# Revalidating Pfirrmann's Magnetic Resonance Image-Based Grading of Lumbar Nerve Root Compromise by Calculating Reliability among Orthopaedic Residents

Arun-Kumar Kaliya-Perumal, MS, Senthil-Kumar Ariputhiran-Tamilselvam, MS,
Chi-An Luo, MD*, Sivaharivelan Thiagarajan, MBBS, Udhayakumar Selvam, MBBS,
Raj-Prabhakar Sumathi-Edirolimanian, MBBS

*Department of Orthopaedic Surgery, Melmaruvathur Adhiparasakthi Institute of Medical Sciences and Research,
Affiliated to the Tamil Nadu Dr MGR Medical University, Tamil Nadu, India,
*Department of Orthopaedic Surgery, Chang Gung Memorial Hospital and Chang Gung University College of Medicine, Taoyuan, Taiwan*

**Background:** Intervertebral disc herniations lead to subsequent compromise of the nerve root. The root can either have a mere contact with the disc material or be pushed aside or compressed. This was earlier graded by Pfirrmann and colleagues. We intend to revalidate this grading system by performing a reliability analysis among orthopaedic residents.

**Methods:** Fifty axial cut magnetic resonance (MR) images of the affected lumbar disc level that belonged to different patients (age, 37.8 ± 10.4 years; 33 males and 17 females) were chosen and given to five orthopaedic residents for grading according to the Pfirrmann's MR image-based grading of lumbar nerve root compromise. Responses were received in the form of categorical variables and reliability was assessed.

**Results:** On doing percentage statistics, we found that 14 images had 100% agreement, 22 had 80% agreement and 14 had 60% agreement. We inferred an overall agreement of 80% ± 15.1%. In addition, interrater reliability was determined by calculating the Fleiss' kappa, which was found to be 0.521, signifying moderate agreement. Intrarater reliability was determined by calculating Cohen's kappa, which was found to be 0.696, signifying substantial agreement.

**Conclusions:** Our residents took only a short time to learn and reproduce this grading system as ratings that proved to be moderately reliable. Even though the value of kappa was slightly lower, reliability was similar to that of the original authors. We think that this grading system can be adopted in day-to-day practice by framing appropriate rules to interpret MR images where the nerve roots are not visible.

**Keywords:** *Intervertebral disc, Nerve root compression, Spinal stenosis, Radiculopathy, Reliability*

Nerve root compromise can be deemed as the direct cause of radicular pain even though the primary pathology is the intervertebral disc herniation. The disc herniation, being a mechanical lesion, tends to compromise the nerve root.[1] In addition, the inflammatory response to the exposed nucleus pulposus also contributes to radicular pain.[2-4] However, some disc herniations do not hinder the nerve root, especially if the herniation is focal and away from the root. Therefore, the amount of nerve root compromise caused

211

Kaliya-Perumal et al. Reliability of Pfirrmann's Grading
Clinics in Orthopedic Surgery • Vol. 10, No. 2, 2018 • www.ecios.org

by disc herniation may vary depending upon the location of herniation.

The nomenclature versions currently in practice for grading nerve root compromise include the grading systems of Pfirrmann et al.[5] and van Rijn et al.[6] These systems are well established, proved to be reliable, and periodically used by various authors. Due to its dichotomized nature, the van Rijn's system demonstrates higher reliability;[7] however, we preferred the Pfirrmann's grading for its itemized nature. The orthopaedic residents at our hospital were aware of this grading system. Even though they did not use it regularly, they were comfortable with the system as it was simple and facile. Hence, we decided to quantify the reliability of Pfirrmann's grading among orthopaedic residents at our institute.

## METHODS

Fifty consecutive magnetic resonance imaging (MRI) studies that belonged to patients who initially presented with lumbar radiculopathy and were diagnosed to have a single level lumbar disc herniation were selected. We evaluated all the sagittal and axial cut MR images of the affected disc level to select one axial cut MR image that portrays maximal herniation in each patient. Therefore, 50 axial cut MR images were selected, which included images depicting concomitant ligamentum flavum thickening and/or facet hypertrophy. We excluded images portraying far lateral disc herniations, spondylolisthesis, infections and neoplasms at the chosen level.

Pfirrmann's grading is based on the evaluation of an axial cut MR image at the level of maximal disc herniation to grade the unilateral traversing nerve root compromise due to the herniating disc. Pfirrmann et al.[5] graded the nerve root compromise into normal (grade 0), contact (grade 1), deviation (grade 2), and compression (grade 3). A calibrating session was held to brief the residents regarding this grading system using Pfirrmann's original work. A set of ten axial cut MR images other than the 50 selected images were chosen for discussion at the calibrating session.

Residents were clearly explained about the normal position of the nerve roots in an axial cut MR image. Concerns were raised about the nerve roots being nonvisible in certain MR images. In addition, there was also a concern about images portraying a broad-based disc herniation causing bilateral nerve root compromise where the contralateral nerve root cannot be used as a reference to differentiate contact (grade 1) or deviation (grade 2). In such circumstan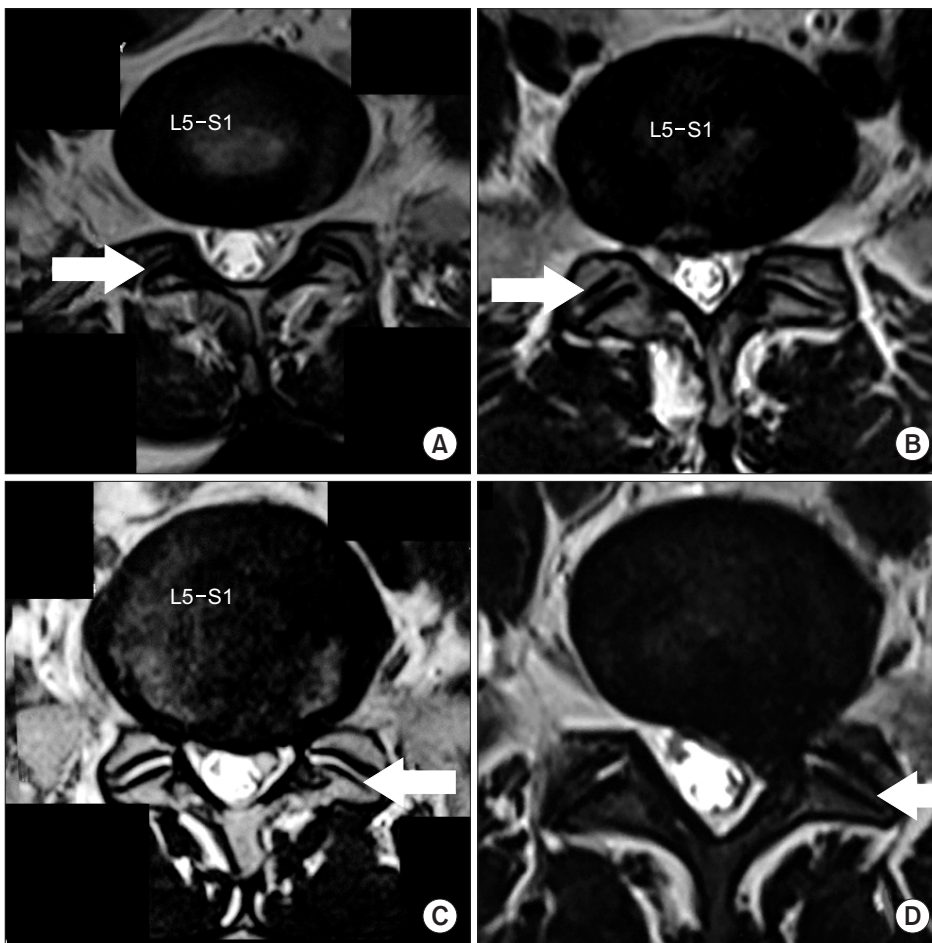ces, we instructed the residents to assume the position of the nerve root with their understanding about its normal location and the four described grades in the Pfirrmann's grading system.

Accordingly, each resident had to grade the 50 MR images for which, they were not put under any obligation regarding time, mainly to reduce instances of fatigue and to maintain the precision of their ratings. Once they were ready with their responses, they were asked to recheck under supervision. Their responses to each MR image as grade 0, 1, 2 and 3 were considered as categorical variables. Interrater reliability of the grading system was determined by calculating the Fleiss' kappa coefficient. The same MR images were shuffled and provided to one of the residents for reassessment after a month and the response was collected. This data was used to determine the intrarater reliability by calculating the Cohen's kappa coefficient.

Statistical analyses were performed using IBM SPSS ver. 23.0 (IBM Corp., Armonk, NY, USA) and Graph Pad Prism 5.0 (GraphPad Software Inc., San Diego, CA, USA). Reliability was assessed by calculating percentage agreement and also the kappa statistic. We interpreted the values of kappa as per the recommendation of Landis and Koch;[8] according to them, a kappa value of 0–0.20 indicates slight agreement, 0.21–0.40 indicates fair agreement, 0.41–0.60 indicates moderate agreement, 0.61–0.80 indicates substantial agreement, 0.81 or higher indicates excellent agreement, and 1.00 indicates absolute agreement.

**Table 1.** Demographics

| Variable | Value |
| --- | --- |
| Axial cut magnetic resonance image | 50 |
| Age (yr), mean ± standard deviation (range) | 37.8 ± 10.4 (22–65) |
| Sex (male:female) | 33 : 17 |
| Level | |
| L2–3 | 2 |
| L3–4 | 1 |
| L4–5 | 27 |
| L5–S1 | 20 |
| Image in each grade as per majority agreement | |
| Grade 0 | 8 |
| Grade 1 | 11 |
| Grade 2 | 15 |
| Grade 3 | 16 |

212

Kaliya-Perumal et al. Reliability of Pfirrmann's Grading
Clinics in Orthopedic Surgery • Vol. 10, No. 2, 2018 • www.ecios.org

**Fig. 1.** Axial cut magnetic resonance images at the affected lumbar disc levels of different patients (arrows indicate the side which was graded). (A) Grade 0 (normal), when there is no contact of disc material with the nerve root. (B) Grade 1 (contact), when there is contact of disc material with the nerve root. (C) Grade 2 (deviation), when there is deviation of the nerve root dorsally. (D) Grade 3 (compression), when the nerve root is compressed between the spinal canal wall and the herniated disc.

Informed consent was obtained from all patients to use their MR images for research purposes without revealing identity and the study was performed abiding by all ethical considerations.

## RESULTS

Consecutive MRI studies of different patients (n = 50; male, 33; female, 17; age, 37.8 ± 10.4 years) were selected. The selected images were predominantly the axial cuts of the affected L4–5 and L5–S1 disc levels depicting maximal herniation; apart from which, there was only one L3–4 and two L2–3 disc levels chosen (Table 1). Images were distributed to five orthopaedic residents and their grading for each image was received within a week. The grade on which the majority of the residents had agreement for a particular image was considered as the actual grade. In this way, our selected images included all four grades of nerve root compromise as described in the Pfirrmann's grading system (Fig. 1).

The percentage of agreement for a grade that was

given by the majority to each image was calculated. This data was used to calculate the overall percentage of agreement (80% ± 15.1%) and grade-wise percentage of agreement (Table 2). Only for 14 of the selected images, 100% agreement was obtained: eight belonged to grade 3, two belonged to grade 2, three belonged to grade 1, and one belonged to grade 0. Among the remaining images, 22 had 80 % agreement and 14 had 60% agreement.

In order to precisely analyse the reliability of this grading system, we calculated the Cohen's kappa between all pairs of residents. This data was used to form the interrater kappa matrix (Table 3). However, the statistical measure for assessing the reliability of agreement between multiple raters is the Fleiss' kappa. We inferred a Fleiss' kappa value of 0.521, which signifies moderate reliability according to the interpretation of kappa by Landis and Koch (Table 4).[8] Even though our kappa coefficient is lower and does not replicate Pfirrmann's original work, the reliability still remains moderate as inferred by Pfirrmann et al.[5]

Intrarater reliability was calculated using the reas-

213

Kaliya-Perumal et al. Reliability of Pfirrmann's Grading
Clinics in Orthopedic Surgery • Vol. 10, No. 2, 2018 • www.ecios.org

sessment data of one of the residents. The chosen resident was consistent with his previous rating for 39 of the images; hence, there was 78% agreement. We calculated the Cohen's kappa between the chosen resident's previous and latest rating; it was found to be 0.696, which signifies substantial agreement according to the interpretation of kappa by Landis and Koch.[8] Therefore, we inferred a moderate interrater and substantial intrarater reliability for the Pfirrmann's MR image-based grading of lumbar nerve root compromise.

## DISCUSSION

Lumbar radiculopathy is the predominant presentation of disc-induced nerve root compromise. A bulging or herniating disc can remain asymptomatic;[9,10] however, radicular

pain may develop if there is any hindrance to the nerve root.[11] Hindrance to the nerve root happens when there is violation of the nerve root area by the displaced disc. The nerve root can either have a mere contact with the disc material or be pushed aside or compressed. This is the baseline of Pfirrmann's MR image-based grading of nerve root compromise. In addition, understanding the normal location of the traversing nerve roots in an axial cut lumbar disc level MR image is a prerequisite to use this grading system; however, far lateral disc herniations affecting the exiting root are not considered.

Despite not regularly used, Pfirrmann's grading is periodically utilized.[12-14] The system incorporates four grades: grade 0 is considered as normal even though there can be a focal disc herniation that does not hinder the nerve root; grade 1 (contact) is when the nerve root is in visible contact with the bulging or herniating disc material but remains in its normal location; grade 2 (deviation) is where the nerve root is displaced dorsally; and grade 3 (compression) is when the nerve root is compressed between the disc material and the wall of the spinal canal. Detailed descriptions of the grades are available in Pfirrmann et al.'s original work.[5]

It is necessary that the nerve roots be clearly visible in the axial cut MR image; however, it may not be possible all the time. This was discussed in our calibrating session as the residents raised this issue looking at some of the axial cut MR images in which the nerve roots were not clearly visible. Another query that was raised by the residents

**Table 2.** Percentage of Agreement

| Grade | Agreement (%) |
|---|---|
| Grade 0 | 75 ± 14.1 (60–100) |
| Grade 1 | 81.8 ± 14 (60–100) |
| Grade 2 | 76 ± 13.5 (60–100) |
| Grade 3 | 85 ± 17.1 (60–100) |
| Overall | 80 ± 15.1 (60–100) |

Values are presented as mean ± standard deviation (range).

**Table 3.** Interrater Kappa Matrix

| Variable | Resident 1 | Resident 2 | Resident 3 | Resident 4 | Resident 5 |
|---|---|---|---|---|---|
| Resident 1 | 1 | 0.614 | 0.454 | 0.425 | 0.587 |
| Resident 2 | 0.614 | 1 | 0.505 | 0.326 | 0.534 |
| Resident 3 | 0.454 | 0.505 | 1 | 0.483 | 0.650 |
| Resident 4 | 0.425 | 0.326 | 0.483 | 1 | 0.651 |
| Resident 5 | 0.587 | 0.534 | 0.650 | 0.651 | 1 |

Interpretation of kappa: 0–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; ≥ 0.81, excellent agreement; and 1.00, absolute agreement.

**Table 4.** Reliability

| Type of reliability | Kappa measure | Value of kappa | Inference |
|---|---|---|---|
| Interrater | Fleiss' kappa | 0.521 | Moderate reliability |
| Intrarater | Cohen's kappa | 0.696 | Substantial reliability |

214

Kaliya-Perumal et al. Reliability of Pfirrmann's Grading
Clinics in Orthopedic Surgery • Vol. 10, No. 2, 2018 • www.ecios.org

during the calibrating session was about broad-based disc herniations where there could be bilateral nerve root compromise. The nerve root on one side cannot be taken as a reference to differentiate grade 1 (contact) or grade 2 (deviation) on the other side. In both situations, residents were instructed to assume the location of the nerve root. This assumption is only subjective and could have biased our results.

Once responses were received, reliability was calculated using percentage statistics and kappa statistic. After proposal of this grading system by Pfirrmann et al.[5] in 2004, its reliability was rechecked by Lurie et al.[15] in 2008. Pfirrmann et al. reported an interobserver kappa of 0.62–0.67; however, Lurie et al. inferred a comparatively low interobserver kappa of only 0.47. Even so, both kappa values can be interpreted as moderate reliability as per the kappa interpretation of Landis and Koch.[8] Similarly, our interobserver Fleiss' kappa value was 0.521 signifying moderate reliability. In addition to interrater reliability, we calculated the intrarater reliability for one of the residents and inferred substantial agreement. This proves that the grading remains consistent.

It should be known that the raters who gave their responses were junior orthopaedic residents and not specialists in this field. Their individual understanding of this grading system may vary. Apart from this, spinal canal stenosis due to hypertrophied facets or a thickened ligamentum flavum could have mislead the assumption of the probable location of a compromised nerve root whenever it was not visible. These factors could have influenced our results; however, if appropriate rules are framed to interpret such MR images, the reliability of this grading system will grow higher and adopting it in day-to-day practice will become feasible.

In conclusion, 50 axial cut MR images at the affected lumbar disc levels were chosen and given to five orthopaedic residents for grading according to the Pfirrmann's MR image-based grading of lumbar nerve root compromise. Responses were received in the form of categorical variables and reliability analysis was done. We inferred moderate interrater and substantial intrarater reliability for this grading system. Moreover, our residents took only a short time to learn and reproduce this grading system as ratings that proved to be reliable and consistent. Therefore, our results prove that the Pfirrmann's MR image-based grading of lumbar nerve root compromise is a valid measure of the radiological severity of nerve root compromise due to the herniated intervertebral disc.

## CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

## ACKNOWLEDGEMENTS

## REFERENCES

1. Takahashi N, Yabuki S, Aoki Y, Kikuchi S. Pathomechanisms of nerve root injury caused by disc herniation: an experimental study of mechanical compression and chemical irritation. Spine (Phila Pa 1976). 2003;28(5):435-41.

2. Goupille P, Mulleman D, Valat JP. Radiculopathy associated with disc herniation. Ann Rheum Dis. 2006;65(2):141-3.

3. Rothman SM, Winkelstein BA. Chemical and mechanical nerve root insults induce differential behavioral sensitivity and glial activation that are enhanced in combination. Brain Res. 2007;1181:30-43.

4. Zhang S, Nicholson KJ, Smith JR, Gilliland TM, Syre PP, Winkelstein BA. The roles of mechanical compression and chemical irritation in regulating spinal neuronal signaling in painful cervical nerve root injury. Stapp Car Crash J. 2013;57:219-42.

5. 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.

6. van Rijn JC, Klemetso N, Reitsma JB, et al. Observer variation in MRI evaluation of patients suspected of lumbar disk herniation. AJR Am J Roentgenol. 2005;184(1):299-303.

7. Li Y, Fredrickson V, Resnick DK. How should we grade lumbar disc herniation and nerve root compression? A systematic review. Clin Orthop Relat Res. 2015;473(6):1896-902.

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

9. Brinjikji W, Luetmer PH, Comstock B, et al. Systematic literature review of imaging features of spinal degeneration

215

Kaliya-Perumal et al. Reliability of Pfirrmann's Grading
Clinics in Orthopedic Surgery • Vol. 10, No. 2, 2018 • www.ecios.org

in asymptomatic populations. AJNR Am J Neuroradiol. 2015;36(4):811-6.

10. Jensen MC, Brant-Zawadzki MN, Obuchowski N, Modic MT, Malkasian D, Ross JS. Magnetic resonance imaging of the lumbar spine in people without back pain. N Engl J Med. 1994;331(2):69-73.

11. Janardhana AP, Rajagopal, Rao S, Kamath A. Correlation between clinical features and magnetic resonance imaging findings in lumbar disc prolapse. Indian J Orthop. 2010;44(3):263-9.

12. Kaliya-Perumal AK, Yeh YC, Luo CA, Joey-Tan KY. Assessment of anteroposterior subpedicular approach and oblique scotty dog subpedicular approach for selective nerve root block. Clin Orthop Surg. 2017;9(1):71-6.

13. Salama AA, Alarabawy RA, Dawoud MM, Zayed HA, Soliman A, El-Tantawy A. Functional disability of occupational-related lumbar disc degeneration: evaluation by magnetic resonance imaging with surgical correlation. Egypt J Radiol Nucl Med. 2017;48(1):189-99.

14. Sung J, Jee WH, Jung JY, et al. Diagnosis of nerve root compromise of the lumbar spine: evaluation of the performance of three-dimensional isotropic T2-weighted turbo spin-echo SPACE sequence at 3T. Korean J Radiol. 2017;18(1):249-59.

15. Lurie JD, Tosteson AN, Tosteson TD, et al. Reliability of magnetic resonance imaging readings for lumbar disc herniation in the Spine Patient Outcomes Research Trial (SPORT). Spine (Phila Pa 1976). 2008;33(9):991-8.