

# Comparison of SDSG and CARDS classifications for L5/S1 lumbar degenerative spondylolisthesis: An independent inter- and intra-observer agreement study

Chongqing Xu (✉ [809663990@qq.com](mailto:809663990@qq.com))

Long Hua Hospital <https://orcid.org/0000-0001-9654-793X>

Mengchen Yin

Long Hua Hospital

Wen Mo

Long Hua Hospital

---

## Research article

**Keywords:** lumbar degenerative spondylolisthesis, SDSG classification, CARDS classification. reliability, agreement study

**DOI:** <https://doi.org/10.21203/rs.3.rs-107455/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

## Background

Lumbar degenerative spondylolisthesis (DS) has been a common disease that make increasing patients suffer from different degrees of low back pain and radicular symptoms. SDSG and CARDS systems are commonly used to classify the disease, and help to make a more detailed treatment plan. The objective of this study is to evaluate the reliability and reproducibility of SDSG and CARDS classifications, and to explore their clinical application value.

## Methods/Design:

All 117 patients with L5/S1 lumbar DS were enrolled. 5 spine surgeons who have certain clinical experience were selected. They determined the classification according to S SDSG and CARDS systems, we used kappa (K) value to check the coefficient consistency for multi-factor and assess the inter- and intra-observer agreement. After 12 weeks, we repeated the analysis.

## Results

The inter-observer reliability and intra-observer reproducibility of SDSG system were substantial with K values of 0.704 and 0.861, while those of CARDS system were substantial with values of 0.620 and 0.878.

## Conclusion

SDSG system has better inter-observer reliability in comparison with CARDS system, and though CARDS system is more intuitive and simpler, it is more likely to produce deviations when using it. Since both SDSG and CARDS systems show substantial agreement and have great significance in surgical strategy of L5/S1 lumbar DS, they can be widely used in clinical practice.

## Background

Lumbar degenerative spondylolisthesis (DS) is defined as anterior displacement of one vertebra over the subjacent vertebra caused by degenerative changes, without an associated disruption of defect in the vertebral ring, which is mainly manifested as lumbosacral pain, sciatic nerve involvement and intermittent claudication. It have been a common disease that mostly occurs in L4/5 and L5/S1 segments. Increasing patients are suffering from different degrees of low back pain and radicular symptoms.[1] However, the pathogenesis, symptoms and imaging manifestations often differentiates in individuals, so it still remains controversial about the choice of treatment strategies.[2, 3] In the past decades, relevant classification of the disease have emerged in an endless stream. Previously, lumbar DS was classified

according to etiology and slip grade, which provided limited clinical value in guiding surgical treatment since the degree of slip rarely exceeds 30%.[4, 5] In addition, classification such as Meyerding system[6] did not consider morphological parameters related to clinical outcomes, for example, disc height or spinopelvic balance. Thus, an appropriate classification of lumbar DS is essential, for it plays a crucial role in determining treatment strategy, especially as a guide for surgical decision making.

With the further study of spine biomechanics and sagittal balance, the understanding of lumbar DS pathogenic factors and natural history have been increasingly comprehensive. At present, the correlation between pelvic incidence (PI) and morbidity of lumbar DS, as well as between spine sagittal balance and progression of lumbar DS have been clarified.[7]

Based on radiographic measurement of slip grade, PI, sacral slope (SS), pelvic tilt (PT) and spinopelvic balance, the Spinal Deformity Study Group (SDSG) developed a classification of L5/S1 lumbosacral spondylolisthesis which included 6 types in 2011,[8] and suggested that for patients with balanced pelvis and spine, fusion can be performed either in situ or in reduction and fixation, while for those with unbalanced pelvis or spine, reduction should be emphasized in order to restore sagittal balance and provide a better biomechanical environment for fusion (Fig. 1).

In 2014, Kepler et al.[9] proposed the Clinical and Radiographic Degenerative Spondylolisthesis classification (CARDS) on the basis of disc space height, sagittal vertebral translation, kyphotic alignment, and lower extremity pain, which included 4 main categories and 12 subtypes. This classification takes both radiographic parameters and clinical manifestations into consideration, so as to provide a more comprehensive evaluation for surgical treatment (Fig. 2).

However, any classification being widely used in clinical evaluation and treatment strategy decision-making should allow communication and easier consultation among specialists and need multiple validations. The purpose of this study is to compare the inter- and intra-observer agreement of the mentioned two systems for classifying L5-S1 lumbar DS, and to analyze their clinical application value.

## **1. Materials And Methods**

### **1.1 Patient Case Selection and Evaluation**

We performed this study in accordance with the principles of the Declaration of Helsinki, and obtained institutional review board approval from our ethics committee. Database records of patients with L5/S1 lumbar DS treated in our hospital were retrospectively collected and analyzed between January 1, 2017, to December 31, 2019. Patients included in the study should have performed posteroanterior and lateral standing radiographs of the entire spine and pelvis showing both femoral heads (including flexion and extension lumbar position). Exclusion criteria were patients with a history or clinical signs of hip, pelvic, or lower limb disorders, and incomplete clinical data or imaging studies. All subjects were required to have complete and available clinical data including demographic characteristics, chief complaint, neurological

function, complications and treatment history. Finally, a total of 117 consecutive patients with L5-S1 lumbar DS were involved in the present study.

2 physicians who collected the cases and treated these patients did not participate in the later statistics and analysis. Another 5 spine surgeons volunteered to be the observers as they were unaware of the patients' identification, treatment history, and original classification used in clinical care. To perform an adequate reliable study, we provided each evaluator essential original literature and pertinent information of cases including leg pain symptoms to assess those cases based on the two classification systems. Face-to-face meetings and evaluation sessions were performed before the agreement study and through which any controversies about the two classifications were discussed until all the observers came to a consensus. Standard imaging reports were available to observers as reference. Each observer used IDC Cygnus Version 1.2 (DICOM image viewing software) for calculating parameters which were needed to classify spondylolisthesis. According to the mentioned classifications, observers respectively assigned each case with a SDSG type and a CARDS type (e.g. SDSG: Type 1; CARDS: A).

We assessed inter-observer reliability by comparing the initial responses of the 5 observers. The intra-observer reproducibility was determined through comparing the two responses to the same case of each observer, which were separated by a 12-week interval, and all cases in the first and second acquisition sessions were displayed randomly so as to minimize the recall bias.

## 1.2 Statistical analysis

All data analyses were performed using Stata Version 16.0 (software for statistics and data science). Considering that the classifications of SDSG and CARDS systems belonged to ordinal data, we used Fleiss's kappa (K) coefficient as well as percentage of agreement to assess inter-observer reliability, while intra-observer reproducibility was evaluated based on the first and second acquisition session for each observer by Cohen's K coefficient and percentage of agreement.[10,11] The K values were expressed with a 95% confidence interval (CI), and the range of the coefficient is between (- 1, 1). Higher values signified better agreement. According to the study by Landis and Koch,[12] levels of agreement for K were divided into five grades, with  $\kappa$  values 0.00 to 0.20 considered 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, near perfect agreement.

## 2 Result

This study totally involved 117 consecutive patients, including 45 males and 72 females, with an average age of 66.3 years (range from 52 to 84 years). All 6 types of SDSG system and 12 types of CARDS system were obtained within those individuals. There were 1170 evaluations made by 5 observers using SDSG classification in 2 assessments ( $117 \times 5 \times 2$ ), including Type 1 (17.1%), Type 2 (30.8%), Type 3 (22.2%), Type 4 (12.8%), Type 5 (10.3%), and Type 6 (6.8%). Similarly, 1170 evaluations of CARDS classification were assigned to Type A (15.4%), B (42.7%), C (36.8%), D (5.1%).

## 2.1 Inter-observer Reliability

According to the reliability analysis, the overall inter-observer agreement of SDSG classification was substantial at 79.4% (74.4%-85.5%) with an associated K value of 0.704 (0.655–0.769) (Table 1), of which 58 (49.6%) cases showed perfect agreement among all 5 observers in the first assessment, while 61 (52.1%) cases in the second assessment. In addition, at least 3 observers showed agreement on 101 (86.3%) cases in the first assessment and 103 (88.0%) cases in the second assessment (Table 2). Besides, the inter-observer agreement of slip grade (low-grade vs. high-grade slips) was near perfect at 89.2% (85.1%-91.7%), with a k value of 0.813 (Table 3).

Table 1  
Inter-observer Reliability of SDSG Classification

Observers*	Cases in Agreement Between Observers							Agreement(%)	K
	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Total		
1-2	15	27	19	11	9	6	87	74.4	0.655
1-3	16	30	19	12	12	6	95	81.2	0.728
1-4	15	28	21	10	11	5	90	76.9	0.671
1-5	17	30	18	12	11	6	94	80.3	0.713
2-3	14	29	21	13	12	6	95	81.2	0.727
2-4	18	32	20	11	13	6	100	85.5	0.769
2-5	15	29	20	11	11	6	92	78.6	0.694
3-4	15	28	19	12	12	6	92	78.6	0.693
3-5	16	28	20	11	12	6	93	79.5	0.707
4-5	14	29	19	12	11	6	91	77.8	0.682
overall	-	-	-	-	-	-	-	79.4	0.704

\*1, 2, 3, 4, 5 represent the 5 observers who participated in the study.

Table 2  
Agreement of SDSG Classification Among Five Observers

	The First Assessment	The Second Assessment
<b>Agreement among all 5 observers</b>	58 (49.6%)	61 (52.1%)
<b>Agreement among at least 3 observers</b>	101 (86.3%)	103 (88.0%)

Table 3  
Agreement Analysis by Slip Grade of SDSG Classification

Slip Grade	Inter-observer Reliability		Intra-observer Reproducibility	
	Agreement (%)	K	Agreement (%)	K
All cases	89.2%	0.813	90.4%	0.873
Low-grade slips	85.3%	0.729	89.5%	0.852
High-grade slips	91.2%	0.864	92.7%	0.911

The inter-observer reliability of CARDS classification was substantial at 72.6% (68.4%-79.5%) with an associated K value of 0.620 (0.572–0.701) (Table 4), of which 50 (42.7%) cases showed perfect agreement among all 5 observers in the first assessment, while 49 (41.9%) cases in the second assessment. Nevertheless, at least 3 observers showed agreement on 89 (76.1%) cases in the first assessment and 86 (73.5%) cases in the second assessment (Table 5). In addition, the K values of 3 items of CARDS system: disc space height, sagittal vertebral translation and kyphotic alignment were 0.618, 0.477, 0.725, respectively (Table 6). There were significant differences between all the above results ( $p < 0.01$ ).

Table 4  
Inter-observer Reliability of CARDS Classification

Observers*	Cases in Agreement Between Observers					Agreement(%)	K
	Type A	Type B	Type C	Type D	Total		
1-2	14	38	32	5	89	76.1	0.663
1-3	13	34	30	6	83	70.9	0.597
1-4	14	39	33	7	93	79.5	0.701
1-5	12	33	30	5	80	68.4	0.572
2-3	11	34	31	6	82	70.1	0.589
2-4	12	34	33	5	84	71.8	0.610
2-5	14	32	32	5	83	70.9	0.596
3-4	13	37	33	7	90	76.9	0.673
3-5	12	34	33	6	85	72.6	0.628
4-5	12	33	30	5	80	68.4	0.573
overall	-	-	-	-	-	72.6	0.620

\*1, 2, 3, 4, 5 represent the 5 observers who participated in the study.

Table 5  
Agreement of CARDS Classification Among Five Observers

	The First Assessment	The Second Assessment
Agreement among all 5 observers	50 (42.7%)	49 (41.9%)
Agreement among at least 3 observers	89 (76.1%)	86 (73.5%)

Table 6  
Agreement Analysis by Three Items of CARDS Classification

Items	Inter-observer Reliability		Intra-observer Reproducibility	
	Agreement (%)	K	Agreement (%)	K
disc space height	76.5%	0.618	90.2%	0.762
vertebral translation	69.2%	0.477	87.6%	0.749
kyphotic alignment	88.4%	0.725	94.7%	0.855

## 2.2 Intra-observer Reproducibility

Reproducibility analysis of the same observer's results after 12 months showed that the intra-observer agreement of SDSG classification was near perfect at 88.2% (84.6%-92.3%) with an average K value of 0.861 (0.823–0.906) (Table 7). The intra-observer agreement of slip grade (low-grade vs. high-grade slips) was near perfect at 90.4% (88.0%-93.7%), with a k value of 0.875 (Table 3).

Table 7  
Intra-observer Reproducibility of SDSG Classification

Observers*	Cases in Agreement Between First and Second Assessment							Agreement(%)	K
	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Total		
1	18	31	22	14	11	7	103	88.0	0.859
2	17	32	22	14	13	7	105	89.7	0.871
3	16	30	23	11	13	6	99	84.6	0.823
4	18	31	20	13	12	7	101	86.3	0.844
5	17	32	23	15	13	8	108	92.3	0.906
overall	-	-	-	-	-	-	-	88.2	0.861

\*1, 2, 3, 4, 5 represent the 5 observers who participated in the study.

The intra-observer reproducibility of CARDS classification ranged from 87.2–94.0% with an average percentage of 90.4%, and the K value was 0.878 (0.835–0.917), which were considered near perfect agreement (Table 8). The K values of 3 items were 0.762, 0.749, 0.855, respectively (Table 6). There were also significant differences between these results ( $p < 0.01$ ).

Table 8  
Intra-observer Reproducibility of CARDS Classification

Observers*	Cases in Agreement Between First and Second Assessment					Agreement(%)	K
	Type A	Type B	Type C	Type D	Total		
1	18	45	34	7	104	88.9	0.862
2	16	44	35	7	102	87.2	0.835
3	19	46	35	8	108	92.3	0.904
4	20	50	33	7	110	94.0	0.917
5	18	47	32	8	105	89.7	0.874
overall	-	-	-	-	-	90.4	0.878

\*1, 2, 3, 4, 5 represent the 5 observers who participated in the study.

### 3. Discussion

Since the incidence of lumbar DS has significantly increased in recent years, classification system and decision-making of the disease have been further improved, providing important clinical value for standardized treatment. With the increasing higher requirements for life quality of patients, surgery has become the major treatment method nowadays. However, it's quite difficult to perform surgeries due to the particularity and complexity of lumbar anatomy.

Various classification systems emerge in an endless stream. At present, the simplest classification of lumbar DS is Meyerding system,[6] which is to grade according to vertebral translation. It should be mentioned that there is not a certain correlation between spondylolisthesis degree and clinical symptoms, and even two cases with the same Meyerding grade may not be homogeneous in nature of disease. Therefore, it can not accurately describe the state and judge the severity of spondylolisthesis to further guide treatment and predict prognosis. Other traditional classification of lumbar DS mainly includes Wiltse and Marchetti classification.[13, 14, 15] Wiltse et al.[13] classified lumbar spondylolisthesis into 6 types based on anatomical and pathological characteristics of imaging, while Marchetti classification[14] is established according to different etiological characteristics of spondylolisthesis, which is more practical. However, these classifications still have the significant defects that they have no quantitative indexes and can not determine the degree of spondylolisthesis, which makes them difficult to be



evaluated and inferior in reproducibility. Thus it is hard to judge the disease process and whether surgical treatment is needed.

The above-mentioned classifications all emphasize on characteristics of slipped vertebrae or bony structures, without considering disc degeneration, spinal-pelvic sagittal balance and clinical symptoms which are regarded as key factors to judge whether lumbar DS will progress.[16, 17, 18] Therefore, none of the above classifications can comprehensively, accurately and objectively reflect the characteristics of different lumbar DS types, and homogeneity comparison can not be performed between them, which leads to great controversy on the clinical treatment standards and prognosis prediction of the disease.

SDSG classification gives spine surgeons a clear definition of spinal-pelvic sagittal balance, and helps them provide targeted treatment for patients.[19, 20] There is always a dispute about whether severe spondylolisthesis needs reduction. According to current study of biomechanics, combined with SDSG classification, specialists have reached a consensus that for patients with imbalanced spine or pelvis, reduction should be emphasized to correct the imbalance as well as the external deformities, and provide a more favorable biomechanical environment for bone graft. For patients with balanced pelvis and spine, either fusion in situ or fusion with reduction and fixation can be used. Our results show that the inter- and intra-observer agreement K value of SDSG classification are 0.704 and 0.861, respectively, which are slightly higher than that of the previous agreement studies by Mac-Thiong et al.[21] (0.65, 0.74) and Bao et al.[22] (0.648, 0.830), indicating relatively better consistency strength. In these studies, the case scope of assessment of SDSG classification covered dysplastic, degenerative and isthmic spondylolisthesis. However, to formulate a case inclusion criterion applicable to both classifications, the cases involved in our study were limited as L5/S1 lumbar DS, while narrower and more specific scope often leads to greater reliability, which may be one of the factors that caused the differences in the results. Moreover, it is worth mentioning that our research has even better reliability of intra-observer reproducibility than those previous to us, for there is an only 1-day or 2-weeks interval between the 2 acquisition sessions in those studies, while too short interval will make observers in the second assessment tend to evaluate according to their recollections of the first assessment, and thus may reduce the reliability of results. The 12-weeks interval in our study may be a more appropriate choice. In addition, we analyzed the agreement of slip grade, and the results show that both inter- and intra-observer K values are high (0.813, 0.875), which may be largely attributed to the accurate measurement of computer-assisted technique, and through which can be seen that the slip grade is not the key factor for the deviation of classifying results between observers. Therefore, we believe that low resolution and clarity of radiographs, and serious osteoporosis of elderly patients make it difficult to judge the bone structure and anatomic landmark, which leads to deviations of sagittal parameter measurements.

As a more recent established classification, CARDS system can provide an ideal treatment plan for patients. For those without clinical symptoms (Type A0), conservative treatment is recommended.[18] On the aspect of surgery, simple decompression can be performed on Type A1 and A2 patients,[23] while internal fixation and fusion is practical in Type B or C patients. For cases of Type D, internal fixation is needed to correct kyphosis deformity and interbody fusion cage is needed to reconstruct anterior column

support, so that physiological lumbar lordosis and fusion rate can be improved as much as possible.[24] Whether the leg pain exists or not is regarded as the clinical index for subtypes, which is also helpful to guide surgical plan. A study published in recent years has confirmed that patients with leg pain as the main symptom before surgery have better postoperative effect than those with back pain as the main symptom.[25] Our results show that the inter- and intra-observer agreement K value of CARDS classification are 0.620 and 0.878. Compare with that of the previous study by Kepler et al.[9] (0.82, 0.83) and Kong et al.[26] (0.837, 0.869), the inter-observer reliability is lower and the intra-observer reproducibility is similar. Since CARDS system was initially proposed based on study subjects with L4/5 DS, however, our study applied this classification to cases with L5/S1 lumbar DS, thus may lead to difference between the results. Another reason may explain this is the relatively Larger sample size, which may reduce the inter-observer deviation and make the result more accurate. In addition, we found that no matter in inter- or intra-observer agreement test, the K values of sagittal vertebral translation are lower than those of disc space and kyphotic alignment (since this study belongs to retrospective analysis, the clinical data is mainly based on the medical records, there is no difference in the judgment of leg pain, thus agreement evaluation of this item is unnecessary), which merely indicates “moderate” agreement. Firstly, the classification requires that any translation longer than 5 mm in neutral, flexion or extension lateral radiographs should be classified as Type C, while in the actual process, observers may have certain marking or measuring deviations. Furthermore, with the multiple measurements, sometimes observers judging by subjective impression is also a factor, which leads to the relatively low agreement.

Both classifications have substantial inter- and intra-observer agreement, while SDSG classification has better inter-observer reliability in comparison with CARDS classification. With regard to sagittal balance parameters, SDSG classification can provide better reference value for surgical strategy. Nevertheless, it doesn't consider the changes on flexion and extension lateral radiograph, and the evaluation of lumbar instability is insufficient, which reduce the guiding value of surgical treatment, and that is the deficiency of SDSG classification. Since the clinical symptoms are often the reasons for DS patients to see a doctor, CARDS classification takes leg pain into account, which makes evaluation of scientific and clinical study more convenient, and that is the advantage of CARDS classification. According to the above, CARDS classification is more intuitive and simpler than SDSG classification. However, it is more likely to produce deviations when using CARDS classification.

The current study has several limitations. Firstly, is the relatively small sample size. Expanding the sample population to include non-operative patients of a wider population, allowing for more meaningful statistical testing on the reliability and reproducibility of these parameters. Secondly, the relatively low resolution and clarity of radiographs. We believe that it may be more accurate in the practical application to observe high resolution radiograph combining with computed tomography (CT) sagittal reconstruction image. Finally, only L5/S1 single segment DS patients were included in this study, and the agreement of two classifications in other segments and backward slipped DS cases were not discussed. Therefore, in future clinical work, high-quality, multicenter, large sample and wide case scope studies should be conducted to provide spine surgeons with the best evidence-based information.

# Conclusion

SDSG system has better inter-observer reliability in comparison with CARDS system, and though CARDS system is more intuitive and simpler, it is more likely to produce deviations when using it. Since both SDSG and CARDS systems show substantial agreement and have great significance in surgical strategy of L5/S1 lumbar DS, they can be widely used in clinical practice. However, we still need more higher-quality, larger samples, and multicenter prospective studies in future work to evaluate whether these classification systems allow better decision-making or prognosis-prediction in individual patients.

# Abbreviations

DS (degenerative spondylolisthesis)

PI (pelvic incidence)

SS (sacral slope)

PT (pelvic tilt)

SDSG (Spinal Deformity Study Group)

CARDS (Clinical and Radiographic Degenerative Spondylolisthesis)

K (kappa)

CI (confidence interval)

CT (computed tomography)

# Declarations

## Ethics approval and consent to participate

The case was reviewed by the Longhua Hospital Ethics Committee and ethical approval was waived as written consent was obtained from the patient.

## Consent for publication

Written patient consent was obtained for publication of all aspects of the case including personal and clinical details and images, which may compromise anonymity.

## Availability of data and material

All supporting data can be provided upon request to the authors.

## Competing interests

All authors read and approved the final manuscript and declare that they have no competing interests.

## Funding

No funding was obtained for this study.

## Authors' contributions

XCQ and YMC are co-first authors of this manuscript. XCQ designed the study and collected the data. YMC did the data analysis. XCQ wrote the manuscript. MW revised the manuscript and decided to submit the manuscript for publication. All authors read and approved the final manuscript.

## Acknowledgements

I want to thank my love, Du Ying, no matter how difficult it is, she never gives up, always cares for me, silently supports me, and gives me courage when I lose confidence. Without her help, understanding, tolerance and support, I believe that the life of a PhD in these three years will be very different.

## References

1. Herman MJ, Pizzutillo PD, Cavalier R. Spondylolysis and spondylolisthesis in the child and adolescent athlete. *Orthop Clin N Am*. 2003;34:461–467.
2. Iguchi T, Wakami T, Kurihara A, et al. Lumbar multilevel degenerative spondylolisthesis: radiological evaluation and factors related to anterolisthesis and retrolisthesis. *J Spinal Disord Tech*. 2002;15(2):93-99.
3. Jacobsen S, Sonne-Holm S, Røvsing H, et al. Degenerative lumbar spondylolisthesis: an epidemiological perspective: the Copenhagen osteoarthritis study. *Spine (Phila Pa 1976)*. 2007;32(1):120-125.
4. Steiger F, Becker HJ, Standaert CJ, et al. Surgery in lumbar degenerative spondylolisthesis: indications, outcomes and complications. A systematic review. *Eur Spine J*. 2014;23:945–973.
5. Matsunaga S, Sakou T, Morizono Y, et al. Natural history of degenerative spondylolisthesis. Pathogenesis and natural course of the slippage. *Spine*. 1990;15:1204–1210.
6. Meyerding HW. Spondylolisthesis. *Surg Gynecol Obstet*. 1932;54:371-377.
7. Labelle H, Roussouly P, Berthonnaud E, et al. Spondylolisthesis, pelvic incidence, and spinopelvic balance: a correlation study. *Spine (Phila Pa 1976)*. 2004;29(18): 2049-2054.
8. Labelle H, Mac-Thiong JM, Roussouly P. Spino-pelvic sagittal balance of spondylolisthesis: a review and classification. *Eur Spine J*. 2011;20(suppl 5):641–646.
9. Kepler CK, Hilibrand AS, Sayadipour A, et al. Clinical and radiographic degenerative spondylolisthesis (CARDS) classification. *Spine J*. 2015;15:1804-1811.

10. Fleiss J. Measuring nominal scale agreement among many raters. *Psycho Bull.* 1971;76:378–381.
11. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas.* 1960;20:37–46.
12. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159–174.
13. Wiltse LL, Newman PH, Macnab I. Classification of spondylolysis and spondylolisthesis. *Clin Orthop Relat Res.* 1976;117:23-29.
14. Marchetti PG, Bartolozzi P. Classification of spondylolisthesis as a guideline for treatment//Bridwell K, DeWald R. *The textbook of spinal surgery.* 2nd ed. Philadelphia: Lippincott-Raven. 1997:1211-1254.
15. Hammerberg KW. New concepts on the pathogenesis and classification of spondylolisthesis. *Spine (Phila Pa 1976).* 2005;30: S4-11.
16. Lamartina C, Zavatsky JM, Petruzzi M, et al. Novel concepts in the evaluation and treatment of high-dysplastic spondylolisthesis. *Eur Spine J.* 2009;18:133-142.
17. Mac-Thiong JM, Wang Z, de Guise JA, et al. Postural model of sagittal spino-pelvic alignment and its relevance for lumbosacral developmental spondylolisthesis. *Spine (Phila Pa 1976).* 2008;33:2316-2325.
18. Matsunaga S, Ijiri K, Hayashi K. Nonsurgically managed patients with degenerative spondylolisthesis: a 10- to 18-year follow-up study. *J Neurosurg.* 2000;93:194-198.
19. Mac-Thiong JM, Labelle H. A proposal for a surgical classification of pediatric lumbosacral spondylolisthesis based on current literature. *Eur Spine J.* 2006;15:1425-1435.
20. Mac-Thiong JM, Labelle H, Parent S, et al. Reliability and development of a new classification of lumbosacral spondylolisthesis. *Scoliosis.* 2008;3:19.
21. Mac-Thiong JM, Duong L, Parent S, et al. Reliability of the Spinal Deformity Study Group Classification of Lumbosacral Spondylolisthesis. *Spine (Phila Pa 1976).* 2012;37:E95-102.
22. Bao H, Yan P, Zhu W, et al. Validation and Reliability Analysis of the Spinal Deformity Study Group Classification for L5-S1 Lumbar Spondylolisthesis. *Spine (Phila Pa 1976).* 2015;40:E1150-1154.
23. Chang HS, Fujisawa N, Tsuchiya T, et al. Degenerative spondylolisthesis does not affect the outcome of unilateral laminotomy with bilateral decompression in patients with lumbar stenosis. *Spine.* 2014;39:400-408.
24. Sengupta DK, Herkowitz HN. Degenerative spondylolisthesis: review of current trends and controversies. *Spine.* 2005;30:71-81.
25. Pearson A, Blood E, Lurie J, et al. Predominant leg pain is associated with better surgical outcomes in degenerative spondylolisthesis and spinal stenosis: results from the Spine Patient Outcomes Research Trial (SPORT). *Spine.* 2011;36:219-229.
26. Kong C, Sun X, Ding J, et al. Comparison of the French and CARDS classifications for lumbar degenerative spondylolisthesis: reliability and validity. *BMC Musculoskelet Disord.* 2019;20:382.

# Figures

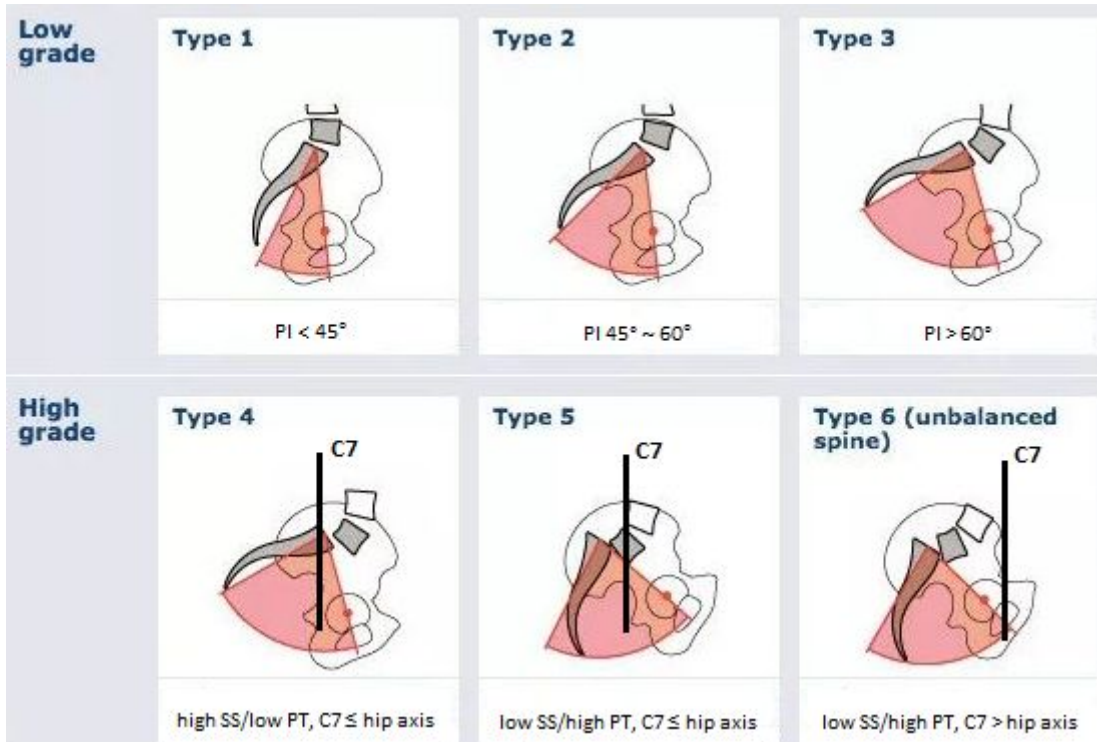


Figure 1

Spinal Deformity Study Group (SDSG) classification of lumbosacral spondylolisthesis.

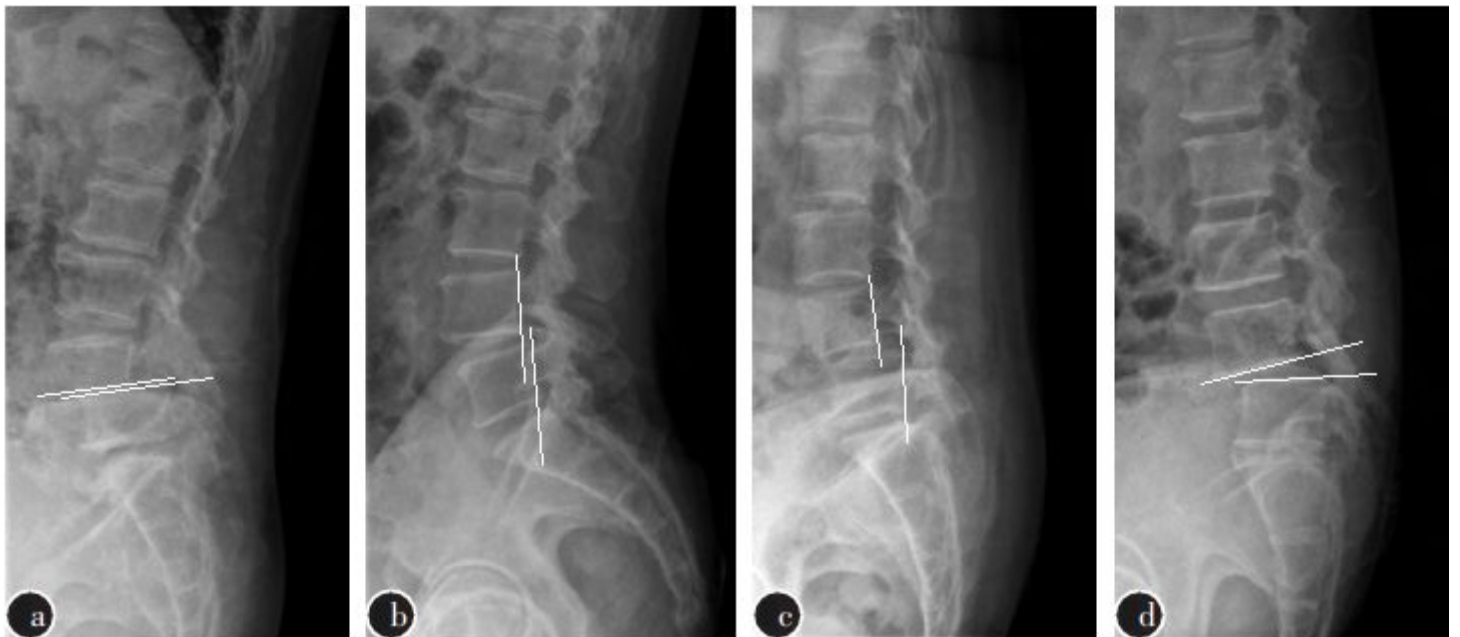


Figure 2

Diagram of CARDS classifications. a: Type A, advanced collapse of L4/5 disc space and contact of adjacent endplates; b: Type B, disc space partially preserved with translation  $< 5\text{mm}$ ; c: Type C, vertebral

translation > 5mm; and d: Type D, kyphotic alignment of L4/5.