Optimizing care in lumbar radiculopathy and neurogenic claudication: from injection to inference, and from clinician to algorithm

Verheijen, E.J.A.

9

External validation of a novel comprehensive grading system for lumbar spinal stenosis

E.J.A. Verheijen, R. Lindberg, W.A. Moojen, G.J. Lycklama à Nijeholt,
B.F.W. van der Kallen, C.L.A. Vleggeert-Lankamp

Submitted

ABSTRACT

Purpose

A standardized grading system for lumbar spinal stenosis (LSS) is lacking. This study evaluated inter-reader agreement and clinical utility of the comprehensive grading system proposed by Miskin et al., assessing central canal stenosis (CCS), foraminal stenosis (FS), lateral recess stenosis (LRS), and facet arthropathy (FA).

Methods

Preoperative MRI data from 155 patients with neurogenic claudication due to LSS were retrospectively analyzed. Two neuroradiologists and one spine surgeon independently graded all lumbar levels using the four-item system. Inter-reader agreement was assessed using weighted Cohen's kappa, exact agreement, and agreement within one grade. Associations between radiological grades and baseline symptoms and postoperative outcomes at 8, 26 and 52 weeks were analyzed using bivariate analyses and multivariable linear regression.

Results

Substantial inter-reader agreement was observed for CCS (kappa = 0.751) and LRS (kappa = 0.619), with exact agreement of 61.2% and 68.7%, respectively. In contrast, agreement was slight for FS ($\kappa = 0.214$) and FA ($\kappa = 0.201$). Agreement was highest for extreme grades across all components. At L4-5 and L5-S1, agreement for CCS was comparable to Miskin et al., whereas agreement for LRS was higher. Higher baseline CCS was independently associated with greater postoperative improvement in pain and physical disability, while correlations for the combined CCS+LRS score were weaker.

Conclusion

The Miskin et al. grading system demonstrates substantial reliability for CCS and LRS, but clinical relevance only for CCS. Limited reliability for FS and FA suggests these components require refinement or substitution. Comparative evaluation with alternative grading systems may further enhance clinical applicability.

INTRODUCTION

Over the past few decades, magnetic resonance imaging (MRI) has become the preferred imaging modality for diagnosing patients suspected of having lumbar spinal stenosis (LSS)[1]. Several grading systems have been proposed to assess the degree of LSS on MRI, typically evaluating the central canal zone, lateral recess, or neuroforamen – regions considered clinically relevant [2-12]. These systems aim to standardize LSS assessment and improve intra- and interdisciplinary communication. In addition, standardized grading is more likely to correlate with clinical symptoms and therapeutic outcomes, thereby aiding clinical decision-making.

Lurie et al. were the first to introduce a grading scale for LSS, comprising four grades (none, mild, moderate or severe) based on compromise of the central zone, lateral recess or neuroforamen, expressed in thirds [2]. Since then, more complex grading tools have been developed. The Schizas classification, which remains one of the most widely used grading systems for LSS, consists of four grades ranging from A (no or minor stenosis) to D (extreme stenosis), with grade A further subdivided into four subgrades to assess central canal stenosis (CCS)[3]. Guen et al. subsequently proposed a four-grade scale ranging from no to severe stenosis [4]. Additionally, Weber et al. and Yuan et al. introduced a modified Schizas scale and a novel system, respectively [5, 6]. Further studies proposed grading systems for lateral recess stenosis (LRS), foraminal stenosis (FS), and facet arthropathy (FA) [7-12]. However, most systems are subject to considerable inter-reader variability and lack correlation with clinical parameters, leading to discrepancies in interdisciplinary communication regarding stenosis severity and limiting widespread clinical adoption.

In 2021, Miskin et al. introduced a novel grading system evaluating four key components of lumbar degeneration: central canal stenosis, foraminal stenosis, lateral recess stenosis and facet arthropathy [13]. This system extends the Schizas classification by incorporating additional regions frequently involved in LSS and, alongside the system proposed by Lurie et al., provides a multifaceted assessment. In the original study, three observers independently assessed MRI scans from 50 patients at the L4-5 and L5-S1 levels, demonstrating fair to substantial inter-reader agreement (Cohen's kappa 0.323 – 0.702). This suggests potential value as a comprehensive grading tool applicable across interdisciplinary clinical settings.

However, external validation is necessary for broader acceptance. The aim of this study is to evaluate inter-reader variability, for L4-5 and L5-S1 combined as

well as for all lumbar levels, and to assess correlations with clinical parameters in order to determine the validity and applicability of this grading system.

MATERIAL AND METHODS

Patient population

Patients were included from the FELIX (Foraminal Enlargement Lumbar Interspinous distraXion) trial, a multicenter randomized study comparing interspinous process implants with conventional decompression surgery for intermittent neurogenic claudication due to LSS [14]. The trial demonstrated similar postoperative outcomes between treatment groups. Patients were aged 40-85 years, had symptoms >3 months unresponsive to conservative therapy, underwent surgery at one or two levels and had a preoperative MRI examination. MRI protocols included T1- and T2-weighted sagittal and axial sequences, obtained with 1.5T or 3.0T scanners, slice thickness 3.0-6.0 mm and interslice gaps of 0.3-5.2 mm.

MRI evaluation and grading

Two experienced neuroradiologists (GLN, BvdK) and one experienced spine-dedicated neurosurgeon (CVL) graded all lumbar levels from L1-2 to L5-S1 using Miskin et al.'s system for CCS, FS, LRS and FA [13]. CCS, FS and FA comprised six possible grades, whereas LRS comprised three possible grades. Both FS and LRS were graded separately for the left and right sides. Reviewers were blinded to operated levels and clinical data, and received a presentation with representative examples beforehand to ensure uniform understanding of the grading scale.

Clinical outcome parameters

The Visual Analogue Scale (VAS) for leg and back pain (0-100 scale)[15], and the modified Roland-Morris Disability Questionnaire (mRMDQ) (0-23 scale) [16] were used as outcome parameters at baseline and at 8, 26 and 52 weeks. Change scores were calculated for each follow-up time point.

Inter-reader agreement analysis

Inter-reader agreement was initially assessed for L4-5 and L5-S1 to match the analysis by Miskin et al. Proportions were calculated for all possible grades. Inter-reader agreement was determined using weighted Cohen's kappa with 95% confidence intervals (CIs) and categorized according to Landis and Koch [17]. Exact and within-one-grade agreement between reviewer pairs were calculated. Agreement across stenosis severity categories was analysed using median

reviewer grades to define ordered severity groups. Analyses were repeated for all lumbar levels combined and for individual levels separately. Missing gradings were omitted pairwise.

Correlation with clinical outcomes

Due to low inter-reader agreement, FA and FS were excluded from correlation analyses with clinical outcomes. Analyses therefore focused on CCS and LRS. Median grades were used for operated levels; in case of two surgical levels, the higher median was selected. In addition to the individual grading items, a combined CCS+LRS score was established by recategorizing CCS into 3 grades and summing with LRS, to assess whether a composite measure of stenosis severity demonstrated stronger correlations with clinical outcomes. Correlations were assessed through bivariate analyses with Spearman's rank correlation coefficient and multivariable linear regression models adjusted for age, gender, body mass index (BMI), duration of symptoms, Hospital Anxiety and Depression Scale (HADS) score and treatment group. A p value ≤ 0.05 was considered statistically significant.

RESULTS

A total of 159 patients were enrolled in the FELIX trial and underwent surgery. Preoperative MRI data were available for 155 patients. Baseline characteristics are presented in Table 1.

Table 1 Baseline characteristics

	N	Mean ± std
Age (yrs)	155	66.0 ± 9.1
Gender (male)	155	52.9% (82)
BMI (kg/m²)	155	27.7 ± 4.6
Smoking	153	
None		74.5% (114)
Sporadically		3.9% (6)
Frequently		21.6% (33)
Duration of symptoms (mos)	154	12.0 (6.0 – 24.0)*
VAS leg pain	152	56.1 ± 23.6
VAS back pain	152	49.4 ± 26.3
mRMDQ	142	14.0 ± 4.8
Randomization group	155	
Standard decompression		51.0% (79)
Interspinous device		49.0% (76)

Table 1 Baseline characteristics (*continued*)

	N	Mean ± std
Surgery at second level (yes)	155	21.3% (33)
Primary surgical level	155	
L2-3		4.5% (7)
L3-4		32.9% (51)
L4-5		60.6% (94)
L5-6		0.6% (1)
L4-S1		1.3% (2)
Secondary surgical level	33	
L3-4		12.1% (4)
L4-5		87.9% (29)

BMI: body mass index; mRMDQ: modified Roland-Morris Disability Questionnaire; VAS: visual analogue scale. *Median (interquartile range)

Inter-reader variability analyses for L4-5 and L5-S1

For CCS, 739 readings were available (L4-5: 432; L5-S1: 307). The majority was graded as normal, followed by severe and moderate-severe stenosis (Figure 1). Inter-reader greement was substantial (kappa 0.764; 95% CI 0.733 – 0.795) (Figure 2), with exact agreement in 61.3% (340/555) and within-one-grade agreement in 88.6% (492/555) (Table 2).

Of the 1475 readings for FS (L4-5: 864; L5-S1: 611), most were graded as normal or mild, without major differences between the left and right sides. Inter-reader agreement was slight (kappa 0.204; 95% CI 0.169 – 0.238), with exact agreement in 39.5% (438/1108) and within-one-grade agreement in 64.4% (714/1108). Agreement did not differ significantly between the left and right sides.

For LRS, 1477 readings were available (L4-5: 865; L5-S1: 612), of which most were graded as normal or nerve root compression, again without differences between the left and right sides. Agreement was substantial (kappa 0.610; 95% CI, 0.572 – 0.648), with exact agreement in 69.2% (766/1107) and within-one-grade agreement in 95.8% (1061/1107). Differences between the left and right sides were minimal.

For FA, 733 readings were available (L4-5: 431; L5-S1: 302), which were relatively evenly distributed across grading categories. Agreement was fair (kappa 0.251; 95% CI 0.207 – 0.295), with exact agreement in 25.2% (138/548) and within-one-grade agreement in 60.8% (333/548). Sensitivity analysis evaluating whether collapsing FA grades into three categories improved inter-reader agreement demonstrated a comparable kappa value (0.255; 95% CI 0.203-0.308).

Across CCS, FS, LRS and FA, agreement was highest in the extreme severity categories (Figures 3 & 4).

Inter-reader variability analyses for all lumbar levels

Agreement across all lumbar levels was comparable to that observed at the L4-5 and L5-S1 levels (Table 3). For FS exact agreement was slightly higher when all levels were included, without meaningful effect on kappa values. Collapsing FA grades modestly increased agreement (kappa 0.337; 95% CI 0.297-0.377).

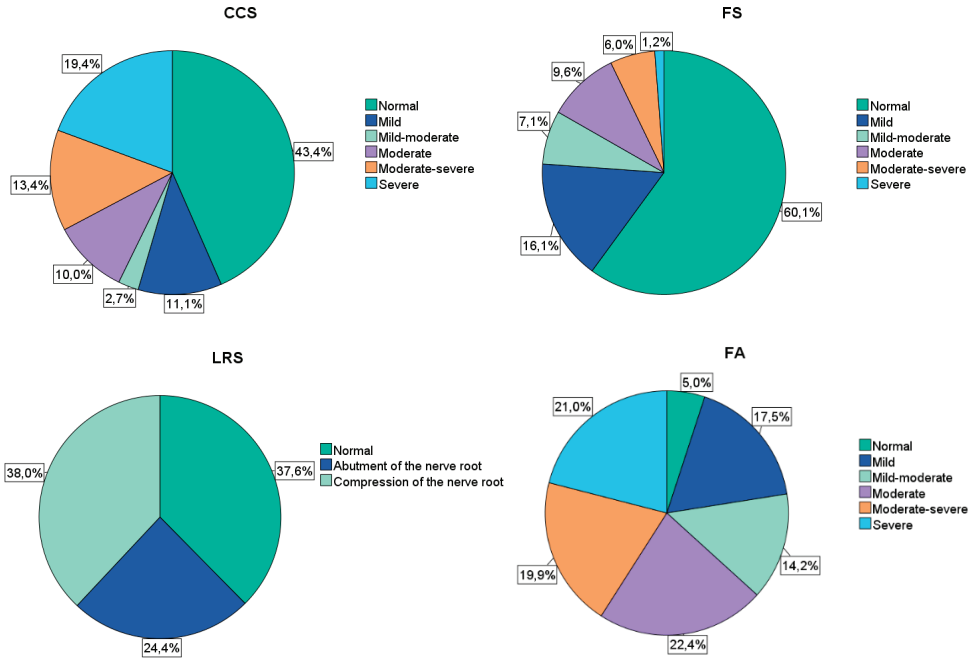

Fig. 1 Distribution of the gradings for CCS, FS, LRS and FA for L4-5 and L5-S1 levels

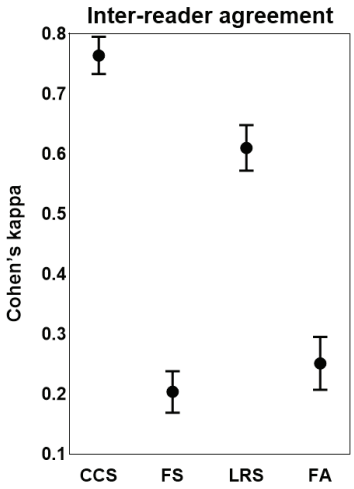

Fig. 2 Box plot depicting the average Cohen's kappa and 95% confidence interval for each item of the scale at L4-5 and L5-S1

Table 2 Inter-reader agreement for CCS, FS, LRS and FA at L4-5 and L5-S1

	Our results			Miskin et al.		
	Kappa (range)	**Exact agreement**	**Agreement within one grade**	**Kappa (95% CI)**	**Exact agreement**	**Agreement within one grade**
CCS	0.764 (0.733-0.795)	61.3%	88.6%	0.702 (0.641 -0.764)	56%	91%
FS	0.204 (0.169-0.238)	39.5%	64.4%	0.544 (0.486 -0.602)	41.5%	78.5%
LRS	0.610 (0.572-0.648)	69.2%	95.8%	0.323 (0.255 -0.392)	34%	90.5%
FA	0.251 (0.207-0.295)	25.2%	60.8%	0.557 (0.495 -0.620)	33%	76%

CCS: central canal stenosis; FA: facet arthropathy; FS: foraminal stenosis; LRS: lateral recess stenosis

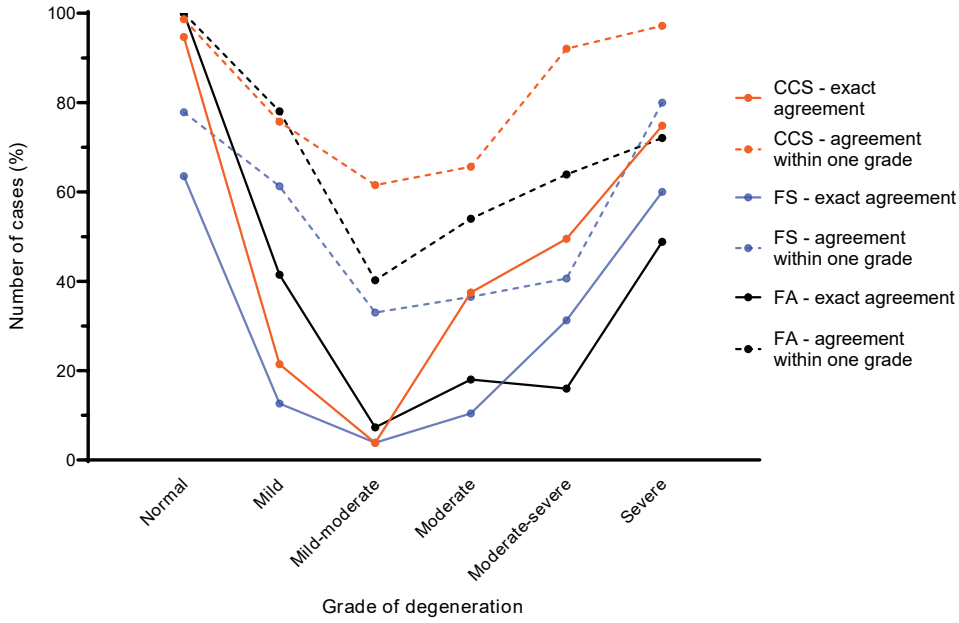

Fig. 3 Distribution of inter-reader agreement by severity of degeneration for CCS, FS and FA.

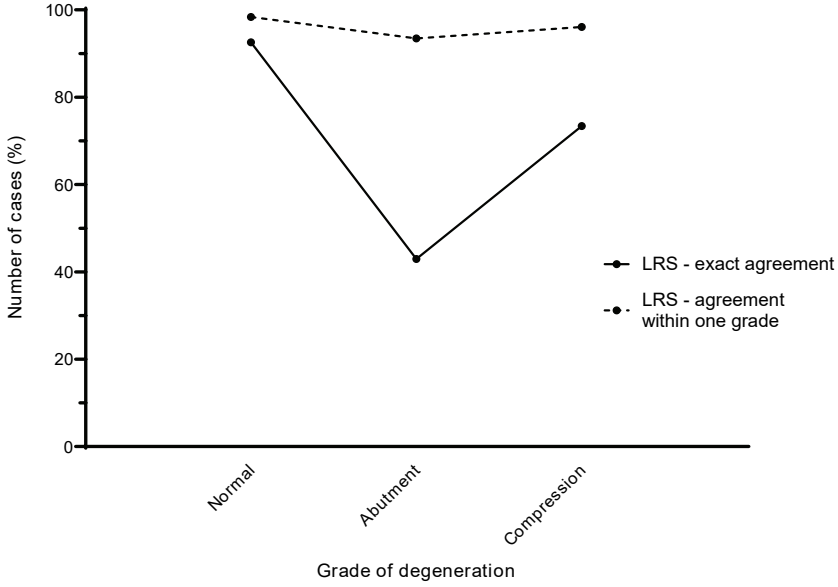

Fig. 4 Distribution of inter-reader agreement by severity of degeneration for LRS.

Table 3 Inter-reader agreement for CCS, FS, LRS and FA for the whole lumbar spine and individual levels

	Whole lumbar spine			L1-2	L2-3	L3-4	L4-5	L5-S1
	Kappa (range)	Exact agreement	Agreement within one grade	Kappa (range)				
CCS	0.751 (0.727-0.776)	61.2% (636/1039)	88.3% (917/1039)	0.524 (0.293-0.754)	0.608 (0.500-0.716)	0.685 (0.629-0.741)	0.639 (0.593-0.686)	0.293 (0.095-0.490)
FS	0.214 (0.187-0.242)	49.0% (1016/2073)	71.4% (1481/2073)	0.237 (-0.033-0.506)	0.071 (-0.068-0.210)	0.131 (0.084-0.177)	0.150 (0.116-0.185)	0.380 (0.279-0.481)
LRS	0.619 (0.592-0.646)	68.7% (1421/2067)	95.4% (1972/2067)	0.585 (0.393-0.777)	0.499 (0.411-0.588)	0.498 (0.438-0.558)	0.371 (0.314-0.429)	0.408 (0.294-0.521)
FA	0.201 (0.102-0.300)	23.6% (241/1020)	62.3% (635/1020)	0.158 (-0.012-0.327)	0.201 (0.102-0.300)	0.243 (0.187-0.299)	0.205 (0.158-0.251)	0.103 (0.014-0.191)

CCS: central canal stenosis; FA: facet arthropathy; FS: foraminal stenosis; LRS: lateral recess stenosis

Correlations between stenosis severity and clinical outcomes

Bivariate analyses for CCS and LRS demonstrated no significant correlations with clinical outcomes at baseline (Table 4). At 8 weeks, weak correlations were observed only between stenosis severity and physical disability (mRMDQ). At 26 and 52 weeks, higher CCS scores were correlated with greater reductions in leg pain, back pain and physical disability, whereas correlations for LRS were weaker and less consistent. The combined CCS+LRS score demonstrated similar, but generally weaker, correlations compared with CCS alone.

Multivariable linear regression analyses demonstrated no significant baseline associations between stenosis severity and clinical parameters (Table 5). However, female sex was independently associated with higher pain scores (VAS leg pain: B = 8.5-8.9, p = 0.024-0.031; VAS back pain: B = 10.2-10.4, p = 0.019-0.021) and greater physical disability (B = 2.8, p<0.001) at baseline. At 8 weeks, higher stenosis severity was associated with greater improvement in physical disability across all grading scores ($p \leq 0.006$). Trend-level correlations were observed between higher CCS and CCS+LRS scores and lower leg pain. At 26 weeks, improvement in physical disability correlated with all three scores ($p \leq 0.015$). Reduction in leg pain correlated only with CCS severity, whereas back pain was associated with CCS and CCS+LRS scores. At 52 weeks, CCS demonstrated the strongest and most consistent associations with all clinical outcomes ($p \leq 0.017$). LRS severity correlated with greater improvement in back pain, showed

no association with leg pain, and demonstrated only a near-significant association with physical disability (p = 0.067). The combined CCS+LRS score was significantly correlated with leg and back pain, whereas the association with physical disability was near-significant (p = 0.077). Treatment group affected outcomes in only a limited number of analyses.

Clinical effect size

Leg pain scores decreased by 5.6 and 7.9 points per one-category increase in CCS severity at 26 and 52 weeks, respectively. For the CCS+LRS score, leg pain decreased by 9.7 points per category at 52 weeks. Back pain scores decreased by 5.8 and 6.5 points per CCS category at 26 and 52 weeks, respectively, compared with decreases of 6.3 and 8.1 points per CCS+LRS category. mRMDQ scores demonstrated smaller changes across all follow-up time points, with decreases ranging from 1.0 to 3.5 points per one-category increase in stenosis severity at 8, 26 and 52 weeks.

Table 4 Bivariate correlations between stenosis scores and clinical parameters

		CCS			LRS			CCS+LRS		
	Follow-up	N	Spearman's ρ	P value	N	Spearman's ρ	P value	N	Spearman's ρ	P value
VAS leg pain	Baseline	152	-0.038	0.646	152	-0.083	0.307	152	-0.021	0.800
	8w	150	-0.144	0.080	150	-0.088	0.285	150	-0.138	0.091
	26w	148	-0.250	0.002*	148	-0.066	0.429	148	-0.184	0.025*
	52w	151	-0.284	< 0.001*	151	-0.129	0.115	151	-0.238	0.003*
VAS back pain	Baseline	152	0.020	0.809	152	-0.031	0.702	152	-0.020	0.805
	8w	149	-0.161	0.051	149	-0.029	0.724	149	-0.065	0.433
	26w	148	-0.280	< 0.001*	148	-0.123	0.135	148	-0.179	0.030*
	52w	151	-0.266	< 0.001*	151	-0.180	0.027*	151	-0.187	0.021*
mRMDQ	Baseline	142	0.036	0.670	142	0.088	0.297	142	0.076	0.369
	8w	133	-0.227	0.009*	133	-0.229	0.008*	133	-0.236	0.006*
	26w	133	-0.194	0.025*	133	-0.190	0.028*	133	-0.144	0.099*
	52w	136	-0.244	0.004*	136	-0.193	0.024*	136	-0.148	0.085

CCS: central canal stenosis; FA: facet arthropathy; FS: foraminal stenosis; LRS: lateral recess stenosis. * P value < 0.05

Table 5 Linear regression analysis with covariates age, gender, BMI, duration of symptoms, HADS at baseline and treatment group.

		CCS			LRS			CCS+LRS		
	Follow-up	B	95% CI	P value	B	95% CI	P value	B	95% CI	P value
VAS leg pain	Baseline	-0.356	-3.111 – 2.399	0.799	--3.766	-12.182 – 4.649	0.378	-0.151	-3.994 – 3.693	0.938
	8w	-3.044	-6.272 – 0.184	0.064	-4.627	-14.564 – 5.309	0.359	-3.919	-1.709 - -8.454	0.090
	26w	-5.587	-11.108 – -0.066	0.047*	-1.237	-18.135 – 15.660	0.885	-4.564	-12.318 – 3.190	0.246
	52w	-7.879	- 11.507 – -4.251	< 0.001*	- 10.773	- 22.469 – 0.924	0.071	- 9.639	- 14.783 – -4.495	< 0.001*
VAS back pain	Baseline	0.809	-2.254 – 3.873	0.602	0.359	-9.060 – 9.779	0.940	0.854	-3.430 – 5.137	0.694
	8w	-2.177	-5.558 – 1.204	0.205	-2.017	-12.537 – -8.504	0.705	-1.807	-6.601 – 2.987	0.457
	26w	-5.818	-9.346 – -2.291	0.001*	-7.603	-18.613 – 3.408	0.174	-6.271	-11.264 – -1.278	0.014*
	52w	-6.488	- 9.924 – -3.051	< 0.001*	-14.332	-25.152 – -3.513	0.010*	-8.066	-12.924 – -3.207	0.001*
mRMDQ	Baseline	0.282	-0.315 – 0.879	0.352	0.854	-0.935 – 2.643	0.347	0.406	-0.425 – 1.237	0.335
	8w	-1.294	-2.106 – -0.483	0.002*	-3.484	-5.926 – 1.043	0.006*	- 1.878	-3.019 – -0.737	0.001*
	26w	-1.081	-1.925 – -0.237	0.013*	-3.244	-5.735 – -0.752	0.011*	-1.468	-2.645 – -0.290	0.015*
	52w	- 0.973	-1.782 – -0.165	0.019*	-2.332	-4.791 – 0.127	0.063	-1.015	-2.161 – 0.131	0.082

CCS: central canal stenosis; FA: facet arthropathy; FS: foraminal stenosis; LRS: lateral recess stenosis. * P value < 0.05

DISCUSSION

This study provides external validation of the LSS grading system proposed by Miskin et al. [13]. Our findings demonstrate substantial inter-reader agreement for CCS and LRS, whereas agreement was slight for FS and fair for FA. These findings are supported by the proportions of exact and within-one-grade agreement. The relatively high agreement for LRS is partly expected due to its three-grade scale, which inherently limits the potential for disagreement. The weighted kappa values for CCS were comparable to those reported by Miskin et al., while agreement for LRS was higher but lower for FS and FA. Overall, the grading system appears reliable for CCS and LRS, but less so for FS and FA.

The lower agreement for FS and FA observed in this study may be attributed to differences in patient populations. Miskin et al. included patients with isolated low back pain, a condition often associated with more pronounced facet degeneration, whereas our cohort consisted of patients with intermittent neurogenic claudication, in whom FA may be less severe or more heterogeneous. Similarly, most patients in our study demonstrated no or mild FS, and fewer than 10% exhibited severe FS, likely due to exclusion of patients requiring fusion surgery in the FELIX trial. Our larger sample size may also have contributed to greater variability, thereby reducing inter-reader agreement. Sensitivity analysis demonstrated improved agreement for FA when grades were collapsed, suggesting that the current FA scale may be overly granular.

Over the years, multiple grading systems have been proposed to assess one or more components of LSS. The most commonly used CCS grading systems are those proposed by Guen et al. and Schizas et al. [3, 4]. Guen et al. reported kappa values of 0.730-0.953, although external validation yielded substantially lower agreement (kappa 0.33-0.40) [18]. Schizas et al. reported a kappa of 0.44, with subsequent studies demonstrating variable inter-reader agreement [19-22]. Two additional CCS grading systems reported kappa values of 0.58-0.76 and 0.681, respectively [5, 6]. For FS, three grading systems have been introduced. Wildermuth et al. reported substantial agreement (kappa 0.62), Lee et al. demonstrated near-perfect agreement (kappa 0.909-0.942), which was externally validated [22], and Sartoretti et al. reported similarly high agreement (kappa 0.886-1.000) [7-9]. For LRS, Pfirrmann et al. proposed a grading scale with substantial agreement (kappa 0.62-0.67), although the cohort consisted of patients with lumbar disc herniation [10]. For FA, two grading scales have been presented. Weishaupt et al. reported a kappa of 0.41 [11], although this was lower in subsequent studies [19, 22], whereas Fujiwara et al. demonstrated a kappa of 0.636 [12], which was substantiated by another study [20]. While some

grading systems demonstrate high inter-reader agreement, most assess only a single anatomical component of LSS, thereby limiting overall clinical utility. Only the system proposed by Lurie et al., similar in concept to that of Miskin et al., incorporates multiple anatomical zones by estimating the proportion of compromised area (normal, ≤1/3, 1/3-2/3, ≥2/3) in the central canal, lateral recess, and neuroforamen. Lurie et al. reported kappa values of 0.73 for CCS, 0.49 for FS, and 0.58 for LRS [2]. In this context, our findings indicate that the Miskin et al. grading system performs comparably to, or better than, most existing grading systems for CCS and LRS, supporting its clinical utility for these components. However, the low agreement observed for FS and FA suggests that these components may require refinement or substitution with alternative, more reliable grading systems.

In the present study, higher baseline stenosis severity, particularly CCS, was independently associated with greater postoperative improvement in pain and physical disability. Although effect sizes per category were modest and below commonly accepted minimally clinical important difference (MCID) thresholds for VAS and RMDQ scores[23], cumulative differences across severity categories approached or exceeded MCID values, indicating potential clinical relevance. These findings were not attributable to a greater capacity for symptomatic improvement in patients with more severe stenosis, as baseline pain and disability did not differ significantly across stenosis severity categories. Consequently, patients with higher baseline CCS severity, as graded using the Miskin et al. scale, appear more likely to benefit from surgery. Importantly, the combined CCS+LRS score did not demonstrate stronger correlations with clinical outcomes than CCS alone.

In contrast, many previously proposed grading systems have not demonstrated meaningful correlations with clinical parameters. Prior studies assessing the Schizas or Weishaupt grading scales reported no significant associations with baseline symptom severity or postoperative outcomes [3, 19, 20], with the exception of one study demonstrating a correlation between Schizas grades and both baseline and postoperative leg pain scores [21]. Yuan et al. reported that their CCS grading system correlated with baseline physical disability but not with postoperative clinical outcomes [6]. Hence, these findings underscore the limited value of many existing grading systems in clinical decision-making.

Strengths and limitations

Strengths of this study include a representative patient population with intermittent neurogenic claudication, a large sample size, variability in MRI scanning parameters, and independent grading by three experienced and blinded review-

ers. Limitations include the exclusion of patients with missing data, potentially introducing selection bias, and the absence of a direct comparison with other grading systems.

Recommendations for future research

While the grading system proposed by Miskin et al. offers a valuable step toward comprehensive LSS assessment, alternative systems, specifically for FS and FA, should not be discounted. A modular approach in which the most reliable grading method is selected for each anatomical region, may be preferable. Accordingly, the grading systems proposed by Miskin et al. for CCS, Lee et al. and Sartoretti et al. for FS, and Fujiwara et al. for FA warrant further external validation. Ultimately, grading systems should be evaluated not only for reproducibility but also for their association with clinical outcomes across all lumbar levels.

Conclusion

The Miskin et al. grading system appears promising for LSS evaluation and clinical decision-making. The substantial interobserver agreement for CCS and LRS, which is comparable to or exceeds that of other grading systems, supports their use in clinical practice and interdisciplinary communication. Moreover, CCS severity correlated with postoperative clinical outcomes and may therefore aid in preoperative counseling. However, limited inter-reader agreement for FS and FA reduces their reliability, and substitution of these components with alternative, more reliable grading systems may improve the overall clinical utility of this approach.

REFERENCES

[1] Andreisek G, Hodler J, Steurer J. Uncertainties in the diagnosis of lumbar spinal stenosis. Radiology. 2011;261(3):681–4.

[2] Lurie JD, Tosteson AN, Tosteson TD, Carragee E, Carrino JA, Kaiser J, et al. Reliability of readings of magnetic resonance imaging features of lumbar spinal stenosis. Spine (Phila Pa 1976). 2008;33(14):1605–10.

[3] Schizas C, Theumann N, Burn A, Tansey R, Wardlaw D, Smith FW, et al. Qualitative grading of severity of lumbar spinal stenosis based on the morphology of the dural sac on magnetic resonance images. Spine (Phila Pa 1976). 2010;35(21):1919–24.

[4] Guen YL, Joon WL, Hee SC, Kyoung-Jin O, Heung SK. A new grading system of lumbar central canal stenosis on MRI: an easy and reliable method. Skeletal Radiology. 2011;40(8):1033–9.

[5] Weber C, Rao V, Gulati S, Kvistad KA, Nygaard Ø P, Lønne G. Inter- and Intraobserver Agreement of Morphological Grading for Central Lumbar Spinal Stenosis on Magnetic Resonance Imaging. Global Spine J. 2015;5(5):406–10.

[6] Yuan S, Zou Y, Li Y, Chen M, Yue Y. A clinically relevant MRI grading system for lumbar central canal stenosis. Clin Imaging. 2016;40(6):1140–5.

[7] Wildermuth S, Zanetti M, Duewell S, Schmid MR, Romanowski B, Benini A, et al. Lumbar spine: quantitative and qualitative assessment of positional (upright flexion and extension) MR imaging and myelography. Radiology. 1998;207(2):391–8.

[8] Lee S, Lee JW, Yeom JS, Kim KJ, Kim HJ, Chung SK, et al. A practical MRI grading system for lumbar foraminal stenosis. AJR Am J Roentgenol. 2010;194(4):1095–8.

[9] Sartoretti E, Wyss M, Alfieri A, Binkert CA, Erne C, Sartoretti-Schefer S, et al. Introduction and reproducibility of an updated practical grading system for lumbar foraminal stenosis based on high-resolution MR imaging. Sci Rep. 2021;11(1):12000.

[10] 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(2):583–8.

[11] Weishaupt D, Zanetti M, Boos N, Hodler J. MR imaging and CT in osteoarthritis of the lumbar facet joints. Skeletal Radiol. 1999;28(4):215–9.

[12] Fujiwara A, Tamai K, Yamato M, An HS, Yoshida H, Saotome K, et al. The relationship between facet joint osteoarthritis and disc degeneration of the lumbar spine: an MRI study. Eur Spine J. 1999;8(5):396–401.

[13] Miskin N, Isaac Z, Lu Y, Makhni MC, Sarno DL, Smith TR, et al. Simplified Universal Grading of Lumbar Spine MRI Degenerative Findings: Inter-Reader Agreement of Non-Radiologist Spine Experts. Pain Med. 2021;22(7):1485–95.

[14] Moojen WA, Arts MP, Brand R, Koes BW, Peul WC. The Felix-trial. Double-blind randomization of interspinous implant or bony decompression for treatment of spinal stenosis related intermittent neurogenic claudication. BMC Musculoskelet Disord. 2010;11:100.

[15] Zanoli G, Strömqvist B, Jönsson B. Visual analog scales for interpretation of back and leg pain intensity in patients operated for degenerative lumbar spine disorders. Spine (Phila Pa 1976). 2001;26(21):2375–80.

[16] 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(17):1899–908; discussion 909.

[17] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–74.
[18] Al-Tameemi HN, Al-Essawi S, Shukri M, Naji FK. Using Magnetic Resonance Myelography to Improve Interobserver Agreement in the Evaluation of Lumbar Spinal Canal Stenosis and Root Compression. Asian Spine J. 2017;11(2):198–203.
[19] Moojen WA, Schenck CD, Lycklama À Nijeholt GJ, Jacobs WCH, Van der Kallen BF, Arts MP, et al. Preoperative MRI in Patients With Intermittent Neurogenic Claudication: Relevance for Diagnosis and Prognosis. Spine (Phila Pa 1976). 2018;43(5):348–55.
[20] Minetama M, Kawakami M, Teraguchi M, Matsuo S, Enyo Y, Nakagawa M, et al. MRI grading of spinal stenosis is not associated with the severity of low back pain in patients with lumbar spinal stenosis. BMC Musculoskelet Disord. 2022;23(1):857.
[21] Yang J, Xiong Y, Hu Y, Huang M, Zhang L, Pu X, et al. The reliability, correlation with clinical symptoms and surgical outcomes of dural sac cross-sectional area, nerve root sedimentation sign and morphological grade for lumbar spinal stenosis. BMC Musculoskelet Disord. 2023;24(1):225.
[22] Banitalebi H, Espeland A, Anvar M, Hermansen E, Hellum C, Brox JI, et al. Reliability of preoperative MRI findings in patients with lumbar spinal stenosis. BMC Musculoskelet Disord. 2022;23(1):51.
[23] Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, 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(1):90–4.