Author: Barzouhi, Abdelilah el
Title: Paradigm shift in MRI for sciatica
Issue Date: 2013-12-03

PARADIGM SHIFT IN MRI FOR SCIATICA

ABDELILAH EL BARZOUHI

ISBN: 978-94-6169-436-2

Cover design by Optima Grafische Communicatie, A.el Barzouhi.
Design and lay-out of thesis by Optima Grafische Communicatie, A. el Barzouhi.
Printed by Optima Grafische Communicatie.

The study presented in this thesis was performed at the department of Neursosurgery of the Leiden University Medical Center.

The study was funded by a grant from the Health Care Efficiency Research program of the Netherlands Organization for Health Research and Development (ZonMw) and dr Eduard Hoelen Foundataion (the Hague).

Financial support for the printing of this thesis was generously provided by Leiden University Medical Center, Medical Center Haaglanden, De Puy Synthesis, Dutch Spine Society, InSpine, Promedics Medical Systems, dr Eduard Hoelen Foundataion, Carl Zeiss B.V., ChipShoft, Lycklama Medical Imaging.

PARADIGM SHIFT IN MRI FOR SCIATICA

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 3 december 2013
klokke 15.00 uur

door

Abdelilah el Barzouhi
geboren te Den Haag in 1987

PROMOTIECOMMISSIE

Promotor:	Prof.dr. W.C. Peul
Co-promotor:	Dr. C.LA.M. Vleggeert-Lankamp
Overige leden:	Prof.dr. A.M. Stiggelbout Prof.dr. B.W. Koes (Erasmus Universiteit Rotterdam) Dr. R.H.M.A. Bartels (Radboud Universiteit Nijmegen)

Voor mijn ouders

CONTENTS

Chapter 1

Introduction & Outline of the Thesis

Sciatica, more accurately called lumbosacral radicular syndrome or sciatica neuralgia, is one of the most common lumbar-spine disorders. Sciatica is generally defined as pain radiating to the leg below the knee following a dermatomal pattern.[1] It is probable that human's upright posture and relative longevity have exposed our species to a special, unwelcome affinity for lumbar disc syndrome and associated sciatica.[2] The prevalence of sciatic symptoms reported in the literature varies considerably ranging from 1.6% in the general population to 43% in a selected working population.[3] The natural history of acute sciatica is in general favorable, although a substantial proportion (up to 30%) continues to have pain for one year or longer.[1, 4, 5] Sciatica is associated with significant short- and sometimes long-term morbidity. This affliction, certainly in the industrialized countries, ranks as one of the most costly and ubiquitous medical problems.[6]

THE LONG WAY TOWARD OVERCOMING THE SCIENTIFIC CONFUSION

In classical literature sciatica has been of great interest to Greco-Roman and Eastern scientists and physicians.[7] The Greek physician Hippocrates (460-370 BC) is generally believed to be the first to describe the treatment for sciatica. Since ancient times many etiological explanations for sciatica have been proposed. Domenico Cotugnio (1736-1822), a skilled Italian physician and anatomist, was the first to really add something new to the description of sciatica in his 1764 seminal paper "De ischiade nervosa commentarius".[7, 8] He explained the sciatic complaints as a consequence of neuritis or edema of the sciatic nerve. As treatment he recommended cautery, saying that he had never seen a failure after this procedure (a triumph that might be explained by the reluctance of patients to return to his care after having suffered unbearable pain during treatment, and so unabling him to really measure the effectiveness of this procedure).[7] For years this inflammation of the sciatic nerve, described as sciatic neuritis, was the origin of pain. In 1857 Rudolf Virchow (1821-1902) described the traumatic rupture of an intervertebral disc, which afterwards became known as "Virchow Tumor".[9] In 1858 the famous German pathologist Hubert von Luschka (1820-1875) discovered at autopsy several instances of asymptomatic herniated lumbar discs, which he erroneously described as cartilaginous tumors of the disc.[10] He speculated that in more advanced cases this finding might produce neurological complaints.

With the introduction of effective anesthesia in the second half of the 19th century it became possible to operate upon the vertebral column and observe anatomic relationships.[7] In 1909, the German neurosurgeon Fedor Krause (1857-1937) and his neurologist colleague Hermann Oppenheim (1858-1919) reported on the removal of an "enchondroma", which in retrospect must have been a ruptured disc.[11] In 1911 Joel Goldthwait (1866-1961) reported on a patient with recurrent sciatica and low back pain, in whom Harvey Cushing (1869-1939) had performed a negative surgical exploration.[12] Despite that no lesion was found Goldwait believed that a "disclocated" disc, not evident at surgery, could have produced sciatica. In 1915 Charles Elsberg (1871-1948) reported "a surgical cure for sciatica" effected by the removal of a piece of ruptured ligamentum of "subflavum" which was compromising a nerve root.[7, 13] In 1929 the famous neurosurgeon Walter Dandy (1886-1946) at John Hopkins found cartilaginous fragments lying loose in the spinal canal which he believed might well produce sciatica by compressing the adjacent nerve roots.[14] He even argued that by their removal, the patient's pain and suffering could be cured. Unfortunately, the importance of this paper went largely unrecognized as he continued to call these disc protrusions tumors.

In 1932 the prominent neurosurgeon William Jason Mixter (1880-1958) (Figure 1) and his orthopaedic colleague Joseph Seaton Barr (1901-1964) questioned whether "enchondromas" were truly the cause of sciatica and set out to review all of the previously diagnosed "enchondromas" at Massachusets General Hospital.[15] They observed that most of these cases were pathologically identical to normal disc material. Mixter later would recall that "this made us certain that we were dealing with a considerable group of lesions previously described as neoplasm, but undoubtedly of traumatic origin."[16] They concluded that enchondromas, Schmorl's nodules, and ruptured intervertebral discs were one and the same, and that the lesion was a common cause of the classic signs and symptoms of sciatica. Their ideas were met with considerable resistance at first. Mixter reports that he had asked for permission from the surgical executive committee at Massachusets General Hospital to present his findings at a meeting of the Massachusetts Medical Society, and "permission was refused on the ground that the subject was far too controversial to be given in such a meeting."[15, 16] In the spring of 1933 Barr did get a chance to present their work to a group at the Brigham Hospital Reunion, but the article essentially failed to spark any interest.[15] Finally, Mixter and Barr's report was read before the New England Surgical Society on September 30, 1933. In their famous publication in the August 2 issue of the New England Journal of Medicine in 1934 they stated[17]:

Figure 1. *Dr. William Jason Mixter, neurosurgeon*

"We conclude from this study: that a herniation of the nucleus pulposus into the spinal canal, or as we prefer to call it, rupture of the intervertebral disc, is a not uncommon cause of symptoms. That the lesion frequently has been mistaken for cartilaginous neoplasm arising from the intervertebral disc… That the treatment of this disease is surgical and that the results obtained are very satisfactory if compression has not been too prolonged".

This landmark report of Mixter and Barr greatly revolutionized medical think at the time, ushering in a greater interest in the lumbar disc as a source of sciatica and in the surgical treatment of such a disorder. In fact the report caused a shift in clinical management from largely conservative to that of surgery, which has come to be known as the "Dynasty of the Disc".[15, 18] Surgery for back and leg pain in association with nerve root compression has become one of the most commonly performed operative procedures worldwide.

REVIVAL OF SCIENTIFIC CONFUSION

Walter Dandy (1886-1946) introduced air myelography in 1918 at the Johns Hopkins Hospital for the diagnosis of space- occupying brain lesions. The difficulties in properly performing this procedure limited its widespread use in the spine. In 1920 the French neurologist and radiologist Jean Sicard (1872-1929) introduced iodinated contrast myelography, allowing the relatively accurate diagnosis of intraspinal pathology.[19] In subsequent decades the accuracy and safety of this diagnostic procedure were greatly improved.

In 1977 the first Magnetic Resonance Imaging (MRI) body scan of a human being was performed.[20, 21] Within a few years spinal MR imaging became available and was rapidly becoming the imaging modality of choice for most spinal disorders. However, the high-resolution images which allowed many investigators to detect an enormous variety of previously unappreciated anatomical variations in patients undergoing diagnostic workups for sciatica, also caused scientific confusion of our understanding of sciatia.[22] For example, in the early nineties of the 20th century several MRI studies showed a high prevalence of disc herniations ranging from 20 to 76% in persons without any symptoms.[23-25] Even in patients who were re-imaged after earlier disc surgery, MRI studies have found herniations in up to 53% of persons who at the time of the re-imaging had no symptoms.[26-28] Despite this scientific confusion, however, MRI is considered the imaging procedure of choice for patients suspected of lumbar disc herniations[23, 29] and is frequently performed in patients with persistent or recurrent symptoms of sciatica.[30] Moreover, abnormal MRI findings frequently result in surgical treatment or other invasive procedures such as epidural injections.[31, 32]

The controversy discussed above challenges our understanding of sciatica and the value of MRI in patients with sciatica. Many anatomical abnormalities detected with high-resolution imaging may not be of clinical consequence but are now exposing patients to interventions with potential risks. Establishing correlations between MRI findings and clinical outcome in patients with sciatica may not only help improve our understanding of the etiology of sciatica, but it may also provide anchor points for new therapeutic approaches or fine-tuning of existing therapeutic strategies. To uncover the relevance of imaging findings it does not only require knowledge regarding their prevalence, but also their behavior of change with time, spectrum of changes and their relation with clinical outcome.[33] As the source for determining the clinical relevance of MR imaging findings data from the Sciatica Trial will be used in this thesis.

The Sciatica Trial is a multicentre prospective randomized controlled trial among patients with 6-12 weeks sciatica and disc herniation on MRI. An early surgery strategy was compared to prolonged conservative care for an additional 6 months followed by surgery for patients who did not improve or who did request it earlier because of aggravating symptoms.[34, 35] The trial showed faster recovery after early surgery as compared to a strategy of prolonged conservative care with surgery if needed, but there were no significant differences in clinical outcomes after one year. The randomized patients were part of a larger group of patients with sciatica who underwent a baseline MRI to assess the eligibility for the sciatica trial. All patients who underwent MRI (regardless of participation in the randomized controlled trial) were followed up for one year. Furthermore, all randomized patients underwent MRI at baseline and after one year. The 12 months evaluation period was selected since postoperative fibrosis usually stabilizes by 6 months, with no further changes at 12 months.[36]

OBJECTIVE AND OUTLINE OF THIS THESIS

The main objective of this thesis is to uncover the relationship between MRI findings and clinical outcome in patients with sciatica. As with any diagnostic study requiring expert reading, interpretation of MRI findings may be inconsistent between examiners. In **chapter 2** results are reported regarding the intra- and inter-observer variation in MRI evaluation among two neuroradiologists and one neurosurgeon who routinely assess spinal MRIs. 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. In **chapter 3** clinical outcome results are reported of patients 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.

The natural history of acute sciatica is in general favorable, although a substantial proportion (up to 30%) receives surgery. **Chapter 4** presents the results of both qualitative and quantitative MRI evaluations in predicting surgery for sciatica in a group of prolonged conservative care patients.

Patients with sciatica frequently complain about associated back pain. **Chapter 5** reports on the MRI differences between patients who suffer from sciatica with disabling back pain as compared to patients who suffered from predominantly sciatica, and on the significance of these MRI differences for prognosis.

Despite being scientifically debated, MRI is frequently repeated in patients with persistent or recurrent symptoms of sciatica. **Chapter 6** reports on the 1-year MRI findings of sciatica patients who were treated with either surgery or conservative treatment, changes of MRI findings over time, and their correlation with clinical outcome.

In the search for causes of associated back pain in patients with sciatica, vertebral endplate signal changes (VESC, also called Modic changes) visualized by MRI have been proposed as a possible cause. VESC are a frequent surgical indication to perform a fixation of two or more vertebrae in the lower spine or replacing the disc by a prothesis. **Chapter 7** reports on VESC findings, changes of VESC findings over time and the correlation between VESC findings and back pain in sciatica.

Gadolinium-enhanced MRI is frequently performed in patients with persistent or recurrent symptoms of sciatica after surgical treatment, as it has been proposed to differentiate between postoperative epidural scar tissue and recurrent disc herniation: scar tissue has a homogenous enhancement pattern while disc herniation usually lacks central enhancement. **Chapter 8** reports on the reliability of enhancement findings, their prevalence and their correlation with clinical outcome.

A synthesis and discussion about the results are given in **chapter 9**. The dissertation is concluded with a summary in **chapter 10.**

REFERENCES

1. Koes BW, van Tulder MW, Peul WC. Diagnosis and treatment of sciatica. BMJ 2007;334:1313-7.
2. Karampelas I, Boev AN, 3rd, Fountas KN, Robinson JS, Jr. Sciatica: a historical perspective on early views of a distinct medical syndrome. Neurosurg Focus 2004;16:E6.
3. Konstantinou K, Dunn KM. Sciatica: review of epidemiological studies and prevalence estimates. Spine (Phila Pa 1976) 2008;33:2464-72.
4. Weber H, Holme I, Amlie E. The natural course of acute sciatica with nerve root symptoms in a double-blind placebo-controlled trial evaluating the effect of piroxicam. Spine (Phila Pa 1976) 1993;18:1433-8.
5. Vroomen PC, de Krom MC, Slofstra PD, Knottnerus JA. Conservative treatment of sciatica: a systematic review. J Spinal Disord 2000;13:463-9.
6. Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. Spine J 2008;8:8-20.
7. Robinson JS. Sciatica and the lumbar disk syndrome: a historic perspective. South Med J 1983;76:232-8.
8. Cotugnio D. De ischiade nervosa commentarius. Neapoli apud frat Simonios 1764.;12 ed. Viennæ:.
9. Virchow R. Untersuchungen über die Entwicklung des Schadelgrunde im Gesunden und Krankhaften Zustande. Reimer, Berlin; 1857.
10. Luschka H. Die Halbgelenke des menschlichen Körpers Reimer, Berlin 1858.
11. Oppenheim H, Krause F. Über Einklemmung bzw. Strangulation der cauda equina. Dtsch Med Wochenschr 1909;35:697-700.
12. Goldthwait J. The lumbosacral articulation. An explanation of many cases of "lumbago", "sciatica" and "paraplegia". Boston Med Surg J 1911;164:365-601.
13. Elsberg C. Pain and other sensory disturbances in diseases of the spinal canal and their surgical treatment. Am J Med Sci Neurosurg 1915;149.
14. Dandy WE. Loose cartilage from intervertebral disk simulating tumor of the spinal cord. Arch Surg 1929;19:660-72.
15. Parisien RC, Ball PA. William Jason Mixter (1880-1958). Ushering in the "dynasty of the disc". Spine (Phila Pa 1976) 1998;23:2363-6.
16. Mixter WJ. Rupture of the intervertebral disk: A short history of its evolution as a syndrome of importance to the surgeon. JAMA 1949;140:278-82.
17. Mixter WJ, Barr JS. Rupture of Intervertebral Disc with Involvement of Spinal Canal. N Engl J Med 1934;211:210-5.
18. Allan DB, Waddell G. An historical perspective on low back pain and disability. Acta Orthop Scand Suppl 1989;234:1-23.
19. Sicard JA, Forestier J. Méthode radiographique d'exploration de la cavité épidurale par la lipiodol. Revue neurologique, Paris 1921;28:1264-6.
20. Damadian R, Goldsmith M, Minkoff L. NMR in cancer: XVI. FONAR image of the live human body. Physiol Chem Phys 1977;9:97-100, 8.
21. Hinshaw WS, Bottomley PA, Holland GN. Radiographic thin-section image of the human wrist by nuclear magnetic resonance. Nature 1977;270:722-3.
22. Beattie PF, Meyers SP, Stratford P, Millard RW, Hollenberg GM. Associations between patient report of symptoms and anatomic impairment visible on lumbar magnetic resonance imaging. Spine (Phila Pa 1976) 2000;25:819-28.

23. 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.
24. 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.
25. Boden SD, Davis DO, Dina TS, Patronas NJ, Wiesel SW. Abnormal magnetic-resonance scans of the lumbar spine in asymptomatic subjects. A prospective investigation. J Bone Joint Surg Am 1990;72:403-8.
26. 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.
27. 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.
28. Van Goethem JW, Van de Kelft E, Biltjes IG, et al. MRI after successful lumbar discectomy. Neuroradiology 1996;38 Suppl 1:S90-6.
29. 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.
30. 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.
31. 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.
32. Lurie JD, Birkmeyer NJ, Weinstein JN. Rates of advanced spinal imaging and spine surgery. Spine (Phila Pa 1976) 2003;28:616-20.
33. Modic MT, Obuchowski NA, Ross JS, et al. Acute low back pain and radiculopathy: MR imaging findings and their prognostic role and effect on outcome. Radiology 2005;237:597-604.
34. 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.
35. 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.
36. 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.

Chapter 2

Magnetic Resonance Imaging interpretation in patients with sciatica who are potential candidates for lumbar disc surgery

Abdelilah el Barzouhi, Carmen L.A.M. Vleggeert-Lankamp, Geert J. Lycklama à Nijeholt, Bas F. Van der Kallen, Wilbert B. van den Hout, Annemieke J.H. Verwoerd, Bart W. Koes, and Wilco C. Peul.

PLoS One. 2013 Jul 10;8(7):e68411.

ABSTRACT

BACKGROUND

Magnetic Resonance Imaging (MRI) is considered the mainstay imaging investigation in patients suspected of lumbar disc herniations. Both imaging and clinical findings determine the final decision of surgery. The objective of this study was to assess MRI observer variation in patients with sciatica who are potential candidates for lumbar disc surgery.

METHODS

Patients for this study were potential candidates (n=395) for lumbar disc surgery who underwent MRI to assess eligibility for a randomized trial. Two neuroradiologists and one neurosurgeon independently evaluated all MRIs. A four point scale was used for both probability of disc herniation and root compression, ranging from definitely present to definitely absent. Multiple characteristics of the degenerated disc herniation were scored. For inter-agreement analysis absolute agreements and kappa coefficients were used. Kappa coefficients were categorized as poor (<0.00), slight (0.00-0.20), fair (0.21-0.40), moderate (0.41-0.60), substantial (0.61-0.80) and excellent (0.81-1.00) agreement.

RESULTS

Excellent agreement was found on the affected disc level (kappa range 0.81-0.86) and the nerve root that most likely caused the sciatic symptoms (kappa range 0.86-0.89). Interobserver agreement was moderate to substantial for the probability of disc herniation (kappa range 0.57-0.77) and the probability of nerve root compression (kappa range 0.42-0.69). Absolute pairwise agreement among the readers ranged from 90-94% regarding the question whether the probability of disc herniation on MRI was above or below 50%. Generally, moderate agreement was observed regarding the characteristics of the symptomatic disc level and of the herniated disc.

CONCLUSION

The observer variation of MRI interpretation in potential candidates for lumbar disc surgery is satisfactory regarding characteristics most important in decision for surgery. However, there is considerable variation between observers in specific characteristics of the symptomatic disc level and herniated disc.

INTRODUCTION

Sciatica is defined as intense leg pain in an area served by one or more spinal nerve roots and is occasionally accompanied by neurological deficit.[1] Sciatica places a heavy burden on public health as it is a major source of lost productivity.[2] The most common cause of sciatica is a herniated disc.[1] Magnetic resonance imaging (MRI) is considered the imaging procedure of choice for patients suspected of lumbar herniated discs.[3,4,5] MRI is indicated in patients with severe symptoms who fail to respond to conservative care for at least 6 to 8 weeks.[1] In these cases surgery as a treatment modality might be considered and MRI is used to assess if a herniated disc with nerve root compression is indeed present. Both imaging and clinical findings determine the final decision of surgery.[6] The important role of MRI in clinical decision making makes a reliable interpretation of lumbar MRI therefore desirable.

Despite remarkable advancements in diagnostic imaging and surgical techniques the results after lumbar disc surgery do not seem to have improved during recent decades: depending upon the used outcome measure, the results of lumbar disc surgery are unsatisfactory in 10 to 40% of the patients.[7,8,9] It has been suggested that the poor outcomes following lumbar disc surgery may be more often due to the errors in diagnosis than the surgical technique or its complications.[6,10] For example, a false-positive diagnosis of nerve root compression on MRI may lead to unwarranted surgery. Therefore, if truly substantial interpretation variability exists among those who routinely interpret spine MRI studies, this would influence treatment decisions with possible negative effects. Unreliable interpretation may also pose research problems when attempting to uncover the relationship between specific imaging characteristics and patient outcomes. Therefore, insight in the interpretation variability of MRI findings among potential candidates for lumbar disc surgery is essential.

The investigators previously reported the results of a randomized controlled trial comparing early surgery with prolonged conservative care for patients with sciatica over one year's follow-up.[11] The randomized patients were part of a larger group that underwent MRI to assess the eligibility for the trial. Within this larger group, we report on the intra- and inter-observer variation in MRI evaluation among two neuroradiologists and one neurosurgeon.

MATERIALS AND METHODS

ETHICS STATEMENT

The medical ethics committees at the nine participating hospitals (Leiden University Medical Center, Medical Center Haaglanden, Diaconessen Hospital, Groene Hart Hospital, Reinier de Graaf Hospital, Spaarne Hospital, Bronovo Hospital, Rijnland Hospital and Lange Land Hospital) approved the protocol. Written informed consent was obtained from all patients.

STUDY POPULATION

Patients for this study were patients with 6 to 12 weeks of sciatic symptoms being so severe that they were eligible for surgery according to their family practitioners and were therefore referred to a neurologist. The attending neurologist subsequently evaluated whether these patients were eligible to participate in the Sciatica Trial: a multicenter 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. 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, severe coexisting disease or if they were not between 18 to 65 years of age. All participants who were not meeting one or more of the aforementioned exclusion criteria underwent MRI. If the MRI showed a disc herniation with nerve root compression correlating with clinical symptoms according to the attending neurologist and neurosurgeon the corresponding patient was eligible to participate in the randomized clinical trial. Thus if a patient did not display a disc herniation according to the neurologist who assessed the MRI at the time of enrollment in the Trial, this patient could not enter the randomized controlled Trial. As the purpose of the current study was to evaluate observer variation among sciatica patients who are surgical candidates for sciatica, MRIs of all patients (regardless of participation in the randomized clinical trial) were again evaluated by independent observers (who did not participate in this study before) to determine observer variation regarding MRI characteristics. Details of the design and study protocol have been published previously.[12]

MRI PROTOCOL AND IMAGE EVALUATION

MRI scans were performed in all 9 participating hospitals using standardized protocols tailored to a 1.5 Tesla scanner. Sagittal T1 and axial T1 spin echo images of the lumbar spine were acquired. In addition, T2 weighted sagittal and axial series were obtained. For research purposes also contrast-enhanced (Gadolinium dithylene triamine penta-acetic acid [DTPA] at a standard dose of 0.1 mmol/kg body weight) T1 fat suppressed sagittal and axial images were obtained.

MR images of all included patients were obtained and saved in an Apple PowerBook PC laptop with an 1.67 GHz G4 processor running open-source OsiriX Medical Image software (Version 3.0.1). Size of the monitor was 15,2 inch, 1280 x 854 pixel resolution.

Two neuroradiologists and one neurosurgeon 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. The readers were able to freely adjust contrast and image brightness and zoom, and were able to compare sagittal and axial images simultaneously. All readings were performed on the same Apple PC laptop. 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.

Each reader received a manual containing definitions of imaging characteristics based on the 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 disc pathology in order to standardize the nomenclature.[13] Pictorial examples were also provided where appropriate, gathered from the literature if available. Vertebral endplate signal changes were defined according to criteria of Modic et al.[14,15] Before beginning the study, the readers met in person to review and refine the standardized definitions in case of ambiguities. After reaching final consensus, standardized case record forms with these final definitions were used to evaluate the images (Table 1). First, all readers had to choose whether the MRI showed an impaired lumbar disc level that may have explained the sciatic complaints of the patients. If so, multiple characteristics of the degenerated disc level and disc herniation were scored. For both the presence of disc herniation and nerve root compression a four point scale was used: "Definite about the presence", "Probable about the presence" if there was some doubt but probability >50%, "Possible about the presence" if there was reason to consider but probability <50%, and "Definite about the absence".

When all three observers finished reading the images they repeated the MRI evaluation for ten percent of the evaluated images to provide intra-observer reliability data. The observers were not aware they were actually evaluating the images for a second time since in advance they were not informed about the conduction of an intra-observer reliability study. The images used for this intra-observer study were randomly selected from the first three-quarter of the evaluated images to minimize possible effects of recent memories. The time period between the first and the second evaluation was at least 2 months for all observers.

Table 1 MRI study variables.

MRI variable	Type	Categories
Disc level that most likely caused the lumbosacral radicular syndrome of the patient	Disc level	1. L2L3 2. L3L4 3. L4L5 4. L5S1 5. Not applicable, all disc levels have a normal disc contour: no disc extension beyond the normal margins of the intervertebral disc space at any disc level
	Disc contour at this disc level	1. Bulging: presence of disc tissue circumferentially (50-100%) beyond the edges of the ring apophyses 2. herniation: localized displacement of disc material beyond the normal margins of the intervertebral disc space
	Certainty about the presence of this disc herniation	1. Definite about the presence: no doubt about the presence 2. Probable about the presence: some doubt but likelihood > 50% 3. Possible about the presence: reason to consider but likelihood < 50% 4. Definite about the absence: no doubt about the absence
	Loss of disc height (distance between the planes of the end-plates of the vertebrae craniad and caudad to the disc) at this disc level	1. Yes 2. No
	Signal intensity of nucleus pulposus on T2 images at this level	1. Hypointensity 2. Normal 3. Hyperintensity
	Vertebral endplate signal changes upper endplate	1. No VESC 2. VESC type I: hypointense in T1-weighted sequences and hyperintense in T2-weighted sequences 3. VESC type II: hyperintense both in T1- and T2-weighted sequences 4. VESC type III: hypointense both in T1- and T2-weighted sequences 5. Mixed VESC type I/II 6. Mixed VESC type II/III
	Vertebral endplate signal changes lower endplate	1. No VESC 2. VESC type I 3. VESC type II 4. VESC type III 5. Mixed VESC type I/II 6. Mixed VESC type I/III
	Spinal canal stenosis	1. Yes 2. No
	Absence of epidural fat adjacent to the dural sac or surrounding the nerve root sheath	1. Yes, completely disappeared 2. Yes, partly disappeared 3. No disappearance
	Place of absence of epidural fat adjacent to the dural sac or surrounding the nerve root sheath	1. 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 2. Foraminal zone: zone between planes passing through the medial and lateral edges of the pedicles 3. Extra-foraminal zone: the zone beyond the sagittal plane of the lateral edges of the pedicles, having no well-defined lateral border

Table 1 (Continued)

MRI variable	Type	Categories
	Presence of impaired discs on other disc levels	1. Yes: presence of disc extension(s) beyond the normal margins of the intervertebral disc space at other disc levels 2. No: absence of disc extension(s) beyond the normal margins of the intervertebral disc space at other disc levels
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
	Location on sagittal view of this disc herniation	1. Disc level: herniated disc between the end-plates of the vertebrae craniad and caudad to the disc 2. Folded upwards: disc tissue beyond the end-plate of the vertebrae craniad to the disc 3. Folded downwards: disc tissue beyond the end-plate of the vertebrae caudad to the disc
	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
	Morphology	1. Protrusion: localized displacement of disc material beyond the intervertebral disc space, with the base against the disc of origin broader than any other imension of the protrusion 2. Extrusion: localized displacement of disc material beyond the intervertebral disc space, with the base agains 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 likelihood > 50% 3. Possible about the presence: reason to consider but likelihood < 50% 4. Definitely no nerve root compression
	If nerve root compression present, which nerve root is affected	1. L3 2. L4 3. L5 4. S1 5. Not applicable, definitely no nerve root compression
	Side nerve root compression	1. Right 2. Left

Table 1 (Continued)		
MRI variable	**Type**	**Categories**
	Nerve root thickness distal to the site of compression	1. Normal 2. Thickened 3. Narrowed
	Flattening of the ventrolateral angle of the dural sac or the emerging root sheath	1. Yes 2. No

STATISTICAL ANALYSIS

To assess the intra- and inter-observer reliability, we used percentages of absolute agreement and kappa coefficients. Percentage of absolute agreement equals the number of cases for which the observers fully agree, proportional to the total number of cases.[16] A common interpretation of good agreement is 80%.[17] However, the absolute percentage of agreement is inadequate, because it does not discriminate between actual agreement and agreement which arises due to chance.[18] A measure which attempts to correct for this is the kappa statistic.[19] In case of ordered data, we calculated weighted kappa scores which is based on the idea that in any ordered scale some possible disagreements are more serious than others.

The kappa statistic is affected by the prevalence of the events[20,21] so that findings with very high or low prevalence lead to very low kappa values, even if the observer agreement is high.[22] Therefore, for both the intra- and inter-observer reliability we only calculated kappa values for findings reported in more than 10% and less than 90% of all reports.[23]

Both weighted and unweighted kappa statistics were computed for all possible pairings of observers. In addition we computed overall unweighted kappa coefficients for multiple raters. When the number of raters is two, the kappa statistic is based on the observed proportion of agreement and the expected proportion of agreement. When there are more than 2 raters, STATA (the program used for all analyses, version 12,0) implemented formulas in its statistical package that can be found in the statistical book of Fleiss and co-authors.[24] While no absolute definitions have been accepted for the interpretation of kappa values, we used guidelines proposed by Landis and Koch for interpretation.[25] Values of less than 0.00 indicated poor; 0.00-0.20 slight; 0.21-0.40 fair; 0.41-0.60 moderate; 0.61-0.80 substantial; and 0.81-1.00 excellent or almost perfect agreement. Value of 0.21-0.60 indicates fair to moderate agreement and a value of 0.41-0.80 indicates moderate to substantial agreement.

In a subanalysis we calculated interobserver agreement when the probability of disc herniation or nerve root compression were dichotomized into "probability> 50%" on one hand and "probability < 50%" on the other hand. In a subanalysis we also calculated interobserver agreement in the patients who were not randomized.

RESULTS

Of the 599 patients screened for the study, 395 patients considered eligible for inclusion underwent MRI of whom 283 patients were randomized and 112 not (Figure 1). Reasons why 112 patients were not randomized was that 70 (63%) did not have a disc herniation according to the neurologist who assessed the MRI in one of the 9 participating centers at the time of enrollment (a visible disc herniation on MRI was a prerequisite to enter the Trial), 31 (28%) patients recovered before the randomization procedure could take place, and 11 (10%) patients refused to be randomized. In total, 283 baseline MRIs of the 283 randomized patients and 106 MRIs of the 112 non-randomized patients could be retrieved, bringing the total to 389 MRIs for the interagreement analysis between the MRI observers of the present study (2 neuroradiologists and one neurosurgeon, all 3 observers did not have participated in the study before).

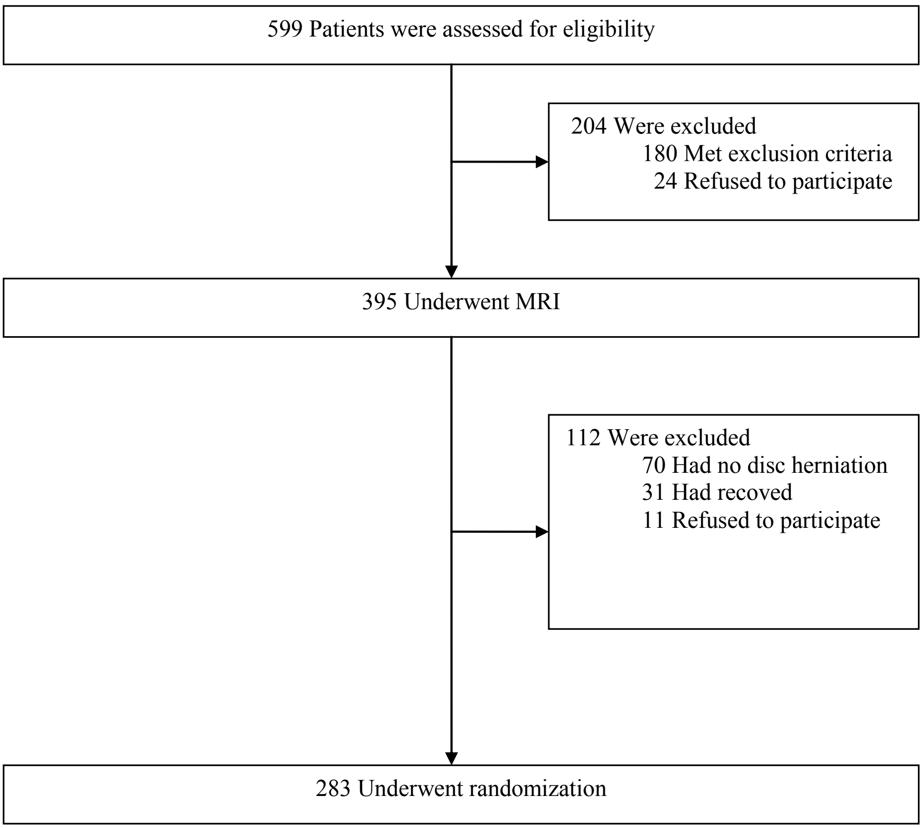


Fig. 1 Flowchart

Table 2 Summary of the interpretation of 389 MRI images.

	Reader A	Reader B	Reader C
Probability of disc herniation			
Definite: no doubt about the presence of disc herniation	299 (76.9)	298 (76.6)	240 (61.7)
Probable: some doubt but probability > 50%	38 (9.8)	28 (7.2)	67 (17.2)
Possible: reason to consider, but probability < 50%	8 (2.1)	4 (1.0)	16 (4.1)
Definitely no disc herniation present	44 (11.3)	59 (15.2)	66 (17.0)
Probability of nerve root compression			
Definite: no doubt about the presence of nerve root compression	222 (57.1)	277 (71.2)	144 (37.0)
Probable: some doubt but likelihood > 50%	97 (24.9)	43 (11.1)	120 (30.8)
Possible: reason to consider, but likelihood < 50%	42 (10.8)	32 (8.2)	64 (16.5)
Definitely no nerve root compression present	28 (7.2)	37 (9.5)	61 (15.7)

Reader A en B represent the two neuroradiologists, while reader C represents the neurosurgeon.
Values are n (%).

The study population had a mean age of 43.2 years with the majority being men (63%). Of the 389 MRIs, there was a definite or probable disc herniation present in 87% of the MRIs according to reader A, in 84% according to reader B and in 79% according to reader C (neurosurgeon) (Table 2).

The interobserver agreement was excellent for the disc level that was assumed to cause the sciatic symptoms of the patient (Table 3). Excellent agreement was also found on the question which nerve root was affected most. With use of a four point scale, interobserver agreement was moderate to substantial for the probability of disc herniation (kappa range 0.57-0.77). When dichotomizing the answers into "probability of disc herniation > 50%" on one hand and "probability of disc herniation < 50%" on the other hand, interobserver agreement was substantial (kappa range 0.67-0.75). With this dichotomized scale all three observers agreed in 88% of the MRIs whether the probability of disc herniation was above or below 50%. With use of a four point scale, interobserver agreement regarding the probability of nerve root compression was moderate to substantial (kappa range 0.42-0.69). In 50 percent of the evaluated MRIs the three observers disagreed on the probability of nerve root compression. The greatest source of reader discrepancy was between the category "definite about the presence" and "probable about the presence", accounting for 58% of all disagreements across all reading pairs. When dichotomizing the answers into "probability of nerve root compression > 50%" on one hand and "probability of nerve root compression < 50%" on the other hand, interobserver agreement among the three readers was substantial (kappa range 0.60-0.80). With this dichotomized scale all three observers agreed in 82% of the MRIs whether the probability of nerve root compression was above or below 50%. In the subgroup consisting of patients who were not randomized, interobserver agreement regarding the probability of nerve root compression was lower than in the total group (Table 4). When dichotomizing the answers into "probability of nerve root compression > 50%" and "probability of nerve root compression < 50%" interob-

Table 3 Agreement among the readers.

	A vs B		A vs C		B vs C		All observers	
	% agreement	kappa	% agreement	kappa	% agreement	kappa	% agreement	multirater kappa
Disc level that is assumed to cause the lumbosacral radicular syndrome ¶	92.0	0.86	88.4	0.81	90.5	0.84	86.4	0.84
Most affected nerve root (including side)	91.0	0.89	88.7	0.86	89.7	0.88	86.1	0.88
Probability of disc herniation (4 categories)ò	88.2	0.77	78.7	0.67	75.6	0.61	72.8	0.57
Probability of disc herniation (2 categories)‡	93.6	0.75	91.8	0.71	90.0	0.67	87.7	0.71
Probability of nerve root compression (4 categories)ò	75.1	0.69	59.9	0.56	57.1	0.51	49.9	0.42
Probability of nerve root compression (2 categories)‡	94.1	0.80	85.4	0.62	84.6	0.60	82.0	0.66

A en B represent the two neuroradiologists, while C represents the neurosurgeon. Analysis with the total number of patients (n=389).
¶ The 5 categories were: 1) L2L3 2) L3L4 3) L4L5 4) L5S1 5) Not applicable, all disc levels have a normal disc contour (no disc extension beyond the normal margins of the intervertebral disc space at any lumbar disc level).
ò The 4 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 >50% 3) "Possible about the presence" if there was reason to consider but the probability was < 50%, and 4) "Definite about the absence" if there was no doubt about the absence.
‡ The categories "Definite and probable about the presence" were combined to one category and the categories "possible about the presence" and "definite about the absence" were also combined to one category.

server agreement was moderate to substantial (kappa range 0.45-0.69). Agreement between the neuroradiologists was higher compared to the agreement between the neurosurgeon and the neuroradiologists.

The interobserver agreement was moderate to substantial for the signal intensity on T2 images; moderate for absence of epidural fat and flattening of the dural sac or the emerging root sheath; and slight for spinal canal stenosis (Table 5). When disc contour was dichotomized into "bulging" and "consideration of herniated disc" absolute agreement among the three observers was 95%.

The interobserver agreement was excellent for side of the disc herniation and location on axial view; and moderate for location on sagittal view, size of disc herniation in relation to spinal canal and disc morphology (Table 6).

Table 4 Agreement among the readers.

	A vs B		A vs C		B vs C		All observers	
	% agree-ment	kappa	% agree-ment	kappa	% agree-ment	kappa	% agree-ment	multi-rater kappa
Disc level that is assumed to cause the lumbosacral radicular syndrome ¶	78.3	0.68	61.3	0.47	70.8	0.59	58.5	0.57
Most affected nerve root (including side)	72.6	0.67	66.0	0.58	69.8	0.61	59.4	0.62
Probability of disc herniation (4 categories)ò	81.1	0.77	69.8	0.61	73.6	0.63	66.0	0.58
Probability of disc herniation (2 categories)‡	87.7	0.75	78.3	0.59	81.1	0.64	73.6	0.65
Probability of nerve root compression (4 categories)ò	61.3	0.65	42.5	0.43	48.1	0.42	36.8	0.32
Probability of nerve root compression (2 categories)‡	84.9	0.69	72.6	0.48	70.8	0.45	64.2	0.52

A en B represent the two neuroradiologists, while C represents the neurosurgeon. Sub analysis of the patients who did not undergo randomization (n=106).
¶ The 5 categories were: 1) L2L3 2) L3L4 3) L4L5 4) L5S1 5) Not applicable, all disc levels have a normal disc contour: no disc extension beyond the normal margins of the intervertebral disc space at any disc level.
ò The 4 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.
‡ The categories "Definite and probable about the presence" were combined to one category and the categories "possible about the presence" and "definite about the absence" were also combined to one category.

Intraobserver agreement regarding the probability of disc herniation and nerve root compression was higher among the neuroradiologists as compared to the neurosurgeon (Table 7). With use of a dichotomized scale absolute intraobserver agreement regarding nerve root compression ranged from 85 to 98%. Intraobserver agreement was substantial for spinal canal stenosis (kappa range 0.61-0.69); moderate to substantial for type of vertebral endplate signal changes (kappa range 0.52-0.74); fair to moderate for loss of disc height (kappa range 0.32-0.48) and flattening of the ventrolateral angle of the dural sac or the emerging root sheath (kappa range 0.30-0.52). Intraobserver agreement regarding the size and morphology of the herniated disc was fair to moderate (for size of the herniated disc kappa range 0.28-0.54, for morphology [extrusion versus protrusion] of the herniated disc kappa range 0.29-0.51).

Table 5 Interobserver agreement regarding characteristics of the impaired disc level.								
	A vs B (n=343)		A vs C (n=329)		B vs C (n=327)		All observers (n=321)	
	% agree-ment	kappa	% agree-ment	kappa	% Agree-ment	kappa	% agree-ment	multi-rater kappa
Disc contour ‡	95.9	*	98.2	*	95.1	*	95.0	*
Loss of disc height ò	97.9	0.86	72.2	0.26	72.4	0.26	71.5	0.31
Signal intensity of nucleus pulposus on T2 images ¶	95.3	0.75	90.4	0.64	90.7	0.57	88.6	0.61
Type of vertebral endplate signal changes upper endplate ‖	75.8	*	83.4	*	84.5	*	72.6	*
Type of vertebral endplate signal changes lower endplate ‖	81.1	*	83.7	*	84.8	*	75.4	*
Spinal canal stenosis ò	63.3	0.21	57.4	0.10	91.3	**	55.1	0.08
Absence of epidural fat adjacent to the dural sac or surrounding the nerve root sheath Ψ	74.0	0.52	74.1	0.54	73.6	0.54	61.7	0.50
Place of absence of epidural fat §	94.4	0.70	96.5	0.72	96.7	0.75	95.3	0.75
Impaired discs on other disc levels ò	93.2	0.79	85.5	0.62	85.4	0.62	82.3	0.68
Nerve root thickness distal to the site of compression ╞	93.5	***	93.5	***	97.5	***	92.1	0.40
Flattening of the ventrolateral angle of the dural sac or the emerging root sheath ò	84.3	0.60	78.7	0.51	78.3	0.46	70.9	0.50

The number between brackets on the first row is the number of patients of which the observers suggested the same disc level as the symptomatic disc level. A en B represent the two neuroradiologists, while C represents the neurosurgeon.
‡ Categories were: bulging disc versus disc herniation.
ò Categories were: yes versus no.
¶ Categories were: 1) Hypointensity 2) Normal 3) Hyperintensity.
‖ Categories were: 1) No vertebral endplate signal changes (VESC) 2) VESC type I 3) VESC type II 4) VESC type III 5) Mixed VESC type I/II 6) Mixed VESC type II/III.
Ψ Categories were: 1) Yes, completely disappeared 2) Yes, partly disappeared 3) No disappearance.
§ Categories were: 1) Sub-articular zone 2) Foraminal zone 3) Extra-foraminal zone.
╞ Categories were: 1) Normal 2) Thickened 3) Narrowed.
* Prevalence of findings too low (< 10% of the reports) to calculate kappa values.
** Prevalence of spinal canal stenosis too low (< 10% of the reports) to calculate kappa values.
*** Prevalence of thickened nerve roots too low (< 10% of the reports) to calculate kappa values.

DISCUSSION

This study showed excellent agreement between observers on the affected disc level (kappa range 0.81-0.86) and the nerve root (kappa range 0.86-0.89) that most likely caused sciatica in patients who were potential candidates for lumbar disc surgery based on clinical grounds. Among the three readers we found also substantial inter- and intra-observer agreement regard-

Table 6 Interobserver agreement regarding characteristics of the disc herniation.

	A vs B (n=314)		A vs C (n=313)		B vs C (n=301)		All observers (n=296)	
	% agree-ment	kappa	% agree-ment	kappa	% agree-ment	kappa	% agree-ment	kappa
Side of disc herniation ╞	98.1	0.96	98.4	0.97	98.0	0.96	97.6	0.97
Location axial view ¶	94.2	0.88	95.5	0.90	96.7	0.93	95.6	0.92
Location sagittal view ‖	73.2	0.55	76.9	0.63	71.3	0.53	61.4	0.56
Size disc herniation in relation to spinal canal (4 categories) §	56.6	0.46	60.6	0.46	64.3	0.50	42.7	0.36
Size disc herniation in relation to spinal canal (2 categories) ‡	82.1	0.55	76.3	0.35	86.3	0.47	71.5	0.44
Protrusion versus extrusion	77.4	0.48	75.0	0.50	73.7	0.44	63.2	0.46

The number between brackets on the first row is the number of patients of which the observers suggested the presence of a disc herniation (on the same disc level). A en B represent the two neuroradiologists, while C represents the neurosurgeon.
╞ Categories were: 1) Right 2) Left 3) Right and left.
¶ Categories were: 1) Central zone 2) Sub-articular zone 3) Foraminal zone 4) Extra-foraminal zone.
‖ Categories were: 1) Disc level 2) Folded upwards 3) Folded downwards.
§ Categories were: 1) Large stenosing: size >75% of the spinal canal 2) Large: size 50-75% of the spinal canal 3) Average: size 25-50% of the spinal canal and 4) Small: size <25% of the spinal canal.
‡ The categories "large stenosing" and "large" were combined to one category and the categories "average" and "small" were also combined to one category.

Table 7 Intraobserver agreement among the three readers based on 40 MRI's.

	Reader A		Reader B		Reader C	
	% agree-ment	kappa	% agree-ment	kappa	% agree-ment	kappa
Level that is assumed to cause the lumbosacral radicular syndrome ¶	97.5	*	90.0	*	87.5	*
Most affected nerve root	90.0	*	82.5	*	80.0	*
Probability of disc herniation (4 categories) ò	95.0	*	92.5	*	70.0	*
Probability of disc herniation (2 categories) ‡	100.0	*	95.0	*	77.5	*
Probability of nerve root compression (4 categories) ò	82.5	*	90.0	*	55.0	*
Probability of nerve root compression (2 categories) ‡	97.5	*	97.5	*	85.0	0.55
Characteristics of the impaired disc level						
Disc contour (consideration of disc herniation vs bulging) ‖	100.0	*	97.2	*	100.0	*
Loss of disc height§	84.6	0.42	77.8	0.32	74.3	0.48
Signal intensity of nucleus pulposus on T2 images Ψ	89.7	0.61	80.6	*	85.7	0.37

Table 7 Intraobserver agreement among the three readers based on 40 MRI's. (Continued)						
	Reader A		Reader B		Reader C	
	% agree-ment	kappa	% agree-ment	kappa	% agree-ment	kappa
Type of vertebral endplate signal changes upper endplate ╞	87.2	0.72	94.4	*	88.6	0.74
Type of vertebral endplate signal changes lower endplate ╞	84.6	0.64	94.4	*	80.0	0.52
Spinal canal stenosis §	84.6	0.69	88.9	0.61	94.3	*
Absence of epidural fat adjacent to the dural sac or surrounding the nerve root sheath ├	84.6	*	69.4	*	77.1	*
Place of absence of epidural fat adjacent to the dural sac or surrounding the nerve root sheath ζ	89.5	*	94.3	*	88.6	*
Impaired discs on other disc levels §	89.7	0.66	94.4	0.82	85.7	0.66
Nerve root thickness distal to the site of compression ╫	82.1	*	97.2	*	88.6	*
Flattening of the ventrolateral angle of the dural sac or the emerging nerve root sheath §	79.5	0.51	83.3	0.52	71.4	0.30
Characteristics the disc herniation						
Side of disc herniation	100.0	1.00	94.3	0.89	100.0	1.00
Location axial view Ω	92.3	*	82.9	*	85.7	*
Location sagittal view Θ	87.2	0.81	82.9	0.71	71.4	0.56
Size disc herniation (4 categories) Ϋ	61.5	0.56	57.1	*	65.7	*
Size disc herniation in relation to spinal canal (2 categories) χ	76.9	0.54	74.3	0.28	85.7	0.37
Protrusion versus extrusion	76.9	0.51	82.9	*	68.6	0.29

Reader A en B represent the two neuroradiologists, while reader C represents the neurosurgeon.
* Since kappa values are afected by the prevalence of events, kappa values were only calculated for findings reported in more than 10% and less than 90% of all reports.
¶ The 5 categories were: 1) L2L3 2) L3L4 3) L4L5 4) L5S1 5) Not applicable, all disc levels have a normal disc contour: no disc extension beyond the normal margins of the intervertebral disc space
ò The 4 categories were: 1) Definite about the presence 2) Probable about the presence 3) Possible about the presence 4) Definite about the absence.
‡ The categories "Definite and probable about the presence" were combined and the categories "possible about the presence" and "definite about the absence" were combined to one category.
║ Categories were: bulging disc versus disc herniation.
§ Categories were: yes versus no.
Ψ Categories were: 1) Hypointensity 2) Normal 3) Hyperintensity.
╞ Categories were: 1) No vertebral endplate signal changes (VESC) 2) VESC type I 3) VESC type II 4) VESC type III 5) Mixed VESC type I/II 6) Mixed VESC type II/III.
├ Categories were: 1) Yes, completely disappeared 2) Yes, partly disappeared 3) No disappearance.
ζ Categories were: 1) Sub-articular zone 2) Foraminal zone 3) Extra-foraminal zone.
╫ Categories were: 1) Normal 2) Thickened 3) Narrowed.
Ω Categories were: 1) Central zone 2) Sub-articular zone 3) Foraminal zone 4) Extra-foraminal zone.
Θ Categories were: 1) Disc level 2) Folded upwards 3) Folded downwards.
Ϋ Categories were: 1) Large stenosing: size >75% of the spinal canal 2) Large: size 50-75% of the spinal canal 3) Average: size 25-50% of the spinal canal and 4) Small: size <25% of the spinal canal.
χ The categories "large stenosing" and "large" were combined to one category and the categories "average" and "small" were also combined to one category.

ing the presence of disc herniation and nerve root compression when the four-point scale was dichotomized into "probability above 50%" and "probability lower than 50%". Therefore, observer variation of MRI interpretation in potential candidates for lumbar disc surgery is satisfactory among spine experts regarding the characteristics most important in the decision for surgery. However, generally moderate agreement was found regarding the characteristics of the impaired disc level and the herniated disc. The moderate agreements may pose a problem when studying the relationships between specific imaging criteria and patient outcome.

Besides herniated discs, the direct evaluation of nerve roots and spinal canal by MRI has been considered an important asset to facilitate decision making in patients with leg and/or back pain.[26,27,28] Unfortunately, no universally accepted imaging criteria exist to define nerve root compression and lumbar spinal stenosis with MRI.[6] The interreader agreement regarding the presence of nerve root compression varies widely between studies. Cihangiroglu and co-authors found fair to substantial agreement (kappa= 0.30-0.63) between two neuroradiologists for classifying nerve root compression, which was dichotomized as absent or present, in 95 patients with low back or radicular pain.[6] Fair to moderate agreement was found for spinal canal stenosis. Van Rijn and co-authors found substantial agreement between two neuroradiologists when evaluating nerve root compression in 59 patients (kappa=0.77).[29] Their kappa is comparable with the agreement between the neuroradiologists in the present study (kappa = 0.80). Sorensen et al. found substantial agreement among two radiologists for classifying disc morphology of herniation (kappa= 0.68) in 50 low-field MRI scans.[30] Jarvik et al. evaluated imaging data from 34 patients with back pain.[31] Agreement between three radiologists for disc morphology was moderate to substantial with weighted kappa values of 0.50 to 0.75 across reader pairs. Interobserver agreement regarding the size and location of the disc herniation has been poorly investigated in previous studies. Characteristics of the disc level of the disc herniation (like signal intensity of the nucleus pulposus, loss of disc height, absence of epidural fat adjacent to the dural sac or surrounding the nerve root sheath, flattening of the dural sac or the emerging root sheath, and nerve root thickness distal to the site of compression) have also been poorly investigated in previous studies

Our results indicate that the assessment of many variables is fairly subjective. However, it is crucial that radiologists and clinicians strive to reduce variability in interpretations as inconsistency in MRI interpretation may lead to alternative treatment options between clinicians and therefore may potentially impact the outcome of patient treatment.[32,33] Previous studies reported that MRI findings play an important role in the decision for surgery.[34,35,36] Carlisle et al. observed that sciatica patients who underwent surgery had larger disc herniations and smaller spinal canals compared to nonoperative patients.[34] Cheng et al. observed that patients with either severe disc herniation or severe spinal stenosis were more likely to be classified as surgical candidates compared to those with mild to moderate findings.[36] Caragee and Kim also observed that patients who underwent surgery had larger disc herniations and smaller sizes of the remaing spinal canal compared to patients who underwent conservative treatment.[35]

Besides that good reliability of imaging data in degenerative disc disease is important from a clinical point of view, it is also important for research purposes attempting to uncover the relationship between specific imaging characteristics and patient outcomes, which unfortunately remains controversial, with several studies showing a high prevalence of disc herniations in persons without any symptoms.[37,38] To gain more insight in the relationship between MRI findings and patient outcomes, those interpreting the images must reliably assess the finding. One reason that a prediction model might lose its predictive power is the incorrect assessment of MRI findings, which causes the inputs in the prediction model to be faulty.[39]

Within the literature, values of agreement on disc degeneration show a high variation depending on the variable investigated.[40] Although a few nomenclatures have been proposed, none has been widely recognized as authoritative or has been widely used in practice. This absence of consensus is greatly related to the multiple controversial aspects of disc abnormalities.[41] As a first step in the attempt to achieve better agreements between observers the language for image interpretation for degenerative disc disease has to be defined. Radiologists and clinicians should strive to define a nomenclature which has the best support among clinicians and radiologists. However, despite the adherence to predefined definitions in the present study, the MRI observers sill only reached moderate agreements regarding many characteristics of the disc level and the herniated disc, which indicate that definitions and the adherence to a well defined nomenclature only is probably not sufficient for reaching substantial to excellent agreements among observers. In addition to defining the language for image interpretation for degenerative disc disease, reading training might be an important next step.[39,42] In support are the results of two reliability studies of The Spine Patient Outcomes Research Trial.[3,5] In one of the two studies the reported agreement on disc morphology was only fair (kappa= 0.24) between the clinicians and radiologists.[5] In another study inter-reader reliability for disc morphology was excellent (kappa= 0.81) between 3 radiologists and 1 orthopedic surgeon.[3] The observation of a much better agreement in the second study might be explained by a better training of the MRI assessors as in that study the MRI assessors, before beginning the study, first evaluated a sample set of images with use of definitions and afterwards they met in person to review each image, enabling them to better streamline the way of interpreting the images.

When comparing kappa coefficients between studies caution should be exercised since there are other factors that can influence the magnitude of the coefficient, especially the number of categories and the prevalence of findings.[43] When the prevalence of findings is very low or high, kappa values also decline, even when the observed agreement remains unchanged.[20,23] However, kappa remains the best available method to measure intra- and inter-observer agreement, in addition to that explained by chance.[23]

We deliberately did not organize an extra meeting in which a sample subset of images was evaluated as the discussion during this meeting might have caused the observers to adjust their diagnostic imaging criteria. This may have led to an overestimation in the interpretation among the three readers compared to the situation as it existed before undertaking the meeting. During the meeting prior to the readings no images were evaluated, only a review of the questions and answers used in the case record forms to assure every reader understands their intended meaning when evaluating the images. If one does not undertake such a meeting this may pose problems when interpreting results as it may well be that a possible low observer agreement may not reflect true low agreement but agreement which arises due to the readers giving a different meaning to the questions or answers. We do not think such a meeting has a similar effect as evaluating together images before beginning the readings as then some observers may adjust their diagnostic criteria according to how other observers are evaluating the images during the meeting, with the consequence that one is not measuring the observer agreement as it existed before undertaking the meeting. Both procedures might lead to improving kappa coefficients, although more negative effects may arise when evaluating images together prior to the readings compared to only reviewing the questions and answers.

Our study has several limitations. An important limitation of the study is the number of observers, in particular the inclusion of only one non-radiologist, which limits the statistical power of the observer variation. Although all analyses were also conducted pairwise, the analyses in which all three observers are included should be carefully interpreted in light of the low statistical power. The inclusion of more observers having the same background, especially the inclusion of one more neurosurgeon in this study, would have strengthened the findings. The concordance found in this study may also 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.[22] Interestingly, however, the agreement between the neuroradiologists was sometimes lower compared to that of the reading pairs containing one of the two neuroradiologists and the neurosurgeon. 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. This might also explain why the observed agreement was lower among the patients who finally were not randomized.[44] However, as the presence of the disc herniations and nerve root compression was defined in different chance categories, the influence on the inter-reader reliability might have been limited. In addition, the use of standardized reporting forms with definitions and multiple choice categories allowed the assessments to be structured far more than possible in general clinical practice which also may have caused an overestimation.[3] Finally, usual reliable statistical packages (STATA, SAS) are only able to calculate unweighted kappa coefficients for multiple raters. However, unweighted kappa coefficients are inappropriate for ordinal scales since they treat all disagreements equally.[43] We encourage the development of statistical software that will solve this problem.

CONCLUSIONS

The observer variation of MRI interpretation in potential candidates for lumbar disc surgery is satisfactory among spine experts with regard to clinically relevant parameters like most affected disc level and nerve root, probability of disc herniation and nerve root compression. However, in general considerable variation between the observers was found regarding specific characteristics of the symptomatic disc level and herniated disc. Therefore, it would be valuable to improve the reliability of image interpretation to subsequently increase our knowledge regarding the etiology, treatment and prevention of back pain and sciatica.

REFERENCES

1. Koes BW, van Tulder MW, Peul WC (2007) Diagnosis and treatment of sciatica. BMJ 334: 1313-1317.
2. Bejia I, Younes M, Zrour S, Touzi M, Bergaoui N (2004) Factors predicting outcomes of mechanical sciatica: a review of 1092 cases. Joint Bone Spine 71: 567-571.
3. Lurie JD, Tosteson AN, Tosteson TD, Carragee E, Carrino JA, et al. (2008) Reliability of magnetic resonance imaging readings for lumbar disc herniation in the Spine Patient Outcomes Research Trial (SPORT). Spine (Phila Pa 1976) 33: 991-998.
4. (1994) Practice parameters: magnetic resonance imaging in the evaluation of low back syndrome (summary statement). Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 44: 767-770.
5. Lurie JD, Doman DM, Spratt KF, Tosteson AN, Weinstein JN (2009) Magnetic resonance imaging interpretation in patients with symptomatic lumbar spine disc herniations: comparison of clinician and radiologist readings. Spine (Phila Pa 1976) 34: 701-705.
6. Cihangiroglu M, Yildirim H, Bozgeyik Z, Senol U, Ozdemir H, et al. (2004) Observer variability based on the strength of MR scanners in the assessment of lumbar degenerative disc disease. Eur J Radiol 51: 202-208.
7. 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: 1168-1171.
8. Korres DS, Loupassis G, Stamos K (1992) Results of lumbar discectomy: a study using 15 different evaluation methods. Eur Spine J 1: 20-24.
9. 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: 1355-1358.
10. Vucetic N, Astrand P, Guntner P, Svensson O (1999) Diagnosis and prognosis in lumbar disc herniation. Clin Orthop Relat Res: 116-122.
11. Peul WC, van Houwelingen HC, van den Hout WB, Brand R, Eekhof JA, et al. (2007) Surgery versus prolonged conservative treatment for sciatica. N Engl J Med 356: 2245-2256.
12. Peul WC, van Houwelingen HC, van der Hout WB, Brand R, Eekhof JA, 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.
13. Fardon DF, Milette PC (2001) 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) 26: E93-E113.
14. Modic MT, Masaryk TJ, Ross JS, Carter JR (1988) Imaging of degenerative disk disease. Radiology 168: 177-186.
15. Modic MT, Steinberg PM, Ross JS, Masaryk TJ, Carter JR (1988) Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging. Radiology 166: 193-199.
16. Lynn MR (1986) Determination and quantification of content validity. Nurs Res 35: 382-385.
17. Brouwer S, Reneman MF, Dijkstra PU, Groothoff JW, Schellekens JM, et al. (2003) Test-retest reliability of the Isernhagen Work Systems Functional Capacity Evaluation in patients with chronic low back pain. J Occup Rehabil 13: 207-218.
18. Brennan P, Silman A (1992) Statistical methods for assessing observer variability in clinical measures. BMJ 304: 1491-1494.
19. Cohen J (1960) A coefficient of agreement for nominal scales. Educ Psychol Meas 20:37–46.

20. Feinstein AR, Cicchetti DV (1990) High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 43: 543-549.
21. Gjorup T (1988) The kappa coefficient and the prevalence of a diagnosis. Methods Inf Med 27: 184-186.
22. Kovacs FM, Royuela A, Jensen TS, Estremera A, Amengual G, et al. (2009) Agreement in the interpretation of magnetic resonance images of the lumbar spine. Acta Radiol 50: 497-506.
23. Arana E, Royuela A, Kovacs FM, Estremera A, Sarasibar H, 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: 809-817.
24. Fleiss JL, Levin B, Paik MC (2003) Statistical Methods for Rates and Proportions. New York: Wiley: Chap 18.
25. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33: 159-174.
26. 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: 1158-1166.
27. Boos N, Rieder R, Schade V, Spratt KF, Semmer N, et al. (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: 2613-2625.
28. van Rijn JC, Klemetso N, Reitsma JB, Bossuyt PM, Hulsmans FJ, et al. (2006) Observer variation in the evaluation of lumbar herniated discs and root compression: spiral CT compared with MRI. Br J Radiol 79: 372-377.
29. van Rijn JC, Klemetso N, Reitsma JB, Majoie CB, Hulsmans FJ, et al. (2005) Observer variation in MRI evaluation of patients suspected of lumbar disk herniation. AJR Am J Roentgenol 184: 299-303.
30. Solgaard Sorensen J, Kjaer P, Jensen ST, Andersen P (2006) Low-field magnetic resonance imaging of the lumbar spine: reliability of qualitative evaluation of disc and muscle parameters. Acta Radiol 47: 947-953.
31. Jarvik JG, Haynor DR, Koepsell TD, Bronstein A, Ashley D, et al. (1996) Interreader reliability for a new classification of lumbar disk disease. Acad Radiol 3: 537-544.
32. 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: 177-184.
33. Ross JS (2010) Babel 2.0. Radiology 254: 640-641.
34. Carlisle E, Luna M, Tsou PM, Wang JC (2005) Percent spinal canal compromise on MRI utilized for predicting the need for surgical treatment in single-level lumbar intervertebral disc herniation. Spine J 5: 608-614.
35. 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: 1650-1660.
36. Cheng F, You J, Rampersaud YR (2010) Relationship between spinal magnetic resonance imaging findings and candidacy for spinal surgery. Can Fam Physician 56: e323-330.
37. Jensen MC, Brant-Zawadzki MN, Obuchowski N, Modic MT, Malkasian D, et al. (1994) Magnetic resonance imaging of the lumbar spine in people without back pain. N Engl J Med 331: 69-73.
38. 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: 403-408.

39. Carrino JA, Lurie JD, Tosteson AN, Tosteson TD, Carragee EJ, et al. (2009) Lumbar spine: reliability of MR imaging findings. Radiology 250: 161-170.
40. Pfirrmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N (2001) Magnetic resonance classification of lumbar intervertebral disc degeneration. Spine (Phila Pa 1976) 26: 1873-1878.
41. Milette PC (2001) Reporting lumbar disk abnormalities: at last, consensus! AJNR Am J Neuroradiol 22: 428-429.
42. Jarvik JG, Deyo RA (2009) Moderate versus mediocre: the reliability of spine MR data interpretations. Radiology 250: 15-17.
43. Sim J, Wright CC (2005) The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 85: 257-268.
44. Kim SW, Yeom JS, Park SK, Chang BS, Lee DH, et al. (2009) Inter- and Intra-observer Reliability of MRI for Lumbar Lateral Disc Herniation. Clin Orthop Surg 1: 34-39.

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.

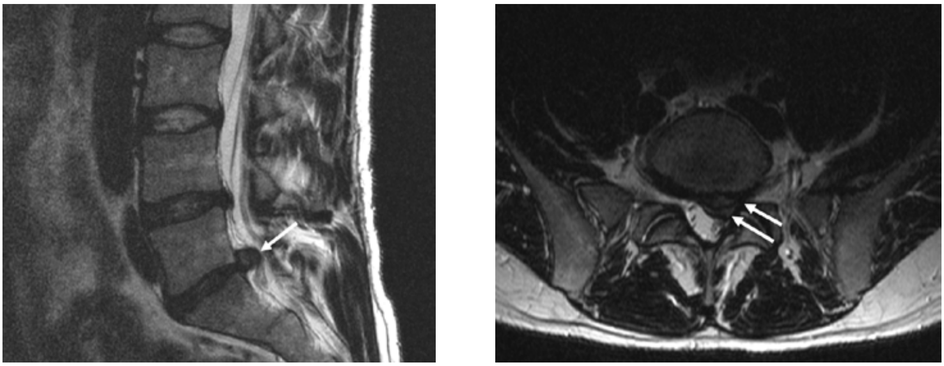
We assumed clinical outcome data to be missing at random and used model-based multiple imputation to impute the outcome values, a method in which the distribution of the observed data is used to construct sets of plausible values for the missing observations (10 imputed datasets). Variables included 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

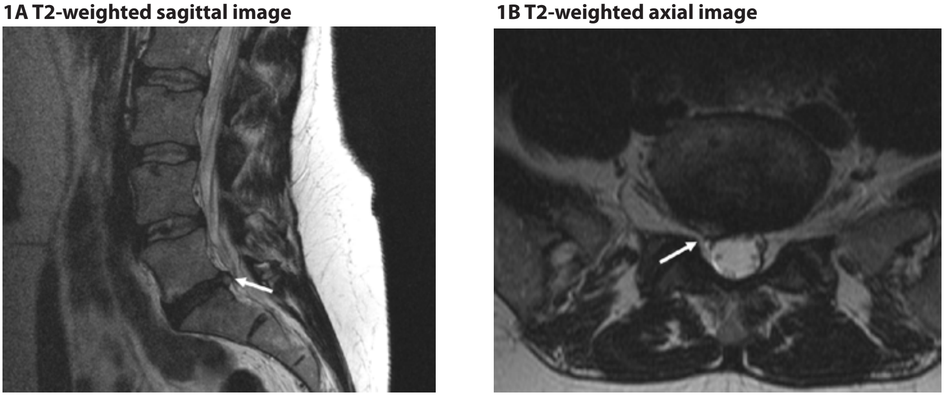
Of the 599 patients screened for the study, 395 patients considered eligible for inclusion underwent MRI of whom 283 patients were randomized.[13] In total, 106 baseline MRIs of the 112 non-randomized patients and 283 MRIs of the 283 randomized patients could be retrieved, bringing the total to 389 MRIs. 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**.**

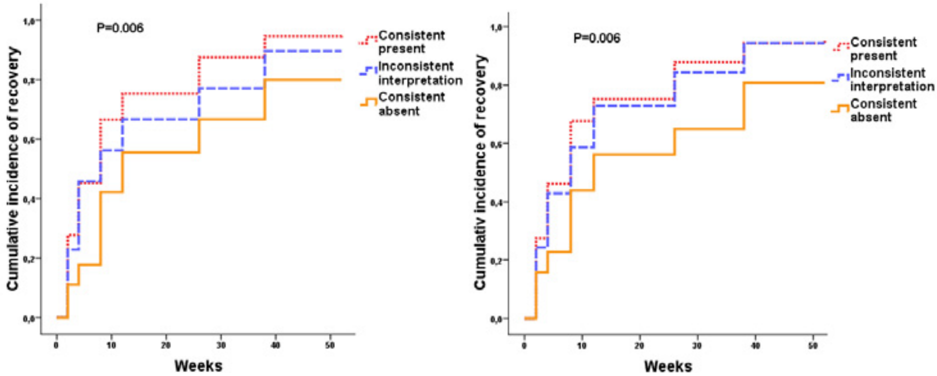
	Patients not randomized				Patients assigned to surgery				Patients assigned to conservative care			
	Consistent present (n=41)	Inconsistent interpretation (n=28)	Consistent absent (n=37)	P Value	Consistent present (n=124)	Inconsistent interpretation (n=13)	Consistent absent (n=4)	P Value	Consistent 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	

Chapter 4

Predictive value of MRI in decision making for disc surgery for sciatica

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

Accepted by Journal of Neurosurgery: Spine

ABSTRACT

OBJECT

In a randomized controlled trial comparing surgery and prolonged conservative treatment for 6-12 weeks sciatica, more than one third of patients assigned to conservative treatment underwent surgery. The objective of this study was to evaluate whether Magnetic resonance imaging (MRI) at baseline could have predicted this delayed surgery.

METHODS

Independently evaluated qualitative and quantitative MRI findings were compared between those who did and those who did not undergo surgery during follow-up in the conservative care group. In addition, area under the receiver operating characteristic (ROC) curve analysis was used to assess how well MRI parameters discriminated those who did and those who did not undergo delayed surgery (0.5-0.7 poor discrimination, ≥0.7 acceptable discrimination).

RESULTS

Of 142 patients assigned to receive prolonged conservative care, 55 (39%) patients received delayed surgery. Of the 55 surgically treated patients 71% had definite nerve root compression at baseline compared to 72% of conservatively treated patients (P=0.76). Large disc herniations (size >50% of spinal canal) were nearly equally distributed between those who did and those who did not undergo surgery (25% vs. 21%, P=0.65). The size of the dural sac was smaller in the surgical compared to the non-surgical group (101.2 vs. 122.9 mm^2, P=0.01). However, the size of the dural sac discriminated poorly between those who did and those who did not undergo delayed surgery (area under ROC curve, 0.62).

CONCLUSION

In patients who suffered from 6 to 12 weeks sciatica MRI at baseline did not distinguish between patients who did and those who did not undergo delayed surgery.

INTRODUCTION

Magnetic resonance imaging (MRI) is widely used in diagnosis and treatment planning of patients with intervertebral disc herniations.[4] It is considered the imaging procedure of choice for patients suspected of lumbar disc herniation[19,29] and is indicated in patients with severe symptoms who fail to respond to conservative care for 6 to 8 weeks.[18] Qualitative MR-findings such as the presence of disc extrusion or severe nerve root compression have indeed been reported to be strongly associated with sciatica.[3] In addition, from MR images the size and shape of disc herniations can be measured accurately, as can the size and proportions of the spinal canal.[5] However, limited data is available concerning the predictive value of both qualitative and quantitative MRI evaluations in assisting clinical decision making for surgical or non-surgical management for sciatica.

The investigators previously reported the results of a randomized controlled trial comparing early surgery with prolonged conservative care for patients with 6 to 12 weeks sciatica over one year's follow-up.[25] Although early surgery achieved more rapid relief of sciatica than conservative care, the clinical outcome results were similar after one year. Despite efforts to the contrary, 39% of the patients assigned to the prolonged conservative treatment group did undergo surgery during the first year after randomization.[24] Reasons for performing delayed surgery were persistent or increasing drug-resistant leg pain and progressive neurological deficit.[24] In a previous study, baseline clinical parameters were tested whether they could have predicted surgery during follow up in this group.[24] Patients with higher pain intensity in the leg or higher disability scores at baseline had a higher risk of undergoing delayed surgery.[24]

The objective of this study was to evaluate the predictive value of qualitative and quantitative MRI assessments for delayed surgery. If early in the course of sciatica specific qualitative and quantitative MRI assessments prove to predict which patients will undergo surgery anyhow during follow-up, this information could be valuable for both patients and physicians as it could enable them to consider early surgery without further delay to reduce the period of suffering.

METHODS

STUDY POPULATION

Patients for this study were participants in the Sciatica Trial: a multicenter randomized controlled trial of patients with 6-12 weeks sciatica. An early surgery strategy was compared to prolonged conservative care for an additional 6 months followed by surgery for patients who did not improve or who did request it earlier because of aggravating symptoms.[25,26] Patients were included only if they had a dermatomal pattern of pain distribution with concomitant neurological disturbances that correlated to the same nerve root being affected on MRI. No minimal disc size was prespecified for entry into the Trial. For the purpose of the present study, the patients who originally were allocated at random to prolonged conservative care were selected as the study cohort. The medical ethics committee at each of the nine participating hospitals approved the protocol. Written informed consent was obtained from all patients. Details of the design and study protocol have been published previously.[25,26]

Treatment

Prolonged conservative treatment was provided by each patient's practitioner. Patients were informed about the favourable prognosis. Prescription of pain medication was allowed and was adjusted according to existing clinical guidelines if necessary. Opiates were frequently prescribed, but no epidural or periradicular corticosteroids were injected. Patients who were fearful of moving were referred to a physiotherapist. Treatment was aimed mainly at resumption of daily activities. However if sciatica was still present at 6 months after randomization, surgery was considered. Persistent or increasing drug-resistant leg pain and progressive neurological deficit were reasons for performing surgery even before 6 months. When patients requested surgery, they were again evaluated by their treating physician and the assigned research nurse, who had to confirm that recovery had not occurred and that the repeated MRI showed an unresolved disc herniation with nerve root compression. Subsequently the neurosurgeon was consulted by the patient and surgery was performed if all the indicators did direct in sciatica resistant to medical treatment.

MRI protocol and Image evaluation

MRI scans were performed in all 9 participating hospitals using standardized protocols tailored to a 1.5 Tesla scanner. Sagittal T1 and axial T1 spin echo images of the lumbar spine were acquired. In addition, T2 weighted sagittal and axial series, and contrast-enhanced (gadolinium) T1 fat suppressed sagittal and axial images were obtained.

Two neuroradiologists (BK and GL) and one neurosurgeon (CV) independently evaluated all MR images. The readers hold senior positions in busy spinal clinics with a focus on advanced spine surgery, and are confronted with spinal MRIs on a daily basis. The readers were not provided any clinical information and have not been involved in the selection or care of the included patients. Definitions of imaging characteristics were based on the 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 disc pathology.[13] Vertebral end plate changes were defined according to criteria of Modic.[20,21] 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 (Table 1).

First, all readers had to choose the disc level with the most severe nerve root compression. At this disc level, a four point scale was used for both the presence of disc herniation and

Table 1 MRI study variables

Disc level	Variable	Category
Disc level with the most severe nerve root compression	Disc level	1. Not applicable: no nerve root compression 2. L2L3 3. L3L4 4. L4L5 5. L5S1
	Disc contour at this level	1. Bulging: presence of disc tissue circumferentially (50-100%) beyond the edges of the ring apophyses 2. Herniation: localized displacement of disc material beyond the normal margins of the intervertebral disc space
	Certainty about the presence of 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
	Loss of disc height at this level	1. Yes 2. No
	Signal intensity of nucleus pulposus on T2 images at this level	1. Hypointensity 2. Normal 3. Hyperintensity
	Certainty about the presence 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 about the absence: no doubt about the absence
	Spinal canal stenosis	1. Yes 2. No
	Disappearance of epidural fat	1. Completely disappeared 2. Partly disappeared 3. No disappearance
	Presence of impaired discs at more than one level	1. Yes 2. No
If a disc herniation is considered	Location	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

Table 1 (Continued)		
Disc level	**Variable**	**Category**
	Side	1. Right 2. Left 3. Right and left
	Size disc herniation in relation to spinal canal	1. Large stenosing, size >75% of the spinal canal 2. Large, size 50-75% of the spinal canal 3. Average, size 25-50% of the spinal canal 4. Small, size <25% of the spinal canal
	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 ranging from definitely present to definitely absent. Clinically relevant characteristics of the disc level and disc herniation were scored.

In addition quantitative measurements were performed by an independent researcher (AB), blinded to the treatment ultimately received and any other clinical information. He was not involved in the clinical treatment of these patients. Scans were examined with attention to the intervertebral disc with the most severe nerve root compression according to the three observers. On T2-weighted axial views the following parameters were quantified (in square millimeters): (i) cross-sectional size of the intervertebral disc prolapse, (ii) basis of the disc herniation, (iii) cross-sectional size of the dural sac, and (iv) cross-sectional size of the spinal canal not occupied by the disc herniation and without ligamentum flavum (Fig. 1). Next two herniation ratio's (HR) were defined: HR 1, which represents the ratio of the size of the herniated disc to the dural sac and HR 2, which represents the ratio of the size of the herniated disc to the remaining size of the spinal canal not occupied by the herniated disc. When a disc herniation was absent no quantitative measurements were performed.

OUTCOME

The occurrence of surgery performed during one-year follow-up was the event of interest. 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),[23] the 100-mm visual-analogue scale (VAS) for leg and back pain (with 0 representing no pain and 100 the worst pain ever experienced),[9] and a 7-point Likert self-rating scale of global perceived recovery given by the question whether the patient experienced recovery, with answers ranging from completely recovered to much worse. Perceived recovery on the 7-point Likert scale for global perceived recovery was used in dichotomized form: "Complete" or "nearly complete disappearance of symptoms" was defined as "perceived recovery", while a score in the remaining five categories was marked as "no recovery".[25,26] Outcome measures were assessed at baseline, 2, 4, 8, 12, 26, 38 and 52 weeks. For the purpose of the present study only outcome data from the baseline measurements and 52 weeks were used.

STATISTICAL ANALYSIS

The majority opinion of the three readers regarding the (qualitative) MRI characteristics (answer independently given by minimum 2 out of 3 readers) was used in the statistical analysis. Patients were categorized in two groups according to the occurrence of surgery performed during the first 12 months after being randomized to prolonged conservative care. Between-group comparisons for both clinical and MRI variables were performed with Student's t-tests for continuous data and Chi-square tests for categorical data. If a variable proved to be significantly different between patients who did and those who did not undergo surgery during follow-up the sensitivity and specificity of this variable was determined by using Receiver operating characteristic (ROC) curve analysis. The area under the ROC curve (AUC) ranges from 0 to 1 and provides a measure of a test's ability to discriminate between those subjects

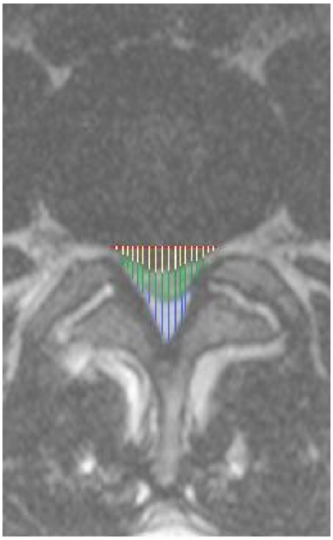


Figure 1 Methods of measuring the different parameters
Red line represents the size of the basis of the disc herniation, the *yellow shaded area* represents the size of the disc herniation, the *green shaded area* represents the size of the dural sac, the *green and blue shaded areas combined* represent the size of the remaining spinal canal.
Herniation ratio 1 (disc herniation in relation to the size of the dural sac) = yellow shaded area/green shaded area.
Herniation ratio 2 (disc herniation in relation to the size of the remaining spinal canal) = yellow shaded area/(green and blue shaded area combined).

who experience the outcome of interest versus those who do not. To derive the AUC value of 2 or more variables combined, these variables were first subjected to a logistic regression model with the occurrence of surgery as the event of interest, and the predicted probability from that model was included in the ROC-curve Analysis.[17] We used the traditional following thresholds for the area under the ROC curve: 0.5 no discrimination; 0.5 to 0.7 poor discrimination; ≥0.7 acceptable discrimination; ≥0.8 excellent discrimination; ≥0.9 outstanding discrimination.[14]

In a subanalysis characteristics were compared between the patients who did not undergo surgery, those who did undergo surgery between 0 and 6 months, and those between 6 and 12 months. The one-way analysis of variance (ANOVA) was used to test for mean differences in continuous data (with Post-hoc analysis using the Bonferroni analysis for the variables which showed a statistically significant difference). A P value of <0.05 was considered statistically significant.

RESULTS

Of 142 patients assigned to receive prolonged conservative care, 55 (39%) patients received surgery after a mean period of 18 weeks (22 [15%] within 3 months, 20 [14%] between 3 and 6 months, 9 [6%] between 6 and 9 months and 4 [3%] after more than 9 months). At baseline, age, gender, duration of sciatica and Body Mass Index and level of the herniated disc were comparable in the "delayed" surgical and non-surgical group. At baseline, in 39 (71%) of 55 surgically treated patients there was no doubt about the presence of nerve root compression compared to 63 (72%) of 87 conservatively treated patients (P=0.76) (Table 2). No significant differences existed in prevalence of Vertebral Endplate Signal Changes between the "delayed" surgical and non-surgical group (29% vs. 40%, P=0.37). Large disc herniations (size >50% of spinal canal) were nearly equally distributed between those who did and those who did not undergo surgery (25% vs 21%, P=0.65). Central or subarticular located disc herniations were also nearly equally distributed between those who did and those who did not undergo surgery (91% vs 90%, P=1.00). Extruded disc herniations were observed in 59% of surgically treated patients compared to 70% of conservatively treated patients (P=0.12).

An example of a patient who had a large disc herniation and definite nerve root compression but who still did not undergo surgery during follow-up is shown in Figure 2.

At baseline, the size of the herniated disc was comparable in the surgical and non-surgical group (76.9 vs. 75.7 mm^2, P=0.86) (Table 3). The size of the dural sac was smaller in the surgical compared to the non-surgical group (101.2 vs. 122.9 mm^2, P=0.01). However, the ratio of the size of the disc herniation to the dural sac was 0.97 for the surgical group compared to a ratio of 0.89 for the non-surgical group (P=0.65). The size of the remaining spinal canal was smaller in the surgical group compared to the non-surgical group (159.4 vs. 189.0 mm^2, P=0.007), although the ratio of the size of the disc herniation to remaining spinal canal was not significantly different between those who did and those who did not undergo surgery (0.57 vs. 0.49, P=0.33).

The mean RDQ score at baseline was higher in the surgical group compared to the non-surgical group (16.9 vs. 13.5, P<0.001). The baseline VAS leg pain was also higher in the surgical group compared to the non-surgical group (63.8 vs. 49.2, P<0.001).

The subanalysis comparing characteristics between patients who did not undergo surgery, those who did undergo surgery between 6 and 12 months, and those between 6 and 12 months,

Table 2 Comparison of baseline characteristics between patients who did and those who did not undergo surgery for sciatica**.** Values are n (%) or means ± SD

	Surgery (n=55)	No surgery (n=87)	P-value
Age at baseline MRI	43.6±10.1	43.2±9.3	0.83
Male gender	39 (71)	58 (67)	0.60
Duration of sciatica in weeks	9.6±2.1	9.5±2.2	0.72
Characteristics of the most impaired disc level			
Disc level			
L3L4 or L4L5	21 (38)	35 (40)	0.81
L5S1	34 (62)	52 (60)	
Presence of disc herniation			0.37
Definite	49 (89)	78 (90)	
Probable	3 (6)	8 (9)	
Possible	1 (2)	0 (0)	
Definite absent	2 (4)	1 (1)	
Presence of nerve root compression			0.76
Definite	39 (71)	63 (72)	
Probable	11 (20)	18 (21)	
Possible	5 (9)	5 (6)	
Definite absent	0 (0)	1 (1)	
Loss of disc height	51 (93)	76 (87)	0.64
Hypo intense signal intensity of nucleus pulposus on T2 images	50 (91)	77 (89)	0.97
Completely disappearance of epidural fat	36 (66)	54 (62)	0.89
Spinal canal stenosis	7 (13)	8 (9)	0.55
Presence of impaired discs at other disc levels	44 (80)	63 (72)	0.49
Characteristics of the herniated disc			
Located on the right side	23 (43)	45 (52)	0.31
Size>50% in relation to spinal canal	13 (25)	18 (21)	0.65
Extrusion	31 (59)	60 (70)	0.12
Central or subarticular located	48 (91)	77 (90)	1.00

Fig. 2A T2-weighted sagittal baseline image

Fig. 2B T2-weighted axial baseline image

Fig. 2C T2-weighted sagittal image after one year follow-up

Fig. 2D T2-weighted axial image after one year follow-up

Figure 2 Sagittal and axial T2 weighted MR images of a patient with sciatica who had a large disc herniation at disc level L4-L5 at baseline, compressing nerve roots L5 bilaterally and narrowing the spinal canal (A and B). This patient did not undergo surgery during the first year and reported complete clinical recovery after one year. Repeated MRI after one year follow-up showed decrease of the herniation at disc level L4-L5 (C and D).

also showed that the three groups only significantly differed in baseline RDQ, VAS-leg pain and size of the dural sac and remaining spinal canal (Table 4).

With surgery as the event of interest, the area under ROC curve for the size of the dural sac was 0.62 (95% Confidence Interval [CI] 0.53-0.72), for the size of the spinal canal 0.62 (95% CI 0.53-0.72), for the VAS of leg pain 0.67 (95% CI 0.58-0.77) and for the RDQ score 0.70 (95% CI 0.61-0.79). Combined the two MRI variables had an area under ROC curve of 0.63 (95% CI 0.53-0.72) compared to 0.72 (95% CI 0.64-0.81) when combining the RDQ and VAS-leg pain. All four variables combined had an area under ROC curve of 0.76 (95% CI 0.68-0.84).

Despite baseline differences, one year after randomization no significant differences were observed between the surgical group and the non-surgical group regarding the clinical outcome scores as assessed by VAS of leg pain, VAS of back pain, RDQ and global perceived recovery (Table 5). One year after randomization a disc herniation was considered (definite, probable or possible present) in 26% of the patients who had undergone surgery compared to 61% of the patients who had undergone non-operative care (P=0.001).

Of the 16 surgical patients who at baseline did not have definite nerve root compression 87.5% reported perceived recovery at one year as compared to 87.2% of the 39 surgical patients who at baseline did have definite nerve root compression (P=0.97).

Table 3 Baseline quantitative MRI measurements and clinical characteristics in the group that underwent surgery and the group that did not undergo surgery for sciatica. Values are n (%) or means ± SD

	Surgery (n=55)	No surgery (n=87)	P-value
Measurements on axial view			
Disc herniation (mm^2)	76.9±37.6	75.7±38.9	0.86
Basis disc herniation (mm)	20.0±5.8	19.3±6.8	0.54
Dural sac (mm^2)	101.2±44.6	122.9±53.8	**0.01**
Remaining spinal canal (mm^2)	159.4±57.0	189.0±65.7	**0.007**
Ratio disc herniation to dural sac	0.97±0.70	0.89±1.15	0.65
Ratio disc herniation to remaining spinal canal	0.57±0.40	0.49±0.52	0.33
Clinical outcomes			
Roland disability score¶	16.9±4.1	13.5±5.0	**<0.001**
Visual-analogue scale of leg pain‡	63.8±23.5	49.2±22.9	**<0.001**
Visual-analogue scale of back pain‡	41.9±31.8	33.4±25.7	0.08

¶ The Roland disability questionnaire for sciatica is a disease-specific disability scale that measures functional status in 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 was indicated on a horizontal 100 mm visual analogue scale, with 0 representing no pain and 100 the worst pain ever experienced.

Table 4 Clinical, qualitative and quantitative MRI evaluations of the two surgical groups and the group that did not undergo surgery for sciatica. Values are n (%) or means ± SD.

	Surgery within 6 months (n=42)	Surgery between 6-12 months (n=13)	No surgery (n=87)	P-value
Clinical outcomes				
Roland disability score	17.4±3.9	15.5±4.6	13.5±5.0	**<0.001** ¶
Visual-analogue scale of leg pain	64.2±24.6	62.6±20.2	49.2±22.9	**0.002** ‡
Visual-analogue scale of back pain	44.2±32.1	34.6±30.8	33.4±25.7	0.13
Characteristics of the most impaired disc level				
Disc level				
L3L4 or L4L5	18 (43)	3 (23)	35 (40)	0.43
L5S1	24 (57)	10 (77)	52 (60)	
Presence of disc herniation				0.43
Definite	37 (88)	12 (92)	78 (90)	
Probable	3 (7)	0 (0)	8 (9)	
Possible	1 (2)	0 (0)	0 (0)	
Definite absent	1 (2)	1 (8)	1 (1)	
Presence of nerve root compression				0.96
Definite	29 (69)	10 (77)	63 (72)	
Probable	9 (21)	2 (15)	18 (21)	
Possible	4 (10)	1 (8)	5 (6)	
Definite absent	0 (0)	0 (0)	1 (1)	
Loss of disc height	40 (95)	11 (85)	76 (87)	0.44
Hypo intense signal intensity of nucleus pulposus on T2 images	39 (93)	11 (85)	77 (89)	0.45
Completely disappearance of epidural fat	27 (64)	9 (69)	54 (62)	0.94
Spinal canal stenosis	5 (12)	2 (15)	8 (9)	0.79
Presence of impaired discs at other disc levels	33 (79)	11 (85)	63 (72)	0.71
Characteristics of the herniated disc				
Located on the right side	19 (46)	4 (33)	45 (52)	0.43
Size>50% in relation to spinal canal	9 (21)	4 (33)	18 (21)	0.64
Extrusion	23 (56)	8 (67)	60 (70)	0.23
Central or subarticular located	36 (88)	12 (100)	77 (91)	0.45

¶ Bonferroni post-hoc analysis showed P<0.001 for the no surgery group compared with the 0-6 months surgical group, P=0.43 for the no surgery group compared with the 6-12 months surgical group, and P=0.64 for the 0-6 months surgical group compared with the 6-12 months surgical group
‡ Bonferroni post-hoc analysis showed P=0.002 for the no surgery group compared with the 0-6 months surgical group, P=0.16 for the no surgery group compared with the 6-12 months surgical group, and P=1.00 for the 0-6 months surgical group compared with the 6-12 months surgical group

Table 5 Clinical and MRI parameters at one year follow-up in the group that underwent surgery and the group that did not undergo surgery for sciatica. Values are n (%) or means ± SD

	Surgery (n=54)	No surgery (n=82)	P-value
Clinical Outcome			
Roland disability score¶	3.2±5.2	3.5±4.9	0.71
VAS leg pain‡	9.5±19.9	10.9±17.0	0.67
VAS back pain‡	13.6±23.2	16.4±20.6	0.47
Perceived recovery one year╫	47 (87)	64 (78)	0.09
Qualitative MRI finings			
Presence of disc herniation			0.001
Definite	6 (11)	22 (27)	
Probable	7 (13)	21 (26)	
Possible	1 (2)	7 (9)	
Definite absent	40 (74)	32 (39)	
Presence of nerve root compression			0.09
Definite	2 (4)	5 (6)	
Probable	4 (7)	6 (7)	
Possible	5 (9)	21 (26)	
Definite absent	43 (80)	50 (61)	
Measurements			
Disc herniation (mm^2)	21.3±30.6	37.3±29.7	0.003
Dural sac (mm^2)	147.7±61.6	141.5±54.4	0.54
Remaining spinal canal (mm^2)	233.1±77.1	211.1±74.7	0.10
Ratio disc herniation to dural sac	0.2±0.4	0.3±0.4	0.07
Ratio disc herniation to remaining spinal canal	0.1±0.2	0.2±0.2	0.02

Clinical outcome data at one year was available for 136 of the 142 patients
¶ The Roland disability questionnaire for sciatica is a disease-specific disability scale that measures functional status in 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 was indicated on a horizontal 100 mm visual analogue scale, with 0 representing no pain and 100 the worst pain ever experienced.
╫ Perceived recovery was defined as complete or nearly complete disappearance of symptoms according to the Likert-7 point scale.

DISCUSSION

Baseline qualitative MRI findings and the size of the disc herniation did not predict future inevitable surgery in patients who were subjected to a wait-and-see policy for sciatica. Patients who did undergo surgery during follow-up had at baseline higher RDQ scores, more intense leg pain and smaller dural sacs and spinal canals compared to patients who did not undergo surgery. The overall results of the current study suggest that MRI is not suitable to distinguish between patients who will and those who will not undergo surgery for sciatica.

The natural history of acute sciatica is in general favourable, with spontaneous resolution of the leg pain within 18 weeks in the overwhelming majority of cases.[25,28,30] When patients fail to recover during conservative care, surgery might be considered. The optimal duration of conservative care is not well known though. The absolute indications for acute surgery of lumbar herniated discs are symptoms of a cauda equina syndrome, presence of acute and severe motor deficits, and intractable pain.[27] However, these absolute indications rarely occur. In all other cases the indications for operation are relative and clear clinical guidelines are lacking. Some studies retrospectively evaluated the MRI differences between patients who did and those who did not undergo surgery. Carlisle et al. observed that sciatica patients who underwent surgery had larger disc herniations and smaller spinal canals compared to nonoperative patients.[4] A limitation of this study is the retrospective case-matched design and surgical case selection that may have been biased towards larger herniations. Cheng et al. retrospectively observed that patients with either severe disc herniation or severe spinal stenosis were more likely to be classified as surgical candidates compared to those with mild to moderate findings.[6] Carragee and Kim also observed that patients who underwent surgery had larger disc herniations and smaller sizes of the remaing spinal canal compared to patients who underwent conservative treatment.[5]

Surgical treatment rates for lumbar discectomy vary widely between countries and even within countries.[8,18] Currently no objective measures are available to determine when to perform surgery for sciatica. The current study only thoroughly assessed the predictive value of MR imaging for disc surgery. However, the decision for surgery does depend on many factors including the pain disability, psychological factors, occupation, expectations, fear of surgery, economic and social considerations, sociocultural preferences and even the preference of the treating surgeon.[8,10,11,15,18,24] The contribution of this present study is that in contrast to the earlier mentioned studies[4,5,7] MR imaging has no value in the prediction of future surgery among patients with sciatica for 6 to 12 weeks. Patients with clear sciatic symptoms and on MRI a large herniated disc with clear nerve root compression might still not undergo surgery. As published earlier, RDQ scores and VAS intensity of leg pain were better able to discriminate between surgical and non-surgical patients,[24] although these variables did also not reach excellent discrimination in the current ROC analysis. Valid tools for appropriate patient selection for disc surgery are therefore still desirable.

We previously reported the 1-year follow-up MRI results of all patients who participated in the randomized clinical trial.[12] At one year follow-up a considerable proportion of patients still had a visible disc herniation on MRI (21% of surgically compared to 60% of conservatively treated patients). Compared to baseline, nerve root compression had disappeared in 82% of surgically treated patients compared to 60% of conservatively treated patients. However, visible MRI abnormalities at one year follow-up did not distinguish between patients with persistent or recurrent symptoms of sciatica from asymptomatic patients. Other studies have reported similar results.[2,16] Jensen did not observe any correlation between improvement in symptoms and improvement of disc herniations and nerve root compression on MRI at 14 months in 154 conservatively treated patients.[16] Bath observed a high incidence (approximately 67%) of extrusions and protrusions 2 years postoperatively.[2] However, postoperative extrusions or protrusions did not correlate with clinical outcomes. A recent systematic review concluded that even in the acute setting of sciatica evidence for the diagnostic accuracy of MRI is not conclusive.[29] This is a well-known paradox in imaging research of sciatica. Although there is poorly evidence that imaging findings relate to clinical symptoms, surgery by means of microsurgical discectomy often proves helpful for these patients.[1,22,25,31] So far there are no studies that assessed the role of MRI in decision making for patients with acute or persistent sciatica, in particular if treatment strategies according to MRI findings lead to different clinical outcomes. Further research is needed to assess the value of MRI in clinical decision making for patients with acute and also in those with persistent or recurrent sciatica.

An important limitation of the current study is that the study population consisted of patients who had severe sciatic symptoms for at least 6 weeks and who were referred by their primay care physician. These patients were willing to undergo surgery, so patients with a clear preference for conservative treatment are underrepresented. Also, surgical treatment rates may have been relatively low because patients were encouraged to persist with the randomized prolonged conservative strategy. Not all patients might have had similar conservative treatments. One may get more information out of a prospective cohort study where patients are treated with similar nonoperative treatment modalities and then followed for a certain time period (for example one year). Baseline MRI findings should then be compared between those who decide to have surgery and those who decide not to have surgery during the follow-up period. Another limitation is that in this study patients already did experience a sciatic pain period of several weeks, before MRI was performed and therefore early anatomical changes might already have been occurred since the acute stage. This makes the assessed MRIs in this study less baseline than in experimental conditions could have been reached. Besides the limitations this is the first study that thoroughly analyzed the predictive value of MR imaging in patients with severe sciatica who were subjected to a wait-and-see policy. Furthermore, all MRI scans were blindly examined by in total four observers who were not involved in the study before.

CONCLUSIONS

MRI showed a poor ability to discriminate between patients who did and those who did not undergo delayed surgery for sciatica. As such the role of MRI remains limited to depict the anatomical features and the level of a herniated disc, necessary for the surgical technical approach, and should not be used as a prognosis tool in the shared decision making discussion for surgery versus wait-and-see.

DISCLOSURE

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. None of the authors does have a conflict of interest with the topic of research published in this manuscript.

REFERENCES

1. Atlas SJ, Keller RB, Wu YA, Deyo RA, Singer DE: Long-term outcomes of surgical and nonsurgical management of sciatica secondary to a lumbar disc herniation: 10 year results from the maine lumbar spine study. Spine (Phila Pa 1976) 30:927-935, 2005
2. 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) 33:273-279, 2008
3. Beattie PF, Meyers SP, Stratford P, Millard RW, Hollenberg GM: Associations between patient report of symptoms and anatomic impairment visible on lumbar magnetic resonance imaging. Spine (Phila Pa 1976) 25:819-828, 2000
4. Carlisle E, Luna M, Tsou PM, Wang JC: Percent spinal canal compromise on MRI utilized for predicting the need for surgical treatment in single-level lumbar intervertebral disc herniation. Spine J 5:608-614, 2005
5. Carragee EJ, Kim DH: 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) 22:1650-1660, 1997
6. Cheng F, You J, Rampersaud YR: Relationship between spinal magnetic resonance imaging findings and candidacy for spinal surgery. Can Fam Physician 56:e323-330
7. Cheng F, You J, Rampersaud YR: Relationship between spinal magnetic resonance imaging findings and candidacy for spinal surgery. Can Fam Physician 56:e323-330, 2010
8. Cherkin DC, Deyo RA, Loeser JD, Bush T, Waddell G: An international comparison of back surgery rates. Spine (Phila Pa 1976) 19:1201-1206, 1994
9. Collins SL, Moore RA, McQuay HJ: The visual analogue pain intensity scale: what is moderate pain in millimetres? Pain 72:95-97, 1997
10. Deyo RA: Back surgery--who needs it? N Engl J Med 356:2239-2243, 2007
11. Deyo RA, Cherkin DC, Weinstein J, Howe J, Ciol M, Mulley AG, Jr.: Involving patients in clinical decisions: impact of an interactive video program on use of back surgery. Med Care 38:959-969, 2000
12. el Barzouhi A, Vleggeert-Lankamp CL, Lycklama a Nijeholt GJ, Van der Kallen BF, van den Hout WB, Jacobs WC, et al: Magnetic resonance imaging in follow-up assessment of sciatica. N Engl J Med 368:999-1007, 2013
13. 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) 26:E93-E113, 2001
14. Hosmer D, Lemeshow S: Applied Logistic Regression. II ed. New York: John Wiley & Sons Inc; 2000.
15. Irwin ZN, Hilibrand A, Gustavel M, McLain R, Shaffer W, Myers M, et al: Variation in surgical decision making for degenerative spinal disorders. Part I: lumbar spine. Spine (Phila Pa 1976) 30:2208-2213, 2005
16. 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 30:98-108, 2007
17. Kleinbaum DG, Klein M: Logistic Regression: A Self-Learning Text (Statistics for Biology and Health) Springer, 2010
18. Koes BW, van Tulder MW, Peul WC: Diagnosis and treatment of sciatica. BMJ 334:1313-1317, 2007

19. Lurie JD, Tosteson AN, Tosteson TD, Carragee E, Carrino JA, Kaiser J, et al: Reliability of magnetic resonance imaging readings for lumbar disc herniation in the Spine Patient Outcomes Research Trial (SPORT). Spine (Phila Pa 1976) 33:991-998, 2008
20. Modic MT, Masaryk TJ, Ross JS, Carter JR: Imaging of degenerative disk disease. Radiology 168:177-186, 1988
21. Modic MT, Steinberg PM, Ross JS, Masaryk TJ, Carter JR: Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging. Radiology 166:193-199, 1988
22. Osterman H, Seitsalo S, Karppinen J, Malmivaara A: Effectiveness of microdiscectomy for lumbar disc herniation: a randomized controlled trial with 2 years of follow-up. Spine (Phila Pa 1976) 31:2409-2414, 2006
23. 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) 20:1899-1908; discussion 1909, 1995
24. Peul WC, Brand R, Thomeer RT, Koes BW: Improving prediction of "inevitable" surgery during non-surgical treatment of sciatica. Pain 138:571-576, 2008
25. Peul WC, van Houwelingen HC, van den Hout WB, Brand R, Eekhof JA, Tans JT, et al: Surgery versus prolonged conservative treatment for sciatica. N Engl J Med 356:2245-2256, 2007
26. Peul WC, van Houwelingen HC, van der Hout WB, Brand R, Eekhof JA, Tans JT, 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 6:8, 2005
27. Postacchini F: Management of herniation of the lumbar disc. J Bone Joint Surg Br 81:567-576, 1999
28. Vroomen PC, de Krom MC, Slofstra PD, Knottnerus JA: Conservative treatment of sciatica: a systematic review. J Spinal Disord 13:463-469, 2000
29. Wassenaar M, van Rijn RM, van Tulder MW, Verhagen AP, van der Windt DA, Koes BW, 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 21:220-227, 2012
30. Weber H, Holme I, Amlie E: The natural course of acute sciatica with nerve root symptoms in a double-blind placebo-controlled trial evaluating the effect of piroxicam. Spine (Phila Pa 1976) 18:1433-1438, 1993
31. Weinstein JN, Lurie JD, Tosteson TD, Skinner JS, Hanscom B, Tosteson AN, et al: Surgical vs non-operative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT) observational cohort. JAMA 296:2451-2459, 2006

Chapter 5

Prognostic value of MRI in sciatica patients in relation to back pain

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

Submitted for publication

ABSTRACT

BACKGROUND

Patients with sciatica frequently do complain about associated back pain. It is not known whether there are prognostic relevant differences in Magnetic Resonance Imaging (MRI) characteristics between sciatica patients with and without disabling back pain.

METHODS

The study population contained patients with sciatica who underwent a baseline MRI to assess eligibility for a randomized trial designed to compare the efficacy of early surgery with prolonged conservative care for sciatica. Blinded evaluated MRI findings were compared between sciatica patients with and without disabling back pain. On the basis of significantly different MRI findings four subgroups were defined that were correlated with perceived recovery at one year: back pain with and without the MRI characteristic, and no back pain with and without the MRI characteristic.

RESULTS

Of 379 included sciatica patients, 158 (42%) had disabling back pain. Of the patients with both sciatica and disabling back pain 68% did reveal a herniated disc with nerve root compression on MRI, compared to 88% of patients with predominantly sciatica ($P<0.001$). The existence of low back pain in sciatica at baseline was negatively associated with perceived recovery at one year (Odds ratio [OR] 0.32, 95% CI 0.18-0.56, $P<0.001$). Sciatica patients with disabling back pain in absence of nerve root compression on MRI at baseline reported less perceived recovery at one year compared to those with predominantly sciatica and nerve root compression on MRI (50% vs 91%, $P<0.001$)

CONCLUSION

Sciatica patients with disabling low back pain reported an unfavorable prognosis at one-year follow-up compared to those with predominantly sciatica. If additionally a clear herniated disc with nerve root compression on MRI was absent, the results were even worse.

INTRODUCTION

Patients with sciatica frequently complain about associated back pain.[1] Sciatica is associated with significant short- and sometimes long-term morbidity. This affliction, certainly in the industrialized countries, ranks as one of the most costly and ubiquitous medical problems.[2] In classical literature sciatica has been of great interest to Greco-Roman and Eastern scientists and physicians.[3] For centuries an inflammation of the sciatic nerve was the origin of pain, described as sciatic neuritis,[4] until 1934 when Mixter and Barr revolutionized the understanding of sciatica into mechanical origin.[3,5] They asserted that sciatica was caused by a herniated disc pressing against a nerve root. Worldwide this mechanical compression theory has been accepted giving rise to a global implementation of disc surgery as the solution to remove the compression on the nerve root and with that resolve the disabling pain problem. However, does this theory still find ground or is it worthwhile to think about the renaissance of the old theory involving inflammation of the nerve root?

This scientific confusion has been caused by the introduction of modern imaging modalities such as Magnetic Resonance Imaging (MRI) which allowed many investigators to detect an enormous variety of previously unappreciated anatomical variations in patients undergoing diagnostic workups for sciatica.[6] For example, several studies show a high prevalence of disc herniations ranging from 20 to 76% in subjects without any symptoms.[7,8] Furthermore, it remains unclear to what extent morphological changes seen on MRI in sciatica patients are associated with back pain, rather than being a representation of irrelevant differences between individuals.[6,7,8] Back pain has been reported to be associated with worse prognosis in patients with sciatica,[9] but one could question whether it is the back pain itself that causes the worse prognosis or the possible MRI anatomical differences between sciatica patients with and without back pain.

The investigators previously reported the results of a randomized controlled trial comparing early surgery with prolonged conservative care for patients with sciatica.[10] 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. In this large group of sciatica patients, we now report on the MRI differences between patients who suffered from sciatica with disabling back pain as compared to patients who suffered from sciatica only. Furthermore we report on the relevance of these MRI differences for prognosis.

MATERIALS AND METHODS

ETHICS STATEMENT

The medical ethics committees at the nine participating hospitals (Leiden University Medical Center, Medical Center Haaglanden, Diaconessen Hospital, Groene Hart Hospital, Reinier de Graaf Hospital, Spaarne Hospital, Bronovo Hospital, Rijnland Hospital and Lange Land Hospital) approved the protocol. Written informed consent was obtained from all patients.

STUDY POPULATION

Patients for this study were patients with intense lumbosacral nerve root pain who underwent a baseline MRI to assess the eligibility for the Sciatica Trial: a multicenter, randomized controlled trial designed to determine whether early surgery results in better outcome compared to a strategy of prolonged conservative treatment with surgery if needed among patients with 6-12 weeks sciatica.[10,11] Patients who had symptoms being so severe that they were eligible for surgery according to their family physicians 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, and severe coexisting disease. Participants who were not meeting one or more of the aforementioned exclusion criteria and had a lumbosacral radicular syndrome lasting between 6-12 weeks underwent an MRI and qualified to be included in this present study (thus for the present study it was not necessary to have a herniated disc visible on MRI). All patients with sciatica who underwent MRI (regardless of participation in the randomized trial) were followed for one year. Details of the design and study protocol have been published previously.[11]

MRI PROTOCOL AND IMAGE EVALUATION

MRI scans were performed in all nine participating hospitals using standardized protocols tailored to a 1.5 Tesla scanner. Sagittal T1 and axial T1 spin echo images of the lumbar spine were acquired. In addition, T2 weighted sagittal and axial series, and contrast-enhanced (gadolinium-DTPA) T1 fat suppressed images were obtained.

Two experienced neuroradiologists (BK and GL) and one neurosurgeon (CV) independently evaluated all MR images. The readers were not provided any clinical information and had not been involved in the selection or care of the included patients.

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 disc pathology.[12] Vertebral Endplate Signal Changes were defined according to criteria of Modic.[13,14] Standardized case record forms with definitions were used to evaluate the images (Appendix Table S1).

First, the blinded readers had to decide which disc level showed the most severe nerve root compression. For both the presence of disc herniation and nerve root compression a four point scale was used: "definite about the presence", "probable about the presence", "possible about the presence" and "definite about the absence". The first two categories were combined and marked as having the abnormality present. The latter two categories were combined and marked as not having the abnormality present. Clinically relevant characteristics of the disc level and disc herniation were scored. Vertebral Endplate Signal Changes were evaluated from L2-L3 through L5-S1.

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 low 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 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 (varying from "minimally improved" to "very much worse") was marked as "no recovery".[10,11] Outcome measures were assessed at baseline, 2, 4, 8, 12, 26, 38 and 52 weeks.

STATISTICAL ANALYSIS

The majority opinion of the three readers regarding the MRI characteristics (answer independently given by minimum 2 out of 3 readers) was used in the statistical analysis. Interobserver agreement regarding the MRI findings was determined by use of absolute percentages of agreement and kappa values (weighted in case of ordered data).

Disabling back pain was defined in the SIPS research group consensus meeting as a VAS for back pain of at least 40, as this cutt-off value is regularly used when the VAS is categorized into favorable and unfavorable outcome.[17,18] Patients with missing VAS-back pain at baseline were excluded. Differences between patients with VAS-back pain of at least 40 and those with a VAS lower than 40 were assessed by using Student's t-test for continuous data and Chi-square tests for categorical data.

Logistic regression was used to determine the association between perceived recovery at one year and disabling back pain at baseline. On the basis of MRI characteristics that proved to be significantly different in proportions between patients with versus those without disabling back pain four subgroups were defined: back pain with and without the MRI characteristic, and no back pain with and without the MRI characteristic. Between group differences in continuous outcome measures (RDQ and VAS pain scores) during the first year were analyzed by repeated measurement analysis of variance.

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 imputation model were age, gender, body-mass index, duration of symptoms, smoking, treatment group, all used MRI variables (Table S1 Appendix), 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.

Table 1 Baseline characteristics stratified by presence of disabling back pain.

Variable	Sciatica with disabling back pain (n=158)	Sciatica with no disabling back pain (n=221)	P-value
Age at baseline MRI	42.8±10.9	43.4±9.6	0.56
Male-sex	92 (58)	147 (67)	0.09
Duration of sciatica (weeks)	9.0±2.4	9.5±3.8	0.11
BMI‖	26.1±4.2	25.9±3.6	0.59
Treatment group			0.09
Non-randomized	48 (30)	50 (23)	
Randomized to early surgery	60 (38)	79 (36)	
Randomized to prolonged conservative care	50 (32)	92 (42)	
Smoking	67 (42)	80 (36)	0.24
Roland disability score for sciatica¶			
Baseline	17.4±3.3	15.0±4.5	<0.001
12 months	4.5±5.9	2.9±4.7	0.004
Visual-analogue scale of leg pain‡			
Baseline	66.6±20.3	60.7±22.7	0.009
12 months	13.7±22.4	8.7±16.5	0.014
Visual-analogue scale of back pain‡			
Baseline	63.3±16.2	12.1±11.6	<0.001
12 months	21.3±26.1	12.2±18.8	<0.001
Perceived recoveryò			
12 months	111 (70)	195 (88)	<0.001

Values are n (%) or means ± SD.
‖ Body-mass index (BMI) is the weight in kilograms divided by the square of the height in meters
¶ The Roland disability questionnaire for sciatica is a disease-specific disability scale that measures functional status in 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 was indicated on a horizontal 100 mm visual analogue scale, with 0 representing no pain and 100 the worst pain ever experienced.
ò Perceived recovery was defined as complete or nearly complete disappearance of symptoms according to the Likert-7 point scale.

Table 2 Comparison of MRI characteristics between sciatica patients with and without disabling back pain at baseline.

	Sciatica with disabling back pain (n=158)	Sciatica with no disabling back pain (n=221)	P-value
MRI characteristic			
Presence of disc herniation	120 (76)	202 (91)	**<0.001**
Presence of nerve root compression	108 (68)	195 (88)	**<0.001**
Presence of Vertebral Endplate Signal Changes at one or more lumbar level ¶	63 (41)	94 (43)	0.91
Type 1	3 (5)	6 (6)	
Type 2	58 (92)	84 (89)	
Type 3	0 (0)	1 (1)	
Mixed Type 1 and 2	2 (3)	3 (3)	
Presence of Schmorl's nodules (herniation of the disc into the vertebral-body endplate) at one or more levels	18 (12)	25 (11)	0.94

Values are n (%)

¶ Vertebral Endplate Signal Changes were defined according to criteria of Modic and their presence was assessed from vertebral endplates L2-L3 through L5-S1. Type 1 lesions: hypointense in T1-weighted sequences and hyperintense in T2-weighted sequences. Type 2 lesions: increased signal on T1 weighted sequences and isointense or slightly hyperintense signal on T2 weighted sequences. Type 3 lesions: hypointense both in T1- and T2-weighted sequences.

RESULTS

Of the 599 patients screened for the study, 395 patients underwent MRI of whom 283 patients were randomized.[10,19] In total, 283 baseline MRI's of the 283 randomized patients and 106 MRI's of 112 non-randomized patients could be retrieved, bringing the total to 389 MRI's. Of the randomized patients 91% depicted a disc herniation with nerve root compression on MRI compared to 49% of the non-randomized patients. Baseline VAS of back pain was not available for 10 (2.6%) patients. Of the 379 eligible patients, 158 (42%) had a VAS of at least 40 with a mean of 63.3 (95% Confidence Interval [CI] 61-66) and 221 (58%) patients had a VAS of back pain of less than 40 with a mean VAS of 12.1 (95% CI 11-14). At baseline, sciatica patients with and without disabling back pain had a statistically significant but clinically small difference in RDQ and VAS-leg pain (17.4 vs. 15.0 and 66.6 vs. 60.7 respectively) (Table 1). Clinical outcome at 52 weeks was missing in 12-13% of patients (Appendix Table S2). Baseline RDQ and VAS for leg and back pain were comparable among patients for whom clinical outcome at 52 weeks was available and those for whom not (P-value range 0.21-0.42).

Substantial agreement was found for the MRI assessed presence of disc herniation (kappa range 0.67-0.75) and nerve root compression (kappa range 0.60-0.80) (Appendix Table S3).

Moderate agreement was found for the size of the disc herniation (kappa range 0.35-0.55) and presence of vertebral endplate signal changes (kappa range 0.49-0.67).

MRI DIFFERENCES WITH AND WITHOUT DISABLING BACK PAIN

Of patients with both sciatica and disabling back pain 76% had a disc herniation on MRI compared to 91% of patients without disabling back pain (P<0.001) (Table 2). Nerve root compression on MRI was observed less frequently in patients with both disabling sciatica and back pain compared to patients with predominantly sciatica (68% vs. 88%, P<0.001). No significant differences existed in prevalence of Vertebral Endplate Signal Changes between sciatica patients with and without disabling back pain (41% vs. 43%, P=0.70).

A comparison of the characteristics of the herniated disc itself between sciatica patients with and without disabling back pain is shown in Table 3**.** Large disc herniations (size >50% of spinal canal) were observed in an equal percentage (18%) between patients with and without disabling back pain. Also, no significant difference existed in extrusions between patients with and without disabling back pain (64% vs. 67%, P=0.66).

CLINICAL OUTCOME IN RELATION TO DISABLING BACK PAIN AND MRI DIFFERENCES

The existence of disabling back pain in sciatica at baseline was negatively associated with perceived recovery at one year (Odds ratio [OR] 0.32, 95% CI 0.18-0.56, P<0.001). This result was consistent with the continuous outcomes RDQ and VAS pain scores (Appendix Figure S1). By contrast, presence of disc herniation on MRI at baseline was positively associated with perceived recovery at one year (OR 3.18, 95% CI 1.6-6.4, P=0.001). Same holds for nerve root compression (OR 4.99, 95% CI 2.7-9.2, P<0.001).

The reported prevalence of perceived recovery at one year was 81% for sciatica patients who had at baseline disabling back pain and nerve root compression, 50% for patients who had at baseline back pain but no nerve root compression, 91% for patients who had at baseline no back pain but depicted nerve root compression on MRI, and 73% for patients who had at baseline no back pain and no nerve root compression (P<0.001) (Table 4). In the stratified analysis according to treatment group the overall trends were comparable with the non-stratified analysis (Appendix Table S4).

In patients with disabling back pain, those who also had nerve root compression on MRI significantly reported more favorable recovery from their back pain at one year compared to those who had not depicted nerve root compression at baseline (Figure 1).

The sensitivity analyses yielded comparable results (with complete case analysis instead of multiple imputation of missing data) (Appendix Table S5).

Table 3 Comparison of the characteristics of the herniated disc on MRI between sciatica patients with and without disabling back pain at baseline. Values are n (%). N=330

	Sciatica with disabling back pain (n=125)	Sciatica with no disabling back pain (n=205)	P-value
Size of disc herniation			
Size > 50% in relation to spinal canal	23 (18)	37 (18)	0.95
Size < 50% in relation to spinal canal	102 (82)	167 (81)	
Not classifiable	0 (0)	1 (1)	
Location of disc herniation			
Central and/or subarticular	111 (89)	183 (89)	0.70
Foraminal and/or extraforaminal	14 (11)	20 (10)	
Not classifiable	0 (0)	2 (1)	
Morphology of disc herniation			
Extrusion	80 (64)	138 (67)	0.66
Protrusion	42 (34)	65 (32)	
Not classifiable	3 (2)	2 (1)	
Loss of disc height at the disc level of the disc herniation			
Yes	112 (90)	186 (91)	0.96
No	10 (8)	17 (8)	
Not classifiable	3 (2)	2 (1)	
Signal intensity of nucleus pulposus on T2 images at the disc level of the disc herniation			
Hypointensity	110 (88)	185 (90)	0.72
Normal	10 (8)	15 (7)	
Hyperintensity	(0)	1 (1)	
Not classifiable	5 (4)	4 (2)	
Presence of Vertebral Endplate Signal Changes at the disc level of the disc herniation ¶			
Type 1	2 (4)	6 (7)	0.70
Type 2	51 (93)	76 (91)	
Type 3	0 (0)	0 (0)	
Mixed Type 1 and 2	2 (4)	2 (2)	

Values are n (%)

¶ Vertebral Endplate Signal Changes were defined according to criteria of Modic. Type 1 lesions: hypointense in T1-weighted sequences and hyperintense in T2-weighted sequences. Type 2 lesions: increased signal on T1 weighted sequences and isointense or slightly hyperintense signal on T2 weighted sequences. Type 3 lesions: hypointense both in T1- and T2-weighted sequences.

Table 4 Clinical outcome measures at one year according to subgroups at baseline. Subgroups defined by the presence of disabling back pain and the presence of a disc herniation or nerve root compression on MRI at baseline. Values are n (%) or means ± SD. N=379

	Clinical outcome at one year			
	Perceived recoveryò	Roland Disability‡	VAS-Leg pain¶	VAS-back pain¶
Subgroups according to back pain and presence of nerve root compression on MRI at baseline				
Back pain and nerve root compression (n=108)	87 (81)	3.6±5.8	11.8±21.7	17.4±23.9
Back pain and no nerve root compression (n=50)	25 (50)	6.4±5.8	17.8±23.5	29.6±28.8
No back pain and nerve root compression (n=195)	177 (91)	2.7±4.4	7.6±14.1	11.4±17.2
No back pain and no nerve root compression (n=26)	19 (73)	4.5±6.6	16.7±27.9	18.7±27.4
Subgroups according to back pain and presence of disc herniation on MRI at baseline				
Back pain and disc herniation (n=120)	90 (75)	4.2±6.2	14.4±23.9	20.0±26.2
Back pain and no disc herniation (n=38)	22 (58)	5.4±5.1	11.6±16.8	25.2±25.8
No back pain and disc herniation (n=202)	181 (90)	2.8±4.5	7.7±14.1	11.6±17.3
No back pain and no disc herniation (n=19)	14 (74)	4.1±6.5	18.8±31.7	18.3±29.9

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.

DISCUSSION

In this study of patients with sciatica who were followed for one year, those with disabling back pain at baseline reported an unfavorable prognosis at one-year follow-up compared to those with predominantly sciatica. If additionally a herniated disc with nerve root compression on MRI was absent, the results were even worse. Herniated discs and nerve root compression on MRI were more prevalent among patients with predominantly sciatica compared to those who suffered from additional disabling back pain. However, vertebral endplate signal changes were equally distributed between those with and without disabling back pain. Remarkably large disc herniations and extruded disc herniations were also equally distributed between the two groups.

Over the past two decades there has been an ongoing scientific debate about the clinical relevance of MRI morphological variations.[7,8] To uncover the relevance of imaging findings, knowledge regarding their prevalence and relation with symptoms in different (sub)groups is needed. However, in most clinical studies, patients with herniated discs have been reported as a single pathological group.[20] Comparable to this study, some researchers have attempted to

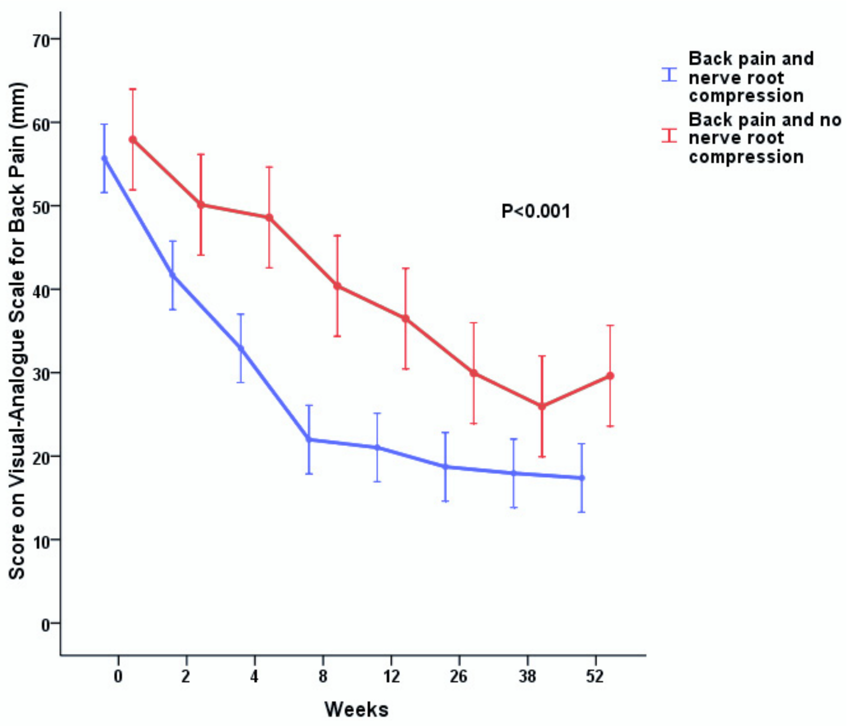


Figure 1 Repeated measurement analysis curve of Mean Scores for back pain on the Visual-Analogue Scale. Sciatica patients with both disabling back pain and nerve root compression on MRI were compared with patients with disabling back pain but who did not depict nerve root compression on MRI at baseline. The vertical bars represent 95% confidence intervals.

identify MRI differences between (sub)groups. MRI differences have been reported between patients with both sciatica and low back pain compared to asymptomatic control subjects,[7] and between sciatica patients compared to low back pain patients.[21] The finding that vertebral endplate signal changes was equally distributed between those with and without disabling back pain was surprising as they are hypothesized to be a causative factor in low back pain.[22,23] The finding that extruded disc herniations and large disc herniations were also equally distributed between the two groups was also surprising as both findings have been reported to correlate with the severity of symptoms in sciatica.[6,7] However, these studies did not compare these findings between sciatica patients with and without back pain. Comparable to this study, Vroomen described a more favourable prognosis for patients with compared to those without nerve root compression on MRI.[24]

The preoccupation with the herniated disc as a source of disabling low back and leg pain has led disc surgery to become one of the most commonly performed operative procedures. However, disc herniations are often seen on imaging studies in patients without symptoms.[7,8] Contrary, in this study, a substantial number of patients without disc herniation or nerve compression suffered from sciatica. The worldwide accepted mechanical compression theory therefore seems not to offer a sufficient explanation for the cause of the disabling back and leg symptoms in sciatica. Some researchers suggested that inflammation of the nerve root may also be a major factor in sciatica.[25,26] Back in time, Cotugnio, an 18th Century Italian physician, explained the sciatic complaints as a consequence of neuritis or edema of the sciatic nerve.[3,4] If this hypothesis is correct, the finding that sciatica patients with back pain less often had a herniated disc compared to patients with predominantly sciatica may be explained by a higher inflammatory component in sciatica patients with back pain. This may also explain why sciatica patients with back pain fared worse compared to patients with predominantly sciatica as the extent of inflammation may be a causative factor in the cases with persistent pain and functional disability.

The results after lumbar disc surgery do not seem to have improved during recent decades. Depending upon the used outcome measure, both classical studies and recent randomized controlled trials show that during longer follow-up treatment results for sciatica are satisfactory in 60 to 85% of the patients.[10,19,27,28,29] The number of proposed interventions, developed by numerous disciplines, is overwhelming. The results of this study indicate that in sciatica subgroups with different prognostic profiles can be identified. A shift from a "one-size fits all" approach, where heterogeneous groups of patients receive broadly similar treatments, towards targeted treatments according to prognostic profiles or specific characteristics, may help to improve the treatment results.[30]

A strength of this study was the blinded MRI assessment and follow-up of all patients with 6-12 weeks sciatica who underwent MRI, regardless of participation in the randomized trial. A limitation of the present study is that the study population consisted of sciatica patients who had severe symptoms and were referred to the neurologists. These patients were willing to undergo surgery, so patients with a clear preference for conservative treatment are under-represented. Some might view the agreement among MRI readers as suboptimal. However, the kappa values are comparable with those found in previous studies [8,31,32] and therefore one might consider them to reflect existing agreement among expert readers in clinical practice.

CONCLUSIONS

Sciatica patients with disabling low back pain reported an unfavorable prognosis at one-year follow-up compared to those with predominantly sciatica. If additionally a clear herniated disc with nerve root compression on MRI was absent, the results were even worse. Further research is needed to identify the reasons behind the different prognostic profiles in sciatica and how to apply new or existing therapeutic strategies accordingly.

REFERENCES

1. Koes BW, van Tulder MW, Peul WC (2007) Diagnosis and treatment of sciatica. BMJ 334: 1313-1317.
2. Dagenais S, Caro J, Haldeman S (2008) A systematic review of low back pain cost of illness studies in the United States and internationally. Spine J 8: 8-20.
3. Robinson JS (1983) Sciatica and the lumbar disk syndrome: a historic perspective. South Med J 76: 232-238.
4. Cotugnio D (1764.) De ischiade nervosa commentarius. Neapoli apud frat Simonios 12 ed. Viennæ:.
5. Mixter WJ, Barr JS (1934) Rupture of Intervertebral Disc with Involvement of Spinal Canal. N Engl J Med 211: 210-215.
6. Beattie PF, Meyers SP, Stratford P, Millard RW, Hollenberg GM (2000) Associations between patient report of symptoms and anatomic impairment visible on lumbar magnetic resonance imaging. Spine (Phila Pa 1976) 25: 819-828.
7. Boos N, Rieder R, Schade V, Spratt KF, Semmer N, et al. (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: 2613-2625.
8. Jensen MC, Brant-Zawadzki MN, Obuchowski N, Modic MT, Malkasian D, et al. (1994) Magnetic resonance imaging of the lumbar spine in people without back pain. N Engl J Med 331: 69-73.
9. Haugen AJ, Brox JI, Grovle L, Keller A, Natvig B, et al. Prognostic factors for non-success in patients with sciatica and disc herniation. BMC Musculoskelet Disord 13: 183.
10. Peul WC, van Houwelingen HC, van den Hout WB, Brand R, Eekhof JA, et al. (2007) Surgery versus prolonged conservative treatment for sciatica. N Engl J Med 356: 2245-2256.
11. Peul WC, van Houwelingen HC, van der Hout WB, Brand R, Eekhof JA, 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.
12. Fardon DF, Milette PC (2001) 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) 26: E93-E113.
13. Modic MT, Masaryk TJ, Ross JS, Carter JR (1988) Imaging of degenerative disk disease. Radiology 168: 177-186.
14. Modic MT, Steinberg PM, Ross JS, Masaryk TJ, Carter JR (1988) Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging. Radiology 166: 193-199.
15. Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, et al. (1995) Assessing health-related quality of life in patients with sciatica. Spine (Phila Pa 1976) 20: 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: 95-97.
17. Peters ML, Sommer M, de Rijke JM, Kessels F, Heineman E, et al. (2007) Somatic and psychologic predictors of long-term unfavorable outcome after surgical intervention. Ann Surg 245: 487-494.
18. Yamashita K, Ohzono K, Hiroshima K (2006) 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) 31: 2602-2608.
19. 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: 1355-1358.
20. Carragee EJ, Han MY, Suen PW, Kim D (2003) Clinical outcomes after lumbar discectomy for sciatica: the effects of fragment type and anular competence. J Bone Joint Surg Am 85-A: 102-108.

21. Modic MT, Obuchowski NA, Ross JS, Brant-Zawadzki MN, Grooff PN, et al. (2005) Acute low back pain and radiculopathy: MR imaging findings and their prognostic role and effect on outcome. Radiology 237: 597-604.
22. Albert HB, Kjaer P, Jensen TS, Sorensen JS, Bendix T, et al. (2008) Modic changes, possible causes and relation to low back pain. Med Hypotheses 70: 361-368.
23. Jensen TS, Karppinen J, Sorensen JS, Niinimaki J, Leboeuf-Yde C (2008) Vertebral endplate signal changes (Modic change): a systematic literature review of prevalence and association with non-specific low back pain. Eur Spine J 17: 1407-1422.
24. Vroomen PC, Wilmink JT, de KM (2002) Prognostic value of MRI findings in sciatica. Neuroradiology 44: 59-63.
25. Chen C, Cavanaugh JM, Ozaktay AC, Kallakuri S, King AI (1997) Effects of phospholipase A2 on lumbar nerve root structure and function. Spine (Phila Pa 1976) 22: 1057-1064.
26. Saal JS, Franson RC, Dobrow R, Saal JA, White AH, et al. (1990) High levels of inflammatory phospholipase A2 activity in lumbar disc herniations. Spine (Phila Pa 1976) 15: 674-678.
27. Jacobs WC, van Tulder M, Arts M, Rubinstein SM, van Middelkoop M, et al. (2011) Surgery versus conservative management of sciatica due to a lumbar herniated disc: a systematic review. Eur Spine J 20: 513-522.
28. Atlas SJ, Keller RB, Wu YA, Deyo RA, Singer DE (2005) Long-term outcomes of surgical and nonsurgical management of sciatica secondary to a lumbar disc herniation: 10 year results from the maine lumbar spine study. Spine (Phila Pa 1976) 30: 927-935.
29. Weber H (1983) Lumbar disc herniation. A controlled, prospective study with ten years of observation. Spine (Phila Pa 1976) 8: 131-140.
30. Konstantinou K, Beardmore R, Dunn KM, Lewis M, Hider SL, et al. (2012) Clinical course, characteristics and prognostic indicators in patients presenting with back and leg pain in primary care. The ATLAS study protocol. BMC Musculoskelet Disord 13: 4.
31. Brant-Zawadzki MN, Jensen MC, Obuchowski N, Ross JS, Modic MT (1995) Interobserver and intraobserver variability in interpretation of lumbar disc abnormalities. A comparison of two nomenclatures. Spine (Phila Pa 1976) 20: 1257-1263; discussion 1264.
32. Jarvik JG, Haynor DR, Koepsell TD, Bronstein A, Ashley D, et al. (1996) Interreader reliability for a new classification of lumbar disk disease. Acad Radiol 3: 537-544.

Table S1 MRI study variables. The three readers (2 neuroradiologists and one neurosurgeon) independently used the same case record form.

MRI variable	Type	Categories
Disc level with the most severe nerve root compression	Disc level	1. L2L3 2. L3L4 3. L4L5 4. L5S1 5. Not applicable, all disc levels have a normal disc contour: no disc extension beyond the normal margins of the intervertebral disc space at any disc level
	Disc contour at this disc level	1. Bulging: presence of disc tissue circumferentially (50-100%) beyond the edges of the ring apophyses 2. Herniation: localized displacement of disc material beyond the normal margins of the intervertebral disc space 3. Not applicable, all disc levels have a normal disc contour: no disc extension beyond the normal margins of the intervertebral disc space at any disc level
	Certainty about the presence of this 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
	Loss of disc height (distance between the planes of the end-plates of the vertebrae craniad and caudad to the disc) at this disc level	1. Yes 2. No
	Signal intensity of nucleus pulposus on T2 images at this level	1. Hypointensity 2. Normal 3. Hyperintensity
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
	Morphology	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 agains 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	Certainty about the presence 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 about the absence: no doubt about the absence
Separate for every end plate from level L2-L3 through L5-S1	Presence of vertebral endplate signal changes (VESC)	1. No VESC 2. VESC type 1: hypointense in T1-weighted sequences and hyperintense on T2-weighted sequences 3. VESC type 2: increased signal on T1 weighted sequences and isointense or slightly hyperintense signal on T2 weighted sequences 4. VESC type 3: hypointense both on T1- and T2-weighted sequences 5. VESC type 1 and 2
	Presence of Schmorl's nodes (herniation of the disc into the vertebral-body end plate)	1. Yes 2. No

Table S2 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=379

	Number of patients (%)
Visual Analogue scale for back pain at 52 weeks¶	
Outcome available at 52 weeks	332 (88)
At least one follow-up examination	37 (10)
Lost to follow-up after baseline examination	10 (3)
Global perceived recovery on a 7-point Likert scale at 52 weeksò	
Outcome available at 52 weeks	330 (87)
At least one follow-up examination	39 (10)
Lost to follow-up after baseline examination	10 (3)
Roland disability questionnaire at 52 weeks‡	
Outcome available at 52 weeks	333 (88)
At least one follow-up examination	36 (9)
Lost to follow-up after baseline examination	10 (3)
Visual Analogue scale for leg pain at 52 weeks¶	
Outcome available at 52 weeks	334 (88)
At least one follow-up examination	35 (9)
Lost to follow-up after baseline examination	10 (3)

¶ 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.
ò Global 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.

Table S3 Interobserver agreement regarding the MRI characteristics. Reader A en B represent the two neuroradiologists, while reader C represents the neurosurgeon. Kappa values and percentages of agreement for the 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	
	% agree-ment	kappa	% agree-ment	kappa	% agree-ment	kappa	% agree-ment	multi-rater kappa
Disc level with the most severe nerve root compression ¶	92.0	0.86	88.4	0.81	90.5	0.84	86.4	0.84
Probability of disc herniation (2 categories) ‡	93.6	0.75	91.8	0.71	90.0	0.67	87.7	0.71
Probability of nerve root compression (2 categories) ‡	94.1	0.80	85.4	0.62	84.6	0.60	82.0	0.66
Presence of vertebral end plate changes ò	73.8	0.49	83.4	0.67	81.0	0.60	69.1	0.58
Presence of Schmorl's nodes ò	80.3	0.25	81.6	0.47	82.6	0.26	72.2	0.33
Characteristics disc herniation								
Location axial view ╞	94.2	0.88	95.5	0.90	96.7	0.93	95.6	0.92
Size disc herniation in relation to spinal canal (2 categories) ║	82.1	0.55	76.3	0.35	86.3	0.47	71.5	0.44
Protrusion versus extrusion	77.4	0.48	75.0	0.50	73.7	0.44	63.2	0.46
Loss of disc height of the disc level ò	97.9	0.86	72.2	0.26	72.4	0.26	71.5	0.31
Signal intensity of nucleus pulposus on T2 images §	95.3	0.75	90.4	0.64	90.7	0.57	88.6	0.61

¶ The 5 categories were: 1) L2L3 2) L3L4 3) L4L5 4) L5S1 5) Not applicable, all disc levels have a normal disc contour: no disc extension beyond the normal margins of the intervertebral disc space at any disc level.
‡ The categories "Definite and probable about the presence" were combined to one category and the categories "possible about the presence" and "definite about the absence" were also combined to one category.
ò Categories were: yes versus no.
╞ Categories were: 1) Central zone 2) Sub-articular zone 3) Foraminal zone 4) Extra-foraminal zone.
║ The categories "large stenosing" and "large" were combined to one category and the categories "average" and "small" were also combined to one category.
§ Categories were: 1) Hypointensity 2) Normal 3) Hyperintensity.

Table S4 Clinical outcome measures at one year stratified according to subgroups at baseline and treatment group. Values are n (%) or means ± SD.

	Clinical outcome at one year			
	Perceived recoveryò	Roland Disability‡	VAS-Leg pain¶	VAS-back pain¶
Patient not randomized				
Back pain and nerve root compression (n=14)	11 (79)	3.4±4.2	5.9±8.5	18.5±21.3
Back pain and no nerve root compression (n=34)	16 (47)	6.3±5.2	14.8±17.3	29.2±2.6
No back pain and nerve root compression (n=33)	31 (94)	2.7±3.9	7.3±13.6	7.5±13.0
No back pain and no nerve root compression (n=17)	11 (65)	5.7±7.1	20.6±32.3	22.8±31.2
Patients assigned to surgery				
Back pain and nerve root compression (n=52)	42 (81)	3.4±6.2	12.9±22.2	15.6±22.2
Back pain and no nerve root compression (n=8)	4 (50)	8.7±7.4	34.7±33.6	36.6±34.4
No back pain and nerve root compression (n=73)	69 (95)	2.4±4.4	6.7±14.5	10.9±18.6
No back pain and no nerve root compression (n=6)	4 (67)	4.3±7.0	11.7±18.0	15.3±20.7
Patients assigned to conservative care				
Back pain and nerve root compression (n=42)	34 (81)	4.1±5.9	13.0±23.7	19.5±26.9
Back pain and no nerve root compression (n=8)	2 (25)	9.8±5.0	31.8±25.4	39.2±32.0
No back pain and nerve root compression (n=89)	77 (87)	3.1±4.5	9.1±14.0	13.1±17.2
No back pain and no nerve root compression (n=3)	3 (100)	1.0±1.7	2.3±2.3	3.7±4.0

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.

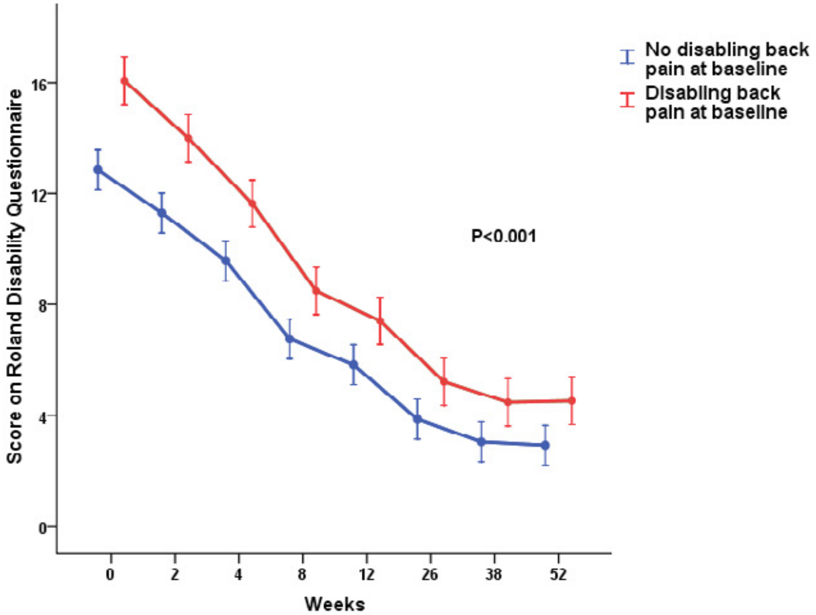
Table S5 Clinical outcome measures at one year according to subgroups at baseline. Subgroups defined by the presence of back pain and disc herniation or nerve root compression on MRI at baseline. *This analysis only included patients with available clinical outcome at one year.* Values are n (%) or means ± SD. N=330

	Clinical outcome at one year			
	Perceived recoveryò	Roland Disability‡	VAS-Leg pain¶	VAS-back pain¶
Subgroups according to back pain and presence of nerve root compression on MRI at baseline				
Back pain and nerve root compression (n=101)	80 (79)	3.8±5.9	12.1±22.0	17.9±24.4
Back pain and no nerve root compression (n=30)	10 (33)	8.3±5.8	22.8±25.6	35.9±30.5
No back pain and nerve root compression (n=176)	161 (91)	2.6±4.4	7.1±13.1	10.9±16.5
No back pain and no nerve root compression (n=23)	16 (70)	4.7±6.9	17.9±28.8	19.7±28.3
Subgroups according to back pain and presence of disc herniation on MRI at baseline				
Back pain and disc herniation (n=111)	82 (74)	4.4±6.3	14.9±24.4	20.8±26.8
Back pain and no disc herniation (n=20)	8 (40)	6.9±5.2	12.9±15.6	28.3±27.1
No back pain and disc herniation (n=185)	167 (90)	2.8±4.5	7.3±13.2	11.3±16.7
No back pain and no disc herniation (n=14)	10 (71)	4.1±7.0	22.3±34.8	21.1±33.2

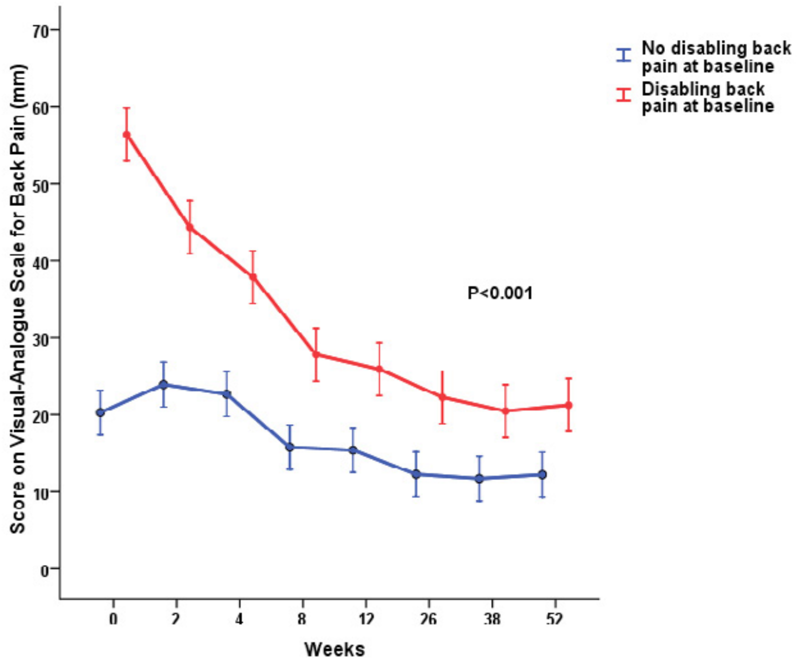
ò Perceived recovery was defined as complete or nearly complete disappearance of symptoms according to the Likert-7 point scale.

‡ 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.

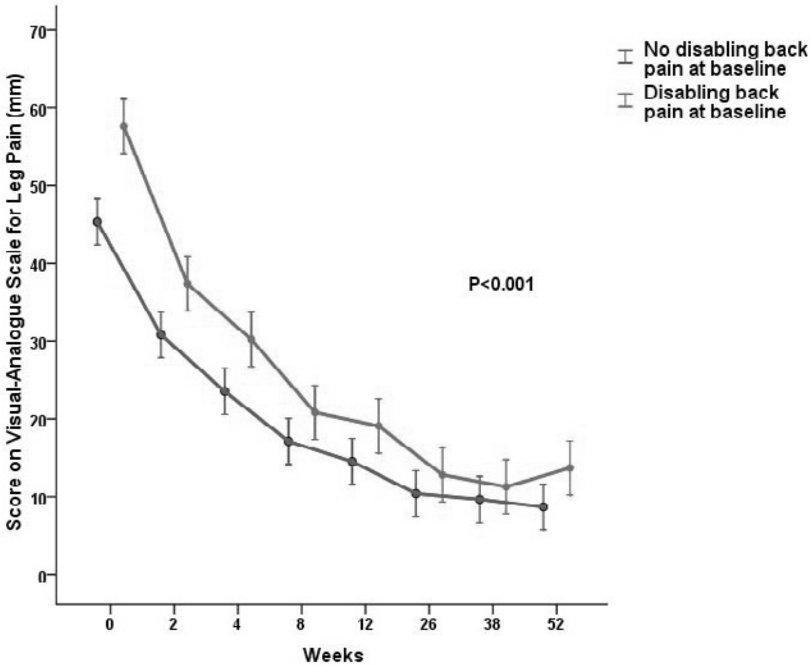
¶ The intensity of pain is indicated on a horizontal 100 mm visual analogue scale (VAS) with 0 representing no pain and 100 the worst pain ever experienced.



S1A Curve for the mean Roland Disability Questionnaire (scores range from 0 to 23, with higher scores indicating worse functional status) in relation to disabling back pain at baseline.



S1B Curve for the mean scores on the visual-analogue scale for intensity of back pain (scale ranges from 0 to 100 mm, with higher scores indicating more intense pain) in relation to disabling back pain at baseline.



S1C Curve for the mean scores on the visual-analogue scale for intensity of leg pain (scale ranges from 0 to 100 mm, with higher scores indicating more intense pain) in relation to disabling back pain at baseline.

Figure S1 Repeated measurement analysis curves of Mean Scores on the Roland Disability Questionnaire (1A), the Visual-Analogue Scale for back pain (1B), and the Visual-Analogue Scale for leg pain (1C) in relation to disabling back pain at baseline. The vertical bars represent 95% confidence intervals.

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).

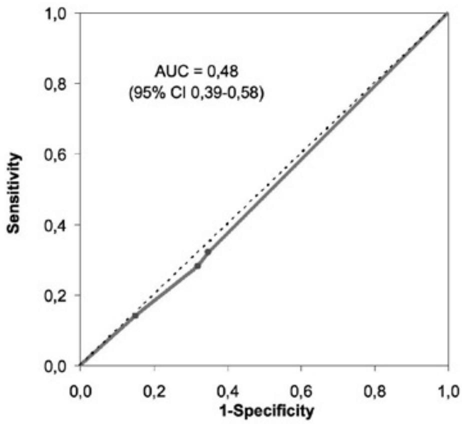
As sensitivity analyses to account for these missing data, we performed analysis as observed (e.g., no imputation, thus depending on the clinical outcome 6 or 7 patients with missing data were excluded from the analysis) and analysis using the last-observation-carried-forward method (depending on the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

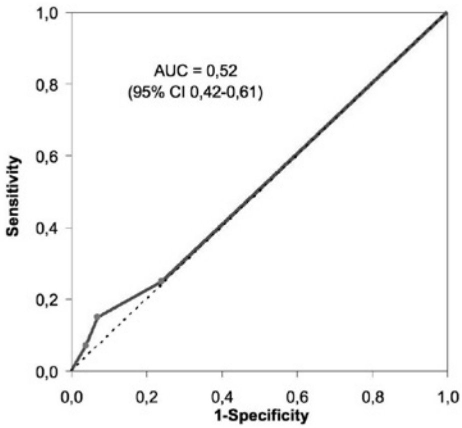
2. Analyses to account for interobserver agreement

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.

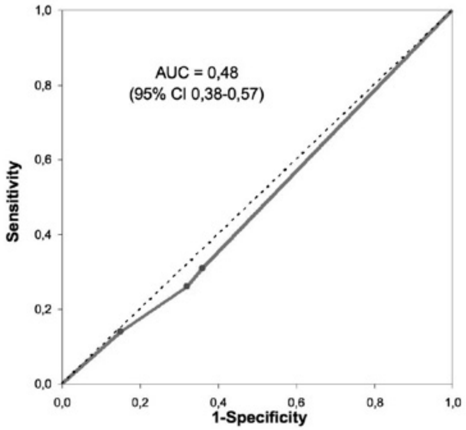


S1A

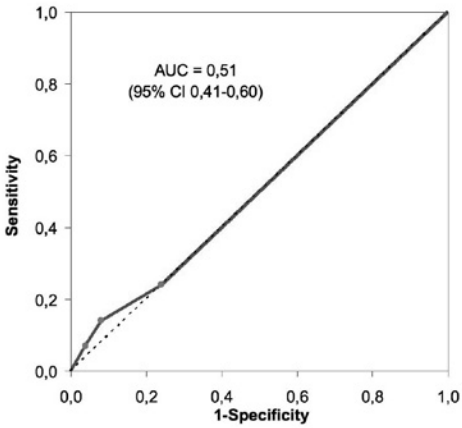


S1B

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.

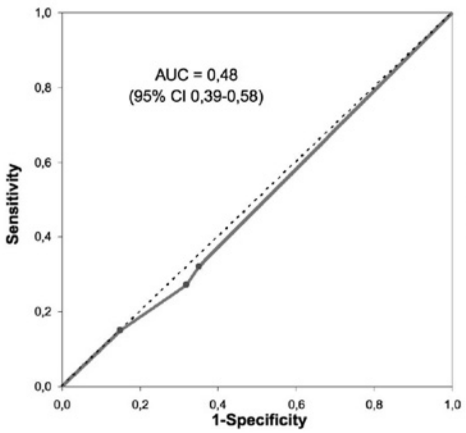


S2A

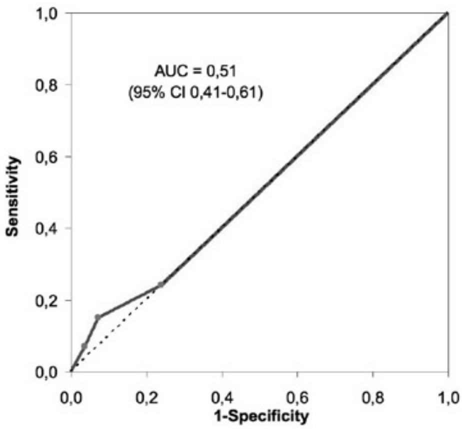


S2B

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) <u>when all participants (n=267) were included.</u>
- *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) <u>when only participants with full agreement [n=200]) were included.</u>*
- 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) <u>when all participant (n=267) were included.</u>
- *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) <u>when only participants with full agreement [n=200]) were included.</u>*

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.

Chapter 7

Back pain's association with vertebral endplate signal changes 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

North American Spine Society Outstanding Paper Award 2013
Will be published in The Spine Journal

ABSTRACT

BACKGROUND CONTEXT

Patients with sciatica frequently experience disabling back pain. One of the proposed causes for back pain is Vertebral Endplate Signal Changes (VESC) as visualized by Magnetic Resonance Imaging (MRI).

PURPOSE

To report on VESC findings, changes of VESC findings over time and the correlation between VESC and disabling back pain in patients with sciatica.

STUDY DESIGN/SETTING

A randomized clinical trial with one year follow-up.

PATIENTS SAMPLE

Patients with 6-12 weeks sciatica who participated in a multicentre randomized clinical trial comparing an early surgery strategy to prolonged conservative care with surgery if needed.

OUTCOME MEASURES

Patients were assessed by means of the 100-mm visual-analogue scale (VAS) for back pain (with 0 representing no pain and 100 the worst pain ever experienced) at baseline and one year. Disabling back pain was defined as a visual analogue scale score of at least 40mm.

METHODS

Patients underwent MRI both at baseline and after one year follow-up. Presence and change of VESC was correlated with disabling back pain using Chi-square tests and logistic regression analysis. This study was supported by a grant from the Netherlands Organisation for Health Research and Development (ZonMW) and the Hoelen Foundation, The Hague.

RESULTS

At baseline 39% of patients had disabling back pain. Of the patients with VESC at baseline 40% had disabling back pain compared to 38% of the patients with no VESC (P=0.67). The prevalence of type 1 VESC increased from 1% at baseline to 35% one year later in the surgical group compared to an increase from 3 to 11% in the conservative group. The prevalence of type 2 VESC decreased from 40 to 29% in the surgical group while remaining almost stable in the conservative group at 41%. The prevalence of disabling back pain at one year was 12% in patients with no VESC at one year, 16% in patients with type 1 VESC, 11% in patients with type 2 VESC and 3% in patients with both type 1 and 2 VESC (P=0.36). Undergoing surgery was associated with increase in the extent of VESC (Odds ratio [OR] 8.6, 95% CI

4.7-15.7, P<0.001). Patients who showed an increase in the extent of VESC after one year did not significantly report more disabling back pain compared to patients who did not show any increase (OR 1.2, 95% CI 0.6-2.6, P=0.61).

CONCLUSION

In this study undergoing surgery for sciatica was highly associated with the development of VESC after one year. However, in contrast with the intuitive feeling of spine specialists, those with and those without VESC reported disabling back pain in nearly the same proportion. Therefore VESC does not seem to be responsible for disabling back pain in patients with sciatica.

INTRODUCTION

Sciatica, more accurately called lumbosacral radicular syndrome, is one of the most common lumbar-spine disorders. The natural history of sciatica is favorable, with spontaneous resolution of the leg pain within 8 weeks in the majority of patients.[1] About 20 to 30% of the patients with sciatica receives surgery.[2] However, contrary to what one might expect given the advancements in diagnostic imaging and surgical techniques, the results after lumbar disc surgery for patients with radiculopathy due to a herniated disc do not seem to have improved during recent decades. Both classical and recent randomized controlled trials demonstrated that during longer follow-up at least 15-35% of the patients has an unsatisfactory outcome.[3-10] One of the most persistent accompanying complaints is chronic low back pain.[9, 11, 12] A considerable proportion of the costs and suffering due to sciatica can be attributed to the minority of patients that continues to experience symptoms like back pain.[12, 13] The identification of determinants of back pain and factors that promote persisting of back pain would be valuable as low back pain increasingly poses an economic burden to industrialized society, mainly in terms of the large number of work days lost.[13-15]

In the search for causes of associated back pain in patients with sciatica, vertebral endplate signal changes (VESC) visualized by Magnetic Resonance Imaging (MRI) have been proposed as a possible cause. In 1988 Modic described three types of signal changes.[16, 17] Type 1 lesions, hypointense on T1-weighted images and hyperintense on T2-weighted images, represent marrow edema, and are associated with an acute process.[16, 18, 19] Type 2 lesions, the most common type, have increased signal on T1 weighted images and isointense or slightly hyperintense signal on T2 weighted images, and represent fatty degeneration of subchondral marrow and are associated with a chronic process.[16, 20] Type 3 lesions, hypointense both on T1- and T2-weighted sequences, are considered to correlate with subchondral bone sclerosis.[16, 21]

The prevalence of VESC varies greatly among studies ranging from less than 1% in adolescents from the Danish general population[22] to 100% in selected patient populations.[23] Some studies observed an association between VESC and back pain,[24-27] while other studies did not observe any association.[28-31] Studies correlating VESC on consecutive MRIs in patients with sciatica are limited, especially studies comparing surgery with conservative treatment for the development of VESC. The determination of the clinical relevance of VESC is meaningful as accompanying endplate changes in patients suffering from radiculopathy due to a disc herniation are a frequent surgical indication to perform, in addition to the usual disc surgery for the radiculopathy, a fixation of two or more vertebrae in the lower spine or replacing the disc by a prothesis.[32-35] Lack of evidence and guideline consensus did result in a global problem of practice variation with regard to spinal surgery.[36, 37]

The investigators previously reported the results of a randomized controlled trial comparing early surgery with prolonged conservative care for patients with sciatica.[38] The trial showed faster recovery after early surgery, but the overall 1-year functional recovery rate was similar.

As the study protocol reported, patients underwent an MRI both at baseline and one year after randomization.[38] We now report on VESC findings, changes of VESC findings over time and the correlation between VESC findings and back pain in sciatica.

METHODS

STUDY POPULATION

Patients for this study were participants in a multicentre randomized trial among patients with 6-12 weeks sciatica (n=283). Patients were included only if they had a dermatomal pattern of pain distribution with concomitant neurological disturbances that correlated to the same nerve root being effected on MRI.[38] An early surgery strategy was compared to prolonged conservative care for an additional 6 months followed by surgery for patients who did not improve or who did request it earlier because of aggravating symptoms. The surgical treatment was standardized in this study (the symptomatic disk herniation was removed by a minimal unilateral transflaval approach with magnification. The goal of surgery was to decompress the nerve root and reduce the risk of recurrent disk herniation by performing an annular fenestration, curettage, and removal of loose degenerated disk material from the disk space with the use of a rongeur). Patients underwent MRI at the time of the initial diagnosis of sciatica and after one year of follow-up.[38] The medical ethics committees at the nine participating hospitals approved the protocol. Written informed consent was obtained from all patients. Details of the design and study protocol were published previously.[38]

MRI PROTOCOL AND IMAGE EVALUATION

MRI scans were performed in all nine participating hospitals using standardized protocols tailored to a 1.5 Tesla scanner. Sagittal T1 and axial T1 spin echo images of the lumbar spine were acquired. In addition, T2 weighted sagittal and axial series, and contrast-enhanced (gadolinium-DTPA) T1 fat suppressed images were obtained.

Two neuroradiologists and one neurosurgeon independently evaluated all MR images according to a predefined protocol (Appendix Table S1). Definitions of imaging characteristics were based on the 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 disc pathology.[39] VESC were defined according to criteria of Modic (as defined in the introduction).[16, 17] The observers graded the extent of VESC using three categories: mild, moderate and severe. As studies did not observe any VESC at level L1-L2,[21] all three observers only evaluated images from L2-L3 through L5-S1. Observers also evaluated the presence of Schmorl's nodes (herniation of the disc into the vertebral-body end plate). The observers were not provided any clinical information and have not been involved in the selection or care of the included patients. Observer experience in reading spine MRIs was

Table 1 Baseline characteristics of the intention-to-treat groups and the as-treated groups. Values are n (%) or means ± SD. N=263

	Intention to treat		As treated	
	Randomized to early surgery (N=129)	Randomized to prolonged conservative care (N=134)	Received surgery (n=168)	Received no surgery (n=95)
Age	41.7±10.0	43.2±9.3	41.9±9.8	43.5±9.3
Male sex	84 (65)	95 (71)	111 (66)	68 (72)
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
Smoking	51 (40)	47 (35)	65 (39)	33 (35)
Vertebral Endplate Signal Changes (VESC)				
No VESC	82 (64)	69 (51)	99 (59)	52 (55)
VESC Type 1	2 (2)	2 (1)	1 (1)	3 (3)
VESC Type 2	44 (34)	62 (46)	67 (40)	39 (41)
VESC Type 3	0 (0)	0 (0)	0 (0)	0 (0)
VESC Type 1 and 2	1 (1)	1 (1)	1 (1)	1 (1)
Suspected disc level and type of displacement on MRI				
L3L4 Herniation	5 (4)	4 (3)	7 (4)	2 (2)
L4L5 Herniation	58 (45)	50 (37)	70 (42)	38 (40)
L4L5 Bulging	2 (2)	1 (1)	3 (2)	0 (0)
L5S1 Herniation	63 (49)	77 (57)	87 (52)	53 (56)
L5S1 Bulging	1 (1)	2 (2)	1 (1)	2 (2)
MRI assessed nerve root compression				
Definite	80 (62)	94 (70)	110 (65)	64 (67)
Probable	35 (27)	29 (22)	42 (25)	22 (23)
Possible	11 (9)	10 (7)	13 (8)	8 (8)
Definitely no root compression	3 (2)	1 (1)	3 (2)	1 (1)
Weeks between baseline and follow-up MRI	53.4±3.1	52.7±3.8	53.0±3.7	53.1±3.2
Roland Disability score ‡ *	16.3±4.4	16.1±4.0	16.7±4.2	15.4±4.1
VAS leg pain in mm § *	66.7±20.1	63.3±21.2	67.1±20.0	61.2±21.5
VAS back pain in mm §	33.6±29.5	30.5±27.1	33.9±30.4	28.7±23.9

Values are n (%) or means ± SD. N= 263.
No significant baseline differences were observed in the intention-to-treat group
* P<0.05 for the difference in the as-treated group
ò Body-mass index is the weight in kilograms divided by the square of the height in meters
‡ 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 2 Detailed description of the alteration of Vertebral Endplate Signal Changes (VESC) types between baseline and one year in the surgical and conservative group.		
	Received surgery (n=168)	Received no surgery (n=95)
No change in VESC type between baseline and one year	55 (33)	78 (82)
Change from no VESC at baseline to		
VESC Type 1 at one year	50 (30)	5 (5)
VESC Type 2 at one year	22 (13)	7 (7)
VESC Type 1 and Type 2 at one year	13 (8)	1 (1)
Change from VESC type 1 at baseline to VESC Type 2 at one year	2 (1)	0 (0)
Change from VESC type 2 at baseline to		
VESC Type 1 at one year	7 (4)	2 (2)
VESC Type 1 and 2 at one year	7 (4)	0 (0)
Different changes: at one endplate change from no VESC to VESC Type 1 and at another endplate change from VESC type 2 at baseline to VESC Type 1 at one year	9 (5)	1 (1)
Otherwise	3 (2)	1 (1)

Values are n (%)

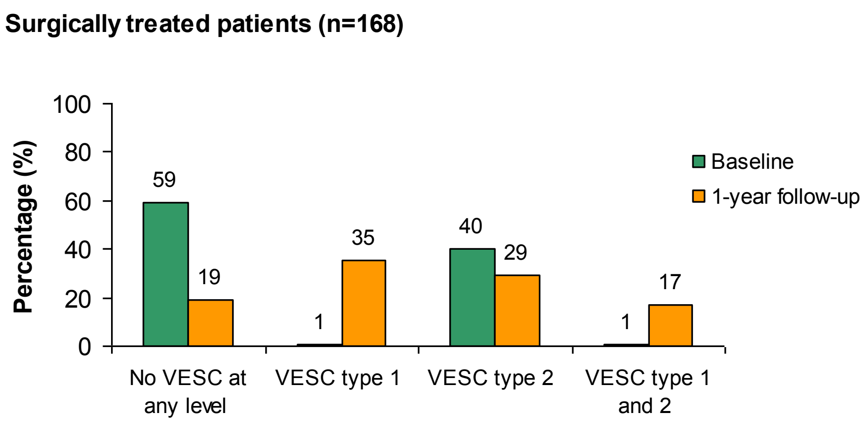
7 and 6 years post-residency for the neuroradiologists and 4 years post-residency for the neurosurgeon. The observers hold senior positions in busy spinal clinics with a focus on advanced spine surgery, and are confronted with spinal MRIs on a daily basis.

OUTCOME

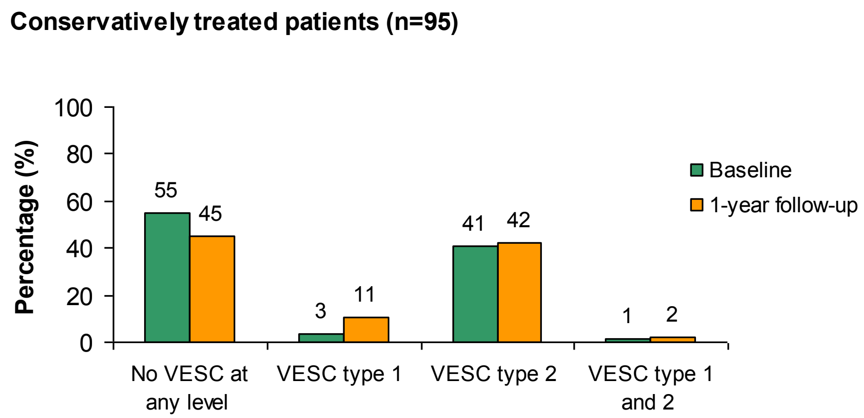
Patients were assessed by means of the 100-mm visual-analogue scale (VAS) for back pain (with 0 representing no pain and 100 the worst pain ever experienced) and a 7-point Likert self-rating scale of global perceived recovery ranging from completely recovered to much worse. The outcome measures were assessed at baseline, 2, 4, 8, 12, 26, 38 and 52 weeks. The patients did not see the results of earlier assessments and were also blinded to the MRI results.

STATISTICAL ANALYSIS

The majority opinion of the two neuroradiologists and neurosurgeon regarding the different MRI characteristics (answer independently given by minimum 2 out of 3 observers) was used in the statistical analysis. In the cases all three observers gave a different answer (e.g. observer A reported no VESC, observer B VESC Type 1 and observer C VESC Type 2), an additional senior neurosurgeon (15 years post-residency experience) independently evaluated the cases of disagreement, and his opinion regarding the VESC type was subsequently used in the statistical analysis. Interreader agreement between the three observers for the baseline and one year follow-up images was assessed with absolute percentages of agreement and kappa coefficients. At the design phase it was pre-specified that kappa values would be calculated only for findings



1A



1B

Figure 1 Prevalence of Vertebral Endplate Signal Changes (VESC) types at baseline and one year later in **1A)** surgically treated patients and **1B)** conservatively treated patients.
Analysis based on the "as-treated" groups.

reported in more than 10 and less than 90% of all reports[40] since the kappa statistic is affected by the prevalence of the events, so that findings with very high or low prevalence lead to very low kappa values, even if the observer agreement is high.[41, 42]

Disabling back pain was defined in the research group consensus meeting as a VAS for back pain of at least 40 mm, as this cutt-off value is regularly used when the VAS is categorized into favorable and unfavorable outcome.[43, 44] Perceived recovery on the 7-point Likert scale was defined as "complete" or "nearly complete recovery". The other (five) categories corresponded to "unsuccessful recovery".

Baseline and follow up characteristics of the surgical and conservative treatment group were compared using Student's t-test for continuous data and Chi-square tests for categorical data. Logistic regression analysis (univariate and multivariate analysis) was used to determine which

Table 3 Factors associated with presence of Vertebral Endplate Signal Changes (VESC) at one or more lumbar levels at baseline (n=263).

Association between general factors and VESC at any lumbar level	Comparison (%)	Univariate analysis			Multivariate analysis		
		OR	95%CI	P-value	OR	95%CI	P-value
VAS back pain	Per additional score	1.00	0.99-1.01	0.67			
Age	Per additional year of age	1.07	1.04-1.10	**<0.001**	1.06	1.03-1.09	**<0.001**
BMI	Per additional unit	1.04	0.97-1.11	0.29			
Gender	Male (68) vs female (32)	0.56	0.33-0.94	**0.029**	0.49	0.27-0.86	**0.013**
Presence of Schmorl's nodes at one or more lumbar level	Yes (11) vs no (89)	2.60	1.18-5.72	**0.017**	2.98	1.28-6.94	**0.012**
Presence of impaired discs at more than one level	Yes (79) vs no (21)	3.09	1.57-6.09	**0.001**	2.50	1.22-5.12	**0.012**

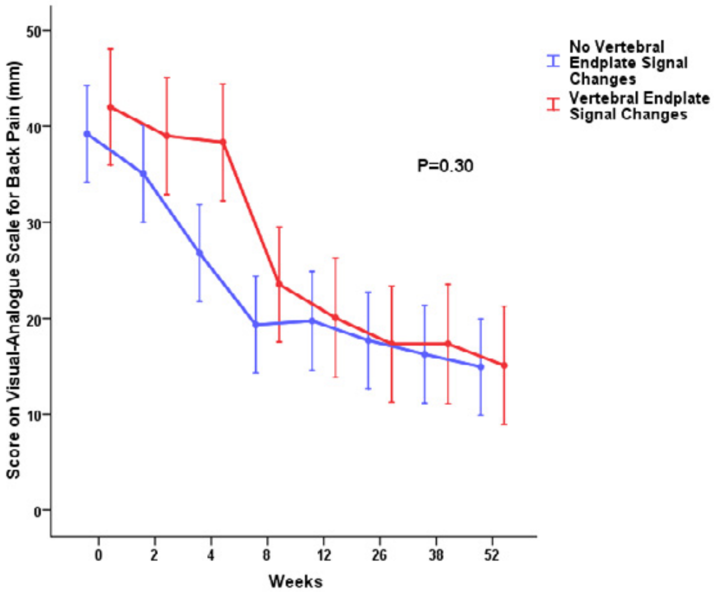
OR denotes odds ratio. CI denotes confidence interval.

baseline factors were associated with the presence of VESC at baseline. Repeated measurement analysis of variance was applied when analyzing differences in mean VAS-back pain during follow-up between patients with and without VESC. Since we specifically wanted to determine the influence of surgical treatment versus conservative treatment on progression of VESC, all analyses were performed according to the per-protocol analysis. As sensitivity analyses, we performed analysis excluding the cases in which all three MRI assessors disagreed on the VESC type. A P value of <0.05 was considered statistically significant.

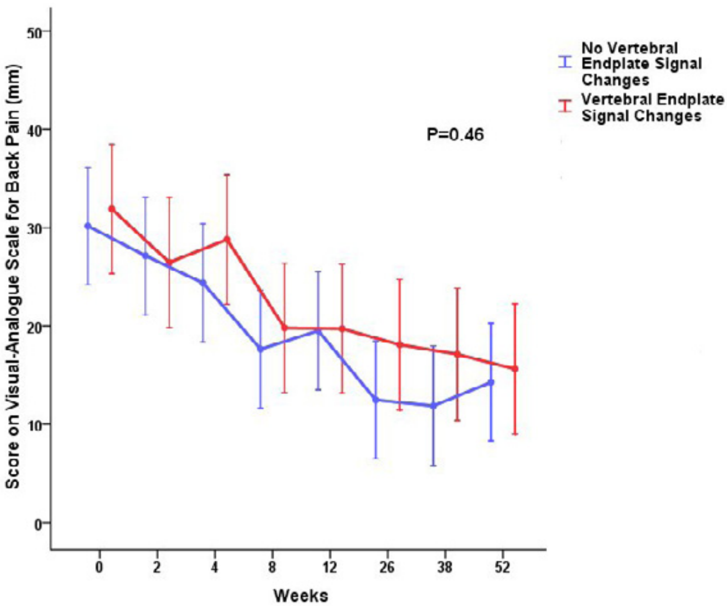
RESULTS

PATIENT CHARACTERISTICS

Of the 599 patients screened for the study, 283 patients were randomized. One year after randomization a second MRI was available for 267 (94.3%) of patients (Appendix Table S2). Baseline characteristics were similar among randomized patients for whom a second MRI was available compared to those for whom not. In total, 263 one-year MRIs could be evaluated properly due to the availability of both T1 and T2 images. Of the 263 patients who were eligible to be analyzed for the current study, 129 patients were randomized to early surgery and 134 to prolonged conservative care. Of the 129 patients randomized to early surgery, 15 recovered before surgery could be performed. Of the 134 patients assigned to prolonged conservative care, 54 eventually received surgery within the first year.

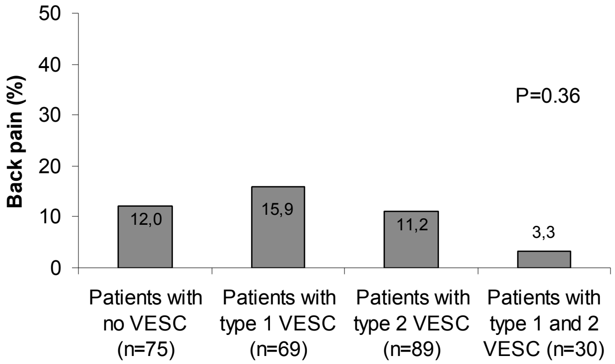


2A Surgically treated patients

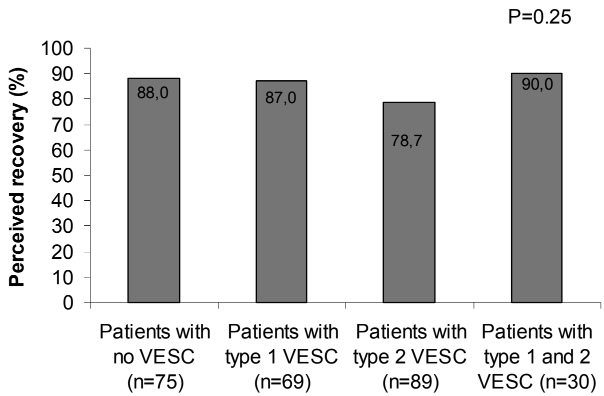


2B Conservatively treated patients

Figure 2 Repeated measurement analysis curve of Mean Scores for back pain on the Visual Analogue Scale. Sciatica patients with and without vertebral endplate signal changes on baseline MRI were compared. The vertical bars represent 95% confidence intervals.



3A



3B

Figure 3A) Disabling back pain *at one year* according to the type of Vertebral Endplate Signal Changes *at one year*. Disabling back pain was defined as a visual analogue scale score of at least 40mm on a scale of 0-100 (with 0 representing no pain and 100 the worst pain ever experienced)
3B) Perceived recovery *at one year* according to the type of Vertebral Endplate Signal Changes *at one year*. Recovery was defined as "complete" or "nearly complete recovery" on the 7-point Likert scale. The other (five) categories corresponded to "No recovery". VESC denotes Vertebral Endplate Signal Changes.

Absolute percentages of pairwise agreement among the three MRI observers for the presence and type of VESC varied from 75 to 99% (Appendix Table S3). As the prevalence of some VESC types were too low (< 10% of the reports) we did not calculate any kappa values for VESC. In 8 baseline MRIs and 9 one-year MRIs (3.4%) all three observers gave a different score regarding VESC type in the same patient (one reader no VESC, one reader VESC type 1 and one reader VESC type 2).

VESC FINDINGS AT BASELINE AND ONE YEAR FOLLOW-UP

At baseline, VESC were observed in 41% of 168 surgically treated patients compared to 45% of 95 conservatively treated patients (P=0.51). At baseline there was no difference in the types of VESC between surgically and conservatively treated patients (P=0.39) (Table 1). When VESC were considered present at baseline, 91% in the conservative treatment group displayed VESC type 2 compared to 97% in the surgical group.

At one year follow-up, 67% of the patients who had undergone surgery altered in VESC type compared to 18% of the patients who were treated conservatively (P<0.001). In the surgical group the most common conversion was from no VESC to VESC type 1, while in the conservative group slightly more conversions were from no VESC to VESC type 2 (Table 2). The prevalence of VESC type 1 increased from a prevalence of 1% at baseline to 35% one year later in surgically treated patients compared to an increase from 3% to 11% in conservatively treated patients (Figure 1). The prevalence of VESC type 2 decreased from 40 to 29% in the surgical group while remaining stable at about 41% in the conservative group.

At one year follow-up, 67% of the patients of surgically treated patients showed an increase in the extent of VESC compared to 19% of conservatively treated patients (P<0.001, Appendix Figure S1). A decrease in the extent of VESC after one year was observed in a minority of patients: 2% of surgically treated patients displayed a decrease in VESC compared to 5% of the conservatively treated patients.

FACTORS ASSOCIATED WITH THE PRESENCE AND CHANGE OF VESC

The presence of VESC at one or more levels at baseline was significantly associated with increasing age, female gender, the presence of Schmorl's nodes and the presence of impaired disc levels at one or more levels (Table 3). Considering only the impaired disc level that was assumed by the observers to cause the lumbosacral radicular syndrome, the presence of VESC at this level was significantly associated with loss of disc height of the same disc level, female gender, presence of VESC at other levels and presence of Schmorl's nodes (Appendix Table S4). Undergoing surgery was significantly associated with increase in the extent of VESC between baseline and one year (OR 8.56, 95% CI 4.67-15.67, P<0.001).

CORRELATION BETWEEN VESC AND CLINICAL OUTCOME (DISABLING BACK PAIN AND RECOVERY)

At baseline, 40% of the patients with VESC had disabling back pain compared to 38% of the patients with no VESC (P=0.67). Of the patients with no or mild VESC at baseline 38% had disabling back pain compared to 41% of the patients with moderate to severe VESC (P=0.75). Patients who were surgically treated and displayed VESC at baseline reported higher VAS back pain scores during the first 8 weeks compared to surgically treated patients who had not displayed VESC at baseline, but after this short-term period the mean VAS back pain scores of these two groups converged (Figure 2).

At one-year follow-up, the prevalence of disabling back pain was 12% in patients with no VESC at one year, 16% in patients with type 1 VESC, 11% in patients with type 2 VESC and 3% in patients with both type 1 and 2 VESC (P=0.36) (Figure 3A). When stratifying according to received treatment during the first year also no significant differences in the prevalence of disabling back pain existed between patients with the different types of VESC (Figure S2, P=0.29 in patients who had undergone surgery and P=0.93 in patients who had not undergone surgery). Of the patients with no or mild VESC at one year 11% had disabling back pain compared to 14% of the patients with moderate or severe VESC (P=0.39). Patients who showed an increase in the extent of VESC between baseline and one year did not significantly report more disabling back pain at one year compared to patients who did not show any increase in the extent of VESC (Odds ratio [OR] 1.21, 95% Confidence Interval [CI] 0.57-2.58, P=0.61).

Of the patients with VESC at one year 84% reported perceived recovery compared to 88% of the patients with no VESC (P=0.36). No significant differences in the prevalence of perceived recovery existed among patients with the different types of VESC (P=0.25) (Figure 3B). In addition, patients who showed an increase in the extent of VESC over one year did not significantly report less recovery compared to patients who did not show any increase in the extent of VESC (OR 1.57, 95% CI 0.79-3.11, P=0.20).

Sensitivity analyses to account for disagreement in VESC type yielded similar results (Figure S3). Also similar results were obtained when the analyses were stratified according to no VESC, VESC at one level and VESC at more than one level (Figure S4).

DISCUSSION

Undergoing disc surgery for sciatica was highly associated with progression in the extent of VESC compared to non-operative care in this study. In one year about two thirds of surgically and one fifth of conservatively treated patients displayed an increase of VESC. However, both at baseline and after one year follow-up, those with and those without VESC reported disabling back pain in nearly the same proportion. In addition, the proportion of patients reporting perceived recovery after one year was also nearly equally distributed between those with and without VESC. Therefore the results do present evidence that VESC are not responsible for disabling back pain in patients with sciatica. This remarkable scientific finding is in contrast with the intuitive intervention-prognostic diagnostic and treatment regimen of spinal physicians.

Studies correlating VESC to back pain in patients with sciatica are limited with conflicting results.[23] VESC have been reported to be associated with low back pain in the general population aged 40 years[26] and in working populations.[27, 45] Two studies did not observe more VESC among chronic low back pain patients compared to control subjects,[46] or between VESC and previous back pain in subjects without current back pain or sciatica.[31] Two earlier studies investigated the correlation between VESC and low back pain in patients treated for lumbar disc herniations, with contradictory results to the present study. Barth et al. evaluated MR images of 84 surgically treated patients for lumbar disc herniations.[47] Unfortunately the VESC were described by only one radiologist and MRI follow-up time ranged from 18 to 29 months.

After exclusion of reoperated patients, pre- and postoperative images were only available for 19 of 32 patients in the microdiscectomy group. Although back pain was significantly associated with progressive endplate changes, the clinical relevance of the association might be limited due to the relatively low observed spearman correlation coefficient (r=0.343).[48] No analysis was presented stratified according to VESC type. Albert et al. evaluated VESC in patients treated conservatively for sciatica.[20] Unfortunately the VESC were also described by only one radiologist. At 14 months follow-up, 60% of patients with VESC had self-reported back pain compared to 20% of patients without VESC. However, the proportion of patients with back pain did not significantly differ between the VESC types. Possibly, the results of the current study are contradictory with these two studies due to the definition of back pain. While they used self-reported back pain as the outcome we defined 'disabling back pain' based on patients' reported VAS for back pain.

The causes of VESC are not known. One theory is that toxic substances produced after damage of a disc invade the endplates and vertebral bones through micro fractures in the endplates and cause an inflammatory reaction.[49] Trauma to a disc by surgery causing the production of irritating substances may therefore accelerate the progression of VESC. The finding in this study of considerably more VESC in surgically treated compared to conservatively treated patients after one-year follow-up supports this theory. The extent of excision of the herniated disc might also be well correlated with the extent of the development of VESC. In support of this hypothesis is the study of Barth et al. who observed that patients who underwent standard discectomy (removal of herniated material plus discal tissue from the intervertebral space) developed significantly more VESC as compared to patients who underwent the less invasive sequestrectomy (only removal of the herniated material).[47]

The results of the present study are in line with previous studies showing a positive association between the prevalence of VESC with increasing age,[19, 21, 23] disc degeneration (loss of disc height, presence of impaired disc at more than one level)[17, 18, 23, 49] and Schmorl's impressions.[49] However, the finding of an association between female gender and VESC differs from previous findings in the literature showing no difference in the prevalence rates in relation to gender.[23] Two studies that examined VESC in unoperated sciatica patients did not provide any information regarding the relation of gender and prevalence of VESC.[20, 21] The current study finding should be confirmed by future research.

The most common VESC at baseline was VESC Type 2, a finding in concordance with previous studies in unoperated sciatica patients.[20, 21] The most common conversion in the surgical group was progression from no VESC at any level to type 1. In the study of Rahm et al. most patients developed VESC type 2 changes after lumbar discectomy.[50] However, contrary to the 12 months time interval between initial and follow-up MRI in this study, their interval varied from 32 to 59 months. In general it is agreed that VESC type 1 are unstable lesions which may convert to type 2 or back to normal.[51] The high observed prevalence of type 1 lesions at 12 months may still represent the more active stage of inflammation following disc surgery and these lesions may convert to type 2 or back to normal over time. Furthermore, the observation that 81% of conservatively treated patients did not convert from one VESC type to another after one year is in concordance with longitudinal studies that investigated the natural course of VESC and have observed that 48% to 86% of people do not convert from one VESC type to another over periods of 14 to 72 months.[17, 19-21, 52]

An important limitation to be considered is that the study population consisted of sciatica patients raising the difficulty of generalizing the results to a population with back pain but without radicular symptoms. In general prevalence of VESC is higher in clinical than in non-clinical populations.[23] However, after one year the overwhelming majority of patients recovered from sciatic symptoms. Still patients exhibiting VESC at one year did not report more back pain compared to patients who did not display any VESC at one year. Also approximately 3% of the cases had three different VESC readings. However, similar results were obtained when those cases were left out of the analyses. Finally, the reported VESC and their relation with back pain were timed only once, one year after randomization. Although seemingly generalizable to other time points it is scientifically uncertain if we would have found comparable results at other moments.

In summary, in this one year follow-up study undergoing surgery for sciatica was highly associated with the development of VESC. However, both at baseline and after one year, those with and those without VESC reported disabling back pain in nearly the same proportion. Therefore the results indicate that VESC are not responsible for disabling low back pain in patients with sciatica and one should be reticent to offer back surgery based on MRI endplate changes.

REFERENCES

1. Vroomen PC, de Krom MC, Slofstra PD, Knottnerus JA. Conservative treatment of sciatica: a systematic review. J Spinal Disord 2000;13:463-9.
2. den Boer JJ, Oostendorp RA, Beems T, Munneke M, Oerlemans M, Evers AW. A systematic review of bio-psychosocial risk factors for an unfavourable outcome after lumbar disc surgery. Eur Spine J 2006;15:527-36.
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. Atlas SJ, Keller RB, Wu YA, Deyo RA, Singer DE. Long-term outcomes of surgical and nonsurgical management of sciatica secondary to a lumbar disc herniation: 10 year results from the maine lumbar spine study. Spine (Phila Pa 1976) 2005;30:927-35.
5. Findlay GF, Hall BI, Musa BS, Oliveira MD, Fear SC. A 10-year follow-up of the outcome of lumbar microdiscectomy. Spine (Phila Pa 1976) 1998;23:1168-71.
6. Korres DS, Loupassis G, Stamos K. Results of lumbar discectomy: a study using 15 different evaluation methods. Eur Spine J 1992;1:20-4.
7. Lequin MB, Verbaan D, Jacobs WC, et al. Surgery versus prolonged conservative treatment for sciatica: 5-year results of a randomised controlled trial. BMJ Open 2013;3.
8. Vucetic N, Astrand P, Guntner P, Svensson O. Diagnosis and prognosis in lumbar disc herniation. Clin Orthop Relat Res 1999:116-22.
9. Weber H. Lumbar disc herniation. A controlled, prospective study with ten years of observation. Spine (Phila Pa 1976) 1983;8:131-40.
10. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonoperative treatment for lumbar disc herniation: four-year results for the Spine Patient Outcomes Research Trial (SPORT). Spine (Phila Pa 1976) 2008;33:2789-800.
11. den Boer JJ, Oostendorp RA, Beems T, Munneke M, Evers AW. Continued disability and pain after lumbar disc surgery: the role of cognitive-behavioral factors. Pain 2006;123:45-52.
12. den Boer JJ, Oostendorp RA, Evers AW, Beems T, Borm GF, Munneke M. The development of a screening instrument to select patients at risk of residual complaints after lumbar disc surgery. Eur J Phys Rehabil Med 2010;46:497-503.
13. Juniper M, Le TK, Mladsi D. The epidemiology, economic burden, and pharmacological treatment of chronic low back pain in France, Germany, Italy, Spain and the UK: a literature-based review. Expert Opin Pharmacother 2009;10:2581-92.
14. Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. Spine J 2008;8:8-20.
15. Woolf AD, Pfleger B. Burden of major musculoskeletal conditions. Bull World Health Organ 2003;81:646-56.
16. Modic MT, Masaryk TJ, Ross JS, Carter JR. Imaging of degenerative disk disease. Radiology 1988;168:177-86.
17. Modic MT, Steinberg PM, Ross JS, Masaryk TJ, Carter JR. Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging. Radiology 1988;166:193-9.
18. Kjaer P, Korsholm L, Bendix T, Sorensen JS, Leboeuf-Yde C. Modic changes and their associations with clinical findings. Eur Spine J 2006;15:1312-9.
19. Jensen TS, Bendix T, Sorensen JS, Manniche C, Korsholm L, Kjaer P. Characteristics and natural course of vertebral endplate signal (Modic) changes in the Danish general population. BMC Musculoskelet Disord 2009;10:81.

20. Albert HB, Manniche C. Modic changes following lumbar disc herniation. Eur Spine J 2007;16:977-82.
21. Kuisma M, Karppinen J, Niinimaki J, et al. A three-year follow-up of lumbar spine endplate (Modic) changes. Spine (Phila Pa 1976) 2006;31:1714-8.
22. Kjaer P, Leboeuf-Yde C, Sorensen JS, Bendix T. An epidemiologic study of MRI and low back pain in 13-year-old children. Spine (Phila Pa 1976) 2005;30:798-823. Jensen TS, Karppinen J, Sorensen JS, Niinimaki J, Leboeuf-Yde C. Vertebral endplate signal changes (Modic change): a systematic literature review of prevalence and association with non-specific low back pain. Eur Spine J 2008;17:1407-22.
24. Braithwaite I, White J, Saifuddin A, Renton P, Taylor BA. Vertebral end-plate (Modic) changes on lumbar spine MRI: correlation with pain reproduction at lumbar discography. Eur Spine J 1998;7:363-8.
25. Weishaupt D, Zanetti M, Hodler J, et al. Painful Lumbar Disk Derangement: Relevance of Endplate Abnormalities at MR Imaging. Radiology 2001;218:420-7.
26. Kjaer P, Leboeuf-Yde C, Korsholm L, Sorensen JS, Bendix T. Magnetic resonance imaging and low back pain in adults: a diagnostic imaging study of 40-year-old men and women. Spine (Phila Pa 1976) 2005;30:1173-80.
27. Kuisma M, Karppinen J, Niinimaki J, et al. Modic changes in endplates of lumbar vertebral bodies: prevalence and association with low back and sciatic pain among middle-aged male workers. Spine (Phila Pa 1976) 2007;32:1116-22.
28. Cvitanic OA, Schimandle J, Casper GD, Tirman PF. Subchondral marrow changes after laser diskectomy in the lumbar spine: MR imaging findings and clinical correlation. AJR Am J Roentgenol 2000;174:1363-9.
29. Sandhu HS, Sanchez-Caso LP, Parvataneni HK, Cammisa FP, Jr., Girardi FP, Ghelman B. Association between findings of provocative discography and vertebral endplate signal changes as seen on MRI. J Spinal Disord 2000;13:438-43.
30. Kokkonen SM, Kurunlahti M, Tervonen O, Ilkko E, Vanharanta H. Endplate degeneration observed on magnetic resonance imaging of the lumbar spine: correlation with pain provocation and disc changes observed on computed tomography diskography. Spine (Phila Pa 1976) 2002;27:2274-8.
31. Jarvik JJ, Hollingworth W, Heagerty P, Haynor DR, Deyo RA. The Longitudinal Assessment of Imaging and Disability of the Back (LAIDBack) Study: baseline data. Spine (Phila Pa 1976) 2001;26:1158-66.
32. Buttermann GR, Heithoff KB, Ogilvie JW, Transfeldt EE, Cohen M. Vertebral body MRI related to lumbar fusion results. Eur Spine J 1997;6:115-20.
33. Chataigner H, Onimus M, Polette A. [Surgery for degenerative lumbar disc disease. Should the black disc be grafted?]. Rev Chir Orthop Reparatrice Appar Mot 1998;84:583-9.
34. Eser O, Gomleksiz C, Sasani M, et al. Dynamic stabilisation in the treatment of degenerative disc disease with modic changes. Adv Orthop 2013;2013:806267.
35. Esposito P, Pinheiro-Franco JL, Froelich S, Maitrot D. Predictive value of MRI vertebral end-plate signal changes (Modic) on outcome of surgically treated degenerative disc disease. Results of a cohort study including 60 patients. Neurochirurgie 2006;52:315-22.
36. Weinstein JN, Lurie JD, Olson PR, Bronner KK, Fisher ES. United States' trends and regional variations in lumbar spine surgery: 1992-2003. Spine (Phila Pa 1976) 2006;31:2707-14.
37. Cherkin DC, Deyo RA, Loeser JD, Bush T, Waddell G. An international comparison of back surgery rates. Spine (Phila Pa 1976) 1994;19:1201-6.
38. Blinded for the review process.

39. 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.
40. Arana E, Royuela A, Kovacs FM, et al. Lumbar spine: agreement in the interpretation of 1.5-T MR images by using the Nordic Modic Consensus Group classification form. Radiology 2010;254:809-17.
41. Gjorup T. The kappa coefficient and the prevalence of a diagnosis. Methods Inf Med 1988;27:184-6.
42. Kovacs FM, Royuela A, Jensen TS, et al. Agreement in the interpretation of magnetic resonance images of the lumbar spine. Acta Radiol 2009;50:497-506.
43. 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.
44. 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.
45. Schenk P, Laubli T, Hodler J, Klipstein A. Magnetic resonance imaging of the lumbar spine: findings in female subjects from administrative and nursing professions. Spine (Phila Pa 1976) 2006;31:2701-6.
46. Kovacs FM, Arana E, Royuela A, et al. Vertebral endplate changes are not associated with chronic low back pain among Southern European subjects: a case control study. AJNR Am J Neuroradiol;33:1519-24.
47. 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.
48. Portney L, Watkins M. Foundations of Clinical Research: Applications to Practice. III ed. Upper Saddle River, NJ: Pearson/Prentice Hall; 2009.
49. Albert HB, Kjaer P, Jensen TS, Sorensen JS, Bendix T, Manniche C. Modic changes, possible causes and relation to low back pain. Med Hypotheses 2008;70:361-8.
50. Rahme R, Moussa R, Bou-Nassif R, et al. What happens to Modic changes following lumbar discectomy? Analysis of a cohort of 41 patients with a 3- to 5-year follow-up period. J Neurosurg Spine 2010;13:562-7.
51. Rahme R, Moussa R. The modic vertebral endplate and marrow changes: pathologic significance and relation to low back pain and segmental instability of the lumbar spine. AJNR Am J Neuroradiol 2008;29:838-42.
52. Mitra D, Cassar-Pullicino VN, McCall IW. Longitudinal study of vertebral type-1 end-plate changes on MR of the lumbar spine. Eur Radiol 2004;14:1574-81.06.

Table S1 MRI study variables

Disc level	Variable	Category
Separate for every end plate from level L2-L3 through L5-S1	Presence of vertebral end plate signal changes (VESC)	1. No VESC 2. VESC type 1: hypointense on T1-weighted sequences and hyperintense on T2-weighted sequences 3. VESC type 2: increased signal on T1 weighted sequences and isointense or slightly hyperintense signal on T2 weighted sequences 4. VESC type 3: hypointense both on T1- and T2-weighted sequences 5. VESC type 1 and 2
	Extent of vertebral endplate signal changes	1. mild 2. moderate 3. severe
	Presence of Schmorl's nodes (herniation of the disc into the vertebral-body end plate)	1. Yes 2. No
From level L1-L2 through L5-S1	Presence of impaired discs at more than one level	1. Yes 2. No
Disc level with the most severe nerve root compression	Disc level	1. Not applicable: no symptomatic disc level 2. L2L3 3. L3L4 4. L4L5 5. L5S1
	Loss of disc height at this level	1. Yes 2. No
	Signal intensity of nucleus pulposus on T2 images at this level	1. Hypointensity 2. Normal 3. Hyperintensity
	Disc contour at this 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.

Table S1 (Continued)		
Disc level	**Variable**	**Category**
	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

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 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 Interobserver agreement regarding the type of Vertebral Endplate Signal Changes (VESC) and other MRI findings. The observers could choose from the following categories: No VESC, VESC type 1, VESC type 2, VESC type 3 and VESC type 1 and 2. A en B represent the neuroradiologists and C represents the neurosurgeon.
A) Interobserver agreement regarding the type of Vertebral Endplate Signal Changes (VESC)
B) Interobserver agreement regarding other MRI findings used in the current study
S3A

	A vs B		A vs C		B vs C		All observers	
	% agree-ment	kappa	% agree-ment	kappa	% agree-ment	kappa	% agree-ment	multi-rater kappa
Type of vertebral endplate signal changes upper endplate L2L3	98	*	98	*	99	*	97	*
Type of vertebral endplate signal changes lower endplate L2L3	97	*	96	*	97	*	96	*
Type of vertebral endplate signal changes upper endplate L3L4	95	*	96	*	97	*	94	*
Type of vertebral endplate signal changes lower endplate L3L4	96	*	96	*	97	*	94	*

Table S3 (Continued)								
	A vs B		A vs C		B vs C		All observers	
	% agree-ment	kappa	% agree-ment	kappa	% agree-ment	kappa	% agree-ment	multi-rater kappa
Type of vertebral endplate signal changes upper endplate L4L5	84	*	85	*	87	*	80	*
Type of vertebral endplate signal changes lower endplate L4L5	84	*	84	*	87	*	79	*
Type of vertebral endplate signal changes upper endplate L5S1	75	*	79	*	77	*	69	*
Type of vertebral endplate signal changes lower endplate L5S1	75	*	81	*	76	*	69	*

* Prevalence of some VESC types too low (< 10% of the reports) to calculate kappa values

S3B								
	A vs B		A vs C		B vs C		All observers	
	% agree-ment	kappa	% agree-ment	kappa	% agree-ment	kappa	% agree-ment	multi-rater kappa
Presence of Schmorl's nodes§	78	0.20	79	0.41	83	0.24	70	0.28
Presence of impaired discs at more than one level§	93	0.80	84	0.60	84	0.60	81	0.66
Characteristic of the impaired disc level that was assumed to cause the sciatic symptoms								
Level¶	97	0.95	99	0.97	98	0.96	97	0.96
Loss of disc height§	99	*	73	0.27	74	0.29	72	0.34
Intensity of nucleus pulposus on T2 images at one or more levelsΨ	96	*	92	*	90	*	89	*
Nerve root compression‡	90	*	90	0.56	90	0.56	89	0.72

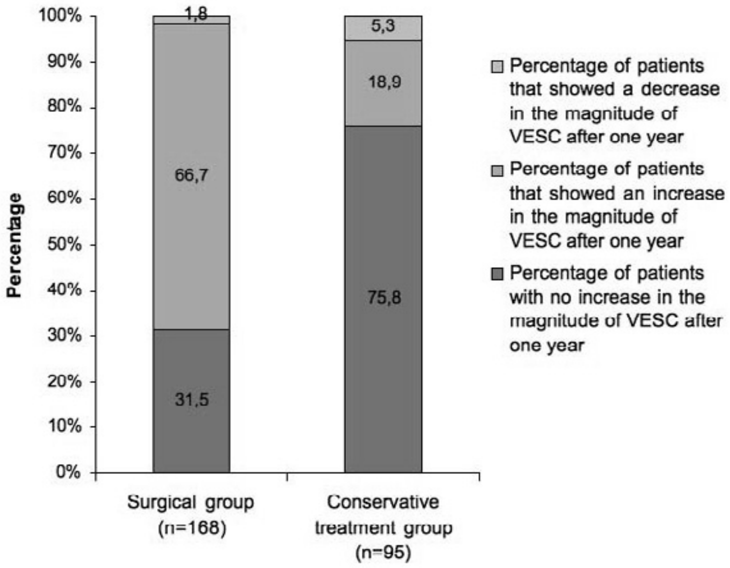
* Since kappa values are afected by the prevalence of events, kappa values were only calculated for findings reported in more than 10% and less than 90% of all reports.
§ Categories were: yes versus no.
¶ The 5 categories were: 1) L2L3 2) L3L4 3) L4L5 4) L5S1
Ψ Categories were: 1) Hypointensity 2) Normal 3) Hyperintensity.
‡ Categories were: probability> 5% vs probability<50%.

Table S4 Uni- and multivariate analysis to determine predictive value on the presence of Vertebral Endplate Signal Changes at the disc level that is assumed to cause the sciatic symptoms (n=263). OR denotes odds ratio. CI denotes confidence interval.

	Comparison (%)	Univariate analysis			Multivariate analysis		
		OR	95%CI	P-value	OR	95%CI	P-value
VAS back pain	Per additional score	1.00	0.99-1.01	0.42			

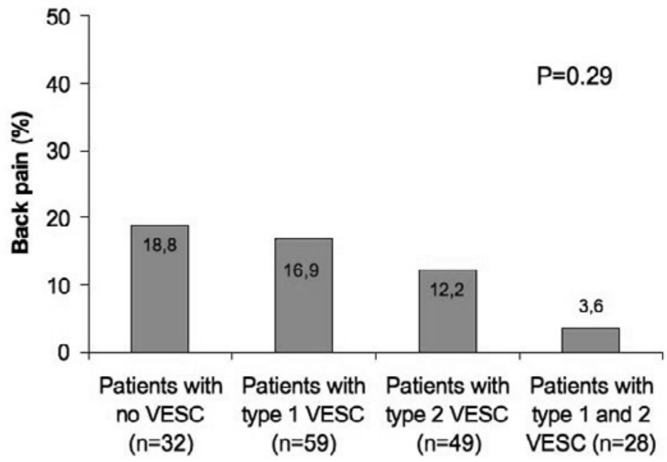
Table S4 (Continued)							
	Comparison (%)	Univariate analysis			Multivariate analysis		
Demographic variables							
Age	Per additional year	1.03	1.00-1.06	0.058			
Gender	Male (68) vs female (32)	0.44	0.26-0.76	**0.003**	0.46	0.26-0.81	**0.007**
MRI characteristics of the disc level that is assumed to cause the sciatic symptoms							
Disc level	L4L5 (44) vs L5S1 (56)	0.75	0.44-1.29	0.30			
Presence of Schmorl's nodes	Yes (5) vs no (95)	3.73	1.18-11.79	**0.025**	3.45	1.04-11.40	**0.042**
Loss of disc height at the disc level	Yes (91) vs no (9)	3.57	1.03-12.32	**0.044**	3.34	0.94-11.81	**0.062**
Signalintensity of nucleus pulposus on T2 images at one or more levels	Hypointens (91) vs normal (9)	2.24	0.73-6.83	0.16			
Presence of nerve root compression on MRI	Probability >50% (90) vs probability <50% (10)	0.98	0.40-2.37	0.96			
Presence of impaired discs at other disc levels	Yes (79) vs no (21)	1.72	0.87-3.40	0.12			
Presence of VESC at other disc levels	Yes (20) vs no (80)	2.12	1.14-3.93	**0.017**	1.99	1.04-3.83	**0.039**

Appendix Figure S1 Progression in the extent of VESC between baseline and one year in patients who underwent surgery and patients who received conservative care

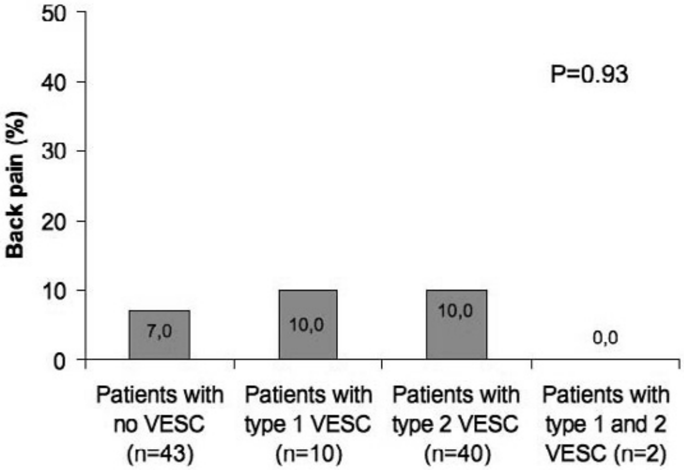


Appendix Figure S2 A) Disabling back pain *at one year* according to the type of Vertebral Endplate Signal Changes at one year in A) patients who underwent surgery during the first year and B) patients who underwent no surgery during the first year. Disabling back pain was defined as a visual analogue scale score of at least 40mm on a scale of 0-100 (with 0 representing no pain and 100 the worst pain ever experienced)

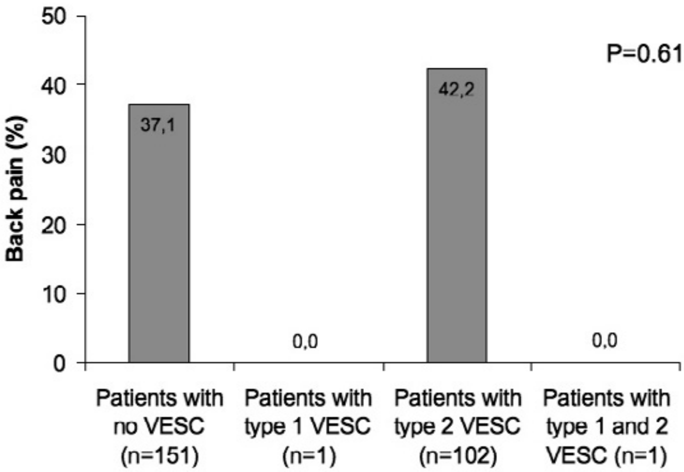
S2A Patients who underwent surgery during the first year



S2B Patients who underwent no surgery (conservative treatment) during the first year

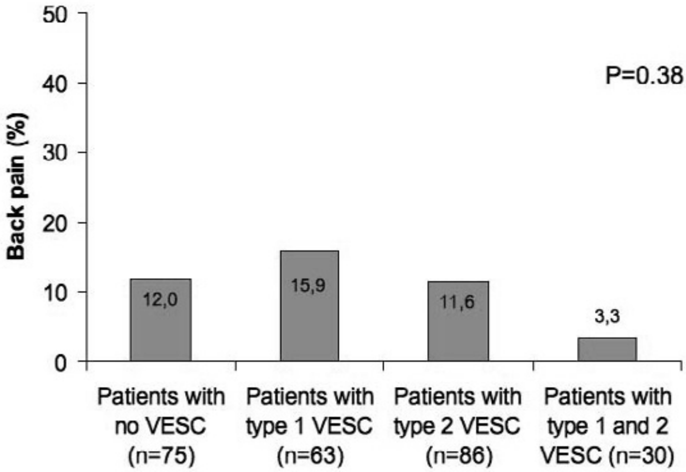


Appendix Figure S3 Disabling back pain at A) baseline and B) one year according to the type of Vertebral Endplate Signal Changes (VESC). This analysis only included patients in whom at least 2 out of the 3 MRI readers gave the same score regarding VESC type.
S3A Baseline (n=255).



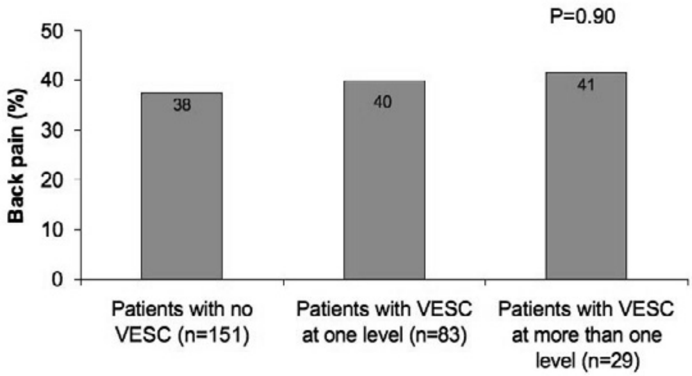
At baseline only 1 patient displayed VESC type 1 and 1 patient VESC type 1 and 2

S3B One-year follow-up (n=254)

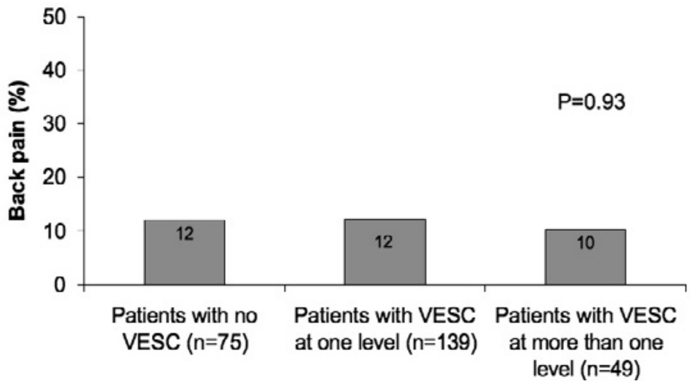


Appendix Figure S4 Disabling back pain at A) baseline and B) one year according to Vertebral Endplate Signal Changes (VESC) at one or more levels.

S4A Baseline



S4B One-year follow-up



Chapter 8

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

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

Submitted for publication

ABSTRACT

BACKGROUND CONTEXT

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

PURPOSE

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

STUDY DESIGN/SETTING

A randomized clinical trial with one year follow-up.

PATIENTS SAMPLE

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

OUTCOME MEASURES

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

METHODS

In total 204 patients underwent Gd-MRI at baseline and after one year. MRI findings were correlated to the outcome measures using The Mann-Whitney U test for continuous data and Fisher's exact tests for categorical data. This study was supported by a grant from the Netherlands Organisation for Health Research and Development (ZonMW) and the Hoelen Foundation, The Hague. None of the authors of this study has any conflict of interest.

RESULTS

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

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

CONCLUSION

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

INTRODUCTION

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

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

The specific objectives of the present study were to evaluate interobserver agreement among experienced readers regarding MRI enhancement findings, and how well enhancement of nerve root and disc herniation are correlated with clinical outcome and neurological findings at baseline and after one year.

METHODS

STUDY POPULATION AND RANDOMIZATION

Patients for this study were participants in a multicentre randomized controlled trial among patients with 6-12 weeks sciatica with a disc herniation on MRI. Patients were only included if they had a dermatomal pattern of pain distribution with concomitant neurologic disturbances that correlated with the same nerve root being affected on MRI. An early surgery strategy was compared to prolonged conservative care for an additional 6 months followed by surgery for patients who did not improve or who did request it earlier because of aggravating symptoms.[22] Patients were excluded if they were presenting with cauda equina syndrome, insufficient strength to move against gravity, identical complaints in the previous 12 months, previous spine surgery, pregnancy, spinal stenosis, spondylolisthesis, or severe coexisting disease.

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

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

The medical ethics committees at the nine participating hospitals approved the protocol. Written informed consent was obtained from all patients. Details of the design and study protocol were published previously.[28]

MRI PROTOCOL AND IMAGE EVALUATION

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

MRI scans were performed in all 9 participating hospitals using standardized protocols tailored to a 1.5 Tesla scanner. Sagittal T1 and axial T1 spin echo images of the lumbar spine were acquired. In addition, T2 weighted sagittal and axial images were obtained. For research purposes also contrast-enhanced (Gadolinium-DTPA at a standard dose of 0.1 mmol/kg body weight) T1 fat suppressed sagittal and axial images were obtained.

Two neuroradiologists (BK and GL) and one neurosurgeon (CV) independently evaluated all MR images. The readers were not provided any clinical information and had not been involved in the selection or care of the included patients. Before the start of the study, the readers met in person to evaluate and refine standardized definitions of imaging characteristics.

After reaching final consensus, standardized case record forms with these final definitions were used (Appendix Table S1). Observer experience in reading spine MRIs was 7 and 6 years post-residency for the neuroradiologists and 4 years post-residency for the neurosurgeon.

First, the blinded readers had to decide on the baseline MRI which disc level showed the most severe nerve root compression. For both the presence of disc herniation and nerve root compression a four point scale was used, ranging from 1 (definitely present) to 4 (definitely absent). The size of the disc herniation was also evaluated. The same disc level thought to cause symptoms at baseline was evaluated on the one-year MRI. On the one-year MRI the readers had also to assess whether scar tissue was present (no, moderate or severe). The readers evaluated the enhancement on the baseline and one year MRI of the following structures using different categories (Appendix Table S1): 1. Disc herniation (if present): no, any edge, complete circumferential or diffuse enhancement, 2. Most affected nerve root: no, mild or strong enhancement and 3. Scar tissue (if present at one year): yes vs no enhancement. Structures were considered enhanced when brighter compared with the precontrast image.

NEUROLOGICAL EXAMINATION

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

OUTCOMES

The outcome measures of the trial were the Roland Disability Questionnaire (RDQ) for Sciatica (scores range from 0 to 23, with higher scores indicating worse functional status),[30] the 100-mm visual-analogue scale (VAS) for leg pain (with 0 representing no pain and 100 the worst pain ever experienced),[31] and a 7-point Likert self-rating scale of global perceived recovery given by the question whether the patient experienced recovery, with answers ranging from completely recovered to much worse. Perceived recovery on the 7-point Likert scale was used in dichotomized form: "Complete" or "nearly complete disappearance of symptoms" was defined as "perceived recovery", while a score in the remaining five categories (varying from "minimally improved" to "very much worse") was marked as "no recovery".[22] These outcome measures were assessed at baseline, 2, 4, 8, 12, 26, 38 and 52 weeks.

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

STATISTICAL ANALYSIS

Interobserver agreement regarding the MRI findings was determined by use of absolute percentages of agreement and kappa values (weighted in case of ordered data). Since the kappa statistic is affected by the prevalence of the events,[32, 33] kappa values were only calculated for findings reported in more than 10% and less than 90% of all reports.[34] Kappa values and percentages of agreement for the enhancement of the structures were also only calculated if the observers marked the same structure as affected (e.g. when there was disagreement about the most affected nerve root in a patient, this patient did not contribute to the interagreement analysis regarding the enhancement of the most affected nerve root). Values of less than 0.00 indicated poor; 0.00-0.20 slight; 0.21-0.40 fair; 0.41-0.60 moderate; 0.61-0.80 substantial; and 0.81-1.00 excellent or almost perfect reliability.[35]

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

RESULTS

Of 599 patients screened for the Trial, 283 patients were randomized.[22] One year after randomization a second MRI was available for 267 (94.3%) patients. However, at baseline 230 (81%) underwent MRI with gadolinium and at one year 245 (87%) patients. No significant differences in patient characteristics existed between patients who underwent Gd-MRI and conventional MRI. In total 204 patients (72%) underwent Gd-MRI both at baseline and one year. Of the 204 patients who were eligible to be analyzed for the present study, 105 patients were randomized to early surgery and 99 to prolonged conservative care. Of the 105 patients randomized to early surgery, 12 patients recovered before surgery could be performed. Of the 99 patients randomized to prolonged conservative care, 36 eventually received surgery within the first year. Thus, during the first year after randomization 129 patients underwent surgery and 75 patients conservative care. Baseline characteristics of the intention-to-treat and the as-treated groups are demonstrated in Table 1.

INTERAGREEMENT ANALYSIS AT BASELINE

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

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

	Intention to treat		As treated	
	Randomized to early surgery (N=105)	Randomized to prolonged conservative care (N=99)	Received surgery (n=129)	Received no surgery (n=75)
Age	42.4±10.4	43.0±9.5	42.1±10.2	43.6±9.4
Male gender	66 (63)	71 (72)	80 (62)	57 (76)
Duration of sciatica in weeks	9.5±2.3	9.5±2.2	9.5±2.3	9.6±2.2
Suspected disc level				
L3L4	5 (5)	2 (2)	6 (5)	1 (1)
L4L5	48 (46)	35 (35)	54 (42)	29 (39)
L5S1	52 (50)	62 (63)	69 (53)	45 (60)
MRI assessed nerve root compression				
Definite	66 (63)	70 (71)	87 (67)	49 (65)
Probable	30 (29)	22 (22)	32 (25)	20 (27)
Possible	8 (8)	6 (6)	9 (7)	5 (7)
Definitely no root compression	1 (1)	1 (1)	1 (1)	1 (1)
Weeks between baseline and follow-up MRI	53.3±2.9	52.7±3.9	52.9±3.6	53.2±3.2
Roland Disability score ‡	16.2±4.3	15.9±3.9	16.2±4.3	15.7±3.8
VAS leg pain in mm §	66.1±20.0	62.0±21.1	65.6±20.5	61.7±20.8
VAS back pain in mm §	33.4±29.0	28.5±25.9	32.7±29.7	28.0±23.2

Values are n (%) or means ± SD.
No significant baseline differences were observed in the intention-to-treat group and the as-treated groups
‡ The Roland Disability Questionnaire for Sciatica measures the functional status of patients with pain in the leg or back. Scores range from 0 to 23, with higher scores indicating worse functional status.
§ The intensity of pain is indicated on a horizontal 100 mm visual analogue scale (VAS) with 0 representing no pain and 100 the worst pain ever experienced.

INTERAGREEMENT ANALYSIS ONE YEAR

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

Table 2 Interobserver agreement regarding MRI findings at baseline.

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

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

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

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

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

MRI FINDINGS

Baseline

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

One year

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

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

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

BASELINE ENHANCEMENT FINDINGS IN RELATION TO CLINICAL DATA

Patients with and without an enhancing herniated disc at baseline showed comparable baseline scores on the RDQ and VAS for leg and back pain (Table 5). At baseline, 80% of the patients with enhancing disc herniation had muscle weakness compared to 62% of the patients with

Table 3 Interobserver agreement regarding MRI findings at one year.

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

A en B represent the two neuroradiologists, while C represents the neurosurgeon. Kappa values and percentages of agreement for the enhancement of the structures were only calculated if the observers marked the same structure as affected (e.g. when there was disagreement about whether at one year a herniated disc was still visible, this case did not contribute to the interagreement analysis regarding the enhancement of the herniated disc).

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

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

╞ Yes vs. no

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

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

Table 4 Differences in 1-year MRI findings between patients with and without surgery during the first year (as-treated). Values are n (%). Total n=204

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

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

Patients with and without enhancement of the herniated disc or affected nerve root at baseline showed comparable scores on RDQ, VAS-leg and Likert scale of global perceived recovery after one year (Table 5). Patients with enhancing nerve roots reported lower VAS-back pain scores at one year compared to patients with no enhancing nerve roots at baseline (9.9

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

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

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

¶ The intensity of pain is indicated on a horizontal 100 mm visual analogue scale (VAS) with 0 representing no pain and 100 the worst pain ever experienced

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

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

Values are n (%) or means ± SD.
‡ The Roland Disability Questionnaire for Sciatica is a disease-specific disability scale that measures the functional status of patients with pain in the leg or back. Scores range from 0 to 23, with higher scores indicating worse functional status
¶ The intensity of pain is indicated on a horizontal 100 mm visual analogue scale (VAS) with 0 representing no pain and 100 the worst pain ever experienced

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

One-year enhancement in relation to one-year clinical data

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

Discussion

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

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

Since the interobserver agreement regarding the enhancement findings was poor to moderate, one could question the added value of correlating enhancement findings with clinical findings. With the exception of one study (kappa=0.66 for nerve root enhancement between two radiologists)[26] no prevailing studies reported on the interobserver agreement with regard to the enhancement findings. Within the radiological literature, values of agreement show a high variation depending on the variable investigated.[36] Even regarding the most involved disc level, important for making treatment decisions, in 3% of the cases disagreement arose in this study, which is in agreement with previous literature.[37] However, it is crucial that radiologists and clinicians strive to reduce variability in interpretations as inconsistency in interpretation may lead to alternative treatment options between clinicians and therefore may impact the outcome of patient treatment.[38, 39] Moreover, to gain more insight in the relationship between specific imaging characteristics and patient outcomes, those interpreting the images must reliably assess the finding. One reason that a prediction model might lose its predictive power is the incorrect assessment of MRI findings, which causes the inputs in the prediction model to be faulty.[40]

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

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

REFERENCES

1. Bejia I, Younes M, Zrour S, Touzi M, Bergaoui N. Factors predicting outcomes of mechanical sciatica: a review of 1092 cases. Joint Bone Spine 2004;71:567-71.
2. Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. Spine J 2008;8:8-20.
3. Koes BW, van Tulder MW, Peul WC. Diagnosis and treatment of sciatica. BMJ 2007;334:1313-7.
4. Vroomen PC, de Krom MC, Slofstra PD, Knottnerus JA. Conservative treatment of sciatica: a systematic review. J Spinal Disord 2000;13:463-9.
5. Weber H, Holme I, Amlie E. The natural course of acute sciatica with nerve root symptoms in a double-blind placebo-controlled trial evaluating the effect of piroxicam. Spine (Phila Pa 1976) 1993;18:1433-8.
6. Cihangiroglu M, Yildirim H, Bozgeyik Z, et al. Observer variability based on the strength of MR scanners in the assessment of lumbar degenerative disc disease. Eur J Radiol 2004;51:202-8.
7. Vucetic N, Astrand P, Guntner P, Svensson O. Diagnosis and prognosis in lumbar disc herniation. Clin Orthop Relat Res 1999:116-22.
8. Peul WC, Brand R, Thomeer RT, Koes BW. Influence of gender and other prognostic factors on outcome of sciatica. Pain 2008;138:180-91.
9. Peul WC, van den Hout WB, Brand R, Thomeer RT, Koes BW. Prolonged conservative care versus early surgery in patients with sciatica caused by lumbar disc herniation: two year results of a randomised controlled trial. BMJ 2008;336:1355-8.
10. Osterman H, Seitsalo S, Karppinen J, Malmivaara A. Effectiveness of microdiscectomy for lumbar disc herniation: a randomized controlled trial with 2 years of follow-up. Spine (Phila Pa 1976) 2006;31:2409-14.
11. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonoperative treatment for lumbar disc herniation: four-year results for the Spine Patient Outcomes Research Trial (SPORT). Spine (Phila Pa 1976) 2008;33:2789-800.
12. Arts MP, Brand R, van den Akker ME, Koes BW, Bartels RH, Peul WC. Tubular diskectomy vs conventional microdiskectomy for sciatica: a randomized controlled trial. JAMA 2009;302:149-58.
13. Bernard TN, Jr. Repeat lumbar spine surgery. Factors influencing outcome. Spine (Phila Pa 1976) 1993;18:2196-200.
14. Fandino J, Botana C, Viladrich A, Gomez-Bueno J. Reoperation after lumbar disc surgery: results in 130 cases. Acta Neurochir (Wien) 1993;122:102-4.
15. Herron L. Recurrent lumbar disc herniation: results of repeat laminectomy and discectomy. J Spinal Disord 1994;7:161-6.
16. Van Goethem JW, Parizel PM, Jinkins JR. Review article: MRI of the postoperative lumbar spine. Neuroradiology 2002;44:723-39.
17. Jonsson B, Stromqvist B. Repeat decompression of lumbar nerve roots. A prospective two-year evaluation. J Bone Joint Surg Br 1993;75:894-7.
18. Grane P, Lindqvist M. Evaluation of the post-operative lumbar spine with MR imaging. The role of contrast enhancement and thickening in nerve roots. Acta Radiol 1997;38:1035-42.
19. Babar S, Saifuddin A. MRI of the post-discectomy lumbar spine. Clin Radiol 2002;57:969-81.
20. Lee YS, Choi ES, Song CJ. Symptomatic nerve root changes on contrast-enhanced MR imaging after surgery for lumbar disk herniation. AJNR Am J Neuroradiol 2009;30:1062-7.
21. Grane P. The postoperative lumbar spine. A radiological investigation of the lumbar spine after discectomy using MR imaging and CT. Acta Radiol Suppl 1998;414:1-23.

22. Peul WC, van Houwelingen HC, van den Hout WB, et al. Surgery versus prolonged conservative treatment for sciatica. N Engl J Med 2007;356:2245-56.
23. Vroomen PC, Van Hapert SJ, Van Acker RE, Beuls EA, Kessels AG, Wilmink JT. The clinical significance of gadolinium enhancement of lumbar disc herniations and nerve roots on preoperative MRI. Neuroradiology 1998;40:800-6.
24. Toyone T, Takahashi K, Kitahara H, Yamagata M, Murakami M, Moriya H. Visualisation of symptomatic nerve roots. Prospective study of contrast-enhanced MRI in patients with lumbar disc herniation. J Bone Joint Surg Br 1993;75:529-33.
25. Jinkins JR, Osborn AG, Garrett D, Jr., Hunt S, Story JL. Spinal nerve enhancement with Gd-DTPA: MR correlation with the postoperative lumbosacral spine. AJNR Am J Neuroradiol 1993;14:383-94.
26. Nygaard OP, Jacobsen EA, Solberg T, Kloster R, Dullerud R. Nerve root signs on postoperative lumbar MR imaging. A prospective cohort study with contrast enhanced MRI in symptomatic and asymptomatic patients one year after microdiscectomy. Acta Neurochir (Wien) 1999;141:619-22; discussion 23.
27. Taneichi H, Abumi K, Kaneda K, Terae S. Significance of Gd-DTPA-enhanced magnetic resonance imaging for lumbar disc herniation: the relationship between nerve root enhancement and clinical manifestations. J Spinal Disord 1994;7:153-60.
28. Peul WC, van Houwelingen HC, van der Hout WB, et al. Prolonged conservative treatment or 'early' surgery in sciatica caused by a lumbar disc herniation: rationale and design of a randomized trial [ISRCT 26872154]. BMC Musculoskelet Disord 2005;6:8.
29. Ross JS, Robertson JT, Frederickson RC, et al. Association between peridural scar and recurrent radicular pain after lumbar discectomy: magnetic resonance evaluation. ADCON-L European Study Group. Neurosurgery 1996;38:855-61; discussion 61-3.
30. Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB. Assessing health-related quality of life in patients with sciatica. Spine (Phila Pa 1976) 1995;20:1899-908; discussion 909.
31. Collins SL, Moore RA, McQuay HJ. The visual analogue pain intensity scale: what is moderate pain in millimetres? Pain 1997;72:95-7.
32. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43:543-9.
33. Gjorup T. The kappa coefficient and the prevalence of a diagnosis. Methods Inf Med 1988;27:184-6.
34. Arana E, Royuela A, Kovacs FM, et al. Lumbar spine: agreement in the interpretation of 1.5-T MR images by using the Nordic Modic Consensus Group classification form. Radiology 2010;254:809-17.
35. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
36. Pfirrmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N. Magnetic resonance classification of lumbar intervertebral disc degeneration. Spine (Phila Pa 1976) 2001;26:1873-8.
37. Lurie JD, Doman DM, Spratt KF, Tosteson AN, Weinstein JN. Magnetic resonance imaging interpretation in patients with symptomatic lumbar spine disc herniations: comparison of clinician and radiologist readings. Spine (Phila Pa 1976) 2009;34:701-5.
38. Mulconrey DS, Knight RQ, Bramble JD, Paknikar S, Harty PA. Interobserver reliability in the interpretation of diagnostic lumbar MRI and nuclear imaging. Spine J 2006;6:177-84.
39. Ross JS. Babel 2.0. Radiology 2010;254:640-1.
40. Carrino JA, Lurie JD, Tosteson AN, et al. Lumbar spine: reliability of MR imaging findings. Radiology 2009;250:161-70.

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

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

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

Table S2A Clinical outcome measures at one year stratified by MRI findings at one year. Values are n (%) or means ± SD.

A) Group that underwent surgery during the first year. Of the 129 surgical patients with both Gd-MRI at baseline and one year, 26 still had a herniated disc at one year.

B) Group that underwent conservative care during the first year. Of the 75 conservative patients with both Gd-MRI at baseline and one year, 45 still had a herniated disc at one year.

S2A

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

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

¶ The intensity of pain is indicated on a horizontal 100 mm visual analogue scale (VAS) with 0 representing no pain and 100 the worst pain ever experienced

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

‡ The Roland Disability Questionnaire for Sciatica is a disease-specific disability scale that measures the functional status of patients with pain in the leg or back. Scores range from 0 to 23, with higher scores indicating worse functional status
¶ The intensity of pain is indicated on a horizontal 100 mm visual analogue scale (VAS) with 0 representing no pain and 100 the worst pain ever experienced

Chapter 9

Synthesis & discussion

Since ancient times many etiological explanations for sciatica have been proposed. In 1934 Mixter and Barr revolutionized the understanding of sciatica when they asserted that sciatica was caused by a herniated disc pressing against a nerve root.[1, 2] Worldwide this mechanical compression theory has been accepted giving rise to a greater interest in the lumbar disc as a source of sciatica and in the surgical treatment of such a disorder, which has come to be known as the "Dynasty of the Disc".[3, 4] Surgery for back and leg pain in association with nerve root compression has become one of the most commonly performed operative procedures worldwide. The mechanical concept of root compression by a herniated disc offers a satisfying explanation for most symptomatic patients. However, the scientific confusion lies in the observation of several Magnetic resonance imaging (MRI) studies showing a high prevalence of disc herniations ranging from 20 to 76% in persons without any symptoms.[5, 6] Nevertheless, MRI is considered the imaging procedure of choice for patients suspected of lumbar disc herniation[5, 7] and is also frequently performed in patients with persistent or recurrent symptoms of sciatica.[8] As such MRI is thus widely used in diagnosis and treatment planning of patients with intervertebral disc herniations.[9] The aim of this thesis was to uncover the relationship between MRI findings and clinical outcome in patients with sciatica. In this chapter the relationship between clinical outcome on one hand and baseline and follow-up MRI findings on the other hand will be placed in a scientific context. Furthermore the limitations and future research directions will be discussed.

HOW IS INTEROBSERVER AGREEMENT AMONG SPINE SPECIALISTS REGARDING MRI FINDINGS AND DOES IT IMPACT CLINICAL OUTCOME?

In patients who suffered from sciatica for 6 to 12 weeks and who were potential candidates for lumbar disc surgery based on clinical grounds interobserver agreement among two neuroradiologists and a neurosurgeon was almost perfect for the affected disc level and the nerve root that most likely caused the sciatic symptoms (chapter 2). Substantial inter- and intra-observer agreement was observed regarding the presence of disc herniation and nerve root compression when the categories were "probability above 50%" and "probability lower than 50%". However, in general moderate agreement was found regarding the more specific characteristics of the impaired disc level (like signal intensity on T2 images and absence of epidural fat) and characteristics of the disc herniation (like its location, size and whether it should be classified as a protrusion or as an extrusion), which indicate that the assessment of many variables is fairly subjective. Within the literature, values of agreement on disc degeneration show a high variation depending on the variable investigated.[10] Although a few nomenclatures for degenerative disc disease have been proposed (like The Combined Task Force nomenclature or the Nordic Modic Consensus Group classification),[11] none has been widely recognized as authoritative or has been widely used in practice. This absence of consensus is greatly related to the multiple controversial aspects of disc abnormalities.[12] However, good reliability of imaging data in degenerative disc disease is important to determine the relationship between specific imaging characteristics and patient outcomes. To gain more insight in this relationship, those interpreting the images should 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.[13] As a first step on the road to determine the relationship between specific imaging characteristics and patient outcomes, radiologists and clinicians should strive to reduce variability in interpretations and adhere to a specific nomenclature for degenerative disc disease.[14] However, despite the adherence to predefined definitions in this thesis, the MRI assessors in this thesis sill only reached moderate agreements regarding many characteristics of the disc level and the herniated disc, which indicate that definitions and the adherence to a well defined nomenclature only is not enough for reaching substantial to excellent agreements. In addition to defining the language for image interpretation for degenerative disc disease, specific training might be an important next step.[13, 14] In support are the results of two reliability studies of The Spine Patient Outcomes Research Trial.[15, 16] In one of the two studies the reported agreement on disc morphology was only fair (kappa= 0.24) between the clinicians and radiologists.[15] In another study inter-reader reliability for disc morphology was excellent (kappa= 0.81) between 3 radiologists and 1 orthopedic surgeon.[16] The observation of a much better agreement in the second study might be explained by a better training of the MRI assessors as in that study the MRI assessors, before beginning the study, first evaluated a sample set of images with use of definitions and afterwards they met in person to review each image, enabling them to better streamline the way of interpreting the images.

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.[17, 18] After 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 in whom independent agreement was reached about the absence of those findings (chapter 3). Thus based on the consistency in MRI interpretation different prognostic profiles could be made in sciatica. These results enable spine physicians to better inform patients with sciatica about their prognosis. If for example a spine surgeon and a radiologist both agree about the presence of a disc herniation the patient can be informed about a likely favorable prognosis, compared to a less favorable prognosis when the spine surgeon and radiologist do disagree about the presence of a disc herniation. 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. This hypothesis is supported by an earlier study that observed that presence of nerve root compression in patients with sciatica is associated with favorable prognosis in primary care patients with sciatica.[19]

CAN MRI HELP TO PREDICT SURGERY FOR SCIATICA?

The natural history of acute sciatica is in general favorable, with spontaneous resolution of the leg pain within 8 weeks in the overwhelming majority of cases.[20, 21] When patients fail to respond to conservative care, surgery might be considered. However, the duration of conservative care is not well defined. Of the patients who were randomized to receive prolonged conservative care in the Sciatica Trial a considerable part of 39% ultimately received surgery during the first year.[22, 23] Qualitative MRI parameters and the baseline size of the disc herniation did not significantly differ between the surgical and non-surgical group (chapter 4). Patients who did undergo surgery during follow-up had at baseline higher RDQ scores, more intense leg pain and smaller dural sacs and spinal canals compared to patients who did not undergo surgery. However, additional Receiver operating characteristic curve analysis showed that the MRI variables have only a poor ability to discriminate between patients who underwent delayed surgery and patients who did not. The overall results suggest that MRI is not suitable to distinguish between patients who will and those who will not undergo surgery for sciatica.

ARE THERE PROGNOSTIC RELEVANT MRI DIFFERENCES BETWEEN SCIATICA PATIENTS WITH AND WITHOUT DISABLING BACK PAIN?

Patients with sciatica frequently complain about associated back pain.[24] Patients with both sciatica and disabling back pain at baseline (defined as a Visual Analogue Scale for back pain of at least 40mm on a 0-100mm scale[25, 26]) reported an unfavorable prognosis at one-year follow-up compared to those with predominantly sciatica (chapter 5). If additionally a clear herniated disc with nerve root compression on MRI was absent, the results were even worse (one-year satisfactory results ranged from 50 to 91%). Herniated discs and nerve root compression on MRI were more prevalent among sciatica patients with compared to those without disabling back pain. However, vertebral endplate signal changes were equally distributed between those with and without disabling back pain. Large disc herniations and extruded disc herniations were also equally distributed between the two groups.

The clinical relevance of MRI morphological variations has been ongoingly debated over the past two decades.[5, 6] MRI differences have been reported between patients with both sciatica and low back pain compared to control subjects without symptoms,[5] and between sciatica patients compared to low back pain patients.[27] However, previous studies did not compare these findings between sciatica patients with and without back pain. Disc herniations are often seen on imaging studies in patients without symptoms.[5, 6] Contrary, in chapter 5 it was shown that a substantial number of patients without disc herniation or nerve root compression suffered from sciatica. Some researchers suggested that inflammation of the nerve root may also be a major factor in sciatica.[28, 29] If this hypothesis is correct, the finding that sciatica patients with back pain less often had a herniated disc compared to patients with predominantly sciatica may be explaind by a higher inflammatory component in sciatica patients with back pain. This may also explain why sciatica patients with back pain fared worse compared to patients with predominantly sciatica as the extent of inflammation may be a causative factor in the cases with persistent pain and functional disability.

Despite remarkable advancements in diagnostic imaging and surgical techniques the results after lumbar disc surgery do not seem to have improved during recent decades: both classical studies and recent randomized controlled trials show that during longer follow-up treatment results for sciatica are satisfactory in 60 to 85% of the patients.[18, 23, 30-33] The number of proposed interventions developed by numerous disciplines including family practice, neurosurgery, orthopedic surgery, neurology, anesthesiology, psychiatry, physical therapy, social work, chiropractics, is overwhelming. Many widely prescribed treatments have no evidence for efficacy. Other effective treatments, which may be of benefit for subsets of patients, are indiscriminately applied. The results in chapter 5 of this thesis indicate that in sciatica subgroups with different prognostic profiles can be identified. A shift from a "one-size fits all" approach, where heterogeneous groups of patients receive broadly similar treatments, towards targeted treatments according to prognostic profiles or specific characteristics, may help to improve the treatment results.[34]

DO ANATOMICAL ABNORMALITIES ON FOLLOW-UP MRIS EXPLAIN WHY PATIENTS EXPERIENCE PERSISTENT OR RECURRENT SYMPTOMS OF SCIATICA?

MRI is considered the imaging procedure of choice for patients suspected of lumbar disc herniation[5, 7] and is frequently performed in patients with persistent or recurrent symptoms of sciatica.[8] Patients with sciatica and symptomatic lumbar disc herniations at baseline who were followed for one year in the Sciatica Trial still showed abnormalities in a considerable percentage after one year: 21% of surgically treated patients still had a herniated disc on MRI at one year compared to 60% of conservatively treated patients (Chapter 6). However, the presence of disc herniation or nerve root compression on MRI at one year follow-up did not distinguish patients with favorable clinical outcomes from those with unfavorable clinical outcomes. The presence of scar tissue was also not associated with patient outcome. The results give rise to a paradox that although imaging findings are not associated with patient outcomes in patients with recurrent or persistent symptoms of sciatica still many studies have shown that 60 to 82% of patients with a recurrent disc herniation on MRI improves after repeat surgery.[35-38] Despite this paradox, the results have implications for both clinicians and patients.

Clinicians should be more cautious in ascribing persistent or recurrent symptoms of sciatica to anatomical abnormalities visible on MRI. Although many physicians are aware that anatomical abnormalities correlate poorly with low back pain, for many it seems intuitively right to repeat MRI in patients with recurrent or persistent symptoms of sciatica who in an earlier stage did show abnormalities in presence of acute sciatic symptoms. Imaging has high costs, not only because of the direct costs of the imaging itself, but also due to the downstream effects such as additional tests, follow-up, referrals and invasive procedures.[39, 40]

For many physicians it is a logical step to perform surgery or other invasive procedures such as epidural injections in case repeat MRI still shows the abnormalities. Increased frequency of lumbar MRI has been shown to be associated with higher rates of spine surgery.[40, 41] However, the real issue is that until now, no better patient outcomes have been demonstrated with this increased use of advanced imaging.[41, 42] For example, in a randomized trial of simple versus advanced imaging for patients with low back pain, patients who received an MRI were twice as often more likely to undergo surgery over the subsequent year than were those undergoing plain radiography.[42] However, clinical outcomes at 1 year were equivalent, despite the difference in surgery rates. Based on the results of this thesis, it may well be that also for patients with recurrent or persistent sciatica follow-up clinical outcomes are simillar between those who undergo and those who do not undergo repeated MRI, rendering it of no benefit in the evaluation of patients with recurrent or persistent symptoms of sciatica.

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. Wanting diagnostic testing is a frequent reason for repeated office visits for patients suffering from chronic back pain.[43] Many patients believe that the more diagnostics tests performed, the higher and better the quality of care.[44] Many physicians admit they succumb to their patients who are asking for spine imaging, even after explaining to the patient that imaging is unnecessary.[45] However, spine imaging may have an adverse effect as telling patients that they have a back imaging abnormality could result in unintended harms related to labeling.[46] In a randomized controlled trial involving patients with acute back pain or sciatica, patients were randomly assigned to whether or not receive their imaging results. Patients who received their imaging results reported less improvements in general health than those who were blinded to their results.[47] The mindset of patients that more imaging testing means better care must be abandoned in favor of a more evidence-based approach.[39] Patient education about the limits of spine imaging may help to bring patient's expectations more in line with the evidence.

Are vertebral endplate signal changes associated with back pain in sciatica?

Patients with sciatica frequently experience disabling back pain. One of the proposed causes for back pain is Vertebral Endplate Signal Changes (VESC) as visualized by MRI. VESC are even a frequent surgical indication to perform a fixation of two or more vertebrae in the lower spine or replacing the disc by a prothesis, resulting in rising back surgery rates. The results in chapter 7 showed that undergoing disc surgery for sciatica was highly associated with progression in the extent of VESC compared to non-operative care. However, both at baseline and after one year follow-up, those with and those without VESC reported disabling back pain (defined as a Visual Analogue Scale for back pain of at least 40mm on a 0-100mm scale[25, 26]) in nearly the same proportion, regardless of having undergone surgery or not. Therefore the results suggest that VESC are not responsible for disabling back pain in patients with sciatica and one should therefore be reticent to offer back surgery based on VESC.

The relevance of VESC is highly debated in current literature.[48] VESC have been reported to be associated with low back pain in the general population aged 40 years[49] and in working populations.[50, 51] Two studies did not observe more VESC among chronic low back pain patients compared to control subjects,[52] or between VESC and previous back pain in subjects without current back pain or sciatica.[53] Two earlier studies investigated the correlation between VESC and low back pain in patients treated for lumbar disc herniations, with contradictory results to the present study.[54, 55] Unfortunately in both studies the VESC were described by only one radiologist. Possibly, results in the current thesis are contradictory with these two studies due to the definition of back pain. While they used self-reported back pain as the outcome, in this thesis 'disabling back pain' was defined according to patients' reported VAS for back pain.

In this thesis VESC Type 2 was the most common VESC at baseline (when patients presented with acute sciatica), a finding in concordance with previous studies in unoperated sciatica patients.[54, 56] The most common conversion in the surgical group was progression from no VESC at any level to type 1. In the study of Rahm et al. most patients developed VESC type 2 changes after lumbar discectomy.[57] However, contrary to the 12 months time interval between initial and follow-up MRI in this thesis, their interval varied from 32 to 59 months. In general it is agreed that VESC type 1 are unstable lesions which may convert to type 2 or back to normal.[58] The high observed prevalence of type 1 lesions at 12 months may still represent the more active stage of inflammation following disc surgery and these lesions may convert to type 2 or back to normal over time. Furthermore, the observation that 81% of conservatively treated patients did not convert from one VESC type to another after one year is in concordance with longitudinal studies that investigated the natural course of VESC and have observed that 48 to 86% of people do not convert from one VESC type to another over periods of 14 to 72 months.[54, 56, 59-61]

Should one give gadolinium-based contrast when imaging patients with sciatica?

Gadolinium (Gd)-enhanced MRI is frequently performed in patients with persistent or recurrent symptoms of sciatica after surgical treatment.[8, 38, 62-64] The MRI assessors (2 experienced neuroradiologists and one neurosurgeon) presented substantial disagreement about gadolinium enhancement in lumbar spine MRIs between observers which was in firm contrast with their excellent agreement about the disc level of the herniated disc and compressed nerve root at baseline (chapter 8). Furthermore, no relationship was observed between enhancement findings and clinical findings at one year. The overall results indicate that Gd-MRI is not more helpful than non-enhanced MRI in the post-treatment evaluation of patients with sciatica.

Since the interobserver agreement regarding the enhancement findings was rather low, one could question the added value of correlating enhancement findings with clinical findings. To truly uncover the meaning of enhancement findings those interpreting the images must reliably assess the enhancement finding as one reason that a prediction model might lose its predictive power is the incorrect assessment of MRI findings, which causes the inputs in the prediction model to be faulty.[13] It is crucial that radiologists and clinicians strive to reduce variability in interpretations as inconsistency in interpretation may lead to alternative treatment options between clinicians and therefore may impact the outcome of patient treatment.[11 65]

This thesis did not demonstrate an added value of Gd-MRI over non-enhanced MRI, which is in contrast with the expected diagnostic value. Given the additional costs of invasive contrast infusion compared to unenhanced MRI scanning, the addition of gadolinium in the postsurgical lumbar spine leads to more confusion at a higher financial reimbursement rate. Combined with the results in chapter 6 and 7 one could question the value of obtaining follow-up MRI at all when evaluating patients with recurrent or persistent symptoms of sciatica.

STRENGTHS AND WEAKNESSES

A strength of the studies in this thesis was that all images were assessed by two experienced neuroradiologists and one neurosurgeon who 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. In addition, no meeting was organized in which a sample subset of images was evaluated as the discussion during this meeting might have caused the observers to adjust their diagnostic imaging criteria. Such a meeting could have led to an overestimation in the agreement among the three readers compared to the situation as it existed before undertaking the meeting. Moreover, the presence of disc herniation and nerve root compression was assessed with a four point scale providing the MRI assessors the opportunity to express their uncertainty. Another strength was that all sciatica patients who underwent MRI at baseline were followed, regardless of participation in the randomized controlled trial. At last, the percentage of patients who underwent MRI at one year follow-up is high (94.3%), especially in light of the observation that the majority of patients was recovered and one therefore may expect less willingness of patients to undergo repeat MRI.

The current thesis has also several shortcomings. Firstly, the interobserver agreement observed 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.[66] However, the agreement between the neuroradiologists did not substantially differ from that of the reading pairs containing one of the two neuroradiologists and the neurosurgeon. Secondly, the use of standardized reporting forms with definitions and multiple choice categories allowed the assessments to be structured far more than possible in general clinical practice which may have caused an overestimation in the interobserver agreements.[16] Thirdly, the study population consisted of sciatica patients who had severe symptoms and were referred to the neurologists. These patients were willing to undergo surgery, so patients with a clear preference for conservative treatment are underrepresented in the current thesis. Fourthly, the reported MRI findings and their relation with clinical outcome was timed only once, one year after randomization. It is uncertain if comparable results at other time points would have been observed.

CURRENT STATUS AND FUTURE PERSPECTIVE

The results of this thesis have placed the relevance of MRI findings in patients with sciatica in a new light and many findings are in contrast with the intuitive feelings and ideas of spinal physicians. The finding that in patients re-imaged one year after treatment for sciatica, anatomical abnormalities visible on MRI did not distinguish patients with persistent or recurrent symptoms of sciatica from asymptomatic patients was remarkable. The same holds for a lack of a correlation between vertebral endplate signal changes and back pain. Gadolinium-enhanced MRI did also not prove more helpful than non-enhanced MRI in the post-treatment evaluation of patients with sciatica.

The role of MRI in patients with sciatica should be rethought, especially in patients with recurrent or persistent symptoms of sciatica. The results of this thesis counteract the intuitive feeling of the necessity of repeating MRI in these patients. The mindset of patients that more imaging testing means better care should be reshaped in favor of a more evidence-based approach. MRI may be repeated only when repeat surgery is considered in presence of unfavorable clinical history and physicical examination. More research is needed to assess the value of MRI in clinical decision making for patients with persistent or recurrent sciatica, in particular if treatment strategies according to MRI findings lead to different clinical outcomes in patients who experience persistent or recurrent symptoms of sciatica. It also remains unclear why symptoms relate poorly to evidence of disc herniation or nerve root compression on MRI. Inflammation of the nerve root may be a more important factor than mechanical compression. It seems that we will make a paradigm shift from mechanical into inflammatory origin, which comes down to going back in time to the 18th century when Cotugnio argued that sciatic complaints are a consequence of neuritis or edema of the sciatic nerve. Insight in the balance of mechanical and inflammatory factors may enable us to solve the paradox of no relationship between presence of a disc herniation on MRI and patient outcome in patients with recurrent or persistent sciatica while surgery is often helpful for these patients.

Prognostic profiles in sciatica vary greatly (ranging from 50 to 91% in this thesis). Reasons behind the different prognostic profiles in sciatica are currently not known. It may well be that the inflammatory component, which is currently not visible or quantifiable, plays also an important role in this observation. More research is needed to identify the reasons behind the different prognostic profiles in sciatica and how to apply new or existing therapeutic strategies accordingly. A shift from a "one-size fits all" approach towards targeted treatments according to prognostic profiles or specific characteristics will probably improve the treatment results.

At last, to thoroughly gain insight in the clinical relevance of imaging findings, good interobserver agreement is a prerequisite, which for some (especially gadolinium-enhancement) findings in this thesis was very low. As earlier mentioned, no nomenclature in the literature has been widely recognized as authoritative or has been widely used in practice. It is worthwile to consider approaches how to reach more consensus and how to subsequently adhere to one nomenclature. Specific imaging training and defining the language for image interpretation by involving the various desciplines may help to attain this goal.

CONCLUSION

In contrast with the intuitive feeling of physcians many worrisome MRI findings do not correlate with patient outcome in patients with sciatica. Physicians should not ascribe persistent or recurrent symptoms of sciatica to the presence of abnormalities visible on MRI. However, many issues remain to be solved, especially the paradox of no association between presence of a disc herniation on MRI and patient outcome in patients with recurrent or persistent sciatica while surgery is often helpful for these patients.

Physicians should inform their patients about the limits of spine imaging in sciatica. One possible strategy is to explain patients of how common worrisome MRI findings are observed in persons who do not have any symptoms. This thesis enables physicians to implement this strategy and in that way reshape the mindset of many patients thinking that knowing imaging findings can only be good.

REFERENCES

1. Mixter WJ, Barr JS. Rupture of Intervertebral Disc with Involvement of Spinal Canal. N Engl J Med 1934;211:210-5.
2. Robinson JS. Sciatica and the lumbar disk syndrome: a historic perspective. South Med J 1983;76:232-8.
3. Allan DB, Waddell G. An historical perspective on low back pain and disability. Acta Orthop Scand Suppl 1989;234:1-23.
4. Parisien RC, Ball PA. William Jason Mixter (1880-1958). Ushering in the "dynasty of the disc". Spine (Phila Pa 1976) 1998;23:2363-6.
5. 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.
6. 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.
7. 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.
8. 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.
9. Carlisle E, Luna M, Tsou PM, Wang JC. Percent spinal canal compromise on MRI utilized for predicting the need for surgical treatment in single-level lumbar intervertebral disc herniation. Spine J 2005;5:608-14.
10. Pfirrmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N. Magnetic resonance classification of lumbar intervertebral disc degeneration. Spine (Phila Pa 1976) 2001;26:1873-8.
11. Ross JS. Babel 2.0. Radiology 2010;254:640-1.
12. Milette PC. Reporting lumbar disk abnormalities: at last, consensus! AJNR Am J Neuroradiol 2001;22:428-9.
13. Carrino JA, Lurie JD, Tosteson AN, et al. Lumbar spine: reliability of MR imaging findings. Radiology 2009;250:161-70.
14. Jarvik JG, Deyo RA. Moderate versus mediocre: the reliability of spine MR data interpretations. Radiology 2009;250:15-7.
15. Lurie JD, Doman DM, Spratt KF, Tosteson AN, Weinstein JN. Magnetic resonance imaging interpretation in patients with symptomatic lumbar spine disc herniations: comparison of clinician and radiologist readings. Spine (Phila Pa 1976) 2009;34:701-5.
16. Lurie JD, Tosteson AN, Tosteson TD, et al. Reliability of magnetic resonance imaging readings for lumbar disc herniation in the Spine Patient Outcomes Research Trial (SPORT). Spine (Phila Pa 1976) 2008;33:991-8.
17. Cihangiroglu M, Yildirim H, Bozgeyik Z, et al. Observer variability based on the strength of MR scanners in the assessment of lumbar degenerative disc disease. Eur J Radiol 2004;51:202-8.
18. Vucetic N, Astrand P, Guntner P, Svensson O. Diagnosis and prognosis in lumbar disc herniation. Clin Orthop Relat Res 1999:116-22.
19. Vroomen PC, Wilmink JT, de KM. Prognostic value of MRI findings in sciatica. Neuroradiology 2002;44:59-63.
20. Vroomen PC, de Krom MC, Slofstra PD, Knottnerus JA. Conservative treatment of sciatica: a systematic review. J Spinal Disord 2000;13:463-9.

21. Weber H, Holme I, Amlie E. The natural course of acute sciatica with nerve root symptoms in a double-blind placebo-controlled trial evaluating the effect of piroxicam. Spine (Phila Pa 1976) 1993;18:1433-8.
22. Peul WC, Brand R, Thomeer RT, Koes BW. Improving prediction of "inevitable" surgery during non-surgical treatment of sciatica. Pain 2008;138:571-6.
23. 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.
24. Koes BW, van Tulder MW, Peul WC. Diagnosis and treatment of sciatica. BMJ 2007;334:1313-7.
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. Modic MT, Obuchowski NA, Ross JS, et al. Acute low back pain and radiculopathy: MR imaging findings and their prognostic role and effect on outcome. Radiology 2005;237:597-604.
28. Chen C, Cavanaugh JM, Ozaktay AC, Kallakuri S, King AI. Effects of phospholipase A2 on lumbar nerve root structure and function. Spine (Phila Pa 1976) 1997;22:1057-64.
29. Saal JS, Franson RC, Dobrow R, Saal JA, White AH, Goldthwaite N. High levels of inflammatory phospholipase A2 activity in lumbar disc herniations. Spine (Phila Pa 1976) 1990;15:674-8.
30. Jacobs WC, van Tulder M, Arts M, et al. Surgery versus conservative management of sciatica due to a lumbar herniated disc: a systematic review. Eur Spine J 2011;20:513-22.
31. Atlas SJ, Keller RB, Wu YA, Deyo RA, Singer DE. Long-term outcomes of surgical and nonsurgical management of sciatica secondary to a lumbar disc herniation: 10 year results from the maine lumbar spine study. Spine (Phila Pa 1976) 2005;30:927-35.
32. Weber H. Lumbar disc herniation. A controlled, prospective study with ten years of observation. Spine (Phila Pa 1976) 1983;8:131-40.
33. 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.
34. Konstantinou K, Beardmore R, Dunn KM, et al. Clinical course, characteristics and prognostic indicators in patients presenting with back and leg pain in primary care. The ATLAS study protocol. BMC Musculoskelet Disord 2012;13:4.
35. Bernard TN, Jr. Repeat lumbar spine surgery. Factors influencing outcome. Spine (Phila Pa 1976) 1993;18:2196-200.
36. Fandino J, Botana C, Viladrich A, Gomez-Bueno J. Reoperation after lumbar disc surgery: results in 130 cases. Acta Neurochir (Wien) 1993;122:102-4.
37. Herron L. Recurrent lumbar disc herniation: results of repeat laminectomy and discectomy. J Spinal Disord 1994;7:161-6.
38. Van Goethem JW, Parizel PM, Jinkins JR. Review article: MRI of the postoperative lumbar spine. Neuroradiology 2002;44:723-39.
39. Chou R, Qaseem A, Owens DK, Shekelle P. Diagnostic imaging for low back pain: advice for high-value health care from the American College of Physicians. Ann Intern Med 2011;154:181-9.
40. Deyo RA. Cascade effects of medical technology. Annu Rev Public Health 2002;23:23-44.
41. Lurie JD, Birkmeyer NJ, Weinstein JN. Rates of advanced spinal imaging and spine surgery. Spine (Phila Pa 1976) 2003;28:616-20.
42. Jarvik JG, Hollingworth W, Martin B, et al. Rapid magnetic resonance imaging vs radiographs for patients with low back pain: a randomized controlled trial. JAMA 2003;289:2810-8.

43. McPhillips-Tangum CA, Cherkin DC, Rhodes LA, Markham C. Reasons for repeated medical visits among patients with chronic back pain. J Gen Intern Med 1998;13:289-95.
44. Carman KL, Maurer M, Yegian JM, et al. Evidence that consumers are skeptical about evidence-based health care. Health Aff (Millwood) 2010;29:1400-6.
45. Campbell EG, Regan S, Gruen RL, et al. Professionalism in medicine: results of a national survey of physicians. Ann Intern Med 2007;147:795-802.
46. Fisher ES, Welch HG. Avoiding the unintended consequences of growth in medical care: how might more be worse? JAMA 1999;281:446-53.
47. Ash LM, Modic MT, Obuchowski NA, Ross JS, Brant-Zawadzki MN, Grooff PN. Effects of diagnostic information, per se, on patient outcomes in acute radiculopathy and low back pain. AJNR Am J Neuroradiol 2008;29:1098-103.
48. Jensen TS, Karppinen J, Sorensen JS, Niinimaki J, Leboeuf-Yde C. Vertebral endplate signal changes (Modic change): a systematic literature review of prevalence and association with non-specific low back pain. Eur Spine J 2008;17:1407-22.
49. Kjaer P, Leboeuf-Yde C, Korsholm L, Sorensen JS, Bendix T. Magnetic resonance imaging and low back pain in adults: a diagnostic imaging study of 40-year-old men and women. Spine (Phila Pa 1976) 2005;30:1173-80.
50. Kuisma M, Karppinen J, Niinimaki J, et al. Modic changes in endplates of lumbar vertebral bodies: prevalence and association with low back and sciatic pain among middle-aged male workers. Spine (Phila Pa 1976) 2007;32:1116-22.
51. Schenk P, Laubli T, Hodler J, Klipstein A. Magnetic resonance imaging of the lumbar spine: findings in female subjects from administrative and nursing professions. Spine (Phila Pa 1976) 2006;31:2701-6.
52. Kovacs FM, Arana E, Royuela A, et al. Vertebral endplate changes are not associated with chronic low back pain among Southern European subjects: a case control study. AJNR Am J Neuroradiol;33:1519-24.
53. Jarvik JJ, Hollingworth W, Heagerty P, Haynor DR, Deyo RA. The Longitudinal Assessment of Imaging and Disability of the Back (LAIDBack) Study: baseline data. Spine (Phila Pa 1976) 2001;26:1158-66.
54. Albert HB, Manniche C. Modic changes following lumbar disc herniation. Eur Spine J 2007;16:977-82.
55. 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.
56. Kuisma M, Karppinen J, Niinimaki J, et al. A three-year follow-up of lumbar spine endplate (Modic) changes. Spine (Phila Pa 1976) 2006;31:1714-8.
57. Rahme R, Moussa R, Bou-Nassif R, et al. What happens to Modic changes following lumbar discectomy? Analysis of a cohort of 41 patients with a 3- to 5-year follow-up period. J Neurosurg Spine 2010;13:562-7.
58. Rahme R, Moussa R. The modic vertebral endplate and marrow changes: pathologic significance and relation to low back pain and segmental instability of the lumbar spine. AJNR Am J Neuroradiol 2008;29:838-42.
59. Jensen TS, Bendix T, Sorensen JS, Manniche C, Korsholm L, Kjaer P. Characteristics and natural course of vertebral endplate signal (Modic) changes in the Danish general population. BMC Musculoskelet Disord 2009;10:81.
60. Mitra D, Cassar-Pullicino VN, McCall IW. Longitudinal study of vertebral type-1 end-plate changes on MR of the lumbar spine. Eur Radiol 2004;14:1574-81.
61. Modic MT, Steinberg PM, Ross JS, Masaryk TJ, Carter JR. Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging. Radiology 1988;166:193-9.
62. Babar S, Saifuddin A. MRI of the post-discectomy lumbar spine. Clin Radiol 2002;57:969-81.

63. Grane P. The postoperative lumbar spine. A radiological investigation of the lumbar spine after discectomy using MR imaging and CT. Acta Radiol Suppl 1998;414:1-23.
64. Grane P, Lindqvist M. Evaluation of the post-operative lumbar spine with MR imaging. The role of contrast enhancement and thickening in nerve roots. Acta Radiol 1997;38:1035-42.
65. Mulconrey DS, Knight RQ, Bramble JD, Paknikar S, Harty PA. Interobserver reliability in the interpretation of diagnostic lumbar MRI and nuclear imaging. Spine J 2006;6:177-84.
66. Kovacs FM, Royuela A, Jensen TS, et al. Agreement in the interpretation of magnetic resonance images of the lumbar spine. Acta Radiol 2009;50:497-506.

Chapter 10

Summary

Sciatica is one of the most common lumbar-spine disorders and ranks, certainly in the industrialized countries, as one of the most costly medical problems. The main objective of this thesis was to uncover the relationship between MRI findings and clinical outcome in patients with sciatica.

The basis of this study was the Sciatica Trial: a multicentre prospective randomized controlled trial among patients with 6-12 weeks sciatica. An early surgery strategy was compared to prolonged conservative care for an additional 6 months followed by surgery for patients who did not improve or who did request it earlier because of aggravating symptoms. The trial showed faster recovery after early surgery as compared to a strategy of prolonged conservative care with surgery if needed, but there were no significant differences in clinical outcomes after one year. The randomized patients were part of a larger group of patients with sciatica who underwent a baseline MRI to assess the eligibility for the Sciatica Trial. All patients who underwent MRI (regardless of participation in the randomized controlled trial) were followed up for one year. Furthermore, the patients who were randomized underwent MRI both at baseline and after one year.

Chapter 1 gives an introduction and some historical facts about sciatica. Despite many etiological explanations being proposed since ancient time, it was until 1934 when Mixter and Barr overcame the scientific confusion and asserted that sciatica was caused by a herniated disc pressing against a nerve root. However, the scientific confusion revived when modern imaging modalities such as MRI were introduced which allowed many investigators to detect an enormous variety of previously unappreciated anatomical variations in patients undergoing diagnostic workups for sciatica. Several MRI studies showed a high prevalence of disc herniations ranging from 20 to 76% in persons without any symptoms. Even in patients who were re-imaged after earlier disc surgery, MRI studies have found herniations in up to 53% of persons who at the time of the re-imaging had no symptoms. Despite the scientific debate MRI is considered the imaging procedure of choice for patients suspected of lumbar disc herniations and is frequently performed in patients with persistent or recurrent symptoms of sciatica. Moreover, abnormal MRI findings frequently result in surgical treatment or other invasive procedures such as epidural injections.

Chapter 2 describes interobserver agreement among two neuroradiologists and one neurosurgeon with regard to MRI characteristics in patients with sciatica who were potential candidates for lumbar disc surgery on clinical grounds. The interobserver agreement was high with regard to clinically relevant parameters like most affected disc level and nerve root, probability of disc herniation and nerve root compression. However, in general considerable variation between the observers was observed regarding specific characteristics of the symptomatic disc level (like signal intensity on T2 images, absence of epidural fat and flattening of the dural sac or the emerging root sheath) and the disc herniation (like its location, size and whether it should be classified as a protrusion or as an extrusion).

Chapter 3 reports on the implications 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.

Of the patients who were randomized to receive prolonged conservative care in the Sciatica Trial a considerable part of 39% ultimately received surgery during the first year. **Chapter 4** evaluates whether qualitative and quantitative MRI assessment could have predicted this delayed surgery. Qualitative MRI parameters and the baseline size of the disc herniation did not significantly differ between the surgical and non-surgical group. Patients who did undergo surgery during follow-up had at baseline smaller dural sacs and spinal canals compared to patients who did not undergo surgery. However, ROC curve analysis showed that these variables have only a poor ability to discriminate between patients who underwent surgery and those who did not.

Chapter 5 reports on the prognostic value of low back pain in relation to MRI findings in patients with sciatica. Patients who had disabling back pain at baseline reported an unfavorable prognosis at one-year follow-up compared to those with predominantly sciatica. If additionally a clear herniated disc with nerve root compression on MRI was absent, the results were even worse. Herniated discs and nerve root compression on MRI were more prevalent among patients with predominantly sciatica compared to those who suffered from additional disabling back pain. However, vertebral endplate signal changes were equally distributed between those with and without disabling back pain. Large disc herniations and extruded disc herniations were also equally distributed between the two groups.

Chapter 6 reports on the 1-year MRI findings of sciatica patients who were treated with either surgery or conservative treatment, changes of MRI findings over time, and their correlation with clinical outcome. At one year follow-up a considerable proportion of patients still had a visible disc herniation on MRI (21% of surgically compared to 60% of conservatively treated patients). However, the presence of disc herniation or nerve root compression on MRI did not distinguish patients with persistent or recurrent symptoms of sciatica from asymptomatic patients.

Chapter 7 reports on Vertebral Endplate Signal Changes (VESC) findings, changes of VESC findings over time and the correlation between VESC findings and back pain in sciatica. Undergoing disc surgery for sciatica was highly associated with progression in the extent of VESC compared to conservative (non-operative) care. In one year about two thirds of surgically and one fifth of conservatively treated patients displayed an increase of VESC. However, both at baseline and after one year follow-up, those with and those without VESC reported disabling back pain in nearly the same proportion. In addition, the proportion of patients reporting perceived recovery after one year was also nearly equal between those with and without VESC.

Chapter 8 reports on the reliability of enhancement findings, their prevalence and their correlation with clinical outcome. The MRI observers (two neuroradiologists and one neurosurgeon) showed excellent agreement regarding the disc level of the herniated disc and the most affected nerve root. Moreover, they agreed on the presence of a disc herniation and nerve root compression. This was in contrast to their agreement on the enhancement of disc herniation and the affected nerve root, which was only fair. One year after treatment, the observers reached substantial agreement regarding the presence of a disc herniation and whether the nerve root was still compressed. However, again, they substantially disagreed on the enhancement of the herniated disc and nerve root. The observers reached moderate to substantial agreement regarding the presence of scar tissue at one year. If they judged scar tissue to be present, they nearly always regarded it as enhanced. At one year no relationship was observed between enhancement of the nerve root or disc herniation and clinical findings.

CONCLUSION

Patients who also suffered from disabling back pain in the acute stage of sciatica reported an unfavorable prognosis at one-year follow-up compared to those with predominantly sciatica. If additionally in the acute stage of sciatica a clear herniated disc with nerve root compression on MRI was absent, the results after one-year follow-up were even worse. MRI-findings in the acute stage were not acceptably able to discriminate between those who did and those who did not undergo surgery. At one year follow-up, anatomical abnormalities visible on MRI did not distinguish patients with persistent or recurrent symptoms of sciatica from patients without symptoms. Vertebral Endplate Signal Changes were not associated with disabling back pain, neither in the acute stage of sciatica nor after one year of follow-up. The reliability of Gadolinium-enhanced MRI findings was low and enhancement findings did not correlate with clinical findings at one-year follow-up.

NEDERLANDSE SAMENVATTING

Het lumbosacraal radiculair syndroom (LSRS, in de volksmond ook wel ischias genoemd) is een van de meest voorkomende lumbale wervelkolom-aandoening en is, zeker in de geïndustrialiseerde landen, een van de meest kostbare medische condities. De belangrijkste doelstelling van dit proefschrift was om inzicht te verwerven in de relatie tussen MRI-bevindingen en klinische uitkomst bij patiënten met LSRS.

De basis van deze studie was de Sciatica Trial: een gerandomiseerd multicentrisch onderzoek bij patiënten met 6-12 weken LSRS. De effectiviteit van vroeg chirurgisch ingrijpen werd vergeleken met die van verlengd afwachtend beleid voor een extra 6 maanden, gevolgd door een operatie voor patiënten die niet herstelden of die er eerder om vroegen als gevolg van verergerende pijn. De studie toonde sneller herstel aan na vroeg chirurgisch ingrijpen in vergelijking met een strategie van verlengd afwachtend beleid met een operatie als dat nodig is, maar na 1 jaar waren er geen significante verschillen in klinische uitkomsten. De gerandomiseerde patiënten maakten deel uit van een grotere groep patiënten met LSRS die een aanvangs (baseline) MRI hadden ondergaan om te beoordelen of de patienten in aanmerking kwamen voor de Sciatica Trial. Alle patiënten die een MRI hebben ondergaan (ongeacht deelname aan de gerandomiseerde studie) werden gedurende een jaar gevolgd. Daarnaast hebben de gerandomiseerde patiënten zowel op baseline als op 1 jaar een MRI ondergaan.

Hoofdstuk 1 bevat een introductie en enkele historische feiten over het LSRS. Ondanks de vele etiologische verklaringen die sinds de oudheid zijn voorgesteld, was het pas in het jaar 1934 dat Mixter en Barr de wetenschappelijke verwarring omtrend LSRS overwonnen en beweerden dat het LSRS wordt veroorzaakt door een hernia die tegen een zenuwwortel aandrukt. De wetenschappelijke verwarring werd echter weer herboren met de introductie van moderne beeldvormings modaliteiten zoals de MRI die vele onderzoekers in staat stelden om een enorme verscheidenheid van niet eerder opgemerkt anatomische variaties te ontdekken bij patiënten die diagnostische procedures voor LSRS ondergingen. Verschillende MRI studies toonden hoge prevalenties aan van discus hernia's, variërend van 20 tot 76% in personen zonder symptomen. Zelfs in patiënten die nogmaals een MRI hebben ondergaan na discus hernia chirurgie, hebben MRI studies hernia's geobserveerd tot in wel 53% van de personen die ten tijde van de herhaalde MRI scan geen symptomen hadden. Ondanks het wetenschappelijk debat wordt MRI als de beeldvormings procedure van eerste keus beschouwd voor patiënten die verdacht worden van een lumbale discus hernia en wordt MRI ook vaak uitgevoerd bij patiënten met aanhoudende of terugkerende symptomen van LSRS. Bovendien leiden afwijkende MRI bevindingen vaak tot chirurgische behandelingen of andere invasieve procedures, zoals epidurale .

Hoofdstuk 2 beschrijft de interbeoordelaars variatie tussen twee neuroradiologen en en een neurochirurg met betrekking tot MRI-kenmerken bij patiënten met LSRS die potentiële kandidaten waren voor een lumbale discushernia operatie op basis van klinische gronden. De interbeoordelaars overeenkomst was hoog met betrekking tot klinisch relevante parameters zoals het meeste aangedane discus niveau en zenuwwortel, en de waarschijnlijkheid van een discus hernia en zenuwwortel compressie. In het algemeen werd er echter een aanzienlijke variatie tussen de beoordelaars waargenomen met betrekking tot specifieke kenmerken van het meest aangedane niveau (zoals signaalintensiteit op T2 sequenties, afwezigheid epiduraal vet en zichtbaarheid wortelzakje) en de discus hernia zelf (zoals locatie, grootte, en of het geclassificieerd moet worden als protrusie of extrusie).

Hoofdstuk 3 rapporteert over de implicaties van interbeoordelaars variatie tussen drie wervelkolomspecialisten voor de 1-jaars uitkomsten in patiënten die potentiële kandidaten waren voor een lumbale discushernia operatie op basis van klinische gronden. Na 1 jaar follow-up werden de meest gunstige klinische uitkomst resultaten gerapporteerd door de patiënten bij wie allerdrie de MRI beoordelaars het onafhankelijk eens waren over de aanwezigheid van een hernia of wortelcompressie, gevolgd door patiënten met inconsistente MRI interpretatie, en tenslotte door degenen bij wie de beoordelaars het eens waren dat er geen discus hernia of wortelcompressie bij aanvang aanwezig was.

Van de patiënten die werden gerandomiseerd voor afwachtend (conservative) beleid in de Sciatica Trial werd uiteindelijk een aanzienlijk deel van 39% geopereerd in het eerste jaar na randomisatie. **Hoofdstuk 4** evalueert of kwalitatieve en kwantitatieve MRI evaluaties deze chirurgie hadden kunnen voorspellen. Kwalitatieve MRI parameters en de aanvangsgrootte van de hernia verschilden niet significant tussen de chirurgische en niet-chirurgische groep. Patiënten die wel geopereerd werden hadden bij aanvang (baseline) kleinere durale zakken en spinale kanalen in vergelijking met patiënten die niet werden geopereerd. Echter, ROC-curve analyse toonde aan dat deze variabelen een slecht vermogen hebben om goed onderscheid te maken tussen patiënten die wel en geen operatie ondergaan voor LSRS.

Hoofdstuk 5 rapporteert over de prognostische waarde van lage rugklachten in relatie tot MRI bevindingen bij patiënten met het LSRS. Patiënten die bij baseline invaliderende rugpijn hadden rapporteerden een ongunstige prognose na een jaar in vergelijking met patienten met alleen overwegend LSRS. Als bovendien een duidelijke discus hernia met zenuwwortel compressie op MRI afwezig was, waren de resultaten nog slechter. Discus hernia's en zenuwwortel compressie op MRI kwamen vaker voor onder patiënten met overwegend LSRS in vergelijking met degenen die naast LSRS ook invaliderende pijn in de rug hadden. Vertebrale eindplaat signaal veranderingen waren evenredig verdeeld tussen degenen met en zonder invaliderende rugpijn. Grote en geëxtrudeerde hernia's waren ook gelijk verdeeld tussen beide groepen.

Hoofdstuk 6 rapporteert over de 1-jaars MRI bevindingen in patiënten met LSRS die werden behandeld met ofwel een operatie of afwachtend (conservatief) beleid, veranderingen van MRI-bevindingen over de tijd, en de correlatie tussen MRI bevindingen en klinische uitkomst. Na een jaar follow-up had een aanzienlijk deel van de patiënten nog steeds een hernia op MRI (21% van de chirurgisch behandelde patiënten vergeleken met 60% van conservatief behandelde patiënten). Echter de aanwezigheid van hernia of zenuwwortel compressie op MRI kon niet discrimineren tussen enerzijds patiënten met aanhoudende of terugkerende symptomen van LSRS en anderzijds patiënten zonder symptomen.

Hoofdstuk 7 rapporteert over Vertebrale Eindplaat Signaal Veranderingen (VESV), veranderingen van VESV over de tijd en de correlatie tussen VESV bevindingen en rugpijn in LSRS. Het ondergaan van een operatie voor LSRS was sterk geassocieerd met progressie in de mate van VESV ten opzichte van conservatieve (niet-operatieve) zorg. In een jaar tijd toonde ongeveer twee derde van chirurgisch en een vijfde van conservatief behandelde patiënten een toename van VESV. Echter, zowel bij aanvang als na een jaar follow-up, was er nauwelijks een proportioneel verschil wat betreft invaliderende rugpijn tussen de personen met en zonder VESV. Daarnaast was de proportie van patienten die volledig van LSRS waren hersteld ook nagenoeg gelijk tussen de personen met en zonder VESV.

Hoofdstuk 8 rapporteert over de interbeoordelaars variatie van Gadolinium-aankleuringen, de prevalentie van Gadolinium aankleuringen, en hun relatie met klinische uitkomst. De MRI beoordelaars (twee neuroradiolgen en 1 neurochirurg) toonden een uitstekende overeenstemming met betrekking tot discus niveau van de hernia en de meest getroffen zenuwwortel. Bovendien een goede overeenstemming aangaande de aanwezigheid van een hernia en wortelcompressie. Dit in tegenstelling tot alleen een redelijke overeenstemming met betrekking tot de aankleuring van de discus hernia en de aangedane zenuwwortel. Een jaar na de behandeling bereikten de beoordelaars substantiële overeenstemming met betrekking of er nog steeds een hernia aanwezig is en of de zenuwwortel nog gecomprimeerd wordt. Zij waren het echter weer aanzienlijk oneens over de aankleuring van de hernia en de zenuwwortel. De beoordelaars bereikten matige tot substantiële overeenstemming met betrekking tot de aanwezigheid van littekenweefsel na een jaar. Als ze beoordeelden dat littekenweefsel aanwezig is, beschouwden ze het bijna altijd als aangekleurd. Na 1 jaar werd geen relatie gevonden tussen aankleuring van de zenuwwortel of discus hernia en klinische bevindingen.

CONCLUSIE

Patiënten die in het acute stadium van LSRS ook invaliderende rugpijn hadden rapporteerden een ongunstige prognose na een jaar in vergelijking met patienten met alleen overwegend LSRS. Als bovendien een duidelijke discus hernia met zenuwwortel compressie op MRI afwezig was, waren de resultaten nog slechter. MRI-bevindingen in het acute stadium van LSRS konden geen onderscheid maken tussen patiënten die wel en geen operatie hebben ondergaan voor LSRS. Na een jaar follow-up konden anatomische afwijkingen zichtbaar op MRI ook geen onderscheid maken tussen enerzijds patiënten met aanhoudende of terugkerende symptomen van LSRS en anderzijds patiënten zonder symptomen. Vertebrale Eindplaat Signaal Veranderingen waren niet geassocieerd met invaliderende rugpijn, noch in het acute stadium van LSRS, noch na een jaar follow-up. De interbeoordelaars variatie met betrekking tot wel of geen

Gadolinium-aankleuring van MRI bevindingen was laag en er werd geen relatie geobserveerd tussen Gadolinium-ankleuring en klinische bevindingen na een jaar follow-up.

CURRICULUM VITAE

Abdelilah el Barzouhi was born on February 10th 1987 in The Hague, the Netherlands. After graduating from secondary school (Gymnasium, Nature & Technique and Nature & Health, Segbroek College, the Hague), he started Medical school in 2005 at the Erasmus University Rotterdam. In 2007, Abdelilah was among the top 10% of medical students and was selected to participate in a special program for medical students organized by the Netherlands' Institute of Health Sciences (NIHES). This program enabled him to combine Medical school with the Master of Science in Clinical Research program. During this program he received his initial training in epidemiology, part of which was spent at the Harvard School of Public Health in Boston, Massachusetts, USA. As part of this master program he participated in research at the department of Epidemiology ErasmusMC (head: Prof.dr. A. Hofman). In 2007 Abdelilah successfully participated in the Erasmus University Honours Program, an interdisciplinary academic training program to explore the boundaries of science together with students from other disciplines.

In 2007 he started doing research at the Department of Internal Medicine of ErasmusMC (supervisors: Dr. B.J. Rijnders and Dr. M.E. Van der Ende). In 2009 he was awarded the Young Investigators Award for one of his research projects.

In 2009 he started the work described in this thesis at the department of Neurosurgery of the Leiden University Medical Center (supervisors: prof. dr. W.C. Peul and Dr. C.L.A.M. Vleggeert-Lankamp).

In 2010 he graduated cum laude from the MSc in Clinical Research at the Netherlands Institute of Health Sciences. Parallel to Medical school and his PhD research, Abdelilah also participated in a master program in Healthcare Management at the institute of Health Policy of the Erasmus University Rotterdam. In 2010 he successfully obtained the MSc degree in Healthcare management.

In the summer of 2010 he participated in the project IMSTAR (Internship for Medical STudents in Action for Refugees) in Jordan with other Dutch medical students from different faculties. In addition to a medical internship in the emergency department of a local state hospital, several refugee camps were visited.

During the opening of the academic year of the Erasmus University Rotterdam in 2012, Abdelilah was awarded with the prestigious Professor Lambers Prize, which is each year presented to one excellent student who has at least two master's degrees obtained at different faculties, and one of the master's degrees should have been assessed as excellent (cum laude).

In August 2012 he completed his medical studies. Subsequently he maintained working on the work described in this thesis. In 2013 he started working as a resident in neurosurgery at the St. Radboud University Medical Center in Nijmegen (head: Dr. R.H.M.A. Bartels).

In October 2013 he received the 2013 Outstanding Paper Award for Surgical Science during the North American Spine Society 28th Annual Meeting in New Orleans, USA.

LIST OF PUBLICATIONS

THIS THESIS

El Barzouhi, A., Vleggeert-Lankamp, C.L.A.M., Lycklama à Nijeholt, G.J., Van der Kallen B.F., Van den Hout, W.B., Jacobs, W.C.H., Koes B.W., Peul W.C. (2013) Magnetic Resonance Imaging in clinical follow-up assessment of sciatica **N Engl J Med** 2013 Mar 14;368(11):999-1007

El Barzouhi, A., Vleggeert-Lankamp, C.L.A.M., Lycklama à Nijeholt, G.J., Van der Kallen B.F., Van den Hout, W.B., Verwoerd, A.J.H., Koes B.W., Peul W.C. (2013) Magnetic Resonance Imaging interpretation in patients with sciatica who are potential candidates for lumbar disc surgery **PLoS One.** 2013 Jul 10;8(7):e68411.

El Barzouhi, A., Vleggeert-Lankamp, C.L.A.M., Van der Kallen B.F., Lycklama à Nijeholt, G.J., Van den Hout, W.B., Verwoerd, A.J.H., Koes B.W., Peul W.C. (2013) Back pain's association with vertebral endplate signal changes in sciatica. **North American Spine Society Outstanding Paper Award for Surgical Science, accepted for publication in The Spine Journal**

El Barzouhi, A., Vleggeert-Lankamp, C.L.A.M., Lycklama à Nijeholt, G.J., Van der Kallen B.F., Van den Hout, W.B., Verwoerd, A.J.H., Koes B.W., Peul W.C. (2013) Prediction of surgery for sciatica with Magnetic Resonance Imaging **Accepted by Journal of Neurosurgery: Spine**

El Barzouhi, A., Vleggeert-Lankamp, C.L.A.M., Van der Kallen B.F., Lycklama à Nijeholt, G.J, Van den Hout, W.B., Verwoerd, A.J.H., Koes B.W., Peul W.C. (2013) Prognostic value of MRI in sciatica patients in relation to back pain S**ubmitted for publication**

El Barzouhi, A., Vleggeert-Lankamp, C.L.A.M., Lycklama à Nijeholt, G.J., Van der Kallen B.F., Van den Hout, W.B., Verwoerd, A.J.H., Koes B.W., Peul W.C. (2013) Prognostic value of inconsistent interpretation of Magnetic Resonance Imaging in patients with sciatica **Submitted for publication**

El Barzouhi, A., Vleggeert-Lankamp, C.L.A.M., Lycklama à Nijeholt, G.J., Van der Kallen B.F., Van den Hout, W.B., Koes B.W., Peul W.C. (2013) Reliability of Gadolinium-enhanced MRI findings and their correlation with clinical outcome in patients with sciatica **Submitted for publication**

OTHER PUBLICATIONS IN PEER REVIEWED JOURNALS

Lopes, V.B ., Hassing, R.J., de Vries-Sluijs, T.E., **El Barzouhi, A.**, Hansen, B.E., Schutten, M., de Man, R.A., van der Ende, M.E. (2012) Long-term response rates of successful hepatitis B vaccination in HIV-infected patients **Vaccine**. 2013 Feb 4;31(7):1040-4

Pingen, A., van der Ende, M.E., Wensing, A., **El Barzouhi, A.**, Simen, B., Schutten, M., Boucher, C. (2012) Deep sequencing does not reveal additional transmitted mutations in patients diagnosed with HIV-1 variants with single nucleoside reverse transcriptase inhibitor resistance mutations. **HIV Med.** 2012 Sep 19

El Barzouhi, A., Elias-Smale, S., Dehghan, A., Vliegenthart, R., Oudkerk, M., Hofman, A., Witteman, J.C.M. (2011) Renal function is related to severity of coronary artery calcification in elderly persons; The Rotterdam Study. **PLoS One**. 2;6(2):e16738

Slobbe, L., Polinder, S., Doorduijn, J.K., Lugtenburg, P.J., **El Barzouhi, A.**, Boersma, E., Van Leeuwen W.B., Rijnders, B.J.A. (2010) Prevention of endoluminal catheter-related bacteraemia with a daily 70%-ethanol lock in haematology patients with long-term central venous catheters. Randomised placebo-controlled trial. **PLoS One.** 26;5(5):e10840

Slobbe, L., **El Barzouhi, A.**, Boersma, E., Rijnders, B.J.A. (2009) Comparison of the roll plate and sonication method to diagnose catheter colonisation and bacteraemia in patients with long-term tunnelled catheters. A randomised prospective study. **J Clin Microbiol**. 47(4):885-8

Slobbe, L., Polinder, S., Doorduijn, J.K., Lugtenburg, P.J., **El Barzouhi, A.**, Steyerberg, E.W., Rijnders, B.J.A. (2008) Outcome and medical costs of patients with invasive aspergillosis and acute myelogenous leukemia-myelodysplastic syndrome treated with intensive chemotherapy: an observational study. **Clin Infect Dis**. 15;47(12):1507-12