Inter- and Intraobserver Agreement of Morphological Grading for Central Lumbar Spinal Stenosis on Magnetic Resonance Imaging

Clemens Weber[1,2,3]   Vidar Rao[1,2]   Sasha Gulati[1,2,3]   Kjell A. Kvistad[4,5]   Øystein P. Nygaard[1,2,3]
Greger Lønne[2,6]

[1] Department of Neurosurgery, Trondheim University Hospital, Trondheim, Norway
[2] Department of Neuroscience, Norwegian University of Science and Technology, Trondheim, Norway
[3] National Advisory Unit on Spinal Surgery, Trondheim University Hospital, Trondheim, Norway
[4] Department of Radiology, Trondheim University Hospital, Trondheim, Norway
[5] Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Norway
[6] Department of Orthopedic Surgery, Innlandet Hospital Trust, Lillehammer, Norway

Address for correspondence Clemens Weber, MD, Department of Neurosurgery, St. Olavs Hospital, Trondheim University Hospital, Postboks 3250 Sluppen, 7005 Trondheim, Norway (e-mail: clemens.weber@gmail.com).

Abstract

Study Design   Validation study of a morphological grading system for central lumbar spinal stenosis.

Objective   To evaluate and validate the inter- and intraobserver agreement of a morphological grading system for central lumbar spinal stenosis on magnetic resonance imaging between neurosurgeons and radiologists.

Methods   Two neurosurgeons and two radiologists independently assessed the morphological grading of lumbar spinal stenosis on pretreatment magnetic resonance imaging of 84 patients. Inter- and intrarater agreements were calculated by comparing the observers' evaluations level to level on the grading method. The results of both clinicians were compared with the assessment of both radiologists.

Results   On axial magnetic resonance images, 189 lumbar disk levels were evaluated for the grade of stenosis. The interobserver agreement between the clinicians was substantial. The interobserver agreement between clinician 1 and both radiologists was substantial, and it was moderate between clinician 2 and both radiologists. The clinicians' intraobserver agreement was almost perfect, and the radiologists' intraobserver agreement was substantial.

Conclusions   The interobserver agreement of this morphological grading for lumbar spinal stenosis was high between both the clinicians and radiologists, whereas the intraobserver agreement was almost perfect. Experienced clinicians may safely evaluate lumbar magnetic resonance images using this morphological grading for central lumbar spinal stenosis.

Keywords

► lumbar spinal stenosis
► magnetic resonance imaging
► morphological grading
► validation study
► neurogenic claudication

Introduction

Management of lumbar spinal stenosis (LSS) can be challenging and requires the integration of patients' symptoms, clinical findings, and results of diagnostic imaging. Magnetic resonance (MR) imaging is the most commonly used imaging modality for diagnosing LSS.[1] Dural sac cross-sectional area (DSCA) is frequently used to quantify the space available for the spinal nerve roots in patients with central LSS. Radiologists can measure these areas using dedicated computer software. The discrepancy between DSCA and symptoms,[2,3] and the fact that area measurements are not always readily available in daily clinical practice, may challenge treating clinicians to rely on their own visual assessment of MR images without any radiologic measurement. A new system, which evaluates the morphology of central LSS on T2-weighted axial MR images, has been introduced by Schizas et al.[4] A morphological grade of central LSS from A to D is determined based on the space available for the nerve rootlets within the cerebrospinal fluid in the dural sac and the presence of epidural fat. The inter- and intraobserver agreements of DSCA and this morphological grading system between radiologists were recently found acceptable. Furthermore, the correlation between both methods was considered strong.[5] For experienced spine surgeons, this morphological grading system may represent a feasible and fast method to evaluate MR images in patients with central LSS. If new methods and gradings are introduced, external validation is required. This morphological grading system presented by Schizas at al has not yet been validated between radiologists and clinicians.

The aim of this study was to validate the inter- and intra-observer agreement of a recently introduced morphological grading system for central LSS on MR imaging between clinicians treating patients with LSS and experienced radiologists.

Materials and Methods

Patient Population

Preoperative MR images from 84 patients included in a national multicenter randomized controlled trial (RCT) comparing two different surgical methods in the treatment of LSS (ClinicalTrials.gov identifier: NCT00546949) were included in the study.[6,7] The mean age was 68 years; 44 (52%) were women. All patients had one- (76%) or two-level (24%) stenosis. The radiologic grading method was not used to determine eligibility in the RCT, and the radiologic evaluation was performed after the primary RCT was completed. The inclusion criteria was: age 50 to 85 years, neurogenic claudication with relief of symptoms by flexion of the lumbar spine, walking distance less than 250 m before symptoms, duration of symptoms > 6 months, and T2-weighted MR images showing central LSS measured by DSCA in one or two levels from L2 to L5 on axial images. Experienced spine surgeons clinically evaluated the patients before inclusion.

Image Evaluation

Images were provided by three university hospitals and three district hospitals, using 1.5-T MR imaging systems and stored as Digital Imaging and Communications in Medicine (DICOM) files. All patients had sagittal T1- and T2-weighted images of the lumbar spine. LSS suspected on sagittal T2-weighted images was confirmed by obtaining axial T2-weighted images at the stenotic levels. Based on a visual assessment of signal-to-noise ratio, image contrast, and the presence of artifacts, the image quality was rated by a neuroradiologist as good in all patients. Two clinicians and two radiologists independently evaluated all images. Both clinicians are consultant neurosurgeons who treat patients with LSS on a regular basis and annually perform 80 to 100 lumbar spine surgeries each. Both radiologists have long experience in evaluating lumbar MR images. One of the radiologists is head of the neuroradiology section at a university hospital, whereas the other is a highly experienced consultant radiologist with a long-standing interest in neuroradiology and orthopedic radiology. All four observers were blinded for patient history, clinical symptoms, and the operated level. Morphological grading (A to D) was scored on axial MR images in the available disk levels between the second and the fifth lumbar vertebra and was based on the cerebrospinal fluid versus rootlet ratio as seen on axial T2-weighted images. This grading method, shown in ►**Fig. 1**, is based on the original publication by Schizas et al.[4] The original publication defined four subgroups of grade A. In our study, we did not use these subgroups because all subgroups of grade A are defined as no or minor stenosis.

In total, 189 axial disk levels were evaluated by all four readers with 37 at level L2–L3, 70 at level L3–L4, and 82 at level L4–L5. For intraobserver agreement analysis, the images of 20 patients with 40 axial disk levels were reevaluated after 2 months.

Statistics

Inter- and intraobserver agreements were evaluated. Linear weighted kappa (κ) was analyzed by comparing the observer's evaluations level to level on the grading method. The interpretation of linear weighted κ is presented in ►**Table 1**.[8] The results of both clinicians were compared with the assessment of both radiologists. The weighted κ was calculated using online freeware, available at http://vassarstats.net/kappa.html.

Ethics

The regional ethic committee for medical research approved the study and all patients gave informed consent.

Results

The interobserver agreement in the morphological grading of LSS assessed by the two clinicians was substantial, and the linear weighted κ was 0.76 (95% confidence interval [CI] 0.69 to 0.83). The interobserver agreement between the two radiologists was substantial with a linear weighted κ of 0.65 (0.56 to 0.74). The interobserver agreement between clinician 1 and both radiologists was substantial, and it was
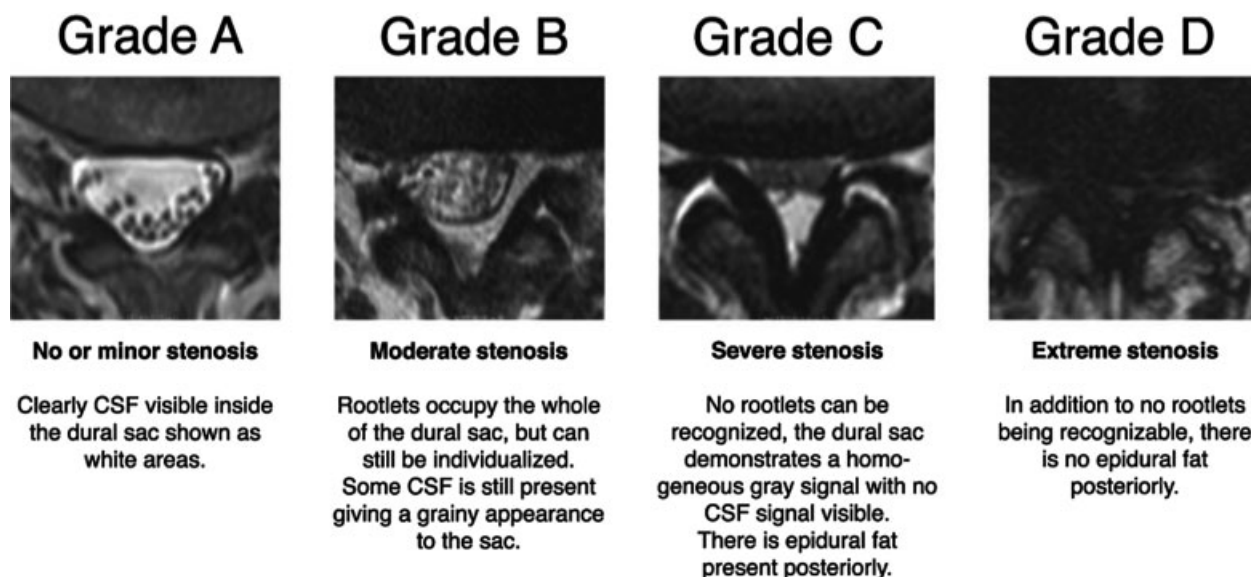
Grade A
No or minor stenosis

Clearly CSF visible inside the dural sac shown as white areas.

Grade B
Moderate stenosis

Rootlets occupy the whole of the dural sac, but can still be individualized. Some CSF is still present giving a grainy appearance to the sac.

Grade C
Severe stenosis

No rootlets can be recognized, the dural sac demonstrates a homogeneous gray signal with no CSF signal visible. There is epidural fat present posteriorly.

Grade D
Extreme stenosis

In addition to no rootlets being recognizable, there is no epidural fat posteriorly.

Fig. 1 Morphological grading of lumbar spinal stenosis according to Schizas et al.[4] Abbreviation: CSF, cerebrospinal fluid.

Table 1 Interpretation of linear weighted κ according to Landis and Koch[8]

κ value	Agreement
<0	< Chance
0.01–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–0.99	Almost perfect

moderate between clinician 2 and both radiologists. The numbers are presented in ►Table 2.

The clinicians' intraobserver agreement was almost perfect, with a linear weighted κ of 0.96 (95% CI 0.89 to 1.00) for clinician 1 and 0.95 (95% CI 0.89 to 1.00) for clinician 2. The radiologists' intraobserver agreement was substantial for radiologist 1 with a linear weighted κ of 0.78 (95% CI 0.65 to 0.92), and almost perfect for radiologist 2 with a linear weighted κ of 0.81 (95% CI 0.68 to 0.94).

Discussion

In this study, we validated inter- and intraobserver agreements of a morphological grading method for central LSS on MR imaging between radiologists and clinicians. We found substantial interobserver agreement between radiologists and clinicians and almost perfect intraobserver agreement for this morphological grading system. Our results show that experienced clinicians may independently evaluate lumbar MR images using this morphological grading system for LSS.

The highest interobserver agreement was found between the clinicians, representing a substantial interobserver agreement. Similarly, a substantial agreement was found between the radiologists. The interobserver agreement between the clinicians and radiologists was moderate. The intraobserver agreement was almost perfect for the evaluated grading method, and the values for the clinicians' intraobserver agreement were even higher than for the radiologists.

When new methods or gradings are introduced, external validations of previous study results are warranted. Our results are in agreement with previous studies evaluating this morphological grading system.[4,5] Furthermore, this

Table 2 Interobserver agreement in the morphological grading of lumbar spinal stenosis

Observers	Linear weighted κ (95% confidence interval)	Interobserver agreement
Clinician 1 vs. clinician 2	0.76 (0.69–0.83)	Substantial
Radiologist 1 vs. radiologist 2	0.65 (0.56–0.74)	Substantial
Clinician 1 vs. radiologist 1	0.63 (0.54–0.71)	Substantial
Clinician 1 vs. radiologist 2	0.67 (0.59–0.75)	Substantial
Clinician 2 vs. radiologist 1	0.58 (0.49–0.66)	Moderate
Clinician 2 vs. radiologist 2	0.59 (051–0.68)	Moderate

grading method seems to have a similar or even higher strength of interobserver agreement compared with other common morphological imaging classifications for degenerative lumbar disk disorders.[9–11] Our interobserver agreement is almost equal to what was recently reported for the four-staged Lee grading. However, the interobserver agreement for clinicians was not tested for the Lee grading.[12]

There are several advantages to using morphological methods for grading of central LSS compared with a quantitative method like DSCA. Morphological grading is performed by a rapid visual assessment, and even though it is a subjective method, it still has high inter- and intraobserver agreements. The evaluation is performed by assessing the axial MR images, and there is no need for more time-consuming computer-based measurements. The angulation of the axial images does not seem to impact morphological grading, whereas the area measurement may be influenced if the image plane of the axial MR images is not perpendicular to the disk space.[13] The combined task forces of the North American Spine Society, the American Society of Spine Radiology, and the American Society of Neuroradiology have recently updated their recommendations for the classification of the degree of spinal canal compromise. A compromise of less than one third of the canal at the evaluated section is classified as "mild," between one and two-thirds as "moderate," and greater than two-thirds as "severe."[14] However, other classifications are widely used and the morphological grading system used in this study is also a practical, objective, and reasonably precise classification.

Based on our study results, an experienced clinician who treats LSS on a regular basis can quickly assess the morphology of LSS on MR images. However, a morphological grading system cannot replace a comprehensive radiologic description of lumbar MR images. The grading system introduced by Schizas can be used to get a quick and easy impression of the morphology of LSS both in clinical and research context. Moreover, a survey among clinicians treating patients with LSS shows that in current clinical practice LSS is likely to be better assessed according to morphology rather than area measurement.[15] Our results show that experienced clinicians also are capable to assess the morphology of LSS on MR imaging with substantial interobserver agreement and almost perfect intraobserver agreement. However, this grading system is only suited to classify central LSS. Lateral recess stenosis or intra- and extra foraminal stenosis with impairment of the exiting nerve root cannot be classified by this system.

The grading system used in our study has not been tested on asymptomatic individuals, and radiologic severity of LSS is not a proven predictor for outcome after surgical treatment of patients with LSS. Moreover, the patient selection for our study is not representative for the standard population of patients with symptomatic LSS as all patients in this study were recruited from an RCT with rigorous inclusion and exclusion criteria. This fact may limit the external validity in our study.

Conclusions

This study validated the inter- and intraobserver agreement of the Schizas morphological grading system for central LSS between radiologists and clinicians. The interobserver agreement was high between both clinicians and radiologists, whereas the intraobserver agreement was almost perfect. Experienced clinicians may safely evaluate lumbar MR images using this morphological grading system for central LSS.

References

1 Sirvanci M, Bhatia M, Ganiyusufoglu KA, et al. Degenerative lumbar spinal stenosis: correlation with Oswestry Disability Index and MR imaging. Eur Spine J 2008;17(5):679–685

2 Lohman CM, Tallroth K, Kettunen JA, Lindgren KA. Comparison of radiologic signs and clinical symptoms of spinal stenosis. Spine (Phila Pa 1976) 2006;31(16):1834–1840

3 Sigmundsson FG, Kang XP, Jönsson B, Strömqvist B. Correlation between disability and MRI findings in lumbar spinal stenosis: a prospective study of 109 patients operated on by decompression. Acta Orthop 2011;82(2):204–210

4 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 (Phila Pa 1976) 2010;35(21):1919–1924

5 Lønne G, Ødegård B, Johnsen LG, Solberg TK, Kvistad KA, Nygaard ØP. MRI evaluation of lumbar spinal stenosis: is a rapid visual assessment as good as area measurement? Eur Spine J 2014;23(6):1320–1324

6 Lønne G, Johnsen LG, Aas E, et al. Comparing cost-effectiveness of X-stop to minimally invasive decompression in lumbar spinal stenosis: a randomized controlled trial. Spine (Phila Pa 1976) 2015

7 Lønne G, Johnsen LG, Rossvoll I, et al. Minimally invasive decompression versus x-stop in lumbar spinal stenosis: a randomized controlled multicenter study. Spine (Phila Pa 1976) 2015;40(2):77–85

8 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–174

9 Kovacs FM, Royuela A, Jensen TS, et al. Agreement in the interpretation of magnetic resonance images of the lumbar spine. Acta Radiol 2009;50(5):497–506

10 Arana E, Royuela A, Kovacs FM, et al. Lumbar spine: agreement in the interpretation of 1.5-T MR images by using the Nordic Modic Consensus Group classification form. Radiology 2010;254(3):809–817

11 Berg L, Neckelmann G, Gjertsen O, et al. Reliability of MRI findings in candidates for lumbar disc prosthesis. Neuroradiology 2012;54(7):699–707

12 Park HJ, Kim SS, Lee YJ, et al. Clinical correlation of a new practical MRI method for assessing central lumbar spinal stenosis. Br J Radiol 2013;86(1025):20120180

13 Henderson L, Kulik G, Richarme D, Theumann N, Schizas C. Is spinal stenosis assessment dependent on slice orientation? A magnetic

resonance imaging study. Eur Spine J 2012;21(Suppl 6): S760–S764

14 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(11):2525–2545

15 Schizas C, Kulik G. Decision-making in lumbar spinal stenosis: a survey on the influence of the morphology of the dural sac. J Bone Joint Surg Br 2012;94(1):98–101