

# Relationship of specific MRI findings to treatment outcomes in patients receiving transforaminal epidural steroid injections

Marco Lechmann<sup>1</sup> · Andrea Roskopf<sup>1</sup> · Christine Ehrmann<sup>1</sup> · Reto Sutter<sup>1</sup> · Christian W. A. Pfirrmann<sup>1</sup> · Cynthia K. Peterson<sup>1</sup>

Received: 5 August 2016 / Revised: 30 August 2016 / Accepted: 12 September 2016 / Published online: 28 September 2016  
© ISS 2016

## Abstract

**Objectives** To determine whether specific MRI findings are related to outcomes after lumbar transforaminal epidural steroid injections (TFESI) and to assess the inter-rater reliability of imaging diagnosis.

**Materials and methods** A prospective outcomes study on 156 consecutive patients with 1-month follow-up outcomes data and MRI within 3 months of TFESI was conducted. Pain levels (numerical rating scale) (NRS) were recorded prior to injection. Overall ‘improvement’ was determined using the Patients Global Impression of Change (PGIC) scale and NRS data were collected at three time points post injection. Two radiologists independently evaluated all images blinded to treatment outcome for reliability of diagnosis. The Chi-square test compared MRI findings for the senior radiologist to ‘improvement’. NRS change scores were compared to MRI findings with the unpaired *t*-test or ANOVA. Kappa and percent agreement assessed inter-rater agreement of diagnosis.

**Results** The only abnormality linked to ‘improvement’ ( $p = 0.03$ ) and higher NRS change scores ( $p = 0.0001$ ) at 1 month was the disc herniation morphology ‘protrusion + sequestration’. Patients with degeneration by osteophytes ( $p = 0.034$ ), grade 3 foraminal nerve root compression ( $p = 0.01$ ) and foraminal/extraforaminal location of herniation ( $p = 0.014$ ) also had higher 1 month NRS change scores. Reliability of diagnosis was ‘fair’ to ‘substantial’ depending on MRI findings.

**Conclusions** Patients with disc protrusion plus sequestration were significantly more likely to report overall improvement and more pain reduction at 1 month. Higher pain reduction was noted in patients with degeneration by osteophytes, grade 3 foraminal nerve root compression, or foraminal/extraforaminal disc herniation location.

**Keywords** Clinical competence · Treatment outcome · Injections · Intraspinal · Epidural

✉ Cynthia K. Peterson  
cynthia.peterson@balgrist.ch

Marco Lechmann  
marco.lechmann@gmx.ch

Andrea Roskopf  
andrea.rosskopf@balgrist.ch

Christine Ehrmann  
christine.ehrmann@gmx.ch

Reto Sutter  
reto.sutter@balgrist.ch

Christian W. A. Pfirrmann  
christian.pfirrmann@balgrist.ch

## Introduction

Seventy percent of people in industrialized countries will suffer from low back pain (LBP) placing huge burdens on the person affected, his/her family, society, health care systems and the wider economy [1–6]. Spinal stenosis and disc herniations are common causes of specific LBP and both can be associated with compression of nerve roots leading to radiculopathy and leg pain in addition to LBP.

Epidural steroid injections have become a well-established procedure for conservative treatment of chronic low back pain caused by radiculopathy [7–10]. The three ways to access the epidural space include transforaminal, interlaminar or a caudal approach [9]; however, although some patients benefit from the treatment, in a substantial portion of patients lumbar transforaminal epidural steroid injections (TFESI) fail to

<sup>1</sup> Department of Radiology, Orthopaedic University Hospital Balgrist, University of Zürich, Forchstrasse 340, 8008 Zürich, Switzerland

provide long-term relief. Researchers and clinicians are often at a loss in determining which patients will respond to this treatment and thus finding predictors for a positive or negative outcome is important [11].

Magnetic resonance imaging (MRI) is easily accessible in most western countries and it is assumed that abnormal findings can be measured more objectively than other patient factors [12]. With MRI, patients can be allocated to subgroups based on specific imaging features. Few trials exist assessing the prognostic power of MRI abnormalities for lumbar TFESI outcome [12–14]; however, because evaluation of MRI findings is still somewhat subjective, a number of classification systems for specific MR features have been developed and used [15–18] in an attempt to make the interpretation more objective and comparisons between studies more reliable. Nevertheless, further studies are needed to not only evaluate the reliability of using these classification systems, but also to assess which particular MRI findings, if any, are predictive of outcome after therapeutic interventions. Therefore, the primary purpose of this study is to determine whether or not specific MRI criteria are related to outcomes after lumbar TFESI in patients with radiculopathy due to intervertebral disc herniation (DH) and/or spinal stenosis. A secondary purpose is to assess the inter-rater reliability of identifying and classifying specific MRI findings.

## Materials and methods

This prospective outcomes study evaluated the MRI scans of 199 patients who had TFESI and returned their outcomes questionnaires; however, 44 of these patients were excluded due to lumbar spine surgery. In all, 156 consecutive patients met the inclusion criteria, returned their outcomes-based postal questionnaires and received imaging-guided lumbar TFESI from experienced radiologists at this orthopedic/rheumatologic university radiology department between June 2009 and February 2014. (From a previous study it is known that 24 % of patients return these postal questionnaires and thus those patients who did not return their questionnaires could not be included in this study [19].)

Ethics approval was received from the cantonal ethics commission and all patients signed informed consent. The inclusion criteria were: (1) patients with MRI confirmed lumbar DH and/or spinal stenosis who had imaging-guided lumbar TFESI and whose MRI scan was done at this hospital within 3 months of their injection, (2) abnormal MRI findings that could be linked to the presenting clinical presentation. The exclusion criteria included (1) clinical and/or imaging findings of myelopathy, (2) previous spinal surgery, (3) injection in more than one NR-level, and (4) spinal fractures, infections, tumors and spondylolytic spondylolisthesis. The radiologists doing the read-outs applied the exclusion criteria to each patient.

## Outcome measures

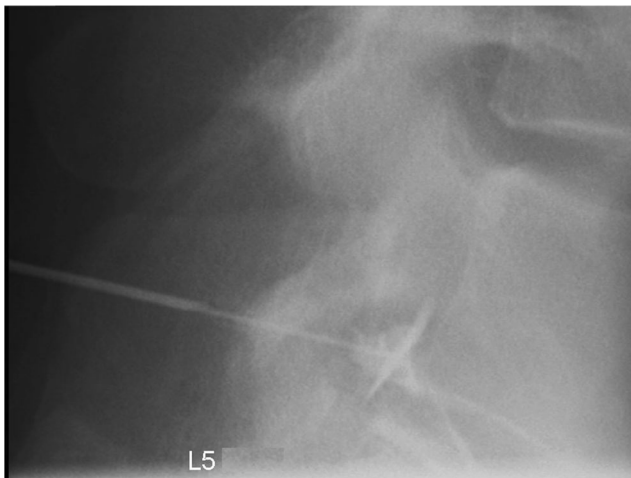
Baseline data collection included assessment of each patient's current pain level using the numerical rating scale for pain (NRS) where 0 = no pain and 10 = the worst pain imaginable. Post-treatment outcome data collection included the NRS for pain and assessment of the patient's overall 'improvement' post-injection using the Patient's Global Impression of Change (PGIC) scale [19–22]. Outcome data were collected at 1 day, 1 week and 1 month after treatment. The PGIC scale includes seven categorical descriptors of 'improvement' including 'much better', 'better', 'slightly better', 'unchanged', 'slightly worse', 'worse', and 'much worse'. The PGIC responses were dichotomized to 'improvement' (yes/no) and 'worsening' (yes/no). Clinically important 'improvement' only included the responses 'much better' or 'better'. 'Improvement' was the primary outcome measure. NRS scores, NRS change scores and 'worsening' (slightly worse, worse, much worse) were secondary outcome measures. Post-injection outcomes were acquired by short questionnaires that were given to the patients immediately after the injection in a pre-paid postal envelope and returned one month after the procedure.

## Lumbar TFESI procedure

All Injections were performed by radiologists in this orthopedic/rheumatology university hospital under fluoroscopic or computed tomographic guidance using sterile conditions (3× skin disinfection, sterile gloves, mask, sterile covering). After the administration of local anaesthesia, a 21 gauge-needle was inserted in a transforaminal approach to the affected nerve root. Prior to injection of 40-mg Kenacort (Triamcinoloni acetonium; Dermapharm AG, Huenenberg AG, Switzerland) and 1-ml Ropivacaine 0.2 % (Naropin; Astra-Zeneca, Södertälje, Sweden), correct position of the needle was confirmed and recorded with contrast medium and imaging (Figs. 1 and 2). Over the data collection time period there were 31 different radiologists or radiology



**Fig. 1** Computed tomography-guided right S1 nerve root block in a 36-year-old male



**Fig. 2** Fluoroscopy-guided left L5 nerve root block in a 51-year-old male

fellows performing the TFESI procedure, including the two radiologists performing the read-outs for this current study.

### Analysis of MRI features

The MR images were evaluated blinded to the clinical outcomes independently by two musculoskeletal (MSK) fellowship trained radiologists with several years of experience. Sagittal T1 and T2-weighted and axial T2-weighted slices were analyzed for each patient. (Fat-suppressed slices were not consistently available for all patients so these were not included.) The following MRI features were evaluated: NR level affected, location (in the axial plane), morphology (type) of disc herniation (DH), severity of NR compression, NR compression due to changes other than DH, and severity of central spinal canal stenosis. A number of classification systems designed to enhance diagnostic consistency of specific MR features were used for the readouts [15–18]. Except for central spinal canal stenosis, MR features were classified as recommended by Ghahreman et al. [14], partly modified, as described in the subsequent and include the recommendations from the combined task forces of the North American Spine Society, American Society of Spine Radiology, and American Society of Neuroradiology [15, 23]:

The morphology of DH was categorized using the classification system of Fardon and Milette [15] (Table 1) with the addition of ‘disc bulges’ in order to distinguish them from ‘broad-based protrusion’. Location of DH in the axial plane was classified based on Fardon et al. [23] as ‘central’, ‘paracentral’ and ‘foraminal/extra foraminal’. Unlike the Fardon et al. classification, DHs with foraminal and/or extraforaminal localization were evaluated as one group.

As broad-based disc herniations may have more than one localization in the axial plane (e.g. ‘paracentral’ and ‘foraminal/extraforaminal’ simultaneously), they were analyzed both separately and as the combination in which they appeared.

**Table 1** DH - Morphology according to Fardon and Milette [15]

Bulge	Disc material overlaps the vertebral body endplates between 50 and 100 % of the circumference. (Not technically a disc herniation)
Protrusion	Broad-based herniation which involves 25–50 % of axial disc/vertebral body endplate circumference
Focal protrusion	Herniation involving <25 % of the axial disc/vertebral body endplate circumference
Extrusion	Herniation where, in at least one plane, any one distance between the edges of the disc material beyond the disc space is greater than the distance between the edges of the base in the same plane
Sequestration	The displaced disc material has lost all continuity with the parent disc

This same principal also applied to the morphology (type) of disc herniation.

The grading system of Pfirrmann et al. [16] was used to assess the severity of NR compression in paracentral DH in the axial planes (Table 2). In cases of foraminal DH, severity of NR compression was analyzed in the sagittal plane and classified according to Lee et al. [17] (Table 3).

Amongst degenerative changes that may affect the NR, the following were analyzed: Spondylolisthesis, osteophytes from either the facets or vertebral body, hypertrophy of facets and ligamentum flavum bulge [14]. Central spinal canal stenosis was graded in the axial plane using the criteria of Schizas et al. [18] (Table 4).

Prior to the MRI data collection, the two radiologists practiced together on 10 randomly selected cases 2 weeks prior to the official data collection to standardize the interpretation

**Table 2** Severity of NR compression in paracentral DH according to Pfirrmann et al. [16]

Grade 0 (normal)	No contact between disc material and nerve root, and the epidural fat layer between the nerve root and the disc material is preserved
Grade 1 (contact)	Visible contact between disc material and nerve root with no epidural fat layer between the two. Normal position without deviation of nerve root
Grade 2 (deviation)	Nerve root displaced dorsally by disc material
Grade 3 (compression)	Compression of the nerve root between disc material and wall of the spinal canal. Nerve root can be flattened or indistinguishable from disc material

**Table 3** Severity of NR compression in foraminal DH according to Lee et al. [17]

Grade 0	No foraminal stenosis
Grade 1 (mild)	Mild foraminal stenosis. Obliteration of perineural fat in two opposing directions (vertical or transverse). Contact of disc material with the superior and inferior or the anterior and posterior portions of the nerve root
Grade 2 (moderate)	Moderate foraminal stenosis. Obliteration of perineural fat in all four directions without morphologic changes in both vertical and transverse directions
Grade 3 (severe)	Severe foraminal stenosis. Collapse or morphologic change of nerve root

criteria prior to reading the included cases independently. These 10 cases were included in the analysis if they met the inclusion criteria.

### Statistical analysis

SPSS version 21.0 (Armonk, New York) was used for the statistical analysis. The PGIC 7-point scale was dichotomized into ‘improvement’ (yes/no) and ‘worsening’ (yes/no). In addition to descriptive statistics of patient age, sex, NR level injected, and frequencies of all evaluated MRI findings identified by each radiologist, the proportion of patients reporting clinically relevant ‘improvement’ (primary outcome) was calculated. Only the findings of the senior radiologist with 7 years of MSK radiological experience were used for comparison with the outcome measures. The Chi-square test compared each MRI finding to ‘improvement’ for each of the data collection time points. Similarly, the Chi-square test was used to compare the proportion of patients reporting ‘worsening’ for each of the MRI findings at each data collection time point.

NRS change scores were calculated (normally distributed data) for each data collection time point. The unpaired Student’s *t*-test was used to compare the mean NRS change scores with MRI findings categorized as ‘present/absent’ for

**Table 4** Grading of severity of central spinal canal stenosis according to Schizas et al. [18]

Grade A	CSF is clearly visible in the dural sac, but its distribution is inhomogeneous
Grade B	The rootlets occupy the whole of the dural sac, but they can still be individualized and some CSF is still present
Grade C	No rootlets can be recognized, the dural sac demonstrating a homogeneous gray signal with no CSF signal visible
Grade D	In addition to no rootlets being recognizable, there is no epidural fat posteriorly

each data collection time point and the ANOVA test was used to compare NRS change scores (normally distributed data) with the MRI findings with more than two options. The Wilcoxon test compared baseline NRS median scores (not normally distributed) to all follow-up scores;  $p < 0.05$  was considered statistically significant. Mean NRS values were also calculated for ease of presentation.

Inter-observer reliability of the two radiologists in identifying the various imaging findings was assessed with the Kappa-statistic, using the scheme of Landis and Koch as well as with percent agreement [24]. The outcomes questionnaires were opened and the data entered by a radiology researcher not involved in performing the injection procedures or the MRI readings (author 6). The statistical analysis was performed together by two authors also not involved in the injection procedures or MRI readings (authors 1 and 6).

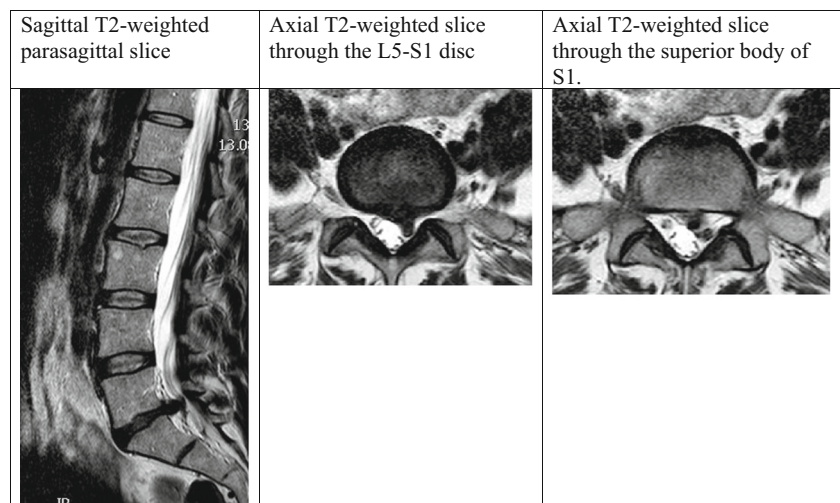
### Results

From the original data set of 199 patients evaluated by the radiologists, 156 patients were included with injections performed between June 2009 and February 2013. The 44 patients excluded by the radiologists during the readings were due to the fact that these patients had undergone lumbar spine surgery. There were 89 (57.1 %) males and 67 (42.9 %) females and the mean age was 55.36 years (SD = 14.91); there was no significant difference in the mean ages between the sexes ( $p = 0.17$ ). The mean age for the male patients was 53.92 (SD = 14.17) years and for the female patients it was 57.27 (SD = 15.74) years.

The only statistically significant relationship between MRI findings and the primary outcome of ‘improvement’ was for DH morphology ( $p = 0.03$ ) at the 1 month data collection time point. The combination of ‘protrusion + sequestration’ ( $n = 20$ ) had the most beneficial outcome with 70.0 % of patients reporting clinically relevant ‘improvement’ (Fig. 3). The worst outcome was for patients with disc bulge only ( $n = 44$ ) where only 31.8 % of patients were ‘improved’.

Table 5 shows the percentage of all patients ‘improved’ or ‘worse’, the NRS mean and NRS change scores for all data collection time points. For unknown reasons, eight patients failed to report their 1-month outcomes in spite of returning their questionnaires. Comparing individual MRI findings with NRS change scores, ‘protrusion plus sequestration’ herniation morphology ( $p = 0.0001$ ); Fig. 3), grade 3 foraminal nerve root compression ( $p = 0.01$ ) (Fig. 4), ‘degeneration by osteophytes’ being present ( $p = 0.034$ ) and ‘foraminal/extraforaminal location of the DH (Fig. 4;  $p = 0.014$ ) were significantly linked to higher NRS change scores at the 1-month time point (Table 6). Patients with paracentral grade 2 findings (deviation of the nerve root) reported significantly

**Fig. 3** Example of paracentral disc protrusion plus left-sided sequestration at L5-S1 in a 45-year-old female



lower levels of pain reduction ( $p = 0.02$ ) at 1 month compared to patients without this finding.

Inter-observer reliability of diagnosing the abnormalities varied depending on the specific MRI findings from ‘fair’ (0.21–0.40) to ‘almost perfect’ (0.81–1.00; Table 7). The lowest inter-observer reliability was found in ‘DH-Location: Central’ and in ‘DH-Classification: Protrusion’. The highest inter-rater reliability was found for ‘DH-Classification: Sequester’ and ‘Severity of Central Canal Stenosis’.

## Discussion

The main purpose of this study was to investigate whether or not specific MRI abnormalities are related to treatment outcomes after lumbar TFESI in patients with lumbar radiculopathy due to intervertebral DH and/or spinal stenosis. When looking at the individual MRI findings compared to ‘improvement’, the senior radiologist found that the disc herniation morphology of ‘protrusion plus sequestration’ was significantly related to an increased likelihood of ‘improvement’ at 1 month with 70 % of patients with this combination of findings reporting improvement. The worst outcome was for patients with only disc bulges where only 31.8 % were

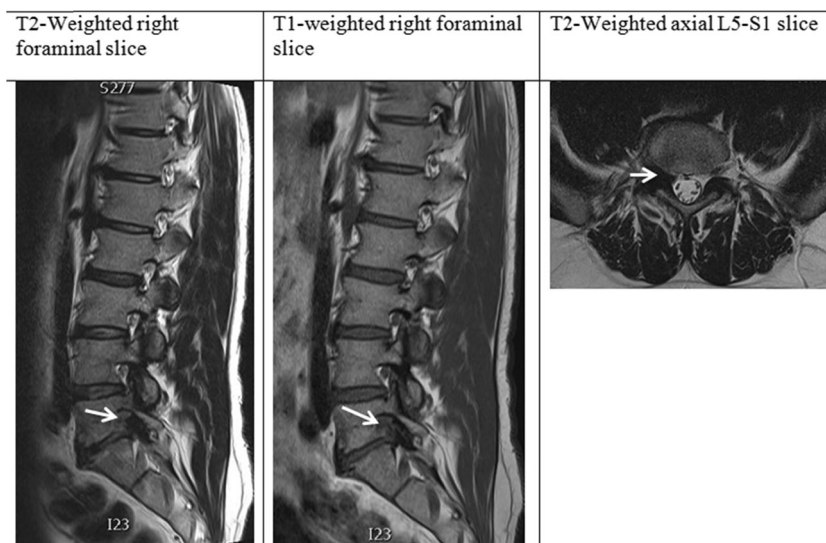
‘improved’. Furthermore, patients with disc ‘protrusion plus sequestration’ also had significantly higher pain change scores (i.e. more pain reduction) and significant relationships for the MRI abnormalities of degeneration by osteophytes being present, foraminal nerve root collapse or morphological change resulting from disc herniation as well as the location of the disc herniation being foraminal/extra foraminal and higher NRS change scores at 1 month were also found. It is not surprising that patients with only disc bulges had worse outcomes as bulges are less likely to cause nerve root compression as compared to actual herniations, and thus the injection of anesthetic and corticosteroid would not be effective [25]. Indeed, patients in this study with foraminal and extra foraminal herniations reported larger reductions in their pain levels at 1 month, consistent with the findings of Janardhana et al., who found that patients with gross foraminal compromise are more likely to have related clinical signs and symptoms [25]. The higher improvement rate for patients with herniations in this location may be due to the fact that these DHs are anatomically closer to the site of injection and therefore the corticosteroid medication is more likely to be able to reach the site of lesion as compared to DHs which are central or subarticular. Choi et al. [13] also found DH location to be significantly linked to a positive response to TFESI.

**Table 5** Outcomes for the various time points. *N* number of patients. *NRS* numerical rating scale; *SD* standard deviation. *NRS* mean-values and *SD* at baseline and at 15 min, 1 day, 1 week and 1 month after transforaminal NR-injection

	Improvement in % (primary outcome)	Worsening in %	Mean NRS score + SD	Mean NRS change score + SD (compared to baseline score)
Pre NRS (baseline) <i>N</i> = 156			6.66 (2.28)	
1 Day <i>N</i> = 156	30.8	8.9	4.23* (2.52)	2.43 (2.39)
1 Week <i>N</i> = 155	45.8	9.0	3.95* (2.55)	2.70 (2.34)
1 Month <i>N</i> = 148	51.4	10.9	3.45* (2.82)	3.14 (2.68)

\* $p < 0.05$  compared to baseline score

**Fig 4** Example of L5-S1 right foraminal plus extraforaminal disc herniation in a 41-year-old male. Also present is grade 3 foraminal nerve root compression on the right (*arrows*)



The severity of nerve root compression was related to the quantity of pain reduction in this current study only for foraminal nerve root compression but not with paracentral compression. Patients having more severe foraminal nerve root compression reported higher levels of pain reduction, whereas this was not the case for patients with paracentral nerve root compression. In fact, patients with paracentral deviation of the nerve root had significantly lower levels of pain reduction compared to patients without this finding. This is similar but not identical to the studies by Choi et al. [13] and Ghahreman and Bogduk [14] who found that the severity of nerve root compression in paracentral as well as foraminal/extraforaminal DH was a significant predictor of a positive response to TFESIs.

### Inter-rater reliability

The secondary purpose of this study was to assess the inter-rater reliability of diagnosing and classifying abnormalities on lumbar spine MRI scans. Although the two experienced raters practiced together using the classification systems prior to official data collection, the results show that the agreement in identifying and classifying lumbar MRI findings depends to a considerable extent on the specific pathologies that were identified. According to the Landis and Koch scoring system [24], the inter-observer reliability in this study ranged from 0.35 ('fair') to 0.81 ('almost perfect'), depending on the pathology or combination of pathologies identified. This variability may seem surprising, considering that the classification systems used were designed to make MRI interpretation more objective and have been studied and recommended by such prestigious groups as the North American Spine Society, the American Society of Spine Radiology, and the American Society of Neuroradiology [15–18]. At least some of this reliability can be explained by the fact that certain classification

systems use more easily detectable anatomic structures/landmarks as reference points/lines, whereas other systems use criteria that cannot be measured as easily or as precisely. Fardon and Milette state that even though the classification of location and morphology of DH as applied in this current study uses clearly defined anatomic structures, these structures are not always as precise as an illustration in an anatomic atlas might depict because as they can be curved or otherwise altered through degenerative processes or postural distortions [15, 23, 26].

In searching for reasons as to why the inter-rater reliability is better for some MRI findings compared to others, it is helpful to compare the number of options for the diagnosis of a particular finding with the level of inter-rater reliability. This can be seen in this study as those classification systems that allowed combinations of several options had a lower reliability than most of the other classification systems where the radiologists were asked to determine only one specific grade.

While the results obtained in this study for the level of inter-rater reliability may at first appear disappointing, they are consistent with what is found in the literature regarding spinal MRI studies. A recent study that examined the inter-rater reliability of assessing degenerative lumbar spine pathologies, found an average inter-rater Kappa coefficient of 0.43 with a range of 0.28 ('fair') to 0.62 ('substantial') [27]. Other studies cited in the paper by Fu et al. showed similar ranges of reliability [27]. It was found that studies that only used binary classification systems showed a higher reliability compared to studies that used classification systems with a grading scale. This supports the findings of this current study, where the binary classification systems achieved better inter-rater reliabilities than the classification systems using more than two options. It is important that referring clinicians understand this phenomenon and therefore do not expect the imaging reports to be 100 % accurate.

**Table 6** Frequency of MRI findings and 1 month NRS change scores per MRI finding. (*p*-value compared to patients without the specific MRI finding.) *DHL* disc herniation location. *NR* nerve root. *NRS* numerical rating scale (for pain). *SD* standard deviation

MRI finding	Quantity in % No. of patients ( <i>N</i> )	NRS change scores at 1 month mean + SD	<i>p</i> -value (baseline NRS – 1 month NRS)
Annulus rupture	38.5 % ( <i>N</i> = 60)	3.20 (2.52)	0.79
Disc herniation	100 % ( <i>N</i> = 156)		
Disc herniation - location:			
Central only	4.5 % ( <i>N</i> = 7)	3.27 (2.55)	0.34
Paracentral only	21.9 % ( <i>N</i> = 34)	3.07 (2.69)	0.48
Foraminal/extraforaminal only	5.2 % ( <i>N</i> = 8)	3.66 (2.56)	0.014 (better)
(DHL various combinations of the above)	42.6 % ( <i>N</i> = 107)		
Disc herniation - morphology:			
Bulge	29.0 % ( <i>N</i> = 45)	3.22 (3.22)	0.78
Protrusion	33.5 % ( <i>N</i> = 52)	3.23 (2.63)	0.72
Focal protrusion	16.1 % ( <i>N</i> = 25)	2.80 (2.92)	0.39
Extrusion alone	0 %		
Sequestration alone	2.6 % ( <i>N</i> = 4)	3.67 (2.65)	0.19
Protrusion + sequestration	12.8 % ( <i>N</i> = 20)	3.1 (2.78)	0.0001 (better)
(Various other combinations of the above)	18.7 % ( <i>N</i> = 29)		
Paracentral NR compression severity:			
Grade 0 (normal)	25.0 % ( <i>N</i> = 39)		
Grade 1 (contact)	11.5 % ( <i>N</i> = 18)	3.19 (2.58)	0.32
Grade 2 (deviation)	12.8 % ( <i>N</i> = 20)	2.13 (2.75)	0.02 (worse)
Grade 3 (compression)	50.6 % ( <i>N</i> = 79)	2.93 (2.51)	0.06 (worse)
Foraminal NR compression severity:			
Grade 0 (normal)	73.7 % ( <i>N</i> = 115)		
Grade 1 (contact from ant. and post. or sup. and inf.)	11.5 % ( <i>N</i> = 18)	3.94 (3.50)	0.14
Grade 2 (contact in all four directions)	7.1 % ( <i>N</i> = 11)	2.70 (2.76)	0.86
Grade 3 (collapse or morphologic change of NR)	7.7 % ( <i>N</i> = 12)	4.91 (1.93)	0.01 (better)
Degeneration by osteophytes	12.8 % ( <i>N</i> = 20)	4.37 (2.37)	0.034 (better)
Facet hypertrophy	76.3 % ( <i>N</i> = 119)	3.17 (2.62)	0.71
Ligamentum flavum hypertrophy	57.7 % ( <i>N</i> = 90)	2.75 (2.53)	0.06
Degenerative spondylolisthesis	3.2 % ( <i>N</i> = 5)	4.67 (3.21)	0.49

**Table 7** Inter-rater reliability in diagnosing the various MRI findings. *DHL* disc herniation location. *CI* confidence interval. *NR* nerve root

MRI finding	Kappa (95 % CI)	Kappa classification	% agreement
Annulus rupture	.65 (.52–.78)	Substantial	84 %
Disc herniation	.66 (.04–1.28)	Substantial	99 %
Disc herniation - location:			
Central	.38 (.21–.53)	Fair	71 %
Paracentral	.49 (.27–.71)	Moderate	89 %
Foraminal/extraforaminal	.66 (.55–.79)	Substantial	83 %
DHL combination	.42 (.32–.53)	Moderate	57 %
Disc herniation - morphology:			
Bulge	.63 (.50–.77)	Substantial	85 %
Protrusion	.35 (.20–.50)	Fair	70 %
Focal protrusion	.45 (.31–.59)	Moderate	75 %
Extrusion	.66 (.05–1.28)	Substantial	99 %
Sequestration	.81 (.69–.92)	Almost perfect	93 %
Paracentral NR compression severity:	.72 (.65–.83)	Substantial	84 %
Foraminal NR compression severity:	.71 (.60–.82)	Substantial	88 %
Severity of central spinal canal stenosis	.80 (.66–.94)	Substantial	95 %

## Limitations

Some of the MRI classification systems used were slightly modified as mentioned in the section ‘Methods’. This is relevant as our results showed a statistically significant NRS reduction after 1 month in patients with foraminal and/or extraforaminal DH location. This study combined those localizations of DH into one group, not allowing further analysis of the two localizations individually. Better localization would also be helpful when looking at ‘degeneration by osteophytes’, as this MRI-feature was also associated with statistically significant NRS reduction at 1 month after lumbar TFESI; however, this study did not distinguish between different localizations of osteophytes.

It should also be pointed out that this current study only included outcome data from patients who had returned their postal questionnaires, had relevant MRI scans within the required dates and did not have lumbar spine surgery. A recent study found that this questionnaire mode of data acquisition could distort the true treatment outcome, as only a portion of the whole group that underwent lumbar TFESI returned the questionnaire. Patients who did return their postal questionnaires tended to report a less beneficial outcome after 1 month [19]; however, the mode of data collection in this current study was the same for all patients. Additionally, patients only returned their postal questionnaires after completing the 1-month outcome data; thus, this same questionnaire also included the outcomes from 1 day and 1 week as well. As to why eight patients returned their questionnaire without completing the 1-month outcomes is however unknown.

This study also did not attempt to compare clinical data to the treatment outcomes either before or after treatment. Future studies could focus on this information. Although the PGIC asks patients to report on their ‘over-all improvement’ including disability and psychosocial aspects of their condition [19–22], this is not the same as having a physician assess their functional status post-injection. Additionally, this study only assessed outcomes up to the 1-month time point. Longer-term outcomes would be very helpful, including the proportion of patients requiring surgery or future injections. Indeed, a similar study was performed at this hospital recently on cervical indirect nerve-root-block patients and found that those with disc extrusions were much more likely to go to surgery by the 1-year post-injection time point [28].

Finally, this study did not look at whether or not the experience of the radiologist or radiology fellow was related to the treatment outcome. It may be logical to think that more experience may be linked to better outcomes, particularly in a teaching university hospital. However, radiology fellows are supervised when learning to perform all interventional procedures.

## Conclusion

Patients with disc protrusion plus sequestration were significantly more likely to report overall improvement and more pain reduction at 1 month. Higher pain reduction was noted in patients with degeneration by osteophytes, grade 3 foraminal nerve root compression, or foraminal/extraforaminal disc herniation location.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Gore M, Sadosky A, Stacey BR, Tai KS, Leslie D. The burden of chronic low back pain: clinical comorbidities, treatment patterns, and health care costs in usual care settings. *Spine*. 2012;37:E668–77.
- van Tulder M, Koes B, Bombardier C. Low back pain. *Best Pract Res Clin Rheumatol*. 2002;16:761–75.
- Burton AK, Balagué F, Cardon G, et al. Chapter 2. European guidelines for prevention in low back pain. *Eur Spine J*. 2006;15(Suppl)2:S136–68.
- Stewart WF, Ricci JA, Chee E, Morganstein D, Lipton R. Lost productive time and cost due to common pain conditions in the US workforce. *JAMA*. 2003;290:2443–54.
- Wieser S, Horisberger B, Schmidhauser S, et al. Cost of low back pain in Switzerland in 2005. *Eur J Health Econ*. 2011;12:455–67.
- Ricci JA, Stewart WF, Chee E, Leotta C, Foley K, Hochberg MC. Back pain exacerbations and lost productive time costs in United States workers. *Spine*. 2006;31:3052–60.
- 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.
- Balagué F, Dudler J. An overview of conservative treatment for lower back pain. *Int J Clin Rheumatol*. 2011;6:281–90.
- Manchikanti L, Buenaventura RM, Manchikanti KN, et al. Effectiveness of therapeutic lumbar transforaminal epidural steroid injections in managing lumbar spinal pain. *Pain Physician*. 2012;15:E199–245.
- Manchikanti L, Nampiaparampil D, Manchikanti K, et al. Comparison of the efficacy of saline, local anesthetics, and steroids in epidural and facet joint injections for the management of spinal pain: a systematic review of randomized controlled trials. *Surg Neurol Int*. 2015;6:S194–235.
- Peterson C, Hodler J. Evidence-based radiology (part 1): is there sufficient research to support the use of therapeutic injections for the spine and sacroiliac joints? *Skeletal Radiol*. 2010;39:5–9.
- Lee JW, Choi SW, Park SH, Lee GY, Kang HS. MR-based outcome predictors of lumbar transforaminal epidural steroid injection for lumbar radiculopathy caused by herniated intervertebral disc. *Eur Radiol*. 2013;23:205–11.
- Choi SJ, Song JS, Kim C, et al. The use of magnetic resonance imaging to predict the clinical outcome of non-surgical treatment for lumbar intervertebral disc herniation. *Korean J Radiol*. 2007;8:156–63.



14. Ghahreman A, Bogduk N. Predictors of a favorable response to transforaminal injection of steroids in patients with lumbar radicular pain due to disc herniation. *Pain Med.* 2011;12:871–9.
15. 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.* 2001;26:E93–113.
16. Pfirrmann CW, Dora C, Schmid MR, Zanetti M, Hodler J, Boos N. MR image-based grading of lumbar nerve root compromise due to disk herniation: reliability study with surgical correlation. *Radiology.* 2004;230:583–8.
17. Lee S, Lee JW, Yeom JS, et al. A practical MRI grading system for lumbar foraminal stenosis. *Am J Roentgenol.* 2010;194:1095–8.
18. Schizas C, Theumann N, Burn A, et al. Qualitative grading of severity of lumbar spinal stenosis based on the morphology of the dural sac on magnetic resonance images. *Spine.* 2010;35:1919–24.
19. Lechmann M, Peterson CK, Pfirrmann CW, Hodler J. Lumbar nerve root injections: a prospective cohort outcomes study comparing age- and gender-matched patients who returned an outcomes-based postal questionnaire with patients who did not return the postal questionnaire. *Skeletal Radiol.* 2013;42:1429–35.
20. Childs JD, Piva SR, Fritz JM. Responsiveness of the numeric pain rating scale in patients with low back pain. *Spine.* 2005;30:1331–4.
21. Newell D, Bolton JE. Responsiveness of the Bournemouth questionnaire in determining minimal clinically important change in subgroups of low back pain patients. *Spine.* 2010;35:1801–6.
22. Farrar JT, Young JP, 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.
23. Fardon DF, Williams AL, Dohring EJ, Murtagh FR, Gabriel Rothman SL, Sze GK. Lumbar disc nomenclature: version 2.0: recommendations of the combined task forces of the North American Spine Society, the American Society of Spine Radiology and the American Society of Neuroradiology. *Spine J.* 2014;14:2525–45.
24. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159–74.
25. Janardhana AP, Rajagopal, Rao S, Kamath A. Correlation between clinical features and magnetic resonance imaging findings in lumbar disc prolapse. *Indian J Orthop.* 2010;44:263–9.
26. Rehman IU, Hamid RS, Akhtar W, Naqi R, Shamim MS. Observer variation in MRI evaluation of patients with suspected lumbar disc herniation and nerve root compression: comparison of Neuroradiologist and Neurosurgeon's interpretations. *J Pak Med Assoc.* 2012;62:826–9.
27. Fu MC, Buerba RA, Long WD, et al. Interrater and intrarater agreements of magnetic resonance imaging findings in the lumbar spine: significant variability across degenerative conditions. *Spine J.* 2015;14:2442–8.
28. Bensler S, Sutter R, Pfirrmann CW, Peterson CK. Long term outcome from CT-guided indirect cervical nerve root blocks and their relationship to the MRI findings: a prospective study. *Eur Radiol.* 2015;25:3405–13.