Spine

SPINE Volume 40, Number 11, pp E653-E660

DIAGNOSTICS

Diagnostic Accuracy of Lumbosacral Spine Magnetic Resonance Image Reading by Chiropractors, Chiropractic Radiologists, and Medical Radiologists

Annemarie de Zoete, DC,* Raymond Ostelo, PT, PhD,†‡ Dirk L. Knol, PhD,§ Paul R. Algra, MD, PhD,¶ Jan T. Wilmink, MD, PhD,‖ Maurits W. van Tulder, PhD,†‡ and MRI Study Group**††

Study Design. A cross-sectional diagnostic accuracy study was conducted in 2 sessions.

Objective. It is important to know whether it is possible to accurately detect "specific findings" on lumbosacral magnetic resonance (MR) images and whether the results of different observers are comparable.

Summary of Background Data. Health care providers frequently use magnetic resonance imaging in the diagnostic process of patients with low back pain. The use of MR scans is increasing. This leads to an increase in costs and to an increase in risk of inaccurately labeling patients with an anatomical diagnosis that might not be the actual cause of symptoms.

Methods. A set of 300 blinded MR images was read by medical radiologists, chiropractors, and chiropractic radiologists in 2 sessions. Each assessor read 100 scans in round 1 and 50 scans in round 2. The reference test was an expert panel.

For all analyses, the magnetic resonance imaging findings were dichotomized into "specific findings" or "no specific findings." For the agreement, percentage agreement and κ values were calculated and for validity, sensitivity, and specificity. Sensitivity analysis was done for classifications A and B (prevalence of 31% and 57%, respectively).

Results. The intraobserver κ values for chiropractors, chiropractic radiologists, and medical radiologists were 0.46, 0.49, and 0.69 for A and 0.55, 0.75, and 0.64 for B, respectively.

The interobserver κ values were lowest for chiropractors (0.28 for A, 0.37 for B) and highest for chiropractic radiologists (0.50 for A, 0.49 for B).

The sensitivities of the medical radiologists, chiropractors, and chiropractic radiologists were 0.62, 0.71, and 0.75 for A and 0.70, 0.74, 0.84 for B, respectively.

The specificities of medical radiologists, chiropractic radiologists, and chiropractors were 0.82, 0.77, and 0.70 for A and 0.74, 0.52, and 0.61 for B, respectively.

Conclusion. Agreement and validity of MR image readings of chiropractors and chiropractic and medical radiologists is modest at best. This study supports recommendations in clinical guidelines against routine use of magnetic resonance imaging in patients with low back pain.

Key words: radiologist, chiropractic, MRI, reliability, validity, lumbar disc herniation, spinal stenosis.

Level of Evidence: 3

Spine 2015;40:E653–E660

From the *Department Health Sciences and EMGO+ Institute for Health and Care Research, Faculty of Earth & Life Science, VU University Medical Center Amsterdam, Amsterdam, the Netherlands; †Department of Epidemiology and Biostatistics, EMGO+ Institute for Health and Care Research, VU University Medical Center Amsterdam, Amsterdam, the Netherlands; ‡Department of Health Sciences, Faculty of Earth & Life Sciences, VU University Amsterdam, Amsterdam, the Netherlands; §Department of Epidemiology and Biostatistics, EMGO+ Institute for Health and Care Research, VU University Medical Center Amsterdam, Amsterdam, the Netherlands; ¶Medical Centre, Alkmaar, Alkmaar, the Netherlands; ‖MRI Centre, Amsterdam, the Netherlands; **Department of Health Science and EMGO+ Institute for Health and Care Research, Faculty of Earth & Life Science, VU University Amsterdam, the Netherlands; and ††Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, the Netherlands. MRI group, they are working mainly in private practice.

Acknowledgement date: November 18, 2014. Acceptance date: March 6, 2015.

The manuscript submitted does not contain information about medical device(s)/drug(s).

No funds were received in support of this work.

Relevant financial activities outside the submitted work: consultancy, grants.

Address correspondence and reprint requests to Annemarie de Zoete, DC, Department of Health Sciences, Faculty of Earth & Life Sciences, VU University Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam; E-mail: a.de.zoete@vu.nl

DOI: 10.1097/BRS.0000000000000896

Use of magnetic resonance imaging in patients with low back pain is increasing because of increased availability, the high degree of anatomical detail, and the assumed reassurance of the patient and the practitioner.[1,2]

Although clinical guidelines[3–5] advise against routinely imaging the lumbar spine, there has been a more than 300% increase in the use of magnetic resonance imaging between 1994 and 2005.[1] Also approximately one-third of Medicare patients with low back pain received a MR before they were treated conservatively.[1]

As low back pain is very common, this increase in imaging not only leads to an increase in direct costs of imaging. In addition, as imaging also has a substantial rate of false positives, this increase in imaging also increases the number of patients inaccurately labeled with an anatomical diagnosis that might not be the actual cause of symptoms. As it has been shown that there is an association between rates of magnetic resonance imaging utilization and back surgery,[1] it might be hypothesized that as a consequence of this increase, back surgery rates may increase.

In most countries, medical practitioners will refer patients for magnetic resonance imaging when they find a clinical indicator of specific or serious pathologies during history taking or physical examination. Medical radiologists prepare formal reports to inform the medical practitioners of the findings. In some countries also, primary care health care providers such as chiropractic radiologists and chiropractors may also be involved in the interpretation of MR images.

In both settings, imaging is recommended only when there is a suspicion of a specific pathology such as infection, malignancy, fracture, herniated disc, or spinal stenosis.[6] If such specific pathology is detected on MR image, either the patient needs referral to a medical specialist or the intervention plan needs to be modified.

So, various health care providers are involved in referring and interpretation of diagnostic imaging in low back pain. But there are some distinct differences in training between these different professions.

Medical radiologists undergo at least 5 years of postgraduate training in diagnostic imaging and often work in hospitals or diagnostic centers. Chiropractors are trained in diagnostic imaging during their undergraduate education and interpret diagnostic imaging in private practice. Chiropractic radiologists are chiropractors who undergo 3 years of postgraduate training in radiology. They teach at chiropractic colleges or work in diagnostic imaging centers. Several studies have investigated the diagnostic accuracy of reading MR image by radiologists and medical specialists,[7–16] but there are no studies that directly compare the performance of chiropractors and chiropractic radiologists with medical radiologists. Therefore, the aim of the current study is to evaluate the intra- and interobserver agreement and validity of MR image interpretation of "specific findings" in the lumbosacral spine by medical radiologists, chiropractic radiologists, and chiropractors.

MATERIALS AND METHODS

The study consisted of 2 parts. First, the intra- and interobserver agreement per professional group (medical radiologists, chiropractic radiologists, and chiropractors) was evaluated and compared. Second, the validity of MR image interpretation by the 3 professional groups was evaluated and compared.

MR Images

Three hundred sets of MR images of the lumbosacral spine of patients referred by primary care clinicians and specialists were retrospectively selected, using a computerized database from a general hospital. Its medical ethical committee approved the study. The magnetic resonance imaging unit used was a 1.5 T Siemens Magnetom Avanto system. Each magnetic resonance imaging study consisted of at least set of T1- and T2-weighted spin-echo sagittal images and T1- and T2-weighted spin-echo axial images from the L3 to L5 vertebral levels angulated through the disc.

Only MR scans of which the diagnostic quality of the image was sufficient and obtained from patients 18 years of age and older were included. The MR scans were blinded for all patient characteristics and clinical data.

In this study, "specific findings" were defined as infections, malignancies, fractures, herniated disc, and central stenosis. The relevance of a specific finding on MR image when the correlate with clinical examination findings, especially in primary care, is that regardless of the type of specific finding, the patient needs referral to a medical specialist or the treatment plan needs to be reassessed. The frequency of the "specific findings" was slightly higher than the actual prevalence in primary care[5] to ensure sufficient power. The prevalence of each abnormality is presented in Table 1. To reflect everyday practice, we aimed to select scans with a range of different stages of each specific finding.

TABLE 1. Prevalence of the "Specific Findings" in the Final Set of MR Images Used in This Study (n = 300)

Diagnosis*	Prevalence No. of MR Images (%)
1. Malignancy	2 (0.7)
2. Infection	3 (1.0)
3. Fracture	7 (2.3)
4. Herniated disc with clear nerve root compression	75 (25.0)
5. Herniated disc with doubtful nerve root compression	60 (20.0)
6. Herniated disc without nerve root compression	89 (29.7)
7. Central stenosis with nerve root compression	23 (7.7)
8. Central stenosis with doubtful nerve root compression	35 (11.7)
9. Central stenosis without nerve root compression	19 (6.3)

Prevalence of classification A is 31% and consists of specific findings 1, 2, 3, 4, and 7. Prevalence of classification B is 57% and consists of specific findings 1, 2, 3, 4, 5, 7, and 8.

**An overrepresentation of these important findings was used as described in the methodology. Different "specific findings" can be present in 1 case. If a "specific finding" is present at more than 1 level, it was counted as 1 "specific finding."*

MR indicates magnetic resonance.

Reference Test

A chiropractic radiologist and a medical radiologist, neither of whom participated in this study as an assessor, dichotomized the magnetic resonance imaging studies into "specific" *versus* "no specific" findings. A second medical radiologist checked a random sample of the "normal" scans, 75 "abnormal" scans to confirm the "specific findings," and the quality of the scans. If consensus could not be reached between the experts (in 6 cases) about "specific findings," the first medical radiologist made the final decision.

As the criteria to define nerve root involvement in disc herniation and central stenosis are not always clear-cut, 2 classifications were applied to the MR images with "specific finding." In classification A malignancy, fracture, infection, herniated disc with definite nerve root involvement, and central stenosis with definite nerve root involvement were classified as a "specific finding" (prevalence 31%). In classification B, herniated disc with doubtful nerve root involvement and central stenosis with doubtful nerve root involvement were also labeled as "specific findings" (prevalence 57%).

Assessors

Six chiropractors, 6 chiropractic radiologists, and 6 medical radiologists were invited to participate.

Design

The MR images were divided in 6 groups of 50 images; the prevalence of specific findings was approximately equal in each group. Each set of 50 images was randomly ordered, differently in both sessions. The images were provided in Dicom files on DVD. The viewing program K-Pacs 1.6.0 was used.

All assessors, who were unaware of the total number of abnormal images, evaluated the images independently. In every professional group, the first 2 assessors read 1 to 100 images, the third and fourth assessors read 101 to 200 images, and the fifth and sixth assessors read 201 to 300 images. After 3 months, assessors reread 50 images of the initial set of 100 images (Table 2).

Rating List

Definitions for "specific findings" were derived from the literature and consensus and collected in a detailed handbook[14,17–19] (see Supplemental Digital Content Appendix, available at: http://links.lww.com/BRS/A972). Each assessor noted the presence of each "specific finding" separately. Scores of each assessor were dichotomized in "abnormal" and "normal."

Analysis

The sample size was calculated according to Sim and Wright.[20] The number of subjects required in a 2-rater study assuming that the null hypothesis value of κ is 0.40 and 0.30 proportion of positive ratings is 255 with a power of 90%. Our pooled κ was based on 3 κ values as each assessor read only 100 images as described previously.

Agreement

We calculated the agreement percentages and Cohen κ. Values of κ less than 0.20 indicate poor agreement, whereas values between 0.20 and 0.40 indicate fair agreement, between 0.40 and 0.60 moderate agreement, between 0.60 and 0.80 substantial agreement, and values exceeding 0.80 indicate an almost perfect agreement.[21]

For intraobserver agreement, the ratings of the same assessor in rounds 1 and 2 were used. The κ values and agreement percentages for both classifications were calculated for all assessors in SPSS 17.0.

To arrive at a κ per professional group, the κ values and standard error of the assessors per professional group were calculated and combined to a weighted κ and 95% confidence interval, using the Stata procedure "metan" as described by DerSimonian and Laird.[22] The assigned weights were the precisions, which is the inverse of the variance or the inverse of the square of the SE. This method is often used in fixed-effects meta-analysis.

The results for the 3 professional groups were pairwise compared (for pair i, j) using the test statistic $z = (\hat{\kappa}_i - \hat{\kappa}_j) / \sqrt{se_i^2 + se_j^2}$.

A *P* value of 0.05 or less was considered significant.

TABLE 2. Design of the Test-Retest Study for Each Group of Raters

	Raters											
	Rater 1		Rater 2		Rater 3		Rater 4		Rater 5		Rater 6	
MR Images	Round 1	Round 2	Round 1	Round 2	Round 1	Round 2	Round 1	Round 2	Round 1	Round 2	Round 1	Round 2
1–50	X	X	X									
51–100	X		X	X								
101–150					X	X	X					
151–200					X		X	X				
201–250									X	X	X	
251–300									X		X	X

MR indicates magnetic resonance.

For the interobserver agreement, the ratings of 2 members of a particular professional group who read the same images (round 1) were used. The pooled interobserver agreement per professional group was calculated in a similar way as the intraobserver agreement; only now the κ values and agreement percentages were calculated for all pairs per professional group. The results for the 3 groups were compared similarly to the intraobserver agreement.

Validity

The 2 × 2 tables for both classifications were produced with the results of round 1 for each assessor. The results were pooled in Stata using procedure metandi, because there is a negative correlation between sensitivity and specificity and, therefore, the pooling of the sensitivity and specificity needs to be performed simultaneously.[23]

The pooled sensitivities and specificities of the different professional groups were compared in the same way as the agreement.

RESULTS

Assessors

A total of 18 assessors initially agreed to participate in the study. One medical radiologist and 1 chiropractic radiologist did not score any of the images because of lack of time. One medical radiologist scored 19 studies and 1 chiropractic radiologist scored 33 studies. Two medical radiologists scored only the first round. All other assessors scored both rounds.

One of the medical radiologists (n = 5) was in his last year of postgraduate training. The experience of the other medical radiologists, working in 3 different hospitals in the Netherlands, ranged from 3 to 32 years (median of 4 yr) and all had a specialization in neuroradiology. The professional experience of the chiropractic radiologists (n = 5) varied from 5 to 25 years (median of 9 yr). One of them worked in Australia and the others in the United States. Five of the chiropractors (n = 6) had graduated from the same college. All of them worked in private practice in the Netherlands for 3 to 8 years.

Agreement

Intraobserver Agreement

For classification A, the medical radiologists demonstrated substantial agreement, whereas the chiropractic radiologists and chiropractors demonstrated moderate agreement; this difference was statistically significant. For classification B, the chiropractic radiologists and medical radiologists demonstrated substantial intraobserver agreement, whereas the chiropractors demonstrated moderate agreement. The difference between the chiropractors and chiropractic radiologists was statistically significant (Table 3).

Interobserver Agreement

All κ values of the chiropractic radiologists and medical radiologists were moderate. The κ values of the chiropractors were fair. For classification A, the chiropractic radiologists

TABLE 3. Intraobserver Agreement (Rounds 1–2) in the Interpretation of Lumbosacral MR Images (n = 300) for Specific Findings in Classifications A and B per Professional Group and the Comparison of the 3 Professional Groups

	Intraobserver Agreement		Comparison of the κ of the 3 Professional Groups		
	Classification A*		Classification A*		
Professional Group	Agreement Percentage (Range)	κ (95% CI)	Chiropractors *vs.* Chiropractic Radiologists (*P*)	Chiropractors *vs.* Medical Radiologists (*P*)	Chiropractic Radiologists *vs.* Medical Radiologists (*P*)
Chiropractors (n = 6)	66%–84%	0.46 (0.35–0.56)	0.68 NS	0.007	0.02
Chiropractic radiologists (n = 4)	64%–84%	0.49 (0.37–0.61)			
Medical radiologists (n = 2)	86%–88%	0.69 (0.56–0.83)			
	Classification B†		Classification B†		
Chiropractors (n = 6)	68%–84%	0.55 (0.45–0.64)	0.006	0.32 NS	0.26 NS
Chiropractic radiologists (n = 4)	86%–92%	0.75 (0.64–0.85)			
Medical radiologists (n = 2)	80%–84%	0.64 (0.48–0.79)			

**Prevalence 31%.*
†Prevalence 57%.
95% CI indicates 95% confidence interval; NS, not significant (P > 0.05).

performed significantly better than the chiropractors; all other comparisons were not statistically significant (Table 4).

Validity

The pooled sensitivities and specificities are presented in Table 5. The sensitivity in our study ranged from 0.62 to 0.87 and specificities ranging from 0.52 to 0.82.

For classification A, there were no statistically significant differences between the groups. For classification B, the specificity of the chiropractic radiologists was statistically significantly lower than that of the medical radiologists, but the sensitivity of the chiropractic radiologists was statistically significantly higher than that of the medical radiologists.

DISCUSSION

This diagnostic accuracy study showed that agreement of the medical radiologists and chiropractic radiologists was higher than that of chiropractors, but overall, the agreement was moderate. The validity was reasonable, but still a substantial number of MR images were misclassified.

Agreement

It was difficult to compare our results with those in other studies, because most other studies have calculated κ values for specific signs at specific levels instead of calculating a κ value for the classification of "specific findings" as we applied in our study. There were also differences in prevalence and patient populations. In most other studies, the patient population was more homogeneous, for example, candidates for surgery.[7–9,24] In our study, the images were obtained from all patients referred for lumbosacral magnetic resonance imaging. The heterogeneity of patients might explain the lower agreement compared with other studies. However, our findings better reflect the situation in primary care.

Two studies included different health care providers.[25,26] Speciale *et al*[25] found that medical radiologists scored better than other health care providers, whereas Lurie *et al*[26] found less difference between different health care providers, which probably more closely matches our outcomes.

Several studies[8,9,14,27,28] have assessed agreement identifying nerve root involvement, caused by herniated discs or central stenosis, showing moderate agreement. Although our study was not limited to these criteria, herniated discs and central stenosis with nerve root involvement were the most common "specific findings" in our study. As such, our results seem consistent with what has been previously published.

Although these previous studies[7–9,14,24–29] were different in design, similar κ values were reported. Combining the results of our study with the published studies seems to suggest that in clinical practice only moderate agreement can be expected.

Validity

The ideal "gold standard" would be to confirm all specific findings by combination of clinical findings, laboratory test,

TABLE 4. Interobserver Agreement in the Interpretation of Lumbosacral MR Images (n = 300) for Specific Findings in Classifications A and B per Professional Group (Round 1) and the Comparison of the 3 Professional Groups

	Interobserver Agreement		Comparison of the κ of the 3 Professional Groups		
	Classification A*		Classification A*		
Professional Group	Agreement Percentage (Range)	κ (95% CI)	Chiropractors *vs.* Chiropractic Radiologists (*P*)	Chiropractors *vs.* Medical Radiologists (*P*)	Chiropractic Radiologists *vs.* Medical Radiologists (*P*)
Chiropractors (n = 6)	62%–71%	0.28† (0.18–0.38)	0.01	0.11	0.57
				NS	NS
Chiropractic radiologists (n = 5)	66%–88%	0.50‡ (0.36–0.64)			
Medical radiologists (n = 5)	69%–90%	0.43‡ (0.27–0.60)			
	Classification B§		Classification B§		
Chiropractors (n = 6)	69%–73%	0.37† (0.28–0.47)	0.17	0.38	0.55
			NS	NS	NS
Chiropractic radiologists (n = 5)	67%–91%	0.49‡ (0.35–0.63)			
Medical radiologists (n = 5)	73%–74%	0.46‡ (0.30–0.62)			

**Prevalence 31%.*
†Combines the results of 3 pairs.
‡Combines the results of 2 pairs.
§Prevalence 57%.
95% CI indicates 95% confidence interval; NS, not significant (P > 0.05).

TABLE 5. Validity of the Interpretation of Lumbosacral MR Images (n = 300) for Specific Findings in Classification A and B per Professional Group in Session 1 and the Comparison of the 3 Professional Groups

	Validity		Comparison of the 3 Professional Groups					
	Classification A*		Classification A*					
			Chiropractors *vs.* Chiropractic Radiologists (*P*)		Chiropractors *vs.* Medical Radiologists (*P*)		Chiropractic Radiologists *vs.* Medical Radiologists (*P*)	
Professional Group	**Pooled Sensitivity (95% CI)**	**Pooled Specificity (95% CI)**	**Pooled Sensitivity**	**Pooled Specificity**	**Pooled Sensitivity**	**Pooled Specificity**	**Pooled Sensitivity**	**Pooled Specificity**
Chiropractors (n = 6)	0.71 (0.51–0.85)	0.70 (0.60–0.79)	0.71 NS	0.35 NS	0.45 NS	0.08 NS	0.12 NS	0.38 NS
Chiropractic radiologists (n = 5)	0.75 (0.66–0.82)	0.77 (0.68–0.84)						
Medical radiologists (n = 5)	0.62 (0.47–0.74)	0.82 (0.74–0.88)						
	Classification B†		**Classification B†**					
Chiropractors (n = 6)	0.79 (0.62–0.90)	0.61 (0.51–0.71)	0.37 NS	0.24 NS	0.54 NS	0.06 NS	0.005	0.0007
Chiropractic radiologists (n = 5)	0.87 (0.83–0.91)	0.52 (0.42–0.63)						
Medical radiologists (n = 5)	0.74 (0.68–0.79)	0.74 (0.67–0.80)						

*Prevalence 31%.
†Prevalence 57%.
95% CI indicates 95% confidence interval; NS, not significant ($P > 0.05$).

results of other diagnostic imaging, surgical results, and an expert panel opinion,[8,9,14,27,30,31] but this was not feasible in our study. The expert opinion was the best possible reference test we could accomplish. As the aim of the study was to compare and contrast the magnetic resonance imaging readings of these 3 professional groups and not to achieve "absolute" validity, the "relative" sensitivities and specificities were calculated.

The sensitivity of the findings in this study ranged from 0.62 to 0.87, with specificity ranging from 0.52 to 0.82. The sensitivities achieved by chiropractors and chiropractic radiologists were higher than those achieved by medical radiologists. However, the medical radiologists reached higher specificities. Chiropractors and chiropractic radiologists, if in doubt about a finding, tend to make a cautious decision to classify the MR image as abnormal in order not to under-report severe pathology. This is the preferred strategy in primary care. Medical radiologists are generally better in identifying true negatives. This means that for medical radiologists, the positive predictive value of reading isolated images is the highest.

Having said that, the results showed that in general there was considerable misclassification in all 3 groups.

Analysis of 2 Classifications

Defining nerve root involvement in the presence of herniated disc or central stenosis is not as clear-cut as compared with some other specific findings.[27] For example, the loss of anterior height of the vertebral body (>20% compared with the posterior height) is a clearly defined criterion for a compression fracture.

Therefore, similar to clinical practice, the assessors had to choose between definite, doubtful, and no nerve root involvement. The results were analyzed for 2 different classifications. The effect of this sensitivity analysis was greater for the κ values of the chiropractors than for the κ values of the other professions. Chiropractors were less certain about classifying nerve root involvement.

Assessors

Because of the considerable time investment by the participants, we decided to invite a convenience sample of health care professionals from our network. Consequently, it may be questionable whether they are an accurate representation of their professions. However, the assessors in all 3 professions worked at different practices and hospitals and a range of years of experience per professional group was accomplished. This variation in work setting and experience is in our opinion an adequate representation of usual practice.[25]

When interpreting sensitivity and specificity, some issues need to be taken into account. First, we used slightly higher prevalence rates to achieve sufficient power. A higher prevalence than in clinical practice may increase the sensitivity and lower the specificity,[32] because the assessors are more likely to score an image as abnormal, although the assessors are unaware of the prevalence of specific findings.[11,12,33,34] Second, all images were obtained from the same hospital thereby decreasing the variability that is expected to be present in everyday clinical practice where MR images will come from different hospitals. Moreover, consensus was reached on the definition of clinical entities and the scoring instructions, which will also lead to less variability as compared with everyday clinical practice.[8,9] In other words, the sensitivity and specificity as presented in this study, though moderate at best, could even be an overestimation of the true sensitivity and specificity in primary care.

In routine clinical practice, the results of MR images will always be used in combination with clinical information and other test results. The effect of withholding clinical information is not clear in the literature; it may have lowered the agreement and validity compared with clinical practice and may have had a greater effect on the results of the chiropractors. Radiologists are more accustomed to receive scarce patient information.[27,30,35–37]

Implication for Clinical Practice

If only moderate agreement and validity of MR image interpretation can be realized in clinical practice beside poor correlation between image findings and clinical examination, magnetic resonance imaging should not be routinely used in patients with low back pain. This is in line with international guidelines.[1,2,38,39] Imaging these patients can result in overdiagnosis of abnormalities, which can lead to an increase in referral to medical specialists, increased surgery rates, and higher costs. Besides, routine imaging is not associated with better patient outcomes and it can expose patients to unnecessary harm.[28]

Although guidelines recommend against the routine use of magnetic resonance imaging in patients with nonspecific low back pain, rates of utilization are still increasing.[4] The results of this study underpin the guidelines and show that implementation of back pain guidelines have the highest priority.

In summary, the agreement and validity of magnetic resonance imaging interpretation of chiropractors and chiropractic and medical radiologists is modest at best. This study supports clinical guidelines in their recommendations against routine use of magnetic resonance imaging in clinical management of patients with low back pain.

➢ Key Points

- ❑ Several health care providers can be involved in the interpretation of MR images of the lumbar spine. Therefore, it is important to evaluate the reliability and validity of MR image interpretation of "specific findings" in the lumbosacral spine.
- ❑ Agreement and validity of MR image readings of chiropractors and chiropractic and medical radiologists is modest at best.
- ❑ This study supports clinical guidelines in their recommendations against routine use of magnetic resonance imaging in clinical management of patients with low back pain.

Acknowledgments

The authors thank Martin Roosenberg, Charles Pfeifle, and Simone Knaap for reviewing an earlier draft of the manuscript.

MRI Study Group: Pauline Byrne, DC; Igor Dijkers, DC; Thomas Soeters, DC; Linda van Viegen, DC; Thomas van der Hof, DC; Winnie de Koning, DC; Cliff Tao, DC, DACBR; Anita Manne, DC, DACBR; Jeffrey Cooley, DC, DACBR; Michael Fergus, DC, DACBR; Rogier Hagenbeek, MD; Henk van Woerden, MD; and Michelle Wessely, DC, DACBR.

Supplemental digital content is available for this article. Direct URL citation appearing in the printed text is provided in the HTML and PDF version of this article on the journal's Web site (www.spinejournal.com).

References

1. Chou R, Deyo RA, Jarvik JG. Appropriate use of lumbar imaging for evaluation of low back pain. *Radiol Clin North Am* 2012;50:569–85.
2. Chou R, Fu R, Carrino JA, et al. Imaging strategies for low-back pain: systematic review and meta-analysis. *Lancet* 2009;373: 463–72.
3. van TM, Becker A, Bekkering T, et al. Chapter 3. European guidelines for the management of acute nonspecific low back pain in primary care. *Eur Spine J* 2006;15(suppl 2):S169-91.
4. Airaksinen O, Brox JI, Cedraschi C, et al. Chapter 4. European guidelines for the management of chronic nonspecific low back pain. *Eur Spine J* 2006;15(suppl 2):S192–300.
5. Chou R, Qaseem A, Snow V, et al. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. *Ann Intern Med* 2007;147:478–91.
6. Bussieres AE, Taylor JA, Peterson C. Diagnostic imaging practice guidelines for musculoskeletal complaints in adults-an evidence-based approach-part 3: spinal disorders. *J Manipulative Physiol Ther* 2008;31:33–88.
7. Carrino JA, Lurie JD, Tosteson AN, et al. Lumbar spine: reliability of MR imaging findings. *Radiology* 2009;250:161–70.
8. Lurie JD, Tosteson AN, Tosteson TD, et al. Reliability of readings of magnetic resonance imaging features of lumbar spinal stenosis. *Spine (Phila Pa 1976)* 2008;33:1605–10.
9. 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.
10. Lurie JD. What diagnostic tests are useful for low back pain? *Best Pract Res Clin Rheumatol* 2005;19:557–75.
11. Stadnik TW, Lee RR, Coen HL, et al. Annular tears and disk herniation: prevalence and contrast enhancement on MR images in the absence of low back pain or sciatica. *Radiology* 1998;206:49–55.
12. Weishaupt D, Zanetti M, Hodler J, et al. MR imaging of the lumbar spine: prevalence of intervertebral disk extrusion and sequestration, nerve root compression, end plate abnormalities, and osteoarthritis of the facet joints in asymptomatic volunteers. *Radiology* 1998;209:661–6.
13. Ledermann HP, Schweitzer ME, Morrison WB, et al. MR imaging findings in spinal infections: rules or myths? *Radiology* 2003;228:506–14.
14. Pfirrmann CW, Dora C, Schmid MR, et al. MR image-based grading of lumbar nerve root compromise due to disk herniation: reliability study with surgical correlation. *Radiology* 2004;230:583–8.
15. Oner FC, Ramos LM, Simmermacher RK, et al. Classification of thoracic and lumbar spine fractures: problems of reproducibility. A study of 53 patients using CT and MRI. *Eur Spine J* 2002;11: 235–45.
16. Haba H, Taneichi H, Kotani Y, et al. Diagnostic accuracy of magnetic resonance imaging for detecting posterior ligamentous complex injury associated with thoracic and lumbar fractures. *J Neurosurg* 2003;99(1 suppl):20–6.
17. 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:E93–113.
18. Fardon DF. Nomenclature and classification of lumbar disc pathology. *Spine (Phila Pa 1976)* 2001;26:461–2.
19. Yochum TR, Rowe LJ. *Essentials of Skeletal Radiology*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2004.
20. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005;85: 257–68.
21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
22. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177–88.
23. Reitsma JB, Glas AS, Rutjes AW, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982–90.
24. Mulconrey DS, Knight RQ, Bramble JD, et al. Interobserver reliability in the interpretation of diagnostic lumbar MRI and nuclear imaging. *Spine J* 2006;6:177–84.
25. Speciale AC, Pietrobon R, Urban CW, et al. Observer variability in assessing lumbar spinal stenosis severity on magnetic resonance imaging and its relation to cross-sectional spinal canal area. *Spine (Phila Pa 1976)* 2002;27:1082–6.
26. Lurie JD, Doman DM, Spratt KF, et al. Magnetic resonance imaging interpretation in patients with symptomatic lumbar spine disc herniations: comparison of clinician and radiologist readings. *Spine (Phila Pa 1976)* 2009;34:701–5.
27. van Rijn JC, Klemetso N, Reitsma JB, et al. Observer variation in MRI evaluation of patients suspected of lumbar disk herniation. *AJR Am J Roentgenol* 2005;184:299–303.
28. el Barzouhi A, Vleggeert-Lankamp CL, Nijeholt GJ, et al. Magnetic resonance imaging in follow-up assessment of sciatica. *N Engl J Med* 2013;368:999–1007.
29. Jarvik JG, Deyo RA. Moderate versus mediocre: the reliability of spine MR data interpretations. *Radiology* 2009;250:15–7.
30. Kuijper B, Beelen A, van der Kallen BF, et al. Interobserver agreement on MRI evaluation of patients with cervical radiculopathy. *Clin Radiol* 2011;66:25–9.
31. 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.
32. Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. *Stat Med* 1997;16:981–91.
33. Boden SD, Davis DO, Dina TS, et al. Abnormal magnetic-resonance scans of the lumbar spine in asymptomatic subjects. A prospective investigation. *J Bone Joint Surg Am* 1990;72:403–8.
34. Jarvik JJ, Hollingworth W, Heagerty P, et al. The Longitudinal Assessment of Imaging and Disability of the Back (LAIDBack) Study: baseline data. *Spine (Phila Pa 1976)* 2001;26: 1158–66.
35. Peterson C. Factors associated with success or failure in radiological interpretation: diagnostic thinking approaches. *Med Educ* 1999;33:251–9.
36. van Rijn JC, Klemetso N, Reitsma JB, et al. Symptomatic and asymptomatic abnormalities in patients with lumbosacral radicular syndrome: clinical examination compared with MRI. *Clin Neurol Neurosurg* 2006;108:553–7.
37. Doubilet P, Herman PG. Interpretation of radiographs: effect of clinical history. *AJR Am J Roentgenol* 1981;137:1055–8.
38. Jarvik JG, Deyo RA. Diagnostic evaluation of low back pain with emphasis on imaging. *Ann Intern Med* 2002;137:586–97.
39. 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.