Evaluation of the Reliability and Reproducibility of the Roussouly Classification for Lumbar Lordosis Types

Avaliação da confiabilidade e reprodutibilidade da classificação de Roussouly para os tipos de lordose lombar

Camila Oda Yamazato[1]   Gustavo Ribeiro[2]   Fabio Chaud de Paula[3]   Ramon Oliveira Soares[3]
Paulo Santa Cruz[4]   Michel Kanas[4]

[1] Department of Orthopedics and Traumatology, Escola Paulista de Medicina, Universidade Federal de São Paulo, São Paulo, SP, Brazil
[2] Hospital Israelita Albert Einstein, São Paulo, SP, Brazil
[3] Complexo Hospitalar Universitário Professor Edgar Santos, São Paulo, SP, Brazil
[4] Spine Outpatient Facility, Sports Traumatology Center, Department of Orthopedics and Traumatology, Escola Paulista de Medicina, Universidade Federal de São Paulo, São Paulo, SP, Brazil

Address for correspondence Camila Oda Yamazato, Department of Orthopedics and Traumatology, Escola Paulista de Medicina, Universidade Federal de São Paulo, Rua Botucatu, 740, Térreo, São Paulo, SP, 04023-900, Brazil (e-mail: camila.yamazato@gmail.com).

Abstract

Objective   The present study aims to determine the intra- and inter-rater reliability and reproducibility of the Roussouly classification for lumbar lordosis types.

Methods   A database of 104 panoramic, lateral radiographs of the spine of male individuals aged between 18 and 40 years old was used. Six examiners with different expertise levels measured spinopelvic angles and classified lordosis types according to the Roussouly classification using the Surgimap software (Nemaris Inc., New York, NY, USA). After a 1-month interval, the measurements were repeated, and the intra- and inter-rater agreement were calculated using the Fleiss Kappa test.

Results   The study revealed positive evidence regarding the reproducibility of the Roussouly classification, with reasonable to virtually perfect (0.307–0.827) intra-rater agreement, and moderate (0.43) to reasonable (0.369) inter-rater agreement according to the Fleiss kappa test. The most experienced examiners showed greater inter-rater agreement, ranging from substantial (0.619) to moderate (0.439).

Conclusion   The Roussouly classification demonstrated good reliability and reproducibility, with intra- and inter-rater agreements at least reasonable, and reaching substantial to virtually perfect levels in some situations. Evaluators with highest expertise levels showed greater intra and inter-rater agreement.

Keywords
► spinal curvatures
► postural balance
► lordosis/classification

Resumo

Objetivo   Determinar a confiabilidade e reprodutibilidade intra- e interavaliadores da classificação dos tipos de lordose lombar de Roussouly.

Métodos   Foram utilizadas 104 radiografias panorâmicas da coluna vertebral em incidência de perfil, de banco de dados, de indivíduos do sexo masculino com idade entre 18 e 40 anos. Utilizando o software Surgimap (Nemaris Inc., Nova York, NY, EUA), seis examinadores com diferentes níveis de experiência aferiram os ângulos espino-pélvicos e classificaram o tipo de lordose de acordo com a classificação de Roussouly. Após um intervalo de 1 mês, as mensurações foram realizadas novamente, sendo calculadas as concordâncias intra- e interavaliadores através do teste Kappa de Fleiss.

Resultados   O estudo demonstrou evidências positivas em relação à reprodutibilidade da classificação de Roussouly, com teste Kappa de Fleiss para concordância intra-avaliador entre razoável à quase perfeita (0,307–0,827), e interavaliador entre moderada (0,43) e razoável (0,369). Os examinadores mais experientes apresentaram maior concordância interavaliador, variando entre substancial (0,619) e moderada (0,439).

Palavras-chave
► curvaturas da coluna vertebral
► equilíbrio postural
► lordose/classificação

Conclusão   A classificação de Roussouly, demonstrou boa confiabilidade e reprodutibilidade, tendo em vista que a concordância intra- e interavaliadores foi considerada no mínimo razoável, podendo atingir os níveis de substanciais à quase perfeitos em algumas situações. Os avaliadores com maior nível de experiência apresentaram maior concordância intra- e interavaliadores.

Introduction

During the last 2 decades, the study of spinopelvic angles and sagittal balance has been increasingly relevant in spinal surgery, mainly for the correction of adult deformities. Reportedly, these measures vary according to age, ethnicity, and biotype in asymptomatic patients, and according to the etiology of the sagittal imbalance in symptomatic patients.[1–3]

Spinal deformities are clinically and radiologically different in adults and adolescents due to the association with degenerative processes, dissimilar patterns, and natural history. Furthermore, spinal imbalance at the sagittal plane may result from fractures or postoperative complications.[4]

Lumbar lordosis severity and pain are inversely correlated; in addition, there is an association between spinopelvic parameters and lumbar lordosis types with disc degeneration, facet overload, spondylolisthesis, chronic lumbar pain, disc herniation, and functional disability.[5–10]

The body uses mechanisms such as increased kyphosis/lordosis of adjacent segments, trunk hyperextension, pelvic anteversion or retroversion, knee flexion, and ankle extension to compensate for sagittal imbalance.[11,12] These mechanisms, along with spinal anatomical parameters and sagittal alignment, must be considered for surgical indication and planning because they affect the postoperative prognosis.[13]

Since this assessment is deemed critical, Roussouly et al.[14] proposed a classification system that divides the lumbar spine into four types according to the lordotic apex and sacral tilt angle. Although this classification has been used for

both research and clinical purposes since its introduction, few studies prove its validation.

The present study aimed to evaluate the reliability and reproducibility of lumbar lordosis classification using the Roussouly et al.[14] system, and to verify if the inter-rater agreement is affected by the expertise level.

Materials and Methods

A total of 104 panoramic radiographs of the spine in lateral view of men aged between 18 and 40 years old was used.

These radiographs belonged to a database that had been used in previous studies for other evaluations. Due to the impossibility of contacting these subjects, who were previously anonymized, the present study was exempted from an informed consent form and it was approved by the Research Ethics Committee under protocol number 3.828.093.

All radiographs were obtained with the same equipment. Patients were asked to stand up, with a straight trunk, upper limbs resting in a support, shoulders at 30° flexion, slightly flexed elbows, and extended knees. Panoramic, lateral radiographs covered from the base of the skull to the proximal region of the femur. Low-quality images that did not allow measurements were excluded from the study.

Using a the Surgimap software, version 2.3.1.5 (Nemaris Inc., New York, NY, USA), 6 evaluators, consisting in 2 spine surgeons with > 5 years of experience (A1 and A2), 2 residents in Spine Surgery (B1 and B2), and 2 residents in Orthopedics and Traumatology (C1 and C2), measured the spinopelvic angles and the sagittal vertical axis (►**Figure 1**). These data
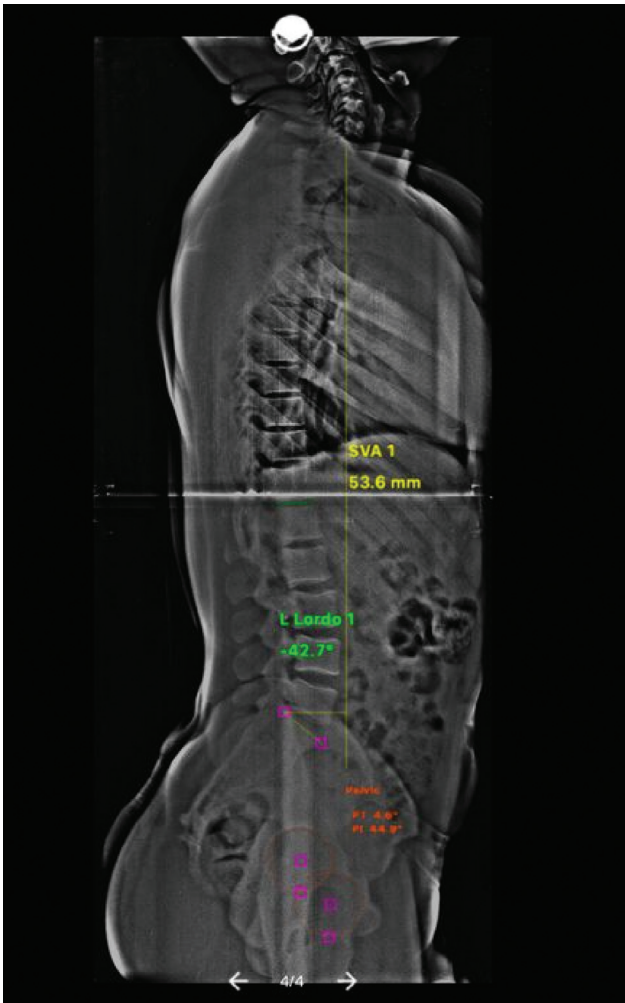
Fig. 1 Example of spinopelvic angles and vertical sagittal axis measurement using the Surgimap software.

were used to classify the type of lumbar lordosis according to Roussouly et al.[15] (►**Figure 2**). After 1 month, the measurements were repeated by the same evaluators to assess intra- and inter-rater agreement.

Statistical Methods

Initially, the results were descriptively analyzed to obtain graphs and frequency tables to characterize the participants of the research. Categorical variables were expressed as absolute frequency and percentage values. Graphs assessed the frequency of variables of interest.

Intra- and inter-rater agreement of the Roussouly classification were determined using the Fleiss kappa test (1981); this is a generalization of the kappa test used when several people evaluate the same sample on a scale with different categories, such as the Roussouly classification, consisting in types 1 to 4. The kappa agreement coefficient ranges from $+1$ (perfect agreement) to 0 (agreement equal to that expected by chance) to -1 (complete disagreement).[16]

The kappa coefficient value was classified according to Landis et al.,[17] as shown in ►**Table 1**.

All analyzes were performed with the statistical software R version 3.3.1 (R Foundation, Vienna, Austria), and the level of significance was set at 5%.[18]

Results

►**Table 2** presents frequency distribution of the lordosis type, attributed by each evaluator at both measurements. Note that type 3 was the most frequently assigned type ($>50\%$) by all evaluators, except for B1, who classified most cases as type 4.

Intra-rater Agreement

The intra-rater agreement analysis revealed that A2 had the best level of agreement, with a virtually perfect

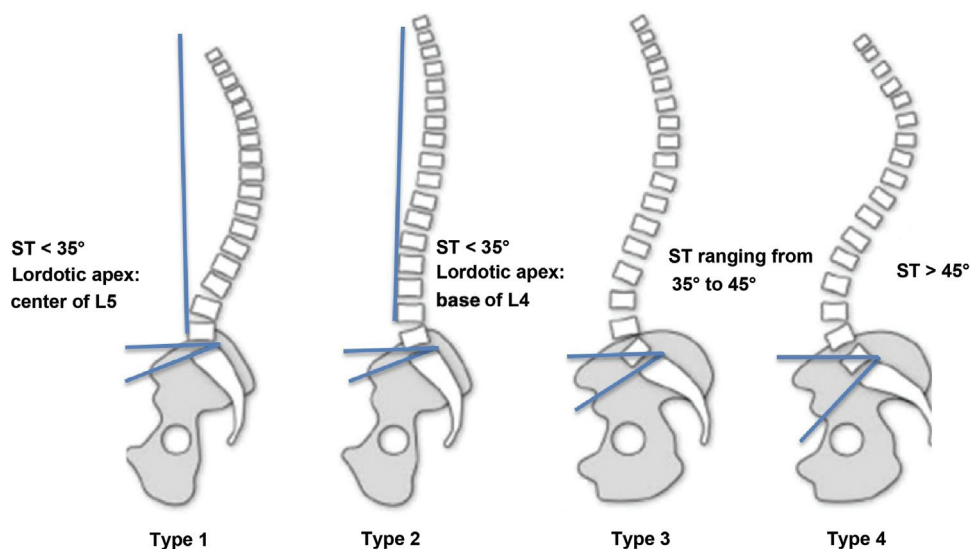**Fig. 2** Lumbar lordosis classification according to Roussouly et al. **Type 1:** The sacral tilt (ST) is $<35°$, and the apex of lumbar lordosis is located at the center of the L5 vertebral body. **Type 2:** The ST is $<35°$, and the apex of lumbar lordosis is located at the base of the L4 vertebral body. **Type 3:** The ST ranges from 35° to 45°. **Type 4:** The ST is $>45°$. (adapted from Roussouly et al.[15]).

Table 1 Kappa coefficient classification according to Landis et al.[17]

Kappa coefficient	Strength of agreement
< 0.00	Poor
0.00–0.20	Weak
0.21–0.40	Reasonable
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Virtually perfect

Fig. 3 Fleiss kappa coefficient for intra-rater agreement.

coefficient (0.827). On the other hand, the lowest level of agreement was obtained by B1, with a reasonable coefficient (0.307). A1 and B2 showed substantial agreement (0.601 and 0.710, respectively), whereas C1 and C2 presented moderate agreement (0.580 and 0.557, respectively). The average intra-rater agreement was 0.597, which was deemed moderate. All values had a p-value < 0.001 (►Figure 3 and ►Table 3).
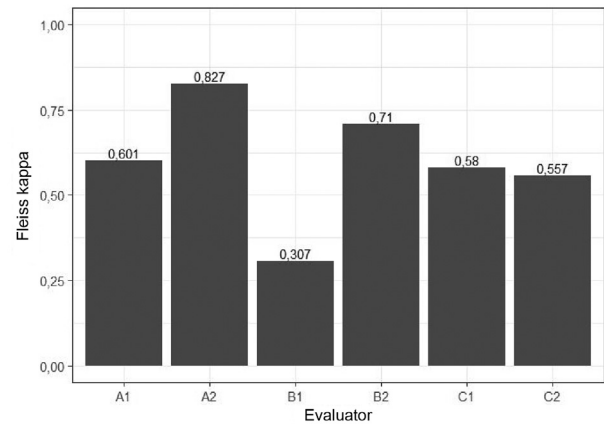
Inter-rater Agreement
In the inter-rater agreement analysis, the general coefficient for the first measurement was 0.43 (moderate); for the second measurement, this value was slightly lower, at 0.369 (reasonable) (►Table 4).

Among evaluators with the same level of expertise, there was a statistically significant agreement ($p < 0.001$) between all groups, and spine surgeons with > 5 years of experience presented the highest level of inter-rater agreement, ranging from substantial at the first measurement (0.619) to moderate at the second measurement (0.439). Spine surgery and orthopedics and traumatology residents showed reasonable levels of inter-rater agreement within their classes (►Table 5).

Discussion

In addition to dictating treatment or providing prognosis, an adequate classification must be reproducible for professionals with different expertise levels. The Roussouly classification for lumbar lordosis was introduced as a tool to analyze the sagittal alignment of the spine while considering pelvic orientation, characterizing an individual biotype.[14]

Roussouly types 1 and 2 have lower sacral tilt (ST) (<35º) and lower angular lordosis, increasing the load at the anterior spine, with a potentially higher association with disc degeneration[3,6,7] and chronic low lumbar pain.[9] Type 3 is the most frequent type in asymptomatic populations, even among different ethnicities and age groups.[1,2] Type 4 has the highest amount of ST (> 45º) and lumbar lordosis, and it is more related to spondylolisthesis and facet overload.[8]

In our study, type 3 lordosis was the most commonly found by the evaluators, which is consistent with previous studies in the asymptomatic population.[1,2] Roussouly considered this type of lordosis as more physiological.[14]

Table 2 Frequency distribution of the Roussouly classification

Measurement	Evaluator	Lumbar lordosis type according to the Roussouly classification			
		1	2	3	4
Measurement 1	A1	9 (9%)	11 (11%)	59 (57%)	25 (24%)
	A2	7 (7%)	12 (12%)	57 (55%)	28 (27%)
	B1	14 (13%)	10 (10%)	23 (22%)	57 (55%)
	B2	8 (8%)	8 (8%)	57 (55%)	31 (30%)
	C1	7 (7%)	8 (8%)	58 (56%)	31 (30%)
	C2	7 (7%)	10 (10%)	59 (57%)	28 (27%)
Measurement 2	A1	6 (6%)	9 (9%)	64 (62%)	25 (24%)
	A2	10 (10%)	9 (9%)	57 (55%)	28 (27%)
	B1	19 (18%)	8 (8%)	30 (29%)	47 (45%)
	B2	10 (10%)	5 (5%)	57 (55%)	32 (31%)
	C1	3 (3%)	8 (8%)	61 (59%)	32 (31%)
	C2	6 (6%)	14 (13%)	62 (60%)	22 (21%)

Evaluators according to expertise: A, spine surgeons; B, spine surgery residents; C, orthopedics and traumatology residents.

Table 3 Fleiss kappa coefficients from ►Figure 1 with confidence intervals and *p-values*

Evaluator	Fleiss kappa coefficients	95%CI	p-value
A1	0.601	(0.462–0.740)	< 0.001*
A2	0.827	(0.738–0.915)	< 0.001*
B1	0.307	(0.163–0.452)	< 0.001*
B2	0.710	(0.586–0.833)	< 0.001*
C1	0.580	(0.440–0.720)	< 0.001*
C2	0.557	(0.407–0.708)	< 0.001*

Abbreviation: CI, confidence interval.
Agreement was significant in all cases at a 5% significance levels ($p < 0.001$).

Table 4 Fleiss kappa test for inter-rater agreement

Evaluator	Fleiss kappa coefficients	95%CI	p-value
Measurement 1	0.430	(0.344–0.516)	< 0.001
Measurement 2	0.369	(0.288–0.451)	< 0.001

The classification requires the measurement of spino-pelvic angles, which can be performed manually using a goniometer on panoramic, lateral radiographs of the spine including the pelvis and femoral heads; however, the Surgimap software has been validated to facilitate measurement.[19]

Even though the Roussouly classification describes objective criteria depending mainly on measurable, well-defined references, a variation of a single degree in ST may change the type of lordosis. As a result, patients with borderline cutoff values (~ 35° or 45°) can receive different ratings from different observers or at different measurements from the same observer. In addition, the definition of the lordotic apex may be doubtful, allowing for divergences between types 1 and 2. Thus, the hypothesis that the greater or lesser presence of spines with these characteristics may affect classification reproducibility is valid.

The present study revealed the good reproducibility of the Roussouly classification, since both the intra- and inter-rater agreements were at least reasonable (> 0.20) according to Fleiss kappa coefficients. The intra-rater agreement ranged from reasonable to virtually perfect, whereas the inter-rater agreement ranged from reasonable to moderate.

Evaluator B1 stood out with the lowest intra- and inter-rater agreement; in addition, he was the only one to find a higher prevalence of type 4 lordosis. These differences may be explained by some divergence in the interpretation of the classification, technical measurement errors, or be inherent to the fact that measurements with close values can be classified as different types.

Experience seems to affect the reproducibility of the classification, since the most experienced evaluators (A1 and A2) showed greater intra- and inter-rater agreement. This finding may be explained by the fact that spinal surgeons have greater familiarity with these measurements and understanding of the spinopelvic angles than residents in training.

Conclusion

The Roussouly classification demonstrated good reliability and reproducibility. Intra- and inter-rater agreements were at least reasonable, ranging from substantial to virtually perfect in some situations. Experts with a higher level of experience showed greater intra- and inter-rater agreement.

Table 5 Fleiss kappa test for inter-rater agreement (peer to peer)

Measurement	Evaluator	A1	A2	B1	B2	C1	C2
Measurement 1	A1	1.000					
	A2	0.619	1.000				
	B1	0.192	0.261	1.000			
	B2	0.488	0.412	0.236	1.000		
	C1	0.565	0.583	0.218	0.434	1.000	
	C2	0.597	0.584	0.196	0.516	0.496	1.000
Measurement 2	A1	1.000					
	A2	0.439	1.000				
	B1	0.222	0.138	1.000			
	B2	0.483	0.458	0.283	1.000		
	C1	0.515	0.539	0.168	0.449	1.000	
	C2	0.440	0.404	0.166	0.496	0.325	1.000

Conflict of Interests
The authors have no conflict of interests to declare.

References

1 Hu P, Yu M, Sun Z, et al. Analysis of global sagittal postural patterns in asymptomatic Chinese adults. Asian Spine J 2016;10 (02):282–288

2 Bassani T, Galbusera F, Luca A, Lovi A, Gallazzi E, Brayda-Bruno M. Physiological variations in the sagittal spine alignment in an asymptomatic elderly population. Spine J 2019;19(11):1840–1849

3 Bae J, Lee SH, Shin SH, Seo JS, Kim KH, Jang JS. Radiological analysis of upper lumbar disc herniation and spinopelvic sagittal alignment. Eur Spine J 2016;25(05):1382–1388

4 Lowe T, Berven SH, Schwab FJ, Bridwell KH. The SRS classification for adult spinal deformity: building on the King/Moe and Lenke classification systems. Spine 2006;31(19, Suppl):S119–S125

5 Schwab FJ, Smith VA, Biserni M, Gamez L, Farcy JP, Pagala M. Adult scoliosis: a quantitative radiographic and clinical analysis. Spine 2002;27(04):387–392

6 Menezes-Reis R, Bonugli GP, Dalto VF, da Silva Herrero CF, Defino HL, Nogueira-Barbosa MH. Association between lumbar spine sagittal alignment and L4-L5 disc degeneration among asymptomatic young adults. Spine 2016;41(18):E1081–E1087

7 Mardare M, Oprea M, Popa I, Zazgyva A, Niculescu M, Poenaru DV. Sagittal balance parameters correlate with spinal conformational type and MRI changes in lumbar degenerative disc disease: results of a retrospective study. Eur J Orthop Surg Traumatol 2016;26(07):735–743

8 Labelle H, Roussouly P, Berthonnaud E, Dimnet J, O'Brien M. The importance of spino-pelvic balance in L5-s1 developmental spondylolisthesis: a review of pertinent radiologic measurements. Spine 2005;30(6, Suppl):S27–S34

9 Chaléat-Valayer E, Mac-Thiong J, Paquet J, et al. Sagittal spinopelvic alignment in chronic low back pain. Eur Spine J 2011;20 (Suppl 5(Suppl 05):634–640

10 Schwab FJ, Blondel B, Bess S, et al. International Spine Study Group (ISSG) Radiographical spinopelvic parameters and disability in the setting of adult spinal deformity: a prospective multicenter analysis. Spine 2013;38(13):E803–E812

11 Barrey C, Roussouly P, Le Huec JC, et al. Compensatory mechanisms contributing to keep the sagittal balance of the spine. Eur Spine J 2013;22(Suppl 6):(Suppl 6):S834–841

12 Le Huec JC, Thompson W, Mohsinaly Y, Barrey C, Faundez A. Sagittal balance of the spine. Eur Spine J 2019;28(09):1889–1905

13 Le Huec JC, Faundez A, Dominguez D, Hoffmeyer P, Aunoble S. Evidence showing the relationship between sagittal balance and clinical outcomes in surgical treatment of degenerative spinal diseases: a literature review. Int Orthop 2015;39(01):87–95

14 Roussouly P, Gollogly S, Berthonnaud E, Dimnet J. Classification of the normal variation in the sagittal alignment of the human lumbar spine and pelvis in the standing position. Spine 2005;30(03):346–353

15 Roussouly P, Berthonnaud E, Dimnet J. [Geometrical and mechanical analysis of lumbar lordosis in an asymptomatic population: proposed classification]. Rev Chir Orthop Repar Appar Mot 2003; 89(07):632–639

16 Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York: Wiley; 1981

17 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(01):159–174

18 R Development Core Team R: a language and environment for statistical computing. R Foundation for Statistical Computing-Vienna, Austria2015. Disponível em: www.Rproject.org

19 Lafage R, Ferrero E, Henry JK, et al. Validation of a new computer-assisted tool to measure spino-pelvic parameters. Spine J 2015;15 (12):2493–2502