Clin Orthop Relat Res (2012) 470:2376–2382
DOI 10.1007/s11999-012-2314-7

Clinical Orthopaedics and Related Research®
A Publication of The Association of Bone and Joint Surgeons®

SYMPOSIUM: LEGG-CALVÉ-PERTHES DISEASE: WHERE DO WE STAND AFTER 100 YEARS?

Reliability and Stability of Three Common Classifications for Legg-Calvé-Perthes Disease

Moon Seok Park MD, Chin Youb Chung MD, Kyoung Min Lee MD, Tae Won Kim MD, Ki Hyuk Sung MD

Published online: 20 March 2012

Abstract

Background To predict the course of Legg-Calvé-Perthes disease (LCPD) and select between treatment options in the early stages, it is critical to have a reliable predictive classification.

Questions/purposes We examined the reliability and stability of three common classification systems for LCPD.

Methods We identified 69 patients with LCPD, who had hip radiographs taken more than twice after the initial presentation with at least a 3-month interval. The Herring lateral pillar, Catterall, and Salter-Thompson classifications were evaluated in terms of reliability and stability. The inter- and intrarater reliability of the classification systems was determined by three orthopaedic surgeons using intraclass correlation coefficients (ICCs). To evaluate the stability of the classification systems, the percentage agreement and ICCs among the initial rating, rating when entering fragmentation, and final rating were used.

Results The interrater reliability was highest in Herring lateral pillar classification (ICC, 0.885) followed by the Catterall and Salter-Thompson classifications (ICC, 0.802 and 0.702, respectively). The percentage agreement and ICC between the initial and final rating were, respectively, 55% and 0.491 for the Herring classification and 48% and 0.378 for the Catterall classification.

Conclusions Our data show the highest reliability of the classification of Herring et al. However, more than 40% of the hip radiographs at the initial presentation, and in particular, most of Herring Group A patients, were upgraded. Therefore, for patients older than 8 years old and graded as Herring Group A initially, surgeons should keep the possibility of surgical treatment in mind.

Level of Evidence Level I, diagnostic study. See Guidelines for Authors for a complete description of levels of evidence.

Introduction

Legg-Calvé-Perthes disease (LCPD) is an osteochondrosis that affects the proximal epiphysis of the femur. The clinical course and prognosis of LCPD vary considerably between different patients. Without proper treatment, LCPD can cause a severe femoral head deformity or subluxation as well as secondary osteoarthritis resulting from aspherical incongruency of the hip. Therefore, an accurate interpretation of radiographs is essential for determining the therapeutic choices in LCPD. These choices include observation, nonsurgical containment using orthotic devices [25], surgical containment procedures including femoral osteotomy [34] and innominate osteotomy [30], and salvage procedures including Chiari osteotomy [5] and shelf acetabuloplasty [18].

Each author certifies that he or she, or a member of their immediate family, has no commercial associations (eg, consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that might pose a conflict of interest in connection with the submitted article.
All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research* editors and board members are on file with the publication and can be viewed on request.
Each author certifies that his or her institution has approved the human protocol for this investigation, that all investigations were conducted in conformity with ethical principles of research, and that informed consent for participation in the study was obtained.

M. S. Park, C. Y. Chung, K. M. Lee, T. W. Kim, K. H. Sung (✉)
Department of Orthopaedic Surgery, Seoul National University Bundang Hospital, 300 Gumi-Dong, Bundang-Gu, Sungnam, Kyungki 463-707, Korea
e-mail: skh1219@naver.com

The first classification of LCPD was described by Waldenström [36] in 1922, who divided it into four stages (initial, fragmentation, reossification, and remodeling stage) representing only the natural course of the disease using the radiographic changes in the femoral head. More recent systems were proposed to classify the radiographic findings in a way that correlates with the disease severity, those of Catterall [4], Salter and Thompson [31], and Herring et al. [15]. In 1971, Catterall [4] suggested a four-group classification based on the degree of epiphyseal involvement and described head-at-risk signs. In 1984, Salter and Thompson [31] proposed a simple and practical two-group classification based on the extent of the subchondral fracture. In 1992, Herring et al. [15] suggested a classification based on the height of the lateral pillar of the involved epiphysis. In 2004, Herring et al. [13] added a further group to their classification, termed the B/C border group, because their multicenter study showed many affected hips have a radiographic appearance that is intermediate between Types B and C.

Stability refers to the extent to which children remain in the same classification level over time [23]. Kuroda et al. [19] reported that the lateral pillar grade at the initial examination did not change 18 months after initial treatment in the patients who received nonweightbearing treatment. On the other hand, Lappin et al. [20] reported that almost three-fourths of Herring Grade A cases and one-third of Herring Grade B cases required upgrading within the first 7 months from the onset of symptoms. Poor stability indicates progression of the disease, that is, the classification is normally upgraded and more aggressive treatment is needed. If the range of stability of these classification systems can be established, it would be possible to predict the course of the disease in the early stages and help determine the treatment plan.

The aim of this study was to evaluate the reliability and stability of the classification system for LCPD.

Materials and Methods

We retrospectively reviewed 114 patients with LCPD. We included only patients with two or more sets of radiographs taken a minimum of 3-month intervals from 2003 to 2011. We excluded 45 patients (1) who visited our hospital only once before surgery such as femoral osteotomy or pelvic osteotomy; and (2) those classified as the reossification and remodeling stage on the Waldenström classification at the initial visit. These exclusions left 69 patients in the study. There were 60 males and nine females. Sixty-six and three patients were unilaterally and bilaterally involved, respectively. The mean age was 6.5 years (range, 2.3–11.3 years) and the mean time between the initial and final ratings was 1.2 ± 0.7 (SD) years. Of the 69 patients, the radiographs that were at reossification and remodeling stages were excluded. A total 379 ratings were made, a mean of 5.5 per child (SD 2.1). The stages on the initial films using the Waldenström stage were as follows: initial stage, 41 (59%); and fragmentation stage, 28 (41%). This retrospective study was approved by the Institutional Review Board at our hospital, which waived the need for informed consent for the study. We obtained all data only from medical record review and hip radiographs.

For the purpose of statistical independence, only one hip from patients with bilateral involvement was included for statistical analysis [24]. Prior precision analysis, which is used in studies concerned with estimating some parameter at a fixed confidence level, was performed to identify the minimal sample size required for the analysis. This study was designed to enable intraclass coefficients (ICCs) of reliability to be calculated at a target value of 0.8. In addition, we used the approximation by Bonett [3]. Accordingly, when we set the 95% confidence interval to 0.2 for three raters, the minimal sample size was calculated to be 36 hips.

We obtained demographic data such as gender and age at each visit from a review of the medical records. Hip AP and frog leg lateral radiographs of each patient were retrieved digitally using a picture archiving and communication system (IMPAX; Agfa, Antwerp, Belgium).

A consensus-building session was held by the four orthopaedic surgeons before radiographic measurements were taken. The original articles by Catterall, Salter and Thompson, and Herring et al. (hereinafter referred to as the lateral pillar classification) were reviewed by all surgeons and a consensus about how to grade was reached on these classification systems as mentioned subsequently. The classification of Herring et al. were determined from the height of the lateral portion of the femoral head as observed on the AP view at the fragmentation stage. The classification consists of four groups: A, B, B/C border, and C. In Group A, height of the lateral pillar was radiographically normal compared with the contralateral hip. In Group B, height of the lateral pillar was between 50 and 100% of the original height [13, 15]. Despite a good understanding of the classification system, even pediatric orthopaedic surgeons might have difficulty in distinguishing Grades A and B. The lateral one-fourth of the head was chosen to represent the lateral pillar [13]. On the other hand, actual measurements of the height showed there were no cases without any collapse of the lateral pillar in our study, ie, there was no completely intact lateral pillar. Therefore, a consensus was established in that a similar height of the lateral pillar to the contralateral hip is the criteria of lateral pillar A. The Catterall classification system was based on the level of capital femoral epiphysis involvement during

the fragmentation stage on both the AP and frog leg lateral views. The classification consisted of four groups: I, II, III, and IV [4]. In Group I, only the anterior part of the epiphysis was affected. In Group II, more of the anterior segment was involved and a central sequestrum was present. In Group IIII, most of the epiphysis was sequestrated and in Group IV, the whole epiphysis was sequestrated. We determined the Salter-Thompson [31] classification system from the extent of the subchondral fracture of the femoral dome observed during the initial stages on the primarily frog leg lateral view. The subchondral fracture might be absent in many patients because it appears briefly in the course of the radiographic changes and is seen in only a minority of patients.

One of the authors (TWK) randomly selected the 36 radiographs for a reliability session. Three examiners (MSP, KML, KHS) with 11, 9, and 7 years of orthopaedic experience including training periods, respectively, assessed the interrater and intrarater reliability of the radiographic classifications. The measurements were performed by the three examiners in two sessions with a 3-week interval between sessions and the order of the radiographs was changed for the second sessions. The examiners were asked to determine the classifications on 36 radiographs and were blinded to the other measurements. They also had the option to state that a radiograph was unclassifiable. All measurements were collected by another orthopaedic surgeon (TKW) who screened the radiographs for reliability.

The stability of the classification was assessed by one of the authors (KHS). He determined the classifications on all radiographs, which were at the initial and fragmentation stages on the Waldenström classification, and examined the change in classifications. Initial grading was performed at the initial visit, and final grading was performed before the reossification stage.

The ICC and their 95% CIs were used to summarize the interrater and intrarater reliability and were calculated in the setting of a two-way random effect model assuming a single measurement and absolute agreement. An ICC value of 1 indicates perfect reliability and an ICC of > 0.8 indicates excellent reliability [9]. We analyzed the stability of the classifications of Herring et al. and Catterall using the percentage agreement and ICCs with 95% CIs among the initial rating, rating at early fragmentation, and final ratings. Statistical analyses were conducted using SPSS for Windows (Version 15.0; SPSS, Chicago, IL, USA).

Results

For the classification of Herring et al., the interrater reliability, the ICC was 0.894 (95% CI, 0.825–0.940, Session 1) and 0.888 (95% CI, 0.782–0.943, Session 2) (Table 1). The intrarater reliabilities were ICC values of 0.907 (95% CI, 0.809–0.953, Examiner 1), 0.951 (95% CI, 0.906–0.975, Examiner 2), and 0.908 (95% CI, 0.770–0.958, Examiner 3) (Table 2). The patients' initial rating by the classification of Herring et al. was as follows: A, 30; B, 30; B/C, one; and C, eight. The percentage agreement and ICC between the initial and final rating were 55% and 0.491, respectively (95% CI, 0.116–0.709) (Table 3). Twenty-five of 30 patients classified as A at the initial rating were reclassified (83%) within the first 4.3 months from the initial grading. Between the rating at early fragmentation and the final rating, the percentage agreement and ICC were 79% and 0.714, respectively (95% CI, 0.507–0.833) (Table 4).

For the Catterall classification, the interrater reliability was ICC values of 0.805 (95% CI, 0.690–0.887, Session 1) and 0.796 (95% CI, 0.676–0.882, Session 2) (Table 1). For intrarater reliability, the ICC was 0.915 (95% CI, 0.840–0.955, Examiner 1), 0.813 (95% CI, 0.655–0.901, Examiner 2), and 0.810 (95% CI, 0.659–0.899, Examiner 3) (Table 2). The initial rating by the Catterall classification was as follows: I, nine; II, 19; III, 13; and IV, 24. The percentage agreement and ICC between the initial and final rating were 48% and 0.378, respectively (95% CI, −0.006

Table 1. Interrater reliability of classification systems for Legg-Calve-Perthes disease

Classification system	ICC	95% CI
Herring (Session 1)	0.894	0.825–0.940
Herring (Session 2)	0.888	0.782–0.943
Catterall (Session 1)	0.805	0.690–0.887
Catterall (Session 2)	0.796	0.676–0.882
Salter-Thompson (Session 1)	0.633	0.461–0.775
Salter-Thompson (Session 2)	0.735	0.589–0.844

ICC = intraclass correlation coefficient.

Table 2. Intrarater reliability of classification systems for Legg-Calve-Perthes disease

Classification system	ICC	95% CI
Herring (Examiner 1)	0.907	0.809–0.953
Herring (Examiner 2)	0.951	0.906–0.975
Herring (Examiner 3)	0.908	0.770–0.958
Catterall (Examiner 1)	0.915	0.840–0.955
Catterall (Examiner 2)	0.813	0.655–0.901
Catterall (Examiner 3)	0.810	0.659–0.899
Salter-Thompson (Examiner 1)	0.823	0.658–0.909
Salter-Thompson (Examiner 2)	0.751	0.562–0.865
Salter-Thompson (Examiner 3)	0.659	0.424–0.811

ICC = intraclass correlation coefficient.

Table 3. Herring classification for initial and final rating of patients with Legg-Calve-Perthes disease

Initial rating	Final rating				
	A	B	B/C	C	Total
A	5	19	3	3	30
B	–	24	3	3	30
B/C	–	–	1	–	1
C	–	–	–	8	8
Total	5	43	7	14	69

Percent agreement, 55%; intraclass correlation coefficient, 0.491 (95% CI, 0.116–0.709).

Table 4. Herring classification for rating at early fragmentation and final rating of patients with Legg-Calve-Perthes disease

Rating at early fragmentation	Final rating				
	A	B	B/C	C	Total
A	2	6	1	–	9
B	–	33	2	4	39
B/C	–	–	4	–	4
C	–	–	–	9	9
Total	2	39	7	13	61

Percent agreement, 79%; intraclass correlation coefficient, 0.714 (95% CI, 0.507-0.833).

Table 5. Catterall classification for initial and final rating of patients with Legg-Calve-Perthes disease

Initial rating	Final rating				
	I	II	III	IV	Total
I	1	2	4	2	9
II	–	–	10	9	19
III	–	–	6	7	13
IV	–	–	–	24	24
Total	1	2	20	42	65

Percent agreement, 48%; intraclass correlation coefficient, 0.378 (95% CI, −0.006 to 0.634).

to 0.634) (Table 5). Twenty-four of 31 patients classified as I, II, and III at the initial rating were reclassified (77%) within the first 4.4 months from the initial grading. Between the rating at early fragmentation and the final rating, the percentage agreement and ICC were 79% and 0.528, respectively (95% CI, 0.274–0.705) (Table 6).

For the Salter-Thompson classification, the intrarater reliability was ICC values of 0.633 (95% CI, 0.461–0.775, Session 1) and 0.735 (95% CI, 0.589–0.844, Session 2) (Table 1). For interrater reliability, the ICC was 0.823 (95% CI, 0.658–0.909, Examiner 1), 0.751 (95% CI, 0.562–0.865,

Table 6. Catterall classification for rating at early fragmentation and final rating of patients with Legg-Calve-Perthes disease

Rating at early fragmentation	Final rating				
	I	II	III	IV	Total
I	–	–	–	–	0
II	–	1	–	2	3
III	–	–	16	10	26
IV	–	–	–	28	28
Total	0	1	16	40	57

Percent agreement, 79%; intraclass correlation coefficient, 0.528 (95% CI, 0.274–0.705).

Examiner 2), and 0.659 (95% CI, 0.424–0.811, Examiner 3) (Table 2). Thirty-one cases (45%) were classified by the Salter-Thompson classification at the initial ratings. The patients' initial rating by the Salter-Thompson classification was as follows: A, 17; B, 14. The percentage agreement and ICC between the initial and final rating were 84% and 0.691, respectively (95% CI, 0.443–0.841).

Discussion

Surgeons typically develop and use classifications to either select between treatment alternatives or to predict the patient's course. Some authors suggest the ability to predict the prognosis is the most important aspect of any classification system [21, 33]. A classification system should be relatively easy to apply and should have acceptable reproducibility and reliability. It is also important that the classification be reasonably stable; that is, the class into which a patient fits does not change over short periods of time. We therefore examined the reliability and stability of three common classification systems for LCPD.

Before discussing the clinical implications of this study, it is important to address the limitations. First, we excluded five patients who visited our outpatient clinic only once and underwent subsequent surgery. Therefore, we focused on less severely involved patients. Second, only one of the authors graded radiographs for stability. It may show better results if more raters are involved in the measurements. However, the interrater reliability from the classifications of Herring et al. and Catterall had ICC values of 0.888–0.894 and 0.796–0.805, respectively. Thus, we believe our study design to measure the stability by using only one examiner can also be justified. Third, 24 children were initially classified as IV by Catterall. Catterall's Group IV represents the upper limit of the classification that could not be upgraded. This may have overestimated the stability of this classification system.

We found the classification of Herring et al. has the highest interrater and intrarater reliabilities, whereas the

Salter-Thompson classification had the lowest. The Herring lateral pillar and Catterall grade at the initial radiographs were upgraded in over 40% of patients.

Previous studies that examined the three common classifications demonstrated highly variable reliabilities (Table 7). The Catterall classification system is used widely but there has been some criticism. Because the classification is based on the epiphyseal involvement at the time of maximum resorption, which occurs relatively late, the grouping will appear to change if the classification is applied too early [35]. An assessment using the Catterall classification is difficult and requires an experienced pediatric orthopaedic surgeon. The distinction between Groups II and III is the difficulty with this classification [6, 12, 26]. Low interobserver and intraobserver reliability (kappa values ranging from 0.44 to 0.67 and percentage agreement ranging from 30 to 53%) has been reported by a number of authors [6, 12, 28, 32, 33] . Others have noted the absence of a prognostic value of the Catterall classification [11, 16]. The interobserver reliability of 0.13–0.55 (ICC) of the head-at-risk signs has also been reported [10]. On the other hand, De et al. [7] reported a reliability of 0.94 (ICC) of the Catterall classification and 0.66–0.85 (ICC) of the head-at-risk signs among pediatric orthopaedic surgeons. The reliability of the Catterall classification could be increased after consensus-building between the

Table 7. Previous studies on the reliability of the classification for Legg-Calve-Perthes disease

Study	Number of subjects	Raters	Classification	Statistics		
				ICC	Weighted kappa	Percentage agreement
Sambandam et al. [32]	44	2 (interobserver)	Herring		0.722	
		2 (intraobserver			0.682	
		2 (interobserver)	Catterall		0.443	
		2 (intraobserver)			0.478	
		2 (interobserver)	Salter-Thompson		0.163	
		2 (intraobserver)			0.503	
Agus et al. [1]	10	18 (interobserver)	Herring	0.388–0.596		
		18 (intraobserver)		0.113–0.495		
		18 (interobserver)	Catterall	0.578–0.620		
		18 (intraobserver)		0.387–0.686		
		18 (interobserver)	Salter-Thompson	0.604		
		18 (intraobserver)		0.576		
Herring et al. [13]	20	6 (interobserver)	Herring	0.70–0.85	0.71–0.79	
		6 (intraobserver)		0.91	0.81	
Akgun et al. [2]	50	3 (interobserver)	Herring		0.526–0.539	
Pietrzak et al. [26]	63	3 (interobserver)	Herring		0.65–0.70	
		3 (intraobserver)			0.48–0.80	
		3 (interobserver)	Catterall		0.28–0.42	
		3 (intraobserver)			0.43–0.61	
De et al. [7]	23	9 (interobserver)	Catterall	0.94		
Wiig et al. [37]	Variable (63–158)	3 (interobserver)	Herring		0.56–0.70	
		3 (interobserver)	Catterall		0.49–0.62	
		3 (interobserver)	Salter-Thompson		0.54–0.63	
Podeszwa et al. [27]	33	5 (interobserver)	Herring		0.510	
		5 (intraobserver)			0.742	
Ritterbusch et al. [28]	78	3 (interobserver)	Herring			72
			Catterall			41
Simmons et al. [33]	40	15 (interobserver)	Catterall		0.49–0.64	
			Salter-Thompson		0.49–0.99	
Christensen et al. [6]	100	4 (interobserver)	Catterall		0.50–0.67	
Hardcastle et al. [12]	69	10 (interobserver)	Catterall			30–53

ICC = intraclass correlation coefficient.

examiners, like in the present study. The Salter-Thompson classification system is simple to use and can be applied earlier in the course of the disease. Kalender et al. [17] reported that the Salter-Thompson classification system was most reliable, but Sambandam et al. [32] showed the lowest interrater and intrarater reliabilities. The major drawback with this system is that it cannot be used when a subchondral fracture can no longer be detected, and it is reportedly present in only 15%–34% of the patients' radiographs [15, 16, 31, 37]. We found 31 patients (45%) with a subchondral fracture seen on radiographs for a short duration. Stability of this classification was highest among three classification systems, but we believe this fact has no clinical implications as a result of the shortcomings of this system. In this study, the interrater reliability of this classification was the lowest among the three classification systems. This is the result of disagreement between the observers when the fracture line was actually present and the borderline cases in whom the subchondral fracture line involves approximately half of the femoral head. The Salter-Thompson classification system is not applicable to all cases but might be helpful in predicting the prognosis. The classification system of Herring et al. is relatively easy to apply and requires only an AP radiograph taken during the fragmentation stage of the disease. This classification was reported to have good to excellent intra- and interobserver reliability [8, 13, 15, 20, 27, 28, 32, 37]. Previous studies reported that the classification of Herring et al. correlates with the final radiographic classification [8, 14–16, 29, 38], and its prognostic value appears to be better than the Catterall classification [28]. Gigante et al. [11] reported that the classification of Herring et al. was not predictive when considered alone but became prognostic when it was related to the age at onset.

Few papers examined the changes in the classification of Herring et al. between the initial and final grade. Kuroda et al. [19] reported that because the lateral pillar grade did not change in patients who received nonweightbearing treatment, the grade determined at the beginning of treatment is believed to be associated with the final radiographic outcome. On the other hand, Lappin et al. [20] noted approximately 75% of Grade A cases and 30% of the Grade B cases required upgrading within the first 7 months from the onset of symptoms. Therefore, predicting the outcome in the early stages of the disease remains difficult. Meurer et al. [22] compared the prognostic value of the classifications of Herring et al. and Catterall and observed the need for upgrading in approximately 30% of cases using both classifications. In the our study, the Herring lateral pillar and Catterall grade at the initial radiographs were upgraded in 45% and 52% of patients, respectively. The ratings at the early fragmentation stage were upgraded in 21% and 21%, respectively. These results indicate the limitations of these two classifications in the early stages of the disease, as Lappin et al. reported. Nevertheless, the classification of Herring et al. and the Catterall classifications could be changed even after fragmentation.

In general, it has been known that patients with Herring Group A and those with Group B with a disease onset before 8 years of age need only symptomatic treatment. However, patients with a Herring Group B or B/C border with a disease onset after 8 years of age need surgical containment [14]. In the present study, 25 of 30 patients classified as A at initial rating were reclassified (83%) within the first 4.3 months. Therefore, even if initial radiographs were graded as Group A, more aggressive treatment could be indicated for hips with a high probability of upgrading. Further study is needed to identify the risk factors of upgrading.

We found the classification of Herring et al. has the highest interrater and intrarater reliability among three classification systems. However, more than 40% of patients, and in particular, most of the Herring Group A patients, needed their classifications upgraded based on initial radiographs. Therefore, for the patients older than 8 years old and graded as Herring Group A initially, surgeons should consider the possibility of the patient needing subsequent surgical treatment.

Acknowledgments We thank Mi Sun Ryu BS for technical support.

References

1. Agus H, Kalenderer O, Eryanlmaz G, Ozcalabi IT. Intraobserver and interobserver reliability of Catterall, Herring, Salter-Thompson and Stulberg classification systems in Perthes disease. *J Pediatr Orthop B*. 2004;13:166–169.
2. Akgun R, Yazici M, Aksoy MC, Cil A, Alpaslan AM, Tumer Y. The accuracy and reliability of estimation of lateral pillar height in determining the Herring grade in Legg-Calve-Perthes disease. *J Pediatr Orthop*. 2004;24:651–653.
3. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med*. 2002;21:1331–1335.
4. Catterall A. The natural history of Perthes' disease. *J Bone Joint Surg Br*. 1971;53:37–53.
5. Chiari K. Medial displacemen osteotomy of the pelvis. *Clin Orthop Relat Res*. 1974;98:55–71.
6. Christensen F, Soballe K, Ejsted R, Luxhoj T. The Catterall classification of Perthes' disease: an assessment of reliability. *J Bone Joint Surg Br*. 1986;68:614–615.
7. De BB, Viel JF, Monnet E, Garnier E, Aubert D. Interobserver reliability in the interpretation of radiologic signs in Legg-Calve-Perthes disease. *J Pediatr Orthop B*. 2002;11:10–14.
8. Farsetti P, Tudisco C, Caterini R, Potenza V, Ippolito E. The Herring lateral pillar classification for prognosis in Perthes disease. Late results in 49 patients treated conservatively. *J Bone Joint Surg Br*. 1995;77:739–742.
9. Fleiss J. *Statistical Methods for Rates and Proportions*. 2nd ed. New York, NY, USA: John Wiley and Sons; 1981.

10. Forster MC, Kumar S, Rajan RA, Atherton WG, Asirvatham R, Thava VR. Head-at-risk signs in Legg-Calve-Perthes disease: poor inter- and intra-observer reliability. *Acta Orthop.* 2006;77:413–417.
11. Gigante C, Frizziero P, Turra S. Prognostic value of Catterall and Herring classification in Legg-Calve-Perthes disease: follow-up to skeletal maturity of 32 patients. *J Pediatr Orthop.* 2002;22: 345–349.
12. Hardcastle PH, Ross R, Hamalainen M, Mata A. Catterall grouping of Perthes' disease. An assessment of observer error and prognosis using the Catterall classification. *J Bone Joint Surg Br.* 1980;62:428–431.
13. 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;86:2103–2120.
14. Herring JA, Kim HT, Browne R. Legg-Calve-Perthes disease. Part II: prospective multicenter study of the effect of treatment on outcome. *J Bone Joint Surg Am.* 2004;86:2121–2134.
15. Herring JA, Neustadt JB, Williams JJ, Early JS, Browne RH. The lateral pillar classification of Legg-Calve-Perthes disease. *J Pediatr Orthop.* 1992;12:143–150.
16. Ismail AM, Macnicol MF. Prognosis in Perthes' disease: a comparison of radiological predictors. *J Bone Joint Surg Br.* 1998;80:310–314.
17. Kalender 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;25:460–464.
18. Kruse RW, Guille JT, Bowen JR. Shelf arthroplasty in patients who have Legg-Calve Perthes disease. A study of long-term results. *J Bone Joint Surt Am*. 1991;73:1338–1347.
19. Kuroda T, Mitani S, Sugimoto Y, Asaumi K, Endo H, Akazawa H, Nakagomi T, Ozaki T. Changes in the lateral pillar classification in Perthes' disease. *J Pediatr Orthop B.* 2009;18:116–119.
20. Lappin K, Kealey D, Cosgrove A. Herring classification: how useful is the initial radiograph? *J Pediatr Orthop.* 2002;22:479–482.
21. Mahadeva D, Chong M, Langton DJ, Turner AM. Reliability and reproducibility of classification systems for Legg-Calve-Perthes disease: a systematic review of the literature. *Acta Orthop Belg.* 2010;76:48–57.
22. Meurer A, Schwitalle M, Humke T, Rosendahl T, Heine J. Comparison of the prognostic value of the Catterall and Herring classification in patients with Perthes disease. *Z Orthop Ihre Grenzgeb.* 1999;137:168–172.
23. Palisano RJ, Cameron D, Rosenbaum PL, Walter SD, Russell D. Stability of the gross motor function classification system. *Dev Med Child Neurol.* 2006;48:424–428.
24. Park MS, Kim SJ, Chung CY, Choi IH, Lee SH, Lee KM. Statistical consideration for bilateral cases in orthopaedic research. *J Bone Joint Surg Am.* 2010;92:1732–1737.
25. Petrie JG, Bitenc I. The abduction weight-bearing treatment in Legg-Perthes' disease. *J Bone Joint Surg Br.* 1971;53: 54–62.
26. Pietrzak S, Napiontek M, Tomaszewski M. Catterall and Herring classifications in assessing Perthes disease—inter- and intra-observer study. *Ortop Traumatol Rehabil.* 2004;6:561–566.
27. 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: 562–565.
28. Ritterbusch JF, Shantharam SS, Gelinas C. Comparison of lateral pillar classification and Catterall classification of Legg-Calve-Perthes' disease. *J Pediatr Orthop.* 1993;13:200–202.
29. Rosenfeld SB, Herring JA, Chao JC. Legg-Calve-Perthes disease: a review of cases with onset before six years of age. *J Bone Joint Surg Am.* 2007;89:2712–2722.
30. Salter RB. The present status of surgical treatment for Legg-Perthes disease. *J Bone Joint Surg Am.* 1984;66:961–966.
31. Salter RB, Thompson GH. Legg-Calve-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;66:479–489.
32. Sambandam NS, Gul A, Shankar R, Goni V. Reliability of radiological classifications used in Legg-Calve-Perthes disease. *J Pediatr Orthop B.* 2006;15:267–270.
33. Simmons ED, Graham HK, Szalai JP. Interobserver variability in grading Perthes' disease. *J Bone Joint Surg Br.* 1990;72:202–204.
34. Than P, Halmai V, Shaikh S, Kranicz J, Bellyei A. Long-term results of derotational femoral varus osteotomy in Legg-Calve-Perthes disease: 26-years follow-up. *Orthopedics*. 2003;26: 487–491.
35. Van Dam BE, Crider RJ, Noyes JD, Larsen LJ. Determination of the Catterall classification in Legg-Calve-Perthes disease. *J Bone Joint Surg Am.* 1981;63:906–914.
36. Waldenström H. The difinite form of the coxa planaa. *Acta Radiol*. 1922;1:384–394.
37. Wiig O, Terjesen T, Svenningsen S. Inter-observer reliability of radiographic classifications and measurements in the assessment of Perthes' disease. *Acta Orthop Scand.* 2002;73:523–530.
38. Wiig O, Terjesen T, Svenningsen S. Prognostic factors and outcome of treatment in Perthes' disease: a prospective study of 368 patients with five-year follow-up. *J Bone Joint Surg Br.* 2008; 90:1364–1371.