Intra- and inter-observer reliability of Laredo classification system in Legg–Calvé–Perthes Disease

Serkan Erkus[1] · Ali Turgut[1] · Onder Kalenderer[1] · Kivanc Yuksel[2]

Abstract

The purpose of the current study was to investigate intra- and inter-observer reliability of arthrographic Laredo classification system in Perthes disease. Forty-seven patients were included in this cross-sectional descriptive study. Patients' age, gender, physical findings and hip arthrographs were collected from hospital medical records. Two different sets of power point slides were prepared in which the order of cases was randomized and blinded. Observers were divided into three groups according to their surgical experience (9 residents, 10 seniors, 10 pediatric orthopedists), and they assessed two times 1 month's intervals. Statistical analysis was performed by using the SPSS v21. Inter- and intra-observer reliabilities were calculated using intraclass correlation coefficient, weighted kappa and percentage agreement. Percentage agreement of Laredo classification was about 50% for all groups (residents, seniors and pediatric orthopedists); intra-observer reliabilities were excellent, excellent and fair, respectively. Inter-observer reliability of Laredo classification for each set was found to be excellent for all groups. Length of experience did not correlated significantly with the level of intra-observer agreement. As a conclusion, our results showed that Laredo's arthrographic classification system's intra-observer reliability is at least at a fair level and inter-observer reliability is at an excellent level. We believe that this classification system is valuable for an orthopedic surgeon who deals with the treatment of Perthes disease.

Keywords Arthrography · Reliability · Laredo classification · Perthes · Children

Introduction

Legg–Calvé–Perthes disease (LCPD) is an idiopathic, avascular necrosis of the proximal femoral epiphysis in growing skeleton [1]. Self-limited osteonecrosis can be resulted in varying degrees of deformation and possible osteoarthritis [2, 3]. There is no consensus on the treatment modality of LCPD in case of stages and age groups [4, 5]. Standard hip radiography is the mostly used diagnostic tool in the evaluation. Stage of the disease, extent of involvement and containment of the femoral head can be assessed with the evaluation of LCPD [1]. Since LCPD is seen in growing skeleton, the cartilage component of both femoral and acetabular side cannot be evaluated by radiography. Within this scope, magnetic resonance imaging (MRI) can be used to evaluate the cartilage component [6–8]. However, both plain radiography and MRI are static methods and thorough evaluation of the congruence of the hip joint requires dynamic method such as arthrographic imaging [8, 9]. In addition, owing to arthrography, secondary changes such as labral pathologies, femoroacetabular impingement, hinge abduction can also be distinguished [6]. Classifying LCPD is as important as other diseases for defining both the disease and guidance of the treatment modality. Catterall, Salter and Thompson, Herring and Stulberg classification systems based on radiography are widely used for LCPD [10–13]. Besides the advantage of arthrography in the choice of treatment, this method is also used in another classification system. Laredo was the first one who described the arthrographic classification which

✉ Onder Kalenderer
  okalenderer@gmail.com

  Serkan Erkus
  s.erkus@ymail.com

  Ali Turgut
  draliturgutort@yahoo.com.tr

  Kivanc Yuksel
  kivanc.yuksel@ege.edu.tr

1   Department of Orthopaedics and Traumatology, Tepecik Training and Research Hospital, İzmir, Turkey

2   Ege University School of Medicine ARGEFAR, İzmir, Turkey

consisted of five types for LCPD in 1992. The classification is based upon relationship with labrum, size and shape of femoral head (Table 1) [14]. There are many studies which investigated the reliability of these classification systems' [15–21], but to our knowledge, there is no study which has evaluated the intra- and inter-observer reliabilities of arthrographic Laredo classification system for LCPD in the current literature. The aim of this study was to report this reliability.

Table 1 Laredo's arthrographic classification of Legg–Calvé–Perthes disease [14]

	Arthrographic view of the hip
Type 1	Normal
Type 2	The femoral head is still spherical, but larger than normal
Type 3	The femoral head is ovoid in shape
Type 4	Large and flattened femoral head, and labrum loses its concavity
Type 5	Femoral head larger than normal and saddle shaped; the labrum is still elevated

Materials and methods

Forty-seven patients with LCPD were included in this cross-sectional descriptive study after the approval of the institutional ethical committee (IRB approval number: 2017-210/17.07). This study was carried out in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. Patients' hip arthrographs were collected from hospital medical records. These 47 patients with unilateral hip involvement, ranging in age from 6 to 10 with complete radiographic follow-up and arthrographic examination were included into the study. Arthrography imaging was performed by a single orthopedic surgeon under general anesthesia in at least two hip positions, particularly in neutral and frog-leg positions. Two different sets of files (Microsoft® Power Point slides) were prepared in which the order of cases was randomized and blinded. The knowledge of age, gender and physical findings of patients was in both sets (Fig. 1).
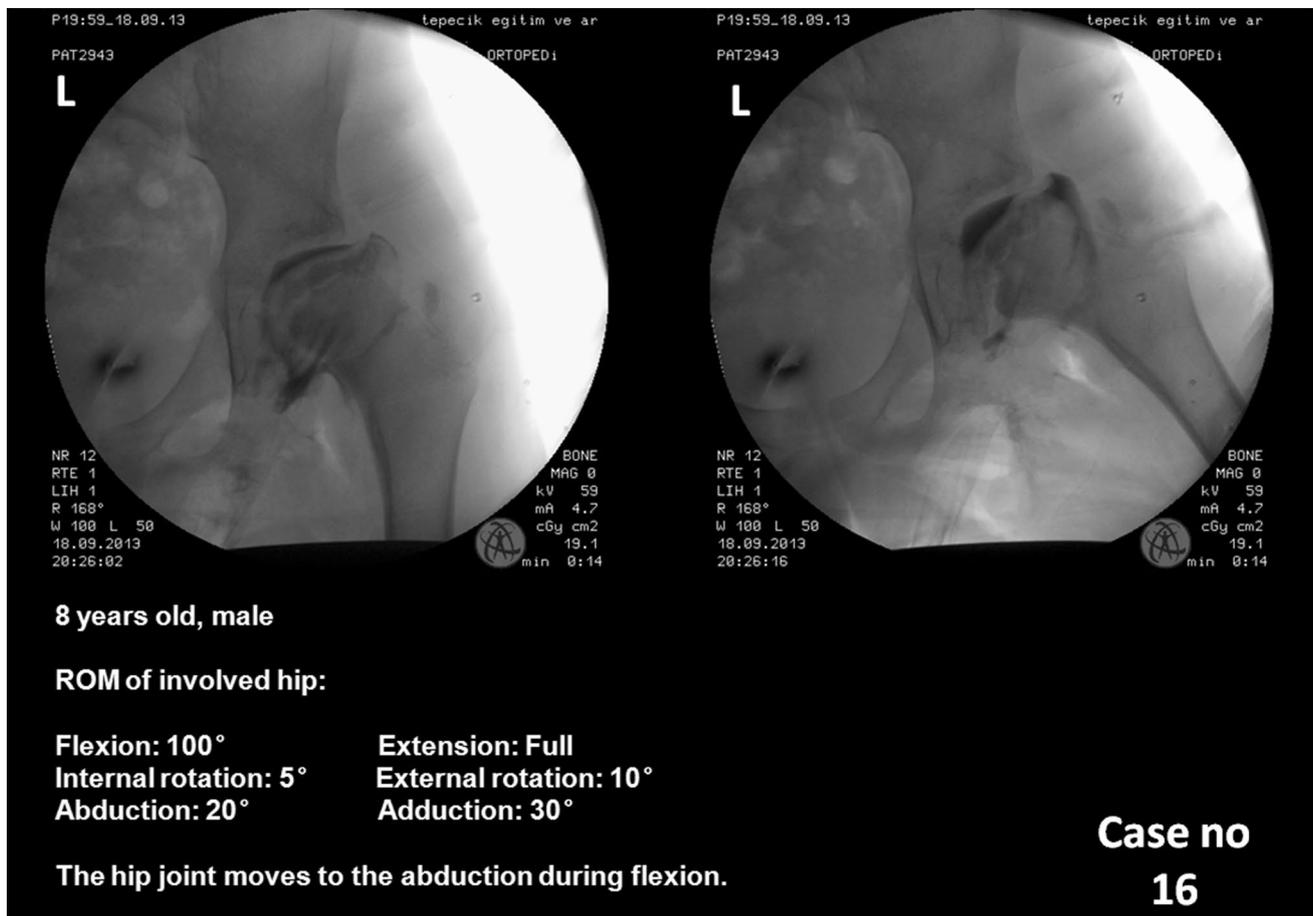

Fig. 1 Presentation of clinical findings and arthrographs of the patient assigned as Case no: 16 in Set 1

Laredo classification was presented to the observers after visual presentation of each patient as a slide.

Observers were divided into three groups according to their surgical experience. The first group consisted of nine residents trained in Department of Orthopedics and Traumatology at the training and research hospital for at least 4 years. In the second group, there were ten seniors working in a general orthopedics department for 3 to 5 years and they were not specifically hip surgeons. Ten attending consultant pediatric orthopedic surgeons from different institutions, who had surgical experience at least 5 years and more in pediatric hip disorders, took part in the third group. Assessments of the sets were performed in random order by each observer on two separate occasions, at least 4 weeks separately. Examples of the user interface typical of each case are shown in Fig. 1. Observers were asked which of the stage was most appropriate for each patient according to Laredo classification system.

Statistical analysis

Statistical analysis was performed by using the SPSS v21 standard statistical version (SPSS Inc., Chicago, Illinois, USA). Intra-observer agreement for all of the observers [comparing the results of the same observer's evaluations at two different times] was determined via percentage agreement and weighted kappa (κ) statistics [22, 23], and inter- and intra-observer reliabilities were calculated using intra-class correlation coefficient (ICC), w2 and p values within 95% confidence intervals [24]. In addition, for each observer, we calculated Spearman coefficients to identify correlations between intra-observer agreement and the length of experience.

A κ value is always between 0 and 1; the higher the value, the better the correlation. The κ and Spearman's rho (ρ) values were graded as slight (0–0.2), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80) and almost perfect (0.81–1) [25]. An ICC value greater than 0.75 is excellent, from 0.75 to 0.60 is good, from 0.59 to 0.40 is fair and below 0.40 is accepted as poor agreement [26].

Results

There were 42 boys and 5 girls with a mean of 7.53 years old (range 6–10). All patients had unilateral involvement (27 right and 20 left).

The mean intra-observer agreement with respect to stage of the experience between sets is presented in Table 2. In addition, percentage agreement was also analyzed by using Spearman's rho to understand if there was a change according to experience. Length of experience (Spearman

Table 2 Mean values of intra-observer agreement

Groups	Percentage agreement (%)	Kappa values (κ)	Agreement strength
Residents (n: 9)	54.8	0.37	Fair
Seniors (n: 10)	54.3	0.35	Fair
Pediatric orthopedists (n: 9)	46.1	0.25	Fair

coefficient, $\rho = -0.237$; $p = 0.224$) did not correlated significantly with the level of intra-observer agreement.

The mean intra-observer reliabilities between sets are shown in Table 3. To find out the correlation between these values and experience, Spearman's rho was measured again. The Spearman coefficient, ρ, was counted up as -0.149; $p = 0.450$.

Inter-observer reliabilities of each group are given in Table 4. In the assessment of both sets, strengths of ICC values were excellent for all groups. Furthermore, the numerical values were increased in the second evaluation.

Discussion

A classification system can be accepted as useful if it is valid, reliable and reproducible [27]. Several classification systems have been described for LCPD such as Catterall [10], Salter and Thompson [11], Herring [2, 13] and Stulberg [12]. Arthrography is useful to evaluate the cartilage structure of hip joint and also classification of LCPD defined by Laredo [14]. Although reliability of Catterall, Salter and Thompson, Herring and Stulberg classifications was evaluated previously [15–21], Laredo classification's reliability was never evaluated. Results of our study showed that percentage agreement of Laredo classification was about 50% for all groups (residents, seniors and pediatric orthopedists); intra-observer reliabilities were excellent, excellent and fair, respectively. And finally, inter-observer reliability of Laredo classification for each set was found to be excellent for all groups.

Various studies have evaluated the reliability of Salter and Thompson classification [15–20]. The average agreement values were found to be fair to good in these studies [19]. Simmons et al. [20] stated that kappa values and percentage agreement were increased to 0.99 and 93% for staff observers; however, these values were nearly two times higher than fellowship and resident observers. These results demonstrated that experience was a positive factor for the reliability of Salter and Thompson classification. Additionally, reliability of Salter and Thompson classification was found to be higher than the reliability of Catterall classification. The reason for this finding is probably due to the

Table 3 Mean intra-observer reliability values

Groups	ICC	%95 CI (min–max)	Strength of values
Residents (n: 9)	0.774	0.727 to 0.813	Excellent
Seniors (n: 10)	0.792	0.751 to 0.827	Excellent
Pediatric orthopedists (n: 9)	0.672	0.603 to 0.729	Good

Table 4 Inter-observer reliability values

Groups	Set I			Set II		
	ICC	%95 CI (min–max)	Comment	ICC	%95 CI (min–max)	Comment
Residents (n: 9)	0.890	0.836–0.932	Excellent	0.894	0.842–0.934	Excellent
Seniors (n: 10)	0.873	0.812–0.921	Excellent	0.905	0.859–0.941	Excellent
Pediatric orthopedists (n: 9)	0.836	0.756–0.898	Excellent	0.849	0.774–0.906	Excellent

simple nature (two stages) of Salter and Thompson classification. This classification system can only be used at the initial stage of LCPD because it uses extent of subchondral fracture line. We think that this is an important drawback of this system.

Catterall classification is being used widely for LCPD [15]. The reliability of this system is noted as moderate to good at initial stage and decreases to poor level at the fragmentation stage. It was also observed that experience was needed for better kappa values [19]. Sambandan et al. stated that identification of metaphyseal changes, subchondral fracture and making difference between involved and noninvolved epiphyseal location were found to be the difficulties about using Catterall classification [18]. In 2008, Wiig et al. simplified the Catterall classification system by staging it into two groups as < or > 50% involvement of capital femoral epiphysis due to poor reliability of four-staged classification [17].

Herring or lateral pillar classification is another widely used system which attaches importance to the height of lateral pillar of the proximal femoral epiphysis [13]. This classification system was also introduced to overcome the poor reliability of Catterall classification [1]. Ritterbusch et al. [21], Wiig et al. [17] and Sambandan et al. [18] all reported better reliabilities than other systems. Kalenderer et al. [15] also investigated whether experience had effect on this classification's reliability and concluded that experience did not affect the reliability at fragmentation stage but did affect at initial stage.

Laredo introduced arthrographic classification of LCPD after evaluating 87 patients' 105 hips in 1992 [14]. This classification system has several advantages such as possibility of dynamic evaluation of the hip joint and usage throughout the course of the disease [28]. Laredo classification has additionally a prognostic value about the progression of LCPD. It was stated that 68% of Laredo class I–II and III cases healed as Stulberg stage I or II and on the other hand 87% of Laredo

class IV or V cases healed as Stulberg stage III and IV [28]. We routinely use arthrography not only for diagnosis of LCPD, but also for decision making of treatment modality. We investigated the intra- and inter-observer reliabilities of Laredo classification and as a result observed that both reliabilities were all at excellent level except intra-observer reliability of pediatric orthopedists which was at good level. Experience did not affect the level of inter-observer reliability. Intra-observer reliability among pediatric orthopedists was observed as lower than senior and resident groups. We could not connect this result for any reason.

The most important limitation of this study is that the reliability of Catterall, Salter and Thompson, Herring classifications on the same patients' radiographs was not evaluated, so comparison of the reliabilities could not be evaluated.

Although percentage agreement and κ statistics were found to be fair for intra-observer reliability, ICC scores showed excellent reliability for each intra- and inter-observer evaluation. As a conclusion, our results showed that Laredo's arthrographic classification system's intra-observer reliability is at least at a fair level and inter-observer reliability is at an excellent level. We believe that this classification system is valuable for an orthopedic surgeon who deals with the treatment of LCPD.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Rampal V, Clément JL, Solla F (2017) Legg–Calvé–Perthes disease: classifications and prognostic factors. Clin Cases Miner Bone Metab 14(1):74–82
2. Herring JA, Kim HT, Browne R (2004) Legg–Calvé–Perthes disease: part I: classification of radiographs with use of the modified lateral pillar and Stulberg classifications. J Bone Joint Surg 86-A:2103–2120
3. Mose K (1980) Methods of measuring in Legg–Calvé–Perthes disease with special regard to the prognosis. Clin Orthop Relat Res 150:103–109
4. Grasemann H, Nicolai RD, Patsalis T, Hovel M (1997) The treatment of Legg–Calvé–Perthes disease. To contain or not to contain. Arch Orthop Trauma Surg 116:50–54
5. Herring JA (1994) The treatment of Legg–Calvé–Perthes disease. Current concepts review. J Bone Joint Surg Am 76:448–458
6. Chaudhry S, Phillips D, Feldman D (2014) Legg–Calvé–Perthes Disease: an overview with recent literature. Bull Hosp Jt Dis 72(1):18–27
7. de Sanctis N, Rondinella F (2000) Prognostic evaluation of Legg–Calvé–Perthes disease by MRI. Part II: pathomorphogenesis and new classification. J Pediatr Orthop 20(4):463–470
8. Jaramillo D, Galen TA, Winalski CS (1999) Legg–Calvé–Perthes disease: MR imaging evaluation during manual positioning of the hip: comparison with conventional arthrography. Radiology 12:519–525
9. Dimeglio A, Canavese F (2011) Imaging in Legg–Calvé–Perthes disease. Orthop Clin N Am 42:297–302
10. Catterall A (1971) The natural history of Perthes' disease. J Bone Joint Surg 53-B:37–53
11. Salter RB, Thompson GH (1984) 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 66:479–489
12. Stulberg SD, Cooperman DR, Wallensten R (1981) The natural history of Legg–Calve–Perthes disease. J Bone Joint Surg 63-A:1095–1108
13. Herring JA, Neustadt JB, Williams JJ, Early JS, Browne RH (1992) The lateral pillar classification of Legg–Calvé–Perthes disease. J Pediatr Orthop 12:143–150
14. Laredo FJ (1992) Legg–Calvé–Perthes disease: arthrographic classification. Rev Bras Orthop 27:7–10
15. Kalenderer O, Agus H, Ozcalabi IT, Ozluk S (2005) The importance of surgeons' experience on intraobserver and interobserver reliability of classifications used for Perthes disease. J Pediatr Orthop 25:460–464
16. Agus H, Kalenderer O, Eryanılmaz G, Ozcalabi IT (2004) Intraobserver and interobserver reliability of Catterall, Herring, Salter–Thompson and Stulberg classification systems in Perthes disease. J Pediatr Orthop 13:166–169
17. Wiig O, Terjesen T, Svenningsen S (2002) Interobserver reliability of radiographic classifications and measurements in the assessment of Perthes disease. Acta Orthop Scand 73:523–530
18. Sambandan NS, Gul A, Shankar R, Goni V (2006) Reliability of radiological classifications used in Legg–Calvé–Perthes disease. J Pediatr Orthop 15:267–270
19. Mahadeva D, Chong M, Langton DJ, Turner AM (2010) Reliability and reproducibility of classification systems for Legg–Calvé–Perthes disease: a systematic review of the literature. Acta Orthop Belg 76:48–57
20. Simmons ED, Graham HK, Szalai JP (1990) Interobserver variability in grading Perthes disease. J Bone Joint Surg 72-B:202–204
21. Ritterbusch JF, Shantaram SS, Gelinas C (1993) Comparison of lateral pillar classification and Catterall classification of Legg–Calvé–Perthes disease. J Pediatr Orthop 13:200–202
22. Fleiss JL (1986) Reliability of measurement. In: Fleiss JL (ed) The design and analysis of clinical experiments. Wiley, New York, pp 1–32
23. Rae G (1988) The equivalence of multiple rater kappa statistics and intraclass correlation coefficients. Educ Psychol Meas 48:367–374
24. Fleiss JL, Cohen J (1973) The equivalence of weighted kappa and intraclass correlation coefficient as measures of reliability. Educ Psychol Meas 33:613–619
25. Altman DB (1999) Practical Statistics for Medical Research. Chapman & Hall, London
26. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174
27. Karanicolas PJ, Bhandari M, Kreder H, Moroni A, Richardson M, Walter SD, Norman GR, Guyatt GH, Collaboration for Outcome Assessment in Surgical Trials (COAST) Musculoskeletal Group (2009) Evaluating agreement: conducting a reliability study. J Bone Joint Surg Am 91(Suppl 3):99–106
28. Milani C, Dobashi ET (2011) Arthrogram in Legg–Calvé–Perthes disease. J Pediatr Orthop 31(Suppl 2):56–62