RBO REVISTA BRASILEIRA DE ORTOPEDIA
www.rbo.org.br

Original Article

Intra and interobserver concordance between the different classifications used in Legg–Calvé–Perthes disease☆

*André Cicone Liggieri**, *Marcos Josei Tamanaha*, *José Jorge Kitagaki Abechain*, *Tiago Moreno Ikeda*, *Eiffel Tsuyoshi Dobashi*

Discipline of Pediatric Orthopedics, Escola Paulista de Medicina, Universidade Federal de São Paulo (UNIFESP), São Paulo, SP, Brazil

ARTICLE INFO

Article history:
Received 2 September 2014
Accepted 26 September 2014
Available online 29 October 2015

Keywords:
Hip/radiography
Legg–Calvé–Perthes disease/classification
Legg–Calvé–Perthes disease/radiography
Osteonecrosis

ABSTRACT

Objective: The aim of this study was to determine the intra and interobserver concordance rates of the Waldenström, Catterall and Herring classifications for Legg–Calvé–Perthes disease.

Methods: One hundred radiographs of the pelvis in anteroposterior and Lauenstein views, from patients with this disease, were selected. The radiographs were classified by four physicians with different levels of experience who had previously been given guidance regarding the classifications used, in order to minimize any bias of interpretation. The radiographs were examined by the same observers at two different times in order to evaluate the intra and interobserver concordance. Reproducibility was assessed using the kappa index.

Results: The concordance analysis was stratified into levels (poor, slight, fair, moderate, good and excellent). The intraobserver analysis showed, for the Waldenström classification, moderate concordance for three examiners and fair for one; for Herring, excellent for one examiner and good for three; and for Catterall, good for all the examiners. The interobserver analysis showed: for the three classification systems, no situations of excellent concordance; for Waldenström, four situations of fair concordance, one moderate and one slight; for Herring, four situations of moderate concordance, one good and one fair; and for Catterall, four situations of moderate concordance and two fair.

Conclusion: The classifications studied are the ones most used for guiding the treatment for Legg–Calvé–Perthes disease, but the degree of intra and interobserver concordance is far from ideal. Complementary staging systems need to be taken into consideration, so that there can be greater certainty regarding the treatment.

☆ Work performed on the database of images of patients attended within the Discipline of Pediatric Orthopedics, Escola Paulista de Medicina, Universidade Federal de São Paulo (UNIFESP), São Paulo, SP, Brazil.

* *Corresponding author.*
E-mail: acliggieri@hotmail.com (A.C. Liggieri).

http://dx.doi.org/10.1016/j.rboe.2015.09.010

Concordância intra e interobservadores das diferentes classificações usadas na doença de Legg–Calvé–Perthes

Palavras-chave:
Quadril/radiografia
Doença de Legg–Calvé–Perthes/classificação
Doença de Legg–Calvé–Perthes/radiografia
Osteonecrose

RESUMO

Objetivo: Determinar o índice de concordância intra e interobservadores das classificações de Waldenström, Catterall e Herring na doença de Legg–Calvé–Perthes.

Métodos: Foram selecionadas 100 radiografias da bacia, nas incidências anteroposterior e de Lauenstein de pacientes portadores da doença. As radiografias foram classificadas por quatro médicos com diferentes níveis de experiência, previamente orientados a respeito das classificações usadas, para minimizar qualquer viés de interpretação. As radiografias foram examinadas pelos mesmos observadores em dois momentos distintos para avaliar as concordâncias inter e intraobservadores. A análise da reprodutibilidade foi avaliada pelo índice de Kappa.

Resultados: A análise de concordância foi estratificada em níveis (ruim, pequena, regular, moderada, boa e excelente) e evidenciou para a concordância intraobservadores: concordância moderada para três examinadores e uma regular para a classificação de Waldenström; excelente para um examinador e boa para três, na classificação de Herring; na classificação de Catterall, a concordância foi considerada boa entre todos os examinadores. Em relação à análise de concordâncias interobservadores foram obtidas: nenhuma concordância excelente para os três sistemas de classificação; quatro regulares, uma moderada e uma pequena para a classificação de Waldenström; quatro moderadas, uma boa e uma regular na classificação de Herring e, pelo sistema de Catterall, quatro concordâncias moderadas e duas regulares.

Conclusão: As classificações estudadas são as mais usadas para guiar o tratamento da DLCP, porém o grau de concordância intra e interobservadores não é ideal e sistemas complementares de estadiamento devem ser levados em consideração, para uma maior assertividade no tratamento.

Introduction

In 1910, Legg–Calvé–Perthes disease (LCPD) was described for the first time. Since then, it has aroused great interest among researchers and has come to be one of the most controversial topics in the orthopedic literature. Several aspects of this clinical entity still remain unexplained, such as its etiology and the best way of treating it in the active phase of the disease.

For a long time, almost all authors concentrated on analyzing the radiographic aspects of LCPD. The evolutionary phases were first described by Waldenström,[1] whose classification was subsequently simplified and correlated with the anatomopathological findings by Jonsäter.[2] Evaluations on the compromising of the nucleus of ossification of the femoral head came to be systematized by Catterall,[3] based on analysis on simple radiographs produced during the phase of maximum fragmentation. With the aim of determining the proportions of lesions during the initial phase or the necrosis phase, Salter and Thompson[4] demonstrated that the size of the subchondral fracture in the Lauenstein view precisely reflected the degree to which the proximal femoral epiphysis was affected by the disease. More recently, Herring et al.[5] proposed a new classification based on the height of the lateral column of the femoral epiphysis. Other classifications have been proposed, but the ones cited above are those that are most used today.

From radiographic analyses on the hips of affected patients, all of these authors developed classifications for use in cases of LCPD and thus sought to systematize the treatment. However, for a given classification to be considered adequate, it needs to be reproducible, i.e. there needs to be inter and intraobserver concordance, and furthermore, the classification system needs to aid in guiding disease outcomes.

Although the treatment of LCPD has been the subject of exhaustive discussions among orthopedists, there is still no clear evidence regarding the best therapeutic method for these patients, and this is not within the scope of the present study.

The objective of this study was to evaluate the intra and interobserver concordance of the classifications of Waldenström,[1] Catterall[3] and Herring et al.,[5] attempt to establish which of them has the greatest degree of reproducibility and thus facilitate therapeutic decision-making.

Materials and methods

This research project was submitted to the research ethics committee of the Brazil Platform and was approved for implementation under the numbers CAAE 33513214.7.0000.5505 and CEP 418466.

The patients were evaluated through analysis on simple radiographs of the pelvis in the anteroposterior and

Lauenstein views. A convenience sample of 100 radiographs of patients with LCPD was collected. These examinations were selected from a database within the Discipline of Pediatric Orthopedics, relating to patients attended at the orthopedics and traumatology outpatient clinic of Hospital São Paulo. The radiographs were selected by two orthopedists who did not participate in the disease classification process, such that good-quality examinations were included and a broad spectrum of lesions was taken into consideration. With the aim of minimizing bias due to difficulties in interpreting the examinations, the observers were provided with an initial explanation of the classification systems used in this study. Furthermore, the protocol used for the data-gathering contained a diagram containing images from the classification systems of Waldenström[1] (as modified by Jonsäter[2]), Catterall[3] and Herring et al.[5]

In order to determine the interobserver concordance, each of the four researchers evaluated the radiographic examinations independently. The different examiners were not allowed to have prior knowledge regarding the patients' histories or any clinical information about how the disease was addressed or treated. The examiners were allowed to take all the time that they needed to evaluate all of the radiographs. After making the classifications, the evaluators were asked to classify all the examinations again, 30 days after the first analysis, without having access to the first round of evaluations.

The participants were instructed not to discuss the classification systems between each other until after all the material to be analyzed in this study had been gathered.

The statistical analysis on the results obtained was performed by a professional within the field of medical statistics. The data collected were analyzed with regard to inter and intraobserver concordance, by means of the kappa index. The tests were interpreted as described by Altman,[6] as "proportional agreement with correction for chance". Kappa is a coefficient of concordance that has values ranging from +1 (perfect concordance), passing through 0 (concordance equal to chance) and going to −1 (complete discordance). There are no definitions regarding which concordance levels should be accepted, but in the study by Svanholm et al.,[7] it was indicated that concordance greater than 0.75 is considered excellent, 0.5–0.75 good and less than 0.5 is poor. However, we used the intervals for the kappa index that are shown in Table 1.

Table 1 – Correlation between the kappa value and the degree of concordance.

Kappa value	Concordance
0	Poor
0–0.20	Low
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Good
0.81–1	Excellent

P, poor; L, low; F, fair; M, moderate; G, good; E, excellent.

Results

Table 2 shows the absolute frequencies of the classifications made by the examiners at the two different evaluation times on the radiographs studied.

Table 3 shows the distribution of the weighted kappa values and the 95% confidence intervals of the intraobserver concordance analysis. In this analysis, moderate agreement for three examiners and fair agreement for one examiner were obtained through application of the Waldenström classification. From the Herring classification, there was an excellent result for one examiner and good agreement for three examiners. In relation to the Catterall classification, all the results presented good agreement.

Table 4 shows the results relating to agreement between the observers according to kappa values with 95% confidence intervals. According to the statistical analysis, no cases of excellent interobserver concordance were found. There were four results in which the concordance was fair, one moderate and one low, in the Waldenström classification. In relation to the Herring classification, four cases of moderate, one good and one fair agreement were obtained. In relation to the Catterall classification, there were four indexes with moderate agreement and two with fair agreement.

Discussion

The challenge for orthopedists in relation to LCPD lies in treating this condition. There has been much discussion about whether there is or is not any definitive possibility of altering what Catterall[3] called the natural history of the disease.

There has also been much discussion regarding the treatment that should be applied. Because of the lack of convincing evidence regarding the effectiveness of therapies, these concepts have been applied over the course of the years, based on each author's experience of diagnosing, classifying and managing LCPD.

We are convinced that, to deal with LCPD correctly, the diagnosis needs to be systematized and the approaches used need to be based especially on classifications that direct us toward appropriate treatment. So far, this has been done on the basis of the classifications assessed in this study.

Therefore, we believe that the first step to be taken, after the diagnosis of LCPD has been established, is to attempt to properly stage it, using classical methodologies based on radiographic analyses and also on magnetic resonance imaging, arthrography and scintigraphy when necessary.

With the aim of dealing with LCPD from a therapeutic point of view, we found that several authors in the literature proposed classifications that would allow this disease to be systematized so as to be able to predict which approach would be the best one to use, with the expectation of thus obtaining better results.

In our study, the Waldenström classification[1] presented concordance that we considered to be inadequate. Since the phases of the disease overlap, lack of knowledge of the length of time over which the disease has evolved makes it more difficult to define the current stage of the disease.

Table 2 – Weighted kappa and 95% CI values for the intraobserver assessment.

Observer	Waldenström	Herring	Catterall
A	0.44 (M)	0.82 (E)	0.79 (S)
	(0.31; 0.58)	(0.74; 0.90)	(0.70; 0.88)
B	0.32 (C)	0.73 (S)	0.69 (S)
	(0.18; 0.45)	(0.63; 0.83)	(0.59; 0.78)
C	0.53 (M)	0.77 (S)	0.72 (S)
	(0.38; 0.69)	(0.67; 0.87)	(0.62; 0.81)
D	0.52 (M)	0.71 (S)	0.65 (S)
	(0.38; 0.65)	(0.62; 0.80)	(0.55; 0.75)

Table 3 – Weighted kappa and 95% CI values for the interobserver assessment.

		Classification		
Obs1	Obs2	Waldenström	Herring	Catterall
A	B	0.30 (C)	0.63 (S)	0.41 (M)
		(0.15; 0.45)	(0.52; 0.74)	(0.31; 0.52)
	C	0.35 (C)	0.49 (M)	0.30 (C)
		(0.20; 0.50)	(0.35; 0.63)	(0.18; 0.42)
	D	0.38 (C)	0.53 (M)	0.32 (C)
		(0.25; 0.52)	(0.40; 0.65)	(0.21; 0.44)
B	C	0.29 (C)	0.41 (M)	0.46 (M)
		(0.14; 0.45)	(0.28; 0.54)	(0.34; 0.57)
	D	0.47 (M)	0.54 (M)	0.47 (M)
		(0.33; 0.61)	(0.41; 0.66)	(0.36; 0.58)
C	D	0.23 (Pq)	0.39 (C)	0.44 (M)
		(0.08; 0.39)	(0.27; 0.51)	(0.32; 0.56)

Initially, the classification system that most polarized the attention of other authors in the orthopedic literature was the one proposed by Catterall in 1971.[3] This author radiographically assessed the behavior of the nucleus of ossification of the femoral head during the progression of the disease, at the phase of maximum fragmentation.

This classification was contested by several researchers who used it. While some authors thought that it was of fundamental importance for indicating the therapy that should be followed and that it had a positive correlation with the final results,[8–10] others criticized it because it is applied at an advanced stage of the disease and has questionable concordance when used by different observers. All of these criticisms were cited by Terjesen et al.[11] Despite the reported discordance between observers, many authors have used this system to guide the therapy that is to be instituted. However, given that we did not find any excellent or good agreement, it is possible that the therapeutic indications may become distorted according to the gradation imposed by this system.

Table 4 – Absolute frequencies of the observers' classifications in the two evaluations.

	Observers							
Classification	A		B		C		D	
	Eval1	Eval2	Eval1	Eval2	Eval1	Eval2	Eval1	Eval2
Waldenström								
1	24	14	14	9	14	10	13	8
2	44	29	35	27	38	32	33	33
3	16	38	23	30	30	40	22	31
4	17	20	29	35	19	19	33	29
Herring								
1	29	29	23	25	26	24	28	23
2	48	45	51	46	55	56	34	39
3	9	12	8	13	14	13	23	21
4	15	15	19	17	6	8	16	18
Catterall								
1	48	42	31	27	24	20	41	35
2	26	31	30	38	31	22	21	21
3	17	18	29	20	27	35	25	28
4	10	10	11	16	19	24	14	17

Because the classification proposed by Salter and Thompson[4] is used at an initial stage of the disease, it may perhaps allow early treatment. However, it presents the limitation that only 25% of the patients with LCPD have subchondral fractures that are recognizable on radiographs.[5,12]

In comparison with the classification of Catterall,[3] that of Herring et al.[5] is easier to interpret. However, since this system too can only be used at the final fragmentation phase, we take the view that it would not be ideal for indicating early treatment.[12] Since the time of the initial description of this classification, which was presented in 1994, and because of the difficulty of defining the patients belonging to group B, another three subgroups were created, which thus allowed the classification to have wider coverage.

In our study, we obtained four moderate, one good and one fair agreement from testing the system of Herring et al.[5] This suggests that there is some difficulty in precisely defining each of the groups and subgroups.

Another important matter is that simple radiography does not precisely mirror what happens to the femoral epiphysis. Cartilaginous tissue, which also shows alterations caused by this disease, is present in greater amounts than bone tissue.[13–15] In this light, in some cases, particularly in the early stages of the disease, studying this cartilage by means of magnetic resonance imaging[2,16,17] and pneumoarthrography of the joint[13,14] may provide greater clarification and more effective guidance for the treatment that is to be instituted.[4,14,15,18–20]

Therefore, we consider that knowledge of the behavior of the cartilaginous structures of the hip is fundamental for staging LCPD and indicating the appropriate treatment.[4,14,15,18–20]

We believe that magnetic resonance imaging currently presents a variety of advantages over other examinations. This is advocated in several studies in which some authors developed their own classifications with the aim of indicating the therapy with a higher proportion of correct choices.

Although we found a certain degree of intraobserver concordance in the three classification systems studied, and interobserver concordance in two of the three systems, our data are not in agreement with what we have observed in the worldwide literature.

Most of the studies observed did not demonstrate this degree of concordance. One likely cause for this divergence may relate to the low average number of patients studied in the literature (40),[21–23] in comparison with the number of patients in the present study (100).

Unfortunately, although most experienced surgeons[24] use the classifications studied here to determine the approaches that they will use, the concordance observed after statistical analysis was not shown to be sufficient for this. Therefore, this should not be the sole factor taken into account in making therapeutic decisions relating to patients with LCPD.

Conclusion

1. The intraobserver concordance analysis with 95% confidence intervals showed the following through the kappa index: moderate concordance for three observers and fair for one, in the Waldenström classification; excellent concordance for one examiner and good for three, in the Herring classification; and good concordance for all the examiners, in the Catterall classification.
2. The interobserver concordance analysis with 95% confidence intervals showed the following through the kappa index after statistical analysis: no excellent concordance for any of the three classification systems; four fair, one moderate and one low agreement for the Waldenström classification; four moderate, one good and one fair agreement for the Herring classification; and four moderate and two fair agreements for the Catterall classification.
3. Although these classifications are the systems most used by orthopedists to treatments for Legg–Calvé–Perthes disease, and although the intra and interobserver concordance indices found in the present study were better than those seen in the worldwide literature, the indices found here are still far from ideal. Therefore, complementary systems for staging the disease should be taken into consideration, in order to have greater precision in treating this disease.

Conflicts of interest

The authors declare no conflicts of interest.

REFERENCES

1. The classic. The first stages of coxa plana by Henning Waldenström, 1938. Clin Orthop Relat Res. 1984;(191):4–7.
2. Jonsäter S. Coxa plana: a histopathologic and arthrographic study. Acta Orthop Scand Suppl. 1953;12:5–98.
3. Catterall A. The natural history of Perthes' disease. J Bone Joint Surg Br. 1971;53(1):37–53.
4. Salter RB, Thompson GH. Legg–Calvé–Perthes disease. The prognostic significance of the subchondral fracture and a two-group classification of the femoral head involvement. J Bone Joint Surg Am. 1984;66(4):479–89.
5. Herring JA, Kim HT, Browne R. Legg–Calve–Perthes disease. Part I: classification of radiographs with use of the modified lateral pillar and Stulberg classifications. J Bone Joint Surg Am. 2004;86(10):2103–20.
6. Altman DG. Some common problems in medical research. In: Practical statistics for medical research. London: Chapman and Hall; 1991. p. 403–9.
7. Svanholm H, Starklint H, Gundersen HJ, Fabricius J, Barlebo H, Olsen S. Reproducibility of histomorphologic diagnoses with special reference to the kappa statistic. APMIS. 1989;97(8):689–98.
8. Kamhi E, MacEwen GD. Treatment of Legg–Calvé–Perthes disease. Prognostic value of Catterall's classification. J Bone Joint Surg Am. 1975;57(5):651–4.
9. Dickens DR, Menelaus MB. The assessment of prognosis in Perthes' disease. J Bone Joint Surg Br. 1978;60(2):189–94.
10. Green NE, Beauchamp RD, Griffin PP. Epiphyseal extrusion as a prognostic index in Legg–Calvé–Perthes disease. J Bone Joint Surg Am. 1981;63(6):900–5.
11. Terjesen T, Wiig O, Svenningsen S. The natural history of Perthes' disease. Acta Orthop. 2010;81(6):708–14.
12. Ismail AM, Macnicol MF. Prognosis in Perthes' disease: a comparison of radiological predictors. J Bone Joint Surg Br. 1998;80(2):310–4.
13. Laredo Filho J, Ishida A, Lourenço AF, Kuwajima SS, Jorge SR. Avaliação radiográfica da cobertura acetabular da cabeça

femoral em pacientes portadores da doença de Legg–Calvé–Perthes unilateral submetidos à osteotomia de Salter. Rev Bras Ortop. 1993;28(5):299–303.
14. Laredo Filho J. Doença de Legg–Calvé–Perthes. Classificação artrográfica. Rev Bras Ortop. 1992;27(1):7–15.
15. Ducou le Pointe H, Haddad S, Silberman B, Filipe G, Monroc M, Montagne JP. Legg–Calvé–Perthes disease: staging by MRI using gadolinium. Pediatr Radiol. 1994;24(2):88–90.
16. Katz JF. Femoral torsion in Legg–Calvé–Perthes disease. J Bone Joint Surg Am. 1968;50(3):473–5.
17. Sales de Gauzy J, Kerdiles N, Baunin C, Kany J, Darodes P, Cahuzac JP. Imaging evaluation of subluxation in Legg–Calvé–Perthes disease: magnetic resonance imaging compared with the plain radiograph. J Pediatr Orthop B. 1997;6(4):235–8.
18. Henderson RC, Renner JB, Sturdivant MC, Greene WB. Evaluation of magnetic resonance imaging in Legg–Calvé–Perthes disease: a prospective, blinded study. J Pediatr Orthop. 1990;10(3):289–97.
19. Bos CF, Bloem JL, Bloem RM. Sequential magnetic resonance imaging in Perthes' disease. J Bone Joint Surg Br. 1991;73(2):219–24.
20. Hiehle JF Jr, Kneeland JB, Dalinka MK. Magnetic resonance imaging of the hip with emphasis on avascular necrosis. Rheum Dis Clin North Am. 1991;17(3):669–92.
21. Mahadeva D, Chong M, Langton DJ, Turner AM. Reliability and reproducibility of classification systems for Legg–Calvé–Perthes disease: a systematic review of the literature. Acta Orthop Belg. 2010;76(1):48–57.
22. Podeszwa DA, Stanitski CL, Stanitski DF, Woo R, Mendelow MJ. The effect of pediatric orthopaedic experience on interobserver and intraobserver reliability of the herring lateral pillar classification of Perthes' disease. J Pediatr Orthop. 2000;20(5):562–5.
23. Rajan R, Chandrasenan J, Price K, Konstantoulakis C, Metcalfe J, Jones S. Legg–Calvé–Perthes: interobserver and intraobserver reliability of the modified Herring lateral pillar classification. J Pediatr Orthop. 2013;33(2):120–3.
24. Kalenderer O, Agus H, Ozcalabi IT, Ozluk S. The importance of surgeons' experience on intraobserver and interobserver reliability of classifications used for Perthes' disease. J Pediatr Orthop. 2005;25(4):460–4.