Intra- and inter-rater Reliability of MRI-based Classifications of Degenerative Changes in the Cervical Spine using Objective Criteria

Shrestha D,[1] Danielsen E,[2] Parajuli B,[1] Karki S,[3] Simkhada S,[3] Grotle M,[4,5] Nygaard OP,[6,7,8] Solberg TK[2,9]

[1]Department of Orthopedics,
Dhulikhel Hospital, Kathmandu University Hospital,
Kathmandu University School of Medical Sciences,
Dhulikhel, Kavre, Nepal.
[2]Department of Clinical Medicine,
Faculty of Health Sciences,
UiT The Arctic University of Norway,
Tromsø, Norway.
[3]Department of Radiology,
Dhulikhel Hospital, Kathmandu University Hospital,
Kathmandu University School of Medical Sciences,
Dhulikhel, Kavre, Nepal.
[4]Department of Research and Innovation,
Division of Clinical Neuroscience,
Oslo University Hospital, Norway.
[5]Department of Rehabilitation Science and Health Technology,
Faculty of Health Science,
Oslo Metropolitan University, Norway.
[6]Department of Neurosurgery,
St. Olavs Hospital, Trondheim, Norway.
[7]Department of Neuromedicine and Movement Science,
Norwegian University of Science and Technology,
Trondheim, Norway.
[8]Norwegian National Quality and Competence Network for Spinal Surgery,
St. Olavs Hospital, Trondheim, Norway.
[9]Department of Neurosurgery and the Norwegian Registry for Spine Surgery (NORspine),
University Hospital of North Norway,
Tromsø, Norway.

Corresponding Author
Bikash Parajuli
Department of Orthopedics,
Dhulikhel Hospital, Kathmandu University Hospital
Kathmandu University School of Medical Sciences,
Dhulikhel, Kavre, Nepal.
E-mail: bikash480@gmail.com

Citation

Shrestha D, Danielsen E, Parajuli B, Karki S, Simkhada S, Grotle M, et al. Intra- and inter-rater Reliability of MRI-based Classifications of Degenerative Changes in the Cervical Spine using Objective Criteria. *Kathmandu Univ Med J.* 2026; 95(2):181-91.

ABSTRACT

Background

Magnetic resonance imaging (MRI) is the modality of choice for the radiological evaluation of neck pain, both with and without neurological symptoms. Magnetic resonance imaging offers a detailed, non-invasive investigation of the spinal cord and nerve roots without exposing patients to radiation.

Objective

To determine the intra- and inter-rater reliability of clinically relevant magnetic resonance imaging classifications of degenerative cervical spine changes in a Nepalese working population.

Method

Standardized magnetic resonance imaging examinations of the cervical segments C2/3 to C7/Th1 were conducted as part of a cross-sectional, case-control study of presumably healthy Nepalese males, i.e., professional porters and white-collar workers. Two radiologists independently assessed nine parameters: Pfirrmann grade, anterior osteophytes, disc-osteophyte complex, neural foraminal stenosis, Schmorl's node, Modic changes, spinal canal stenosis, scoliosis, and kyphosis. The same parameters were reassessed after a one-week interval. Gwet's agreement coefficient (AC1) was the main agreement measure. Cohen's kappa and tetrachoric correlation were also reported to improve interpretability of the results.

Result

Intra-rater reliability for both raters ranged from substantial to almost perfect agreement (AC1: 0.63 to 0.98). Rater II showed somewhat higher agreement than rater I. Inter-rater reliability was substantial to almost perfect for all findings (AC1: 0.62 to 0.90), except for spinal canal stenosis (AC1: 0.51), which had moderate agreement.

Conclusion

The intra- and inter-rater reliability of the magnetic resonance imaging findings classifying degenerative changes in the cervical spine were acceptable among presumably healthy workers.

KEY WORDS

Degenerative cervical changes, Magnetic resonance imaging, Reliability, Nepal

INTRODUCTION

Magnetic resonance imaging (MRI) is the modality of choice for the radiological evaluation of neck pain, both with and without neurological symptoms.[1,2] MRI offers a detailed, non-invasive investigation of the spinal cord and nerve roots without exposing patients to radiation.[1] Common findings on MRI are degenerative changes, often referred to as cervical spondylosis, characterized by intervertebral disc degeneration and herniation, osteophyte formation, and hypertrophy of the facet joints and ligaments.[3] While these changes often increase with age, some studies suggest that physical strain also plays a role.[4] Moreover, most people are asymptomatic, but impingement of neural structures secondary to spondylotic changes may cause pain and/or neurological impairment.[3]

Diagnostic accuracy of MRI is crucial for research, diagnosis, and treatment of neck disorders.[1,2] However, there are few established grading systems for MRI classifications of degenerative cervical changes, and the documentation of their reliability is sparse.[5-10] Convergence of evidence is difficult due to variations in the use of grading systems for degenerative changes.[11] The classification systems should be easy to understand and practical to be relevant in clinical practice, as well as reliable for assessing healthy individuals with variable prevalences of MRI findings.[12,13]

The aim of this study was to determine the intra- and inter-rater reliability of frequently used and clinically relevant classification systems for cervical degeneration in a Nepalese working population.

METHODS

The participants of this reliability study were part of a cross-sectional, case-control study of 50 professional porters and an age-matched group (± 2 years) of 50 white-collar workers to assess if the severity of cervical degeneration is associated with more neck pain-related disability. The study was conducted in Nepal between June 2023 and July 2023. We included porters from the Kathmandu Valley being daily exposed to heavy loads on the cervical spine to ensure variation in the exposure to physical strain to the cervical spine among the participants and, possibly, more variability in the prevalence of degenerative changes seen on MRI. Females were not included due to the male predominance among professional porters in Nepal.

For the purpose of this reliability study, we randomly selected MRI scans of 52 of the 100 participants, i.e., 26 porters and 26 white-collar workers. This number was based on an a priori sample size calculation, assuming a kappa of 0.7 and a drop-out rate of 5%, a precision of ± 0.2, a prevalence of 0.4, and using 95% confidence level (CI).[14,15]

The Guidelines for Reporting Reliability and Agreement Studies (GRAAS) were used to report the study.[11] The grammar of the manuscript was reviewed and improved using OpenAI's Chat GTP, but it was not used for data analysis or manuscript writing.

MRI technique

MR images of the cervical segments C2/3, C3/4, C4/5, C5/6, C6/7 and C7/Th1 were acquired using a 1.5 T Ingenia MRI (Philips Medical Systems, Eindhoven, Netherlands) with the following fast spin echo sequences:

- 11 ms echo time (TE) and 400-750 ms repetition time (TR) for T1-weighted sagittal images.

- 100 ms TE and 2500-4500 ms TR for T2-weighted sagittal images.

- Shortest ms TE and 3000-5000 ms TR for T2-weighted axial images.

- The slice thickness was 3 mm.

- The in-plane resolution was 0.75 × 0.9 mm for the sagittal plane and 0.7 × 0.9 mm for the axial plane.

MRI assessments

Two radiologists, with 13 years (rater I) and 2 years (rater II) of experience in evaluating MR images of the cervical spine assessed all cases, blinded to each other's ratings, participant group adherence, and clinical data. The quality of the images was judged before the assessments. To avoid misinterpretation of the study protocol and diagnostic terms, the radiologists reviewed the MRI images of 20 cases together. The remaining 52 cases were assessed independently for the inter-rater analyses and re-evaluated after one week for the intra-rater analyses (Fig. 1). We chose to dichotomize variables according to previous recommendations in the literature to improve interpretability and practicality in clinical use. Table 1 shows details regarding the radiological assessments, which included the Pfirrmann grade (≥ 3: no/yes), anterior osteophytes (no/yes), disc-osteophyte complex (no/yes: bulging, protrusion, extrusion, sequestration), neural foraminal stenosis in the axial plane (Kim grading system ≥ 1: no/yes), facet joint osteoarthritis (Weishaupt grading system ≥ 1: no/yes), Schmorl's nodes (no/yes), Modic changes (no/yes), spinal canal stenosis (Kang grading system ≥ 2: no/yes), kyphosis (no/yes) and scoliosis (no/yes).[6,16-28]

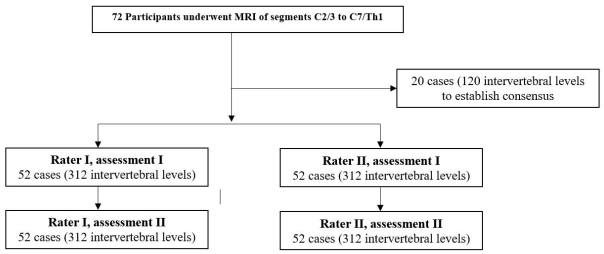


Figure 1. Flow chart

To estimate and interpret intra- and inter-rater reliability, we assessed chance-corrected agreement using Gwet's agreement coefficient (AC1) and Cohen's kappa, latent trait modeling with tetrachoric correlation, and observed agreement (percentage agreement).[13,29-33] Gwet's AC1 was used as the main agreement measure because it provides more stable estimates and is less affected by marginal probabilities and prevalence.[29,32,34] If prevalence is low, Cohen's kappa values tend to be disproportionally low, even zero, despite high percentage agreement ("Kappa Paradox").[31,35] However, since Cohen's kappa is the most frequently reported measure in previous studies, we still chose to report it for comparability. Assuming that a latent trait forms the basis of the MRI findings reported as category variables, we also calculated tetrachoric correlations for dichotomous variables.[30] The rationale was that this analysis separates a difference in threshold (calibration) between the two raters from other, less consistent sources of disagreement.

We used both radiologists' repeated assessments to estimate the prevalence of each MRI finding at each spinal level, and the most senior radiologist's judgment to assess participant prevalence (any level in one individual). Variables with less than eight positive findings were only assessed with percentage agreement, and not with Gwet's AC1, Cohen's kappa, or tetrachoric correlation.[36]

The interpretation of the intra- and inter-rater analyses were in accordance with the classification by Landis and Koch, where < 0.00 indicates poor agreement, 0.00 to 0.20 slight agreement, 0.21 to 0.40 fair agreement, 0.41 to 0.60 moderate agreement, 0.61 to 0.80 substantial agreement, and 0.81 to 1.00 almost perfect agreement.[37]

All statistical analyses were performed using STATA 17 (STATACorp, College Station, Texas, USA). Informed written consent was obtained from all participants. The study protocol was approved by the Institutional Review Committee (IRC) in Nepal (IRC, KUSMS: 74/23) and from the Regional Committees for Medical and Health Research Ethics (REK) in Norway (REK-nord: 611979). The study was conducted in accordance with the Declaration of Helsinki.[38]

RESULTS

Of the 52 cases assessed by the two raters, the porters' mean age (standard deviation (SD)) was 40 years (SD ± 8.3), and the white-collar workers' mean age was 40 years (SD ± 8.4). Both raters assessed six cervical spine levels (C2-Th1) at two-time intervals, resulting in a total of 1248 spine levels assessed (2 radiologists x 6 levels x 52 participants x 2 time points). An additional 208 assessments (2 x 52 x 2) of the entire cervical spine were evaluated for spinal kyphosis and scoliosis. No images were excluded due to low quality.

Prevalence of MRI findings

The sum prevalence of positive MRI findings at all levels

Table 1. Description of magnetic resonance imaging (MRI) parameters assessed at each level of the cervical spine (C2-Th1).

MRI parameters assessed at each level of the cervical spine	Description
Pfirrmann grade ≥ 3 (no/yes)[16]	
Grade I & II	Homogenous, bright white or inhomogeneous with or without horizontal bands.
Grade III	Inhomogeneous, grey.
Grade IV	Inhomogeneous, grey to black.
Grade V	Inhomogeneous, black.
Anterior osteophyte (no/yes)[17]	
No	No anterior osteophyte.
Yes	The presence of anterior osteophyte.
Disc-osteophyte complex (no/yes: bulging, protrusion, extrusion, sequestration)[17-19]	
No	No disc-osteophyte complex.
Bulging	Disc tissue extending beyond the edges of the ring apophyses, throughout the circumference of the disc.
Protrusion	Disc material extending beyond less than 25% of the disc space, with the maximum measure of the displaced disc material is less than the measure of the displaced material at the disc space of origin, measured in the same plane.
Extrusion	Displaced disc material greater than the base of the displaced disc material at the disc space of origin, measured in the same plane.
Sequestration	Defined as displaced disc material that has completely lost continuity with the disc of origin.
Neural foraminal stenosis according to Kim grading system ≥ 1 (no/yes)[20]	
Grade 0	The narrowest width of the neural foramen greater than the width of the extraforaminal nerve root at the level of the anterior margin of the superior articular process.
Grade I	The narrowest width of the neural foramen is 51-100% of the width of the extraforaminal nerve root at the level of the anterior margin of the superior articular process.
Grade II	The width of the neural foramen is the same as or less than 50% of the width of the extraforaminal nerve roots.
Facet joint osteoarthritis according to Weishaupt grading system: ≥ 1 (no/yes)[21]	
Grade 0	Normal facet joint width (~2-4 mm)
Grade I	Facet joint narrowing, small osteophytes and/or articular process hypertrophy
Grade II	Facet joint space narrowing, moderate osteophytes articular process. hypertrophy and/or small sub-articular bone erosions.
Grade III	Facet joint space narrowing, large osteophytes, severe articular process hypertrophy, subarticular bone erosions or subchondral cyst formation.

Schmorl's nodes (no/yes)[22]	
No	No Schmorl's nodes.
Yes	Disc material displaced into the adjacent bony endplate of the vertebra.
Modic changes (no/yes: type I, type II, type III)[23,24]	
No	No Modic changes.
Type I	Decreased signal intensity on T1-weighted spin-echo images and increased signal intensity on T2-weighted images.
Type II	Increased signal intensity on T1-weighted images and isointense or slightly increased signal intensi-ty on T2-weighted images.
Type III	Decreased signal intensity on T1-weighted images and decreased signal intensity on T2-weighted images.
Spinal canal stenosis according to Kang grading system ≥ 2 (no/yes)[6]	
Grade 0	No narrowing of the spinal canal
Grade I	Obliteration of more than 50% of the subarachnoid space, without any sign of cord deformity.
Grade II	Central spinal canal stenosis with spinal cord deformity but no spinal cord signal change.
Grade III	Increased T2 signal intensity of the spinal cord at the level of compression.
MRI parameters assessed for the whole cervical spine for each participant	
Kyphosis (no/yes)[25,26]	
No	No kyphosis.
Yes	Kyphotic angle in the cervical spine defined as at least 1°. Measured using the Jackson physiological stress line method.
Scoliosis (no/yes)[27]	
No	No scoliosis.
Yes	A lateral spinal curvature with a Cobb angle > 10°.

assessed by both raters was classified as Pfirrmann grade ≥ 3 (85.6%), followed by disc-osteophyte complex (26.3%), spinal canal stenosis (16.4%), neural foraminal stenosis (8.7%) and anterior osteophytes (4.6%) (Table 2). Appendix 1 shows the MRI findings at each cervical segment, and Appendix 2 shows common degenerative changes in the cervical spine on MRI.

Intra-rater reliability

The intra-rater reliability measured by Gwet's AC1 for both raters ranged from 0.63 (substantial) to 0.98 (almost perfect) (Table 3). Cohen's kappa ranged from 0.50 (moderate) to 0.94 (almost perfect) (Table 3). Tetrachoric correlation ranged from 0.77 to 1.00. Rater II showed somewhat higher intra-rater agreement than rater I. For the variables with less than eight positive findings, the percentage agreement for facet joint osteoarthritis, Schmorl's node, Modic changes and scoliosis ranged from 96.2 to 98.1.

Inter-rater reliability

The inter-rater reliability measured by Gwet's AC1 ranged from 0.51 (moderate) to 0.90 (almost perfect) (Table 3). Cohen's kappa ranged from 0.48 (moderate) to 0.71 (substantial). Tetrachoric correlation ranged from 0.71 to 1.00. The inter-rater agreement according to Gwet's AC1 was lowest for spinal canal stenosis (0.51), the Pfirrmann grade (0.62), and neural foraminal stenosis (0.69). Substantial agreement in defining kyphosis (0.66) was associated with a perfect tetrachoric correlation (1.00). The percentage agreement for facet joint osteoarthritis, Schmorl's node, Modic changes and scoliosis ranged from 94.2 to 98.1.

DISCUSSIONS

Judging by our main agreement measure, Gwet's AC1, intra-rater agreement for the assessment of cervical spine degeneration ranged from substantial to almost perfect, while inter-rater agreement ranged from moderate to almost perfect. The assessments of spinal canal stenosis had the lowest (moderate) agreement. Notably, there appears to be no prior evaluation of the reliability of anterior osteophytes, the disc-osteophyte complex, or the Jackson physiological stress line, meaning that these findings should be critically appraised in future studies.

We found substantial to almost perfect intra- and inter-rater reliability for the assessment of anterior osteophytes (AC1: 0.90) and disc-osteophyte complex (AC1: 0.76). The use of the disc-osteophyte complex composite variable is controversial but has been recommended when it is difficult to distinguish between disc herniation and a posterior osteophyte in the cervical spine.[18,19] This is often the case if no gradient-echo, low flip-angle, thin-section axial scans are available, as in our study.

Inter-rater reliability (AC1: 0.66) was substantial for assessing kyphosis using the Jackson physiological stress line method, which is a simplified version of the Harrison posterior tangent method.[25] The latter, more complex method, as well as the four-line Cobb method, have shown substantial to almost perfect reliability (intraclass correlation coefficient ranging from 0.7 to 1.0).[25,39]

As expected, Cohen's kappa generally had lower agreement estimates, especially when the prevalence of positive findings was low. This is illustrated by the moderate (0.48) inter-rater agreement of the kyphosis variable, whereas the tetrachoric correlation was perfect (1.00). This discrepancy indicates that disagreement was caused mainly by systematic differences in thresholds for defining kyphosis between the raters, rather than by other random factors. An advantage of assessing tetrachoric correlation is that thresholds can be calibrated through feedback and training to improve agreement. According to our results, the Pfirrmann grade and spinal canal stenosis could also be improved by calibration.

The inter-rater reliability for the Pfirrmann grade (Cohen's kappa 0.50) was lower than reported by Chen et al.

Table 2. **The sum prevalence of all positive magnetic resonance imaging (MRI) findings at each spinal level by both raters.**

Cervical spine level	Pfirrmann grade ≥ 3	Anterior osteo-phytes	Dis-costeophyte complex	Neural foraminal stenosis	Facet joint osteoarthritis	Schmorl's nodes	Modic changes	Spinal canal stenosis
C2/3	177 (16.6)	0 (0.0)	15 (4.6)	0 (0.0)	0 (0.0)	0 (0.0)	1(16.7)	4 (2.0)
C3/4	198 (18.5)	5 (8.6)	37 (11.3)	9 (8.3)	0 (0.0)	0 (0.0)	0 (0.0)	20 (9.8)
C4/5	197 (18.5)	15 (1.2)	62 (18.9)	15 (13.8)	1 (50.0)	0 (0.0)	2 (33.3)	41(20.1)
C5/6	199 (18.6)	21 (36.2)	113 (34.5)	50 (45.9)	1 (50.0)	1 (100.0)	3 (50.0)	87 (42.6)
C6/7	180 (16.8)	15 (1.2)	92 (28.1)	34 (31.2)	0 (0.0)	0 (0.0)	0 (0.0)	50 (24.5)
C7/Th1	117 (11.0)	2 (3.5)	9 (2.7)	1 (0.9)	0 (0.0)	0 (0.0)	0 (0.0)	2 (1.0)
Prevalence: n (%)	1068/1248= (85.6)	58/1248= (4.6)	328/1248= (26.3)	109/1248= (8.7)	2/1248= (0.2)	1/1248= (0.08)	6/1248= (0.5)	204/1248= (16.4)

Table 3. **Prevalence and intra- and inter-rater reliability of degenerative changes in the cervical spine on magnetic resonance imaging in 52 participants.**

		Intra-rater reliability	Inter-rater reliability
Pfirrmann grade	**Rater I**	**Rater II**	
Prevalence[a]: 52 (100%)			
Gwet's AC1[b] (95% CI[c])	0.72 (0.57 – 0.88)	0.63 (0.41 – 0.85)	0.62 (0.45 – 0.80)
Cohen's weighted kappa[d] (95% CI)	0.64 (0.46 – 0.82)	0.61 (0.39 – 0.82)	0.50 (0.32 – 0.69)
Polychoric[e] (SE)[f]	0.91 (0.06)	0.86 (0.09)	0.82 (0.11)
Percentage agreement (95% CI)	78.9 (67.4 – 90.3)	80.8 (69.7 – 91.9)	71.5 (58.4 – 83.9)
Anterior osteophytes			
Prevalence: 9 (17.3%)			
Gwet's AC1 (95% CI)	0.82 (0.67 – 0.96)	0.98 (0.92 – 1.00)	0.90 (0.79 – 1.00)
Cohen's kappa[g] (95% CI)	0.51 (0.18 – 0.84)	0.92 (0.77 – 1.00)	0.71 (0.43 – 0.98)
Tetrachoric[h] (SE)	0.77 (0.15)	1.00 (0.00)	0.93 (0.07)
Percentage agreement (95% CI)	86.5 (76.9 – 96.1)	98.1 (94.2 – 1.0)	92.3 (84.8 – 99.8)
Disc-osteophyte complex			
Prevalence: 37 (71.2%)			
Gwet's AC1(95% CI)	0.80 (0.63 – 0.96)	0.97 (0.92 – 1.00)	0.76 (0.59 – 0.94)
Cohen's kappa (95% CI)	0.73 (0.52 – 0.94)	0.94 (0.81 – 1.00)	0.57 (0.32 – 0.84)
Tetrachoric (SE)	0.92 (0.06)	1.00 (0.00)	0.88 (0.10)
Percentage agreement (95% CI)	88.5 (79.5 – 97.4)	98.1 (94.2 – 1.00)	84.6 (74.5 – 94.8)
Neural foraminal stenosis			
Prevalence: 20 (38.5%)			
Gwet's AC1 (95% CI)	0.85 (0.70 – 1.00)	0.94 (0.84 – 1.00)	0.69 (0.48 – 0.89)
Cohen's kappa (95% CI)	0.84 (0.69 – 0.99)	0.91 (0.78 – 1.00)	0.62 (0.39 – 0.85)
Tetrachoric (SE)	1.00 (0.00)	0.99 (0.01)	0.86 (0.09)
Percentage agreement (95% CI)	92.3 (84.8 – 99.8)	96.2 (90.8 – 1.00)	82.7 (72.1 – 93.3)
Spinal canal stenosis			
Prevalence: 30 (57.7)			
Gwet's AC1(95% CI)	0.71 (0.52 – 0.91)	0.89 (0.75 – 1.00)	0.51 (0.26 – 0.75)
Cohen's kappa (95% CI)	0.67 (0.46 – 0.88)	0.88 (0.75 – 1.00)	0.50 (0.25 – 0.74)
Tetrachoric (SE)	0.92 (0.06)	0.99 (0.02)	0.71 (0.13)
Percentage agreement (95% CI)	84.6 (74.5 – 94.8)	94.2 (87.7 – 1.00)	75.0 (62.8 – 87.2)
Kyphosis			
Prevalence: 8 (15.4%)			
Gwet's AC1 (95% CI)	0.88 (0.77 – 0.99)	0.93 (0.82 – 1.00)	0.66 (0.44 – 0.87)
Cohen's kappa (95% CI)	0.50 (0.13 – 0.88)	0.92 (0.81 – 1.00)	0.48 (0.24 – 0.72)
Tetrachoric (SE)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)

Percentage agreement (95% CI)	90.4 (82.1 – 98.7)	96.2 (90.8 – 1.00)	78.9 (67.4 – 90.3)
Facet joint osteoarthritis			
Prevalence: 0 (0.0%)			
Percentage agreement (95% CI)	98.1 (94.2 – 1.00)	98.1 (94.2 – 1.00)	98.1 (94.2 – 1.00)
Schmorl's node			
Prevalence: 1 (1.9%)		Prevalence 0 (ratings do not vary)	
Percentage agreement (95% CI)	98.1 (94.2 – 1.00)		98.1 (94.2 – 1.00)
Modic changes			
Prevalence: 3 (5.8%)		Prevalence 0 (ratings do not vary)	
Percentage agreement (95% CI)	96.2 (90.8 – 1.00)		94.2 (87.7 – 1.00)
Scoliosis			
Prevalence: 1 (1.9%)		Prevalence 0 (ratings do not vary)	
Percentage agreement (95% CI)	98.1 (94.2 – 1.00)		98.1 (94.2 – 1.00)

[a]Prevalence of positive findings in one individual at any level, rated by the most experienced radiologist. [b]Gwet's agreement coefficient. [c]95% confidence interval. [d]Cohen's weighted kappa for ordinal values. [e]Polychoric correlation. [f]Standard error. [g]Cohen's kappa for nominal values. [h]Tetrachoric correlation.

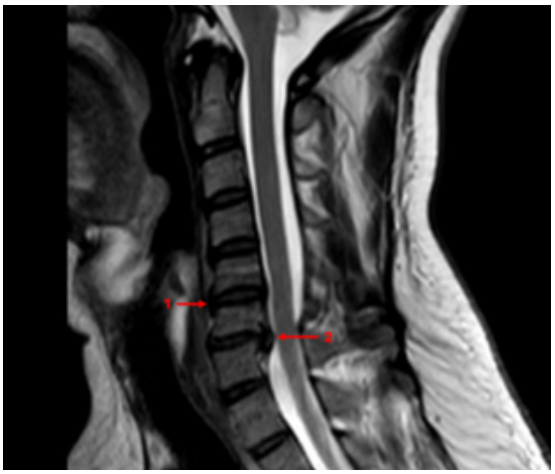
Appendix I. Degenerative cervical changes at each level on magnetic resonance imaging (MRI) assessed by both raters.

MRI findings	Rater I		Rater II	
	First assessment (N=52)	Second assessment (N=52)	First assessment (N=52)	Second assessment (N=52)
Pfirrmann C2/3: n (%)				
Grade ≤ 2	12 (23.1)	12 (23.1)	6 (11.5)	1 (1.9)
Grade 3	25 (48.1)	37 (71.1)	38 (73.1)	44 (84.6)
Grade 4	15 (28.8)	3 (5.8)	8 (15.4)	7 (13.5)
Grade 5	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Pfirrmann C3/4: n (%)				
Grade ≤ 2	4 (7.7)	6 (11.5)	0 (0.0)	0 (0.0)
Grade 3	28 (53.8)	30 (57.7)	43 (82.7)	43 (82.7)
Grade 4	20 (38.5)	16 (30.8)	9 (17.3)	9 (17.3)
Grade 5	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Pfirrmann C4/5: n (%)				
Grade ≤ 2	3 (5.8)	5 (9.6)	2 (3.9)	1 (1.9)
Grade 3	23 (44.2)	21 (40.4)	34 (65.4)	39 (75.0)
Grade 4	26 (50.0)	26 (50.0)	16 (30.7)	12 (23.1)
Grade 5	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Pfirrmann C5/6: n (%)				
Grade ≤ 2	2 (3.9)	4 (7.7)	1 (1.9)	2 (3.9)
Grade 3	19 (36.5)	21 (40.4)	30 (57.7)	33 (63.5)
Grade 4	27 (51.9)	26 (50.0)	21 (40.4)	17 (32.6)
Grade 5	4 (7.7)	1 (1.9)	0 (0.0)	0 (0.0)
Pfirrmann C6/7: n (%)				
Grade ≤ 2	6 (11.5)	12 (23.1)	5 (9.6)	5 (9.6)
Grade 3	21 (40.4)	21 (40.4)	31 (59.6)	34 (65.4)
Grade 4	24 (46.2)	19 (36.5)	16 (30.8)	13 (25.0)
Grade 5	1 (1.9)	0 (0.0)	0 (0.0)	0 (0.0)
Pfirrmann C7/Th1: n (%)				
Grade ≤ 2	33 (63.5)	36 (69.2)	28 (53.8)	27 (51.9)
Grade 3	14 (26.9)	14 (26.9)	24 (46.2)	25 (48.1)
Grade 4	5 (9.6)	2 (3.9)	0 (0.0)	0 (0.0)
Grade 5	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)

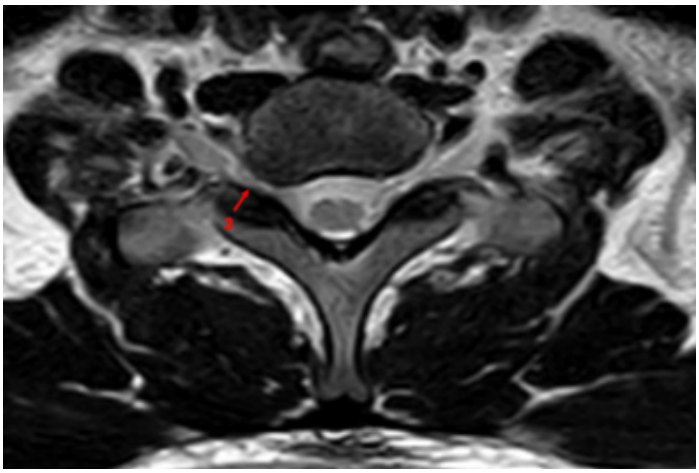
Anterior osteophytes C2/3: n (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Anterior osteophytes C3/4: n (%)	1 (1.9)	1 (1.9)	2 (3.9)	1 (1.9)
Anterior osteophytes C4/5: n (%)	5 (9.6)	3 (5.8)	3 (5.8)	4 (7.7)
Anterior osteophytes C5/6: n (%)	5 (9.6)	5 (9.6)	5 (9.6)	6 (11.5)
Anterior osteophytes C6/7: n (%)	5 (9.6)	3 (5.8)	4 (7.8)	3 (5.8)
Anterior osteophytes C7/Th1: n (%)	0 (0.0)	0 (0.0)	1 (9.2)	1 (1.9)
Disc-osteophyte complex C2/3: n (%)				
No	48 (92.3)	52 (100.0)	46 (88.5)	47 (90.3)
Protrusion	3 (5.8)	0 (0.0)	2 (3.8)	2 (3.9)
Extrusion	1 (1.9)	0 (0.0)	4 (7.7)	3 (5.8)
Sequestration	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Disc-osteophyte complex C3/4: n (%)				
No	43 (82.7)	46 (88.4)	41 (78.9)	41 (78.8)
Protrusion	5 (9.6)	4 (7.7)	6 (11.5)	4 (7.7)
Extrusion	4 (7.7)	2 (3.9)	5 (9.6)	7 (13.5)
Sequestration	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Disc-osteophyte complex C4/5: n (%)				
No	38 (73.1)	42 (80.8)	33 (63.5)	33 (63.5)
Protrusion	8 (15.4)	6 (11.5)	8 (15.4)	6 (11.5)
Extrusion	6 (11.5)	4 (7.7)	11 (21.1)	13 (25.0)
Sequestration	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Disc-osteophyte complex C5/6: n (%)				
No	29 (55.8)	25 (48.0)	20 (38.5)	21 (40.4)
Protrusion	11 (21.1)	24 (46.2)	20 (38.5)	16 (30.7)
Extrusion	12 (23.1)	3 (5.8)	12 (23.0)	15 (28.9)
Sequestration	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Disc-osteophyte complex C6/7: n (%)				
No	30 (57.7)	33 (63.4)	24 (46.2)	29 (55.7)
Protrusion	5 (9.6)	11 (21.2)	13 (25.0)	11 (21.2)
Extrusion	16 (30.8)	8 (15.4)	15 (28.8)	12 (23.1)
Sequestration	1 (1.9)	0 (0.0)	0 (0.0)	0 (0.0)
Disc-osteophyte complex C7/Th1: n (%)				
No	49 (94.2)	51 (98.1)	50 (96.2)	49 (94.2)
Protrusion	3 (5.8)	1 (1.9)	1 (1.9)	1 (1.9)
Extrusion	0 (0.0)	0 (0.0)	1 (1.9)	3 (3.9)
Sequestration	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Foraminal stenosis C2/3: n (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Foraminal stenosis C3/4: n (%)	5 (9.6)	4 (7.7)	0 (0.0)	0 (0.0)
Foraminal stenosis C4/5: n (%)	5 (9.6)	6 (11.5)	2 (3.9)	2 (3.9)
Foraminal stenosis C5/6: n (%)	11 (21.2)	17 (32.7)	10 (19.2)	12 (23.1)
Foraminal stenosis C6/7: n (%)	10 (19.2)	11 (21.2)	8 (15.4)	5 (9.6)
Foraminal stenosis C7/Th1: n (%)	1 (1.9)	0 (0.0)	0 (0.0)	0 (0.0)
Facet arthritis C2/3: n (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Facet arthritis C3/4: n (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Facet arthritis C4/5: n (%)	0 (0.0)	1 (1.9)	0 (0.0)	0 (0.0)
Facet arthritis C5/6: n (%)	0 (0.0)	0 (0.0)	1 (1.9)	0 (0.0)
Facet arthritis C6/7: n (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Facet arthritis C7/Th1: n (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Schmorl's nodes C2/3: n (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Schmorl's nodes C3/4: n (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Schmorl's nodes C4/5: n (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)

Schmorl's nodes C5/6: n (%)	1 (1.9)	0 (0.0)	0 (0.0)	0 (0.0)
Schmorl's nodes C6/7: n (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Schmorl's nodes C7/Th1: n (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Modic changes C2/3: n (%)				
No	52 (100.0)	51 (98.1)	52 (100.0)	52 (100.0)
Type I	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Type II	0 (0.0)	1 (1.9)	0 (0.0)	0 (0.0)
Type III	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Modic changes C3/4: n (%)				
No	52 (100.0)	52 (100.0)	52 (100.0)	52 (100.0)
Type I	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Type II	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Type III	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Modic changes C4/5: n (%)				
No	51 (98.1)	51 (98.1)	52 (100.0)	52 (100.0)
Type I	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Type II	1 (1.9)	1 (1.9)	0 (0.0)	0 (0.0)
Type III	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Modic changes C5/6: n (%)				
No	51 (98.1)	50 (96.1)	52 (100.0)	52 (100.0)
Type I	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Type II	1 (1.9)	2 (3.9)	0 (0.0)	0 (0.0)
Type III	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Modic changes C6/7: n (%)				
No	51 (98.1)	50 (96.2)	52 (100.0)	52 (100.0)
Type I	0 (0.0)	1 (1.9)	0 (0.0)	0 (0.0)
Type II	1 (1.9)	1 (1.9)	0 (0.0)	0 (0.0)
Type III	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Modic changes C7/Th1: n (%)				
No	52 (100.0)	52 (100.0)	52 (100.0)	52 (100.0)
Type I	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Type II	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Type III	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Spinal canal stenosis C2/3: n (%)	**0 (0.0)**	**0 (0.0)**	**2 (3.9)**	**2 (3.9)**
Spinal canal stenosis C3/4: n (%)	**6 (11.5)**	**4 (7.7)**	**4 (7.7)**	**6 (11.5)**
Spinal canal stenosis C4/5: n (%)	**13 (25.0)**	**12 (23.1)**	**8 (15.4)**	**8 (15.4)**
Spinal canal stenosis C5/6: n (%)	**21 (40.4)**	**29 (55.8)**	**17 (32.7)**	**20 (38.5)**
Spinal canal stenosis C6/7: n (%)	**12 (23.1)**	**17 (32.7)**	**12 (23.1)**	**9 (17.3)**
Spinal canal stenosis C7/Th1: n (%)	**0 (0.0)**	**0 (0.0)**	**1 (1.9)**	**1 (1.9)**
Kyphosis: n (%)	**8 (15.4)**	**3 (5.8)**	**19 (36.5)**	**21 (40.4)**
Scoliosis: n (%)	**1 (1.9)**	**0 (0.0)**	**0 (0.0)**	**0 (0.0)**

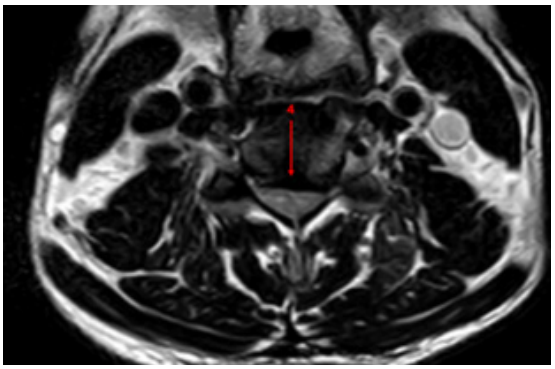
Appendix 2. **Representative cases of common degenerative cervical spine changes on magnetic resonance imaging.**



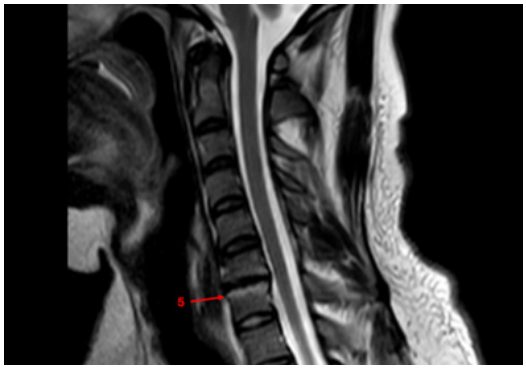
Sagittal imaging of the cervical spine shows (1) anterior osteophytes at C5/6 and Pfirrmann grade III degeneration, and (2) posterior osteophytes at C6/7 and Pfirrmann grade III degeneration.



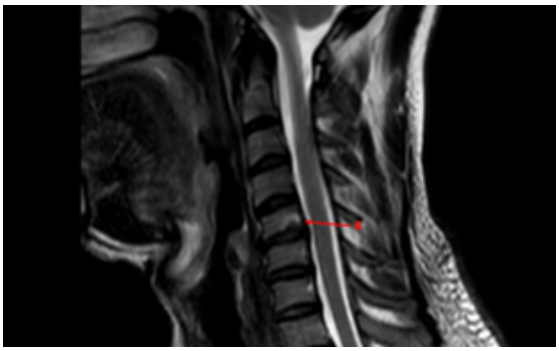
Foraminal stenosis at C7/Th1 according to the Kim grading system (≥ 1; no/yes).



Focal compressive cord myelopathic changes with spinal canal stenosis at C4/5 according to the Kang grading system (≥ 2; no/yes).



Modic changes type II at C6/7.



Schmorl's nodes at C5/6.

(0.85).[40,41] However, they assessed a different age group, used a 3-Tesla MRI, and manually outlined the annulus fibrosus and nucleus pulposus areas for morphological differentiation. These methodological differences may explain much of the variance in results. The Pfirrmann grade was originally developed for the lumbar spine.[16] Even if the alternative Suzuki classification was developed for the cervical spine, it has not shown more than equal reliability.[8]

Substantial inter-rater reliability (Cohen's kappa 0.62) was found for neural foraminal stenosis on axial images, consistent with findings by Kim et al.[5,20] Oblique sagittal images were not available in the present study, however, strong concordance and inter-rater reliability between neural foraminal grading systems based on axial and oblique sagittal images have been reported previously.[42,43]

The inter-rater reliability was moderate (Cohen's kappa 0.51) for spinal canal stenosis. This is lower than values reported by Kang et al. (0.56 to 0.66) for the same dichotomization of the variable.[6] In contrast, Park et al. attained an overall Cohen's kappa value of 0.93.[7] This variation could be attributed to differences in prevalence, since diagnostic accuracy is highly dependent on the prevalence of the condition being studied.

Our study illustrates that reporting more than just Kappa statistics can be advantageous. The inclusion of precise information related to prevalence, Gwet's AC1, and tetrachoric correlation could improve the interpretation of results and comparisons across studies.[13] However, the interpretation of Gwet's AC1 according to the classification by Landis and Koch has been criticized, because it may overestimate agreement in some situations.[29,35,44] However,

to date, no better alternative exists to this classification system.

Only male participants were included. Thus, the prevalence of MRI findings might differ among women, limiting the external validity of the study. The purpose of the present study, however, was not to assess prevalence between sexes but rather to assess agreement within and between raters.

We used dichotomized MRI classifications because they tend to be more relevant and practical in clinical use and may help reduce differences between assessments by specialized radiologists and clinicians. Moreover, dichotomized variables tend to perform better in reliability analyses because the demand for finer discrimination is less.[13] Since we made such simplifications (re-categorizations), we also calculated tetrachoric correlations to ensure that agreement on the underlying traits of the variables was sufficient. For all variables, these correlations were high.

The two radiologists had vastly different clinical experience, which could potentially be a source of bias. However, inter-rater reliability ranged from substantial to almost perfect, indicating that the MRI classifications could work despite different experience levels. The only exception was spinal canal stenosis, which showed moderate agreement, possibly due to the classification system itself or differences in the radiologists' clinical experience. Moreover, the tetrachoric correlation ranged from 0.77 – 1.0 for intra-rater reliability and 0.71 – 1.0 for inter-rater reliability, meaning that agreement could be improved through training. Rater II (2 years of experience) had somewhat higher agreement than rater I (13 years of experience). A possible explanation could be closer adherence to the MRI manual (standardization) by rater II. While including additional raters could lead to more stable assessments, this would not necessarily improve inter-rater reliability, but possibly external validity.

The absence of axial gradient echo sequences is another limitation of the study. T2 axial images alone are reliable for detecting encroachment but less specific for distinguishing disc protrusion from osteophytes, as both may appear hypointense.[45] The decision to combine them into a "disc-osteophyte complex" is pragmatic, but reduces specificity.[45]

Schmorl's nodes, Modic changes, and scoliosis had fewer than eight observations. This also includes facet joint osteoarthritis, which is likely underestimated on MRI.[46] For these variables, there were too few positive findings to allow any meaningful assessment of chance agreement. Even if the percentage agreement for these parameters were high, larger studies with more statistical power will have to provide information about intra- and inter-rater agreement.

CONCLUSION

The inter-rater reliability was substantial to moderate for all magnetic resonance imaging findings assessing degenerative cervical changes in presumably healthy individuals.

ACKNOWLEDGEMENTS

We would like to thank all the participants in the research project, Manoj Tamang and Inger Sperstad.

REFERENCES

1. Expert Panel on Neurological Imaging:; McDonald MA, Kirsch CFE, Amin BY, Aulino JM, Bell AM, Cassidy RC, et al. ACR Appropriateness Criteria® Cervical Neck Pain or Cervical Radiculopathy. *J Am Coll Radiol*. 2019 May;16(5S):S57-S76. doi: 10.1016/j.jacr.2019.02.023. PMID: 31054759.
2. Brown BM, Schwartz RH, Frank E, Blank NK. Preoperative evaluation of cervical radiculopathy and myelopathy by surface-coil MR imaging. *AJR Am J Roentgenol.* 1988;151(6):1205-12. doi:10.2214/ajr.151.6.1205
3. Theodore N. Degenerative Cervical Spondylosis. *N Engl J Med.* 2020;383(2):159-168. doi:10.1056/NEJMra2003558
4. Binder AI. Cervical spondylosis and neck pain. *BMJ.* 2007;334(7592):527-31. doi:10.1136/bmj.39127.608299.80
5. Meacock J, Schramm M, Selvanathan S, Currie S, Stocken D, Jayne D, et al. Systematic review of radiological cervical foraminal grading systems. Neuroradiology. 2021 Mar;63(3):305-16.
6. Kang Y, Lee JW, Koh YH, Hur S, Kim SJ, Chai JW, et al. New MRI grading system for the cervical canal stenosis. *AJR Am J Roentgenol.* 2011 Jul;197(1):W134-40. doi: 10.2214/AJR.10.5560. PMID: 21700974.
7. Park HJ, Kim SS, Chung EC, Lee SY, Park NH, Rho MH, et al. Clinical correlation of a new practical MRI method for assessing cervical spinal canal compression. *AJR Am J Roentgenol.* 2012 Aug;199(2):W197-201. doi: 10.2214/AJR.11.7599. PMID: 22826421.
8. Urbanschitz L, Bensler S, Merat S, Lenz CG, Eid K. How Should We Grade Cervical Disk Degeneration? A Comparison of Two Popular Classification Systems. *Spine Surg Relat Res*. 2021;5(6):347-351. doi:10.22603/ssrr.2021-0048
9. Gan G, Harkey P, Hemingway J, Hughes DR, Duszak R, Jr. Changing Utilization Patterns of Cervical Spine Imaging in the Emergency Department: Perspectives From Two Decades of National Medicare Claims. *J Am Coll Radiol.* 2016;13(6):644-8. doi:10.1016/j.jacr.2016.02.003
10. Fu MC, Webb ML, Buerba RA, Neway WE, Brown JE, Trivedi M, et al. Comparison of agreement of cervical spine degenerative pathology findings in magnetic resonance imaging studies. *Spine J.* 2016 Jan 1;16(1):42-8. doi: 10.1016/j.spinee.2015.08.026. Epub 2015 Aug 17. PMID: 26291398.
11. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol.* 2011 Jan;64(1):96-106. doi: 10.1016/j.jclinepi.2010.03.002. Epub 2010 Jun 17. PMID: 21130355.
12. Lachin JM. The role of measurement reliability in clinical trials. *Clin Trials*. 2004;1(6):553-566. doi:10.1191/1740774504cn057oa
13. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb).* 2012;22(3):276-82.

14. Okada E, Ichihara D, Matsumoto M, Chiba K, Toyama Y, Momoshima S, et al. Modic changes in the cervical spine. *J Bone Joint Surg*. British volume. 2012;94(5):678-83. doi:10.1302/0301-620x.94b5.28519

15. WN A. Sample size calculator (web). 7 April, 2024. http://wnarifin.github.io

16. Pfirrmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N. Magnetic resonance classification of lumbar intervertebral disc degeneration. *Spine* (Phila Pa 1976). 2001;26(17):1873-8. doi:10.1097/00007632-200109010-00011

17. van der Kraan PM, van den Berg WB. Osteophytes: relevance and biology. *Osteoarthritis Cartilage*. 2007;15(3):237-44. doi:10.1016/j.joca.2006.11.006

18. Yousem DM. Point: Don't Call Me a DOC! *AJNR Am J Neuroradiol*. 2016;37(12):2180-2180. doi:10.3174/ajnr.A4927

19. Counterpoint: Appropriately Ambiguous-The Disc Osteophyte Complex. AJNR Am J Neuroradiol. 2016 Dec;37(12):2181. doi: 10.3174/ajnr.A4928. Epub 2016 Aug 25. PMID: 27561836; PMCID: PMC7963872.

20. Kim S, Lee JW, Chai JW, Yoo HJ, Kang Y, Seo J, et al. A New MRI Grading System for Cervical Foraminal Stenosis Based on Axial T2-Weighted Images. *Korean J Radiol*. 2015 Nov-Dec;16(6):1294-302. doi: 10.3348/kjr.2015.16.6.1294. Epub 2015 Oct 26. PMID: 26576119; PMCID: PMC4644751.

21. Kalichman L, Suri P, Guermazi A, Li L, Hunter DJ. Facet orientation and tropism: associations with facet joint osteoarthritis and degeneratives. *Spine* (Phila Pa 1976). 2009;34(16):E579-85. doi:10.1097/BRS.0b013e3181aa2acb

22. Moustarhfir M, Bresson B, Koch P, Perozziello A, Barreau G, Schouman-Claeys E, et al. MR imaging of Schmorl's nodes: Imaging characteristics and epidemio-clinical relationships. *Diagn Interv Imaging.* 2016 Apr;97(4):411-7. doi: 10.1016/j.diii.2016.02.001. Epub 2016 Mar 3. PMID: 26947187.

23. Modic MT, Steinberg PM, Ross JS, Masaryk TJ, Carter JR. Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging. *Radiology.* 1988;166(1 Pt 1):193-9. doi:10.1148/radiology.166.1.3336678

24. Rahme R, Moussa R. The modic vertebral endplate and marrow changes: pathologic significance and relation to low back pain and segmental instability of the lumbar spine. *AJNR Am J Neuroradiol*. 2008;29(5):838-42. doi:10.3174/ajnr.A0925

25. Cohn PL, Spiker WR, Radliff KE, Vaccaro AR. Cervical Sagittal Balance: What Is Normal and What Is the Effect on Reconstruction Outcomes. AOSpine Masters Series. Georg Thieme Verlag KG; 2015. http://www.thieme-connect.de/products/ebooks/lookinside/10.1055/b-0035-120943

26. Scheer JK, Tang JA, Smith JS, Acosta FL Jr, Protopsaltis TS, Blondel B, et al. Cervical spine alignment, sagittal deformity, and clinical implications: a review. *J Neurosurg Spine*. 2013 Aug;19(2):141-59. doi: 10.3171/2013.4.SPINE12838. Epub 2013 Jun 14. PMID: 23768023.

27. Trobisch P, Suess O, Schwab F. Idiopathic scoliosis. *Dtsch Arztebl Int.* 2010;107(49):875-83; quiz 884. doi:10.3238/arztebl.2010.0875

28. Hebelka H, Gunterberg V, Lagerstrand K, Brisby H. Clinical outcome and MRI appearance in a group of chronic low back pain patients more than 10 years after discography evaluation and consideration for surgery. *BMC Musculoskelet Disord.* 2023;24(1):138. doi:10.1186/s12891-023-06242-y

29. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Med Res Methodol.* 2013;13(1):61. doi:10.1186/1471-2288-13-61

30. Hripcsak G, Heitjan DF. Measuring agreement in medical informatics reliability studies. *J Biomed Inform.* 2002;35(2):99-110. doi:https://doi.org/10.1016/S1532-0464(02)00500-2

31. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol*. 1993;46(5):423-429. doi:https://doi.org/10.1016/0895-4356(93)90018-V

32. Silveira PSP, Siqueira JO. Better to be in agreement than in bad company : A critical analysis of many kappa-like tests. *Behav Res Methods*. 2023;55(7):3326-3347. doi:10.3758/s13428-022-01950-0

33. Moon KP, Suh KT, Lee JS. Reliability of MRI findings for Symptomatic Extraforaminal Disc Herniation in Lumbar Spine. *Asian Spine J.* 2009;3(1):16-20. doi:10.4184/asj.2009.3.1.16

34. Zec S, Soriani N, Comoretto R, Baldi I. High Agreement and High Prevalence: The Paradox of Cohen's Kappa. *Open Nurs J.* 2017;11:211-218. doi:10.2174/1874434601711010211

35. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol.* 1990;43(6):543-9. doi:10.1016/0895-4356(90)90158-l

36. Jensen RK, Jensen TS, Grøn S, Frafjord E, Bundgaard U, Damsgaard AL, et al. Prevalence of MRI findings in the cervical spine in patients with persistent neck pain based on quantification of narrative MRI reports. *Chiropr Man Therap.* 2019 Mar 6;27:13. doi: 10.1186/s12998-019-0233-3. PMID: 30873276; PMCID: PMC6402131.

37. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-74.

38. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. 2013;310(20):2191-4. doi:10.1001/jama.2013.281053

39. Harrison DE, Harrison DD, Cailliet R, Troyanovich SJ, Janik TJ, Holland B. Cobb method or Harrison posterior tangent method: which to choose for lateral cervical radiographic analysis. *Spine* (Phila Pa 1976). 2000;25(16):2072-8. doi:10.1097/00007632-200008150-00011

40. Chen C, Huang M, Han Z, Shao L, Xie Y, Wu J, et al. Quantitative T2 magnetic resonance imaging compared to morphological grading of the early cervical intervertebral disc degeneration: an evaluation approach in asymptomatic young adults. *PLoS One*. 2014 Feb 3;9(2):e87856. doi: 10.1371/journal.pone.0087856. PMID: 24498384; PMCID: PMC3912130.

41. Miyazaki M, Hong SW, Yoon SH, Morishita Y, Wang JC. Reliability of a magnetic resonance imaging-based grading system for cervical intervertebral disc degeneration. *J Spinal Disord Tech*. 2008;21(4):288-92. doi:10.1097/BSD.0b013e31813c0e59

42. Park HJ, Kim SS, Lee SY, Park NH, Chung EC, Rho MH, et al. A practical MRI grading system for cervical foraminal stenosis based on oblique sagittal images. *Br J Radiol.* 2013 May;86(1025):20120515. doi: 10.1259/bjr.20120515. Epub 2013 Feb 14. PMID: 23410800; PMCID: PMC3635796.

43. Kim W, Ahn KS, Kang CH, Kang WY, Yang KS. Comparison of MRI grading for cervical neural foraminal stenosis based on axial and oblique sagittal images: Concordance and reliability study. *Clin Imaging*. 2017;43:165-9. doi:10.1016/j.clinimag.2017.03.008

44. Vach W, Gerke O. Gwet's AC1 is not a substitute for Cohen's kappa - A comparison of basic properties. *MethodsX.* 2023;10:102212. doi:10.1016/j.mex.2023.102212

45. Bender YY, Diederichs G, Walter TC, Wagner M, Liebig T, Rickert M, et al. Differentiation of Osteophytes and Disc Herniations in Spinal Radiculopathy Using Susceptibility-Weighted Magnetic Resonance Imaging. *Invest Radiol.* 2017 Feb;52(2):75-80. doi: 10.1097/RLI.0000000000000314. PMID: 27548342.

46. Shur N, Corrigan A, Agrawal K, Desai A, Gnanasegaran G. Radiological and Radionuclide Imaging of Degenerative Disease of the Facet Joints. *Indian J Nucl Med*. 2015;30(3):191-8. doi:10.4103/0972-3919.15852.