA, Lurie JD, Tosteson AN, et al. (2009) Lumbar spine: reliability of MR imaging findings. Radiology 250(1):161-170
31. Jarvik JG, Deyo RA (2009) Moderate versus mediocre: the reliability of spine MR data interpretations. Radiology 250(1):15-17
32. Kim SW, Yeom JS, Park SK, et al. (2009) Inter- and Intra-observer Reliability of MRI for Lumbar Lateral Disc Herniation. Clin Orthop Surg 1(1):34-39

Table S1 Outcome measurements available at 52 weeks after baseline MRI. The mentioned outcome measures were assessed at baseline, 2, 4, 8, 12, 26, 38, and 52 weeks. Values are n (%). Total n=389

	Number of patients (%) Total (n=389)
Global perceived recovery on a 7-point Likert scale at 52 weeksò	
Outcome available at 52 weeks	335 (86)
At least one follow-up examination	40 (10)
Lost to follow-up after baseline examination	14 (4)
Roland disability questionnaire at 52 weeks‡	
Outcome available at 52 weeks	338 (87)
At least one follow-up examination	37 (10)
Lost to follow-up after baseline examination	14 (4)
Visual Analogue scale for leg pain at 52 weeks¶	
Outcome available at 52 weeks	338 (87)
At least one follow-up examination	36 (9)
Lost to follow-up after baseline examination	15 (4)
Visual Analogue scale for back pain at 52 weeks¶	
Outcome available at 52 weeks	336 (86)
At least one follow-up examination	38 (10)
Lost to follow-up after baseline examination	15 (4)

ò Perceived recovery 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 with 0 representing no pain and 100 the worst pain ever experienced.

Table S2 Clinical outcome measures at one year stratified by consistency in MRI interpretation regarding the presence of disc herniation and nerve root compression at baseline**.** *This analysis only included patients with available clinical outcome at one year (n=335).* Values are n (%) or means ± SD.

	Presence of a herniated disc at baseline‖			P Value	Presence of nerve root compression at baseline§			P Value
	Consistent Present (n=269)	Inconsistent interpretation (n=40)	Consistent Absent (n=26)		Consistent present (n=242)	Inconsistent interpretation (n=56)	Consistent Absent (n=37)	
Perceived recoveryò	228 (85)	30 (75)	14 (54)	<0.001	210 (87)	47 (84)	15 (41)	<0.001
Roland Disability‡	3.2±5.2	4.7±6.1	6.1±6.3	0.004	3.1±5.0	3.0±4.7	13.3±19.9	<0.001
VAS-Leg pain¶	9.9±18.3	14.1±23.1	14.9±23.5	0.10	9.0±17.2	10.5±19.5	14.5±21.0	<0.001
VAS-back pain¶	14.4±20.8	18.6±25.9	26.9±30.3	0.005	13.3±19.9	14.5±21.0	34.2±31.2	<0.001

‖ Of the 335 patients the three observers disagreed in 40 patients about the presence of a herniated disc, agreed in 269 patients about its presence and agreed in 26 patients about its absence.
§ Of the 335 patients, the three observers disagreed in 56 patients about the presence of nerve root compression, agreed in 242 patients about its presence and agreed in 37 patients about its absence
ò Perceived recovery 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 S3 Time to perceived recovery within the first year according to consistency in MRI interpretation at baseline. Perceived recovery was defined as "complete" or "nearly complete disappearance of symptoms" on the 7-point Likert scale. *This analysis only included patients with available clinical outcome at one year (n=335).* HR denotes hazard ratio. CI denotes confidence interval.

	HR	95%CI	P-value
Presence of disc herniation			0.04
Consistent present	1.80	1.13-2.88	0.01
Inconsistent interpretation	1.58	0.90-2.76	0.11
Consistent absent		Reference group	
Presence of nerve root compression			0.04
Consistent present	1.66	1.12-2.47	0.01
Inconsistent interpretation	1.58	1.00-2.51	0.05
Consistent absent		Reference group	