

Reliability of the modified lateral pillar classification for Legg Calvé Perthes disease performed by a large group of international paediatric orthopaedic surgeons

Jennifer C. Laine^{1,2}
 Susan A. Novotny^{1,3}
 Stefan Huhnstock⁴
 Andrew J. Ries¹
 John E. Tis⁵
 Wudbhav N. Sankar⁶
 Chan-Hee Jo⁷
 Harry K. W. Kim^{7,8}
 International Perthes Study Group⁹

Abstract

Purpose: The modified lateral pillar classification (mLPC) is used for prognostication in the fragmentation stage of Legg Calvé Perthes disease. Previous reliability assessments of mLPC range from fair to good agreement when evaluated by a small number of observers with pre-selected radiographs. The purpose of this study was to determine the inter-observer and intra-observer reliability of mLPC performed by a group of international paediatric orthopaedic surgeons. Surgeons self-selected the radiograph for mLPC assessment, as would be done clinically.

Methods: In total, 40 Perthes cases with serial radiographs were selected. For each case, 26 surgeons independently selected a radiograph and assigned mLPC and 21 raters re-evaluated the same 40 cases to establish intra-observer reliability. Rater performance was determined through surgeon consensus using the mode mLPC as 'gold standard'. Inter-observer and intra-observer reliability data were analysed using weighted kappa statistics.

Results: The weighted kappa for inter-observer correlation for mLPC was 0.64 (95% confidence interval: 0.55 to 0.74) and was 0.82 (range: 0.35 to 0.99) for intra-observer correlation. Individual surgeon's overall performance varied from 48% to 88% agreement. Surgeon mLPC performance was not influenced by years of experience ($p = 0.51$). Radiograph selection did not influence gold standard assignment of mLPC. There was greater agreement on cases of mild B hips and severe C hips.

Conclusions: mLPC has low good inter-observer agreement when performed by a large number of surgeons with varied experience. Surgeons frequently chose different radiographs, with no impact on mLPC agreement. Further refinement is needed to help differentiate hips on the border of group B and C.

Level of evidence: III

Cite this article: Laine JC, Novotny SA, Huhnstock S, Ries AJ, Tis JE, Sankar WN, Jo C-H, Kim HKW, International Perthes Study Group. Reliability of the modified lateral pillar classification for Legg Calvé Perthes disease performed by a large group of international paediatric orthopaedic surgeons. *J Child Orthop* 2020;14: 529-536. DOI: 10.1302/1863-2548.14.200055

Keywords: lateral pillar classification; Legg Calvé Perthes disease; paediatric hip; inter-observer reliability; intra-observer reliability

Introduction

Patients with Legg Calvé Perthes disease present heterogeneously, in various stages and with a spectrum of severity. Extent of femoral head involvement in the active stages of the disease has been linked to the long-term prognosis of the hip. Multiple classification systems have been

¹ Gillette Children's Specialty Healthcare, Saint Paul, Minnesota, USA

² University of Minnesota Department of Orthopaedic Surgery, Minneapolis, Minnesota, USA

³ University of Minnesota Rehabilitation Science Graduate Program, Minneapolis, Minnesota, USA

⁴ Department for Children's Orthopedics and Reconstructive Surgery, Division of Orthopaedic Surgery, Oslo University Hospital, Norway

⁵ Department of Orthopaedic Surgery, John Hopkins School of Medicine, Baltimore, Maryland, USA

⁶ Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

⁷ Center for Excellence in Hip Disorders, Texas Scottish Rite Hospital for Children, Dallas, Texas, USA

⁸ Department of Orthopaedic Surgery, University of Texas Southwestern Medical Center, Dallas, Texas, USA

⁹ See appendix

developed in efforts to assess the severity of disease and to help predict outcomes. In order to be effective, a classification system should be reproducible and reliable, allowing for comparisons between patients and between studies. Catterall¹ classified femoral head severity into four groups during the fragmentation stage based on location and extent of avascularity. Reports of inter-observer reliability for this classification have been variable, ranging from fair to good.²⁻⁵ Salter and Thompson⁶ utilized the subchondral fracture line seen in the initial stage of disease to classify femoral head involvement. One practical limitation of this classification is that not all patients demonstrate this sign, making it applicable only to a subset of patients.

Unable to find good agreement with the Catterall classification system, Herring et al developed the lateral pillar classification (LPC) which was reported in 1992.⁷ They classified hips into three groups during the fragmentation stage of disease. The LPC divides the femoral head into three anatomic sectors and classifies hips based on the involvement of the lateral pillar alone on an anterior-posterior (AP) view. Group A has no lateral pillar involvement, Group B maintains greater than 50% of the original lateral pillar height, and Group C has > 50% collapse of the lateral pillar. Herring reported agreement 78% of the time between observers and an inter-observer kappa value of 0.52.⁷ Subsequent studies have shown better correlation of the LPC with radiographic outcome than the Catterall classification,^{2,8} and moderate to good inter-observer and intra-observer reliability of the LPC.^{2,3,8-10}

Due to difficulties finding consensus in classifying a specific group of hips that were more severe than typical Group B hips and less severe than most Group C hips, Herring described a modification of the LPC (mLPC) in 2004, adding a B/C border group.¹¹ The B/C group was defined as hips with any of the following radiographic findings: 1) very narrow lateral pillar maintaining > 50% height, 2) lateral pillar with very little ossification but > 50% original height, or 3) lateral pillar is depressed relative to central pillar but maintaining 50% of the original height. Herring reported good inter-observer agreement of the mLPC with a modified weighted kappa of 0.71,¹¹ but subsequent reports of inter-observer agreement have not been able to reproduce this level of agreement, finding only fair reliability (weighted kappa values of 0.39-0.40).^{2,12} Additionally, previous studies of reliability of the mLPC have pre-selected the fragmentation radiographs for rating and have included only a small group of observers (five to six).^{2,12} This does not reproduce the clinical environment in which surgeons with a wide variety of experience sort through serial radiographs, selecting the appropriate radiograph for assigning the mLPC in order to assess severity and predict prognosis.

The purpose of this study was to determine the inter-observer and intra-observer reliability of the mLPC performed

by a large group of international paediatric orthopaedic surgeons with varied clinical experience (two years to 43 years). To simulate clinical practice, each surgeon was asked to self-select the radiograph perceived most appropriate for mLPC assessment. A secondary purpose of the study was to determine if either radiograph selection or years of clinical experience affected rater performance.

Materials and methods

Radiographic assessment

In total, LCPD cases were selected from an international database of prospectively collected radiographs for long-term study of the disease. The appropriate number of cases was determined based on previous evaluations of mLPC reliability in the literature.^{2,11,12} Informed consent was obtained from all individual participants included in the study. Based on the inclusion criteria of the international database, all patients had onset of disease between their 6th and 11th birthdays and were diagnosed in the early stages of disease (modified Waldenstrom I-IIa).^{13,14} All the selected cases had AP and frog lateral radiographs at approximately three-month intervals from diagnosis until, at minimum, two-year follow-up, with a mean of 17.6 ± 3 (range 12 to 26) radiographs per case. All patients had also undergone perfusion magnetic resonance imaging (MRI) shortly after diagnosis, demonstrating at least 50% femoral head involvement.

An international group of 26 paediatric orthopaedic surgeons (24 centres, five countries: China, India, Norway, Sweden, United States) with varied orthopaedic experience (two years to 43 years) reviewed the images. All the observers have a specific interest in Legg Calvé Perthes disease and were familiar with the mLPC. A tutorial on the mLPC was given prior to radiographic review. The 26 observers were given the serial AP and frog lateral radiographs for the 40 cases as printed 8.5 x 11-inch images. The observers worked independently. Each observer self-selected the radiographic image that they perceived had maximal lateral pillar collapse in the fragmentation stage. They documented the radiograph selected and the mLPC assigned. The most frequently assigned lateral pillar for each case (i.e. the mode answer) was considered to be the 'gold standard' answer. The most frequently chosen radiograph that resulted in a gold standard answer for each case was considered the 'gold standard' radiograph. For example, for Case 14, the mode mLPC assignment was Group C. The 'gold standard' radiograph was the most commonly selected radiograph for Case 14 that resulted in the Group C classification.

For the determination of intra-observer reliability, 21 of the 26 paediatric orthopaedic surgeons completed two additional rounds of radiographic assessments, using a

secure, web-based application (Research Electronic Data Capture, REDCap).^{15,16} The decision to perform two additional rounds, rather than one, was due to methodological change in which mLPC was assessed using digital radiographs rather than printed radiographs. All other aspects of the study (e.g. training, cases and documentation procedures) remained consistent with the initial study.

The two web-based rounds were performed at a minimum of two weeks, but a maximum of one month, between assessments.

Statistical methods

Data management and statistical analysis were performed using SAS software, version 9.4 (SAS Institute, Cary, North Carolina, USA) and R 3.4 (R Core Development Team, 2017, R Foundation for Statistical Computing, Vienna, Austria). Rater agreement was assessed using intraclass correlation coefficients for continuous variables and weighted kappa statistics for categorical variables. Quadratic weighted kappa coefficients were utilized for the analysis in order to take account of the degree of disagreement, along with 95% confidence interval based on 1000 bootstrap sample. Linear weighted kappa coefficients were also utilized for consistency with existing literature. Consistent with previous mLPC reliability literature, kappa values were assessed based on the guidelines suggested by Landis and Koch,¹⁷ with a kappa value of 0 to 0.20 representing slight agreement, 0.21 to 0.40 fair agreement, 0.41 to 0.60 moderate agreement,

0.61 to 0.80 good agreement and 0.81 to 1.00 almost perfect agreement. A 95% confidence interval was reported for all measures and significance was set at $p < 0.05$.

Results

Primary outcome – rater agreement, inter-observer and intra-observer reliability

In total, 26 surgeons completed mLPC assessment of 40 Perthes cases. Individual surgeon's performance varied from 48% to 88% agreement with the mLPC gold standard (Fig. 1). The majority of surgeons rated the 40 patients according to the mLPC as follows: no patients were Lateral Pillar Group A, 20 patients were Group B, three patients were Group B/C border and 17 were Group C.

The weighted kappa for inter-observer agreement was 0.64 (95% confidence interval: 0.55 to 0.74), whether quadratic weighting or linear weighting were utilized. This indicates low good agreement. Agreement among the surgeons was highest and similar for Groups B (71.5 ± 17.8%) and C (70.6 ± 20.3%), and lowest for the B/C border Group (42.3 ± 0.0%) (Fig. 2). There were two cases with 100% agreement; both hips were severe Group C

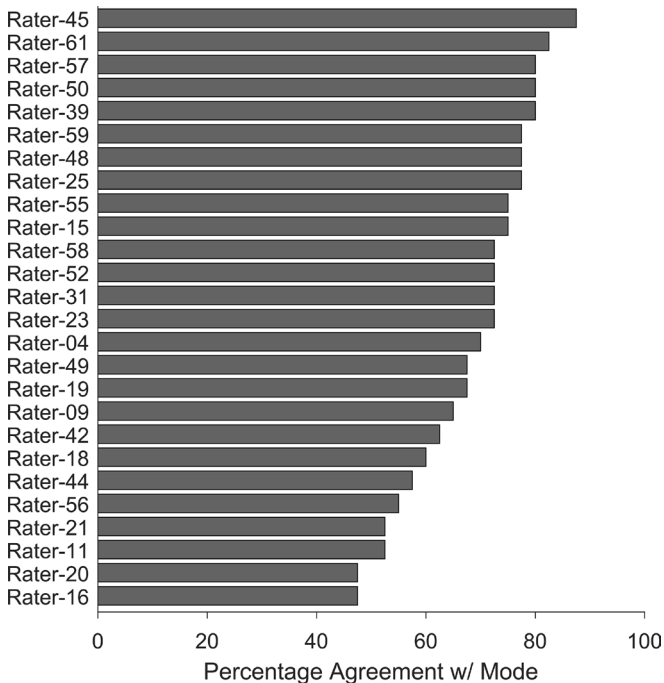


Fig. 1 Percent agreement of each individual rater's mLPC relative to the group mode (i.e. gold standard) mLPC. The top five performer raters are raters 45, 61, 57, 50, 39.

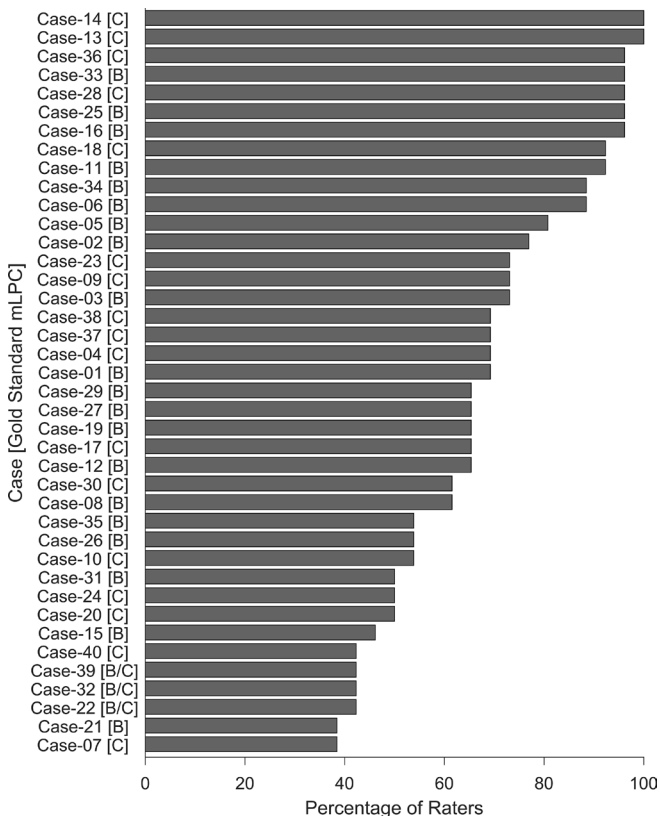


Fig. 2 Percent rater agreement in mLPC on a case-by-case basis. The bars represent the percent of raters who were in agreement for the case. The mode mLPC for each case is denoted in brackets next to the case number along the y-axis.

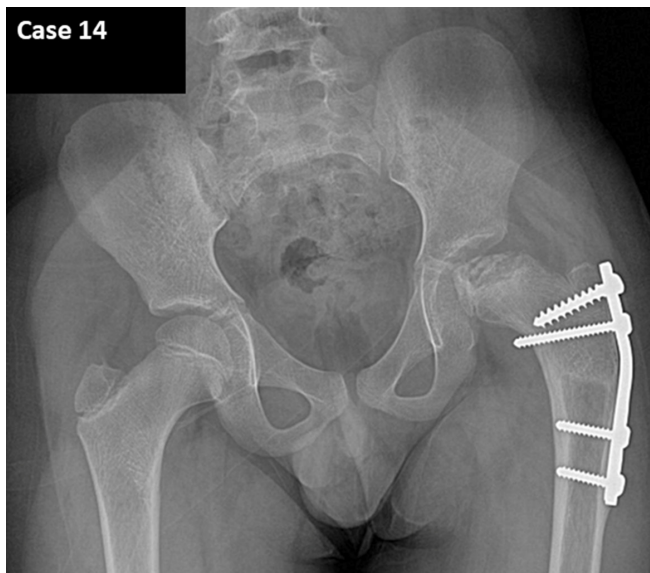


Fig. 3 Radiograph of Case 14, with 100% rater agreement of a Group C hip.

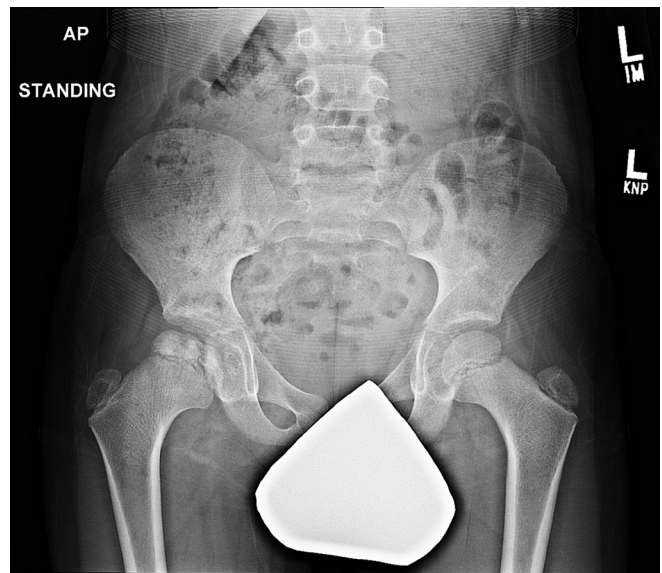


Fig. 4 Radiograph of Case 25, with 90% rater agreement of a mild Group B hip.



a. Case 21



b. Case 7

Fig. 5 Radiographs of two cases with the lowest rater agreement (38%) for mLPC. **(a)** For this case (21), the mode mLPC was Group B, and the top five highest-performing raters assigned three different classifications (group B, B/C border, and C) and selected four different radiographs. **(b)** For this case (7), the mode mLPC was Group C, and the top-five highest raters reported three different classifications (B, B/C border, C) and three different radiographs were selected.

(Fig. 3). Five cases had 90% agreement, all of which were either mild Group B or severe Group C (Fig. 4). For the B/C border Group, there were only three cases, and agreement was 42% (Fig. 2). Other cases with low inter-observer agreement ($\leq 50\%$) were either severe Group B hips or mild Group C (i.e. hips at the border of B and C; Fig. 5).

In all, 21 surgeons completed two additional rounds of mLPC assessments of the 40 Perthes cases for the purpose of determining intra-observer reliability. Average weighted kappa for intra-observer agreement was 0.82 (range: 0.35 to 0.99). Of the 21 surgeons, 20 had a weighted kappa of

0.67 or higher, indicating good or almost perfect agreement. One surgeon had fair agreement.

Secondary outcomes – radiographic selection and experience level

Radiograph selection did not influence gold standard assignment of mLPC (Fig. 6). There were no cases for which all 26 observers selected a single radiograph to assess mLPC. For 95% of cases, the raters selected three or more different radiographs, which spanned a period of six months or more. Even for the two cases with 100% mLPC agreement, only 42% of observers used the same radio-

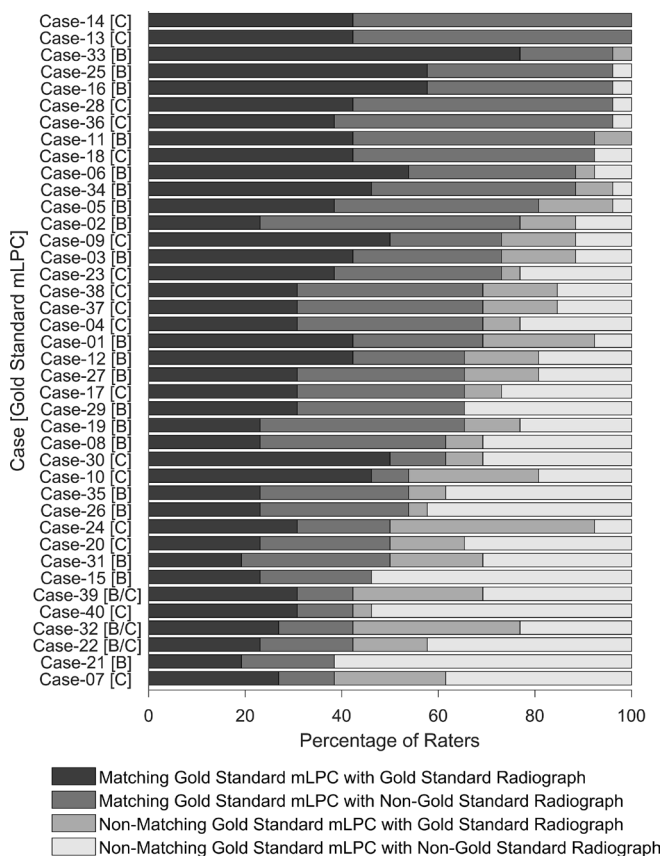


Fig. 6 Percent of raters that selected the gold standard mLPC and/or radiograph. The black bars represent the raters who assigned the same mLPC as the mode of the group (i.e. the gold standard mLPC) and, used the most commonly selected radiograph (gold standard radiograph). The dark-grey bars represent raters who selected the correct mLPC as the gold standard, but selected a different radiograph to reach this conclusion. The medium-grey bars represent the group of raters who selected the gold standard radiograph, but did not assign the gold standard mLPC. The light-grey bars represent the group of raters who did not select the gold standard mLPC nor the gold standard radiograph.

graph to assign mLPC. Surgeons' mLPC performance was not influenced by their years of clinical experience in paediatric orthopaedics ($p = 0.51$).

Discussion

Legg Calvé Perthes disease is a heterogeneous disease with varied presentations and outcomes, coupled with wide-ranging opinions on appropriate treatments for the disease. It is important that paediatric orthopaedic surgeons can reliably classify severity of disease radiographically. An easy to use, reliable classification system is essential to assist with determining clinical prognosis and to allow for consistent comparison of patients and studies. In this study, we demonstrate that the mLPC has low

good inter-observer agreement (weighted kappa = 0.64 with 95% confidence interval of 0.54 to 0.74) among a large group of international paediatric orthopaedic surgeons with varied duration of clinical experience. The intra-observer agreement was good or almost perfect for 20 of 21 raters, with an overall range between 0.35 to 0.99 (weighted kappa). Allowing surgeons to independently select the radiograph for rating, as is done in the clinical setting, did not appear to affect the mLPC assigned.

Various groups have investigated the reliability of the original lateral pillar classification which consists of three groups (Group A, B and C), and demonstrated moderate to good agreement, with weighted kappa values ranging from 0.49 to 0.72.^{2,3,7,9,10} With the exception of Herring's initial study of the LPC which included 16 observers,⁷ subsequent studies have used a smaller number of raters, ranging from two to five observers. Podeszwa et al⁹ included five observers of different levels of experience, spanning from a third-year orthopaedic resident to a paediatric orthopaedic surgeon with 25 years of experience, and concluded that level of experience did not affect inter-rater reliability for LPC.

Since the publication of the original LPC paper, Herring, et al identified a group of hips which challenged the LPC, a specific group for which there was limited consensus. These hips had certain radiographic features and tended to be more severe than typical Group B hips and milder than Group C hips.¹¹ Aiming for greater agreement with the classification, Herring et al modified the LPC by adding a B/C border Group.¹¹ To evaluate the reliability of this modification, three staff paediatric orthopaedic surgeons and three paediatric orthopaedic fellows reviewed 20 AP radiographs of hips in the fragmentation stage. Mean agreement per radiograph was 81% (range: 50% to 100%), with a quadratic weighted kappa of 0.71.¹¹ This was an improvement over the initial report of agreement and reliability of the original classification, with 78% agreement and a kappa of 0.52 (authors did not specify whether this was weighted).⁷

Two subsequent studies have assessed the reliability of the mLPC, but have not reproduced the same level of reliability. Rajan et al reported only fair agreement (linear weighted kappa 0.39), using six observers from a single institution to rate 35 cases with pre-selected radiographs which showed greatest lateral pillar involvement for review.¹² Huhnstock et al also reported fair agreement (linear weighted kappa 0.40), using four paediatric orthopaedic surgeons and one radiologist from three different hospitals to assess 42 cases with pre-selected radiographs.² The current study, in contrast, using a group of 26 international paediatric orthopaedic surgeons from 24 different centres, with experience ranging from 2 years to 43 years, had low good reliability for the mLPC with a weighted kappa of 0.64, using both linear and quadratic weighting.

Unlike the previous three studies of the mLPC, our study asked surgeons to self-select what they believed was the appropriate radiograph in fragmentation from a series of radiographs for each patient, starting from first diagnostic radiograph to a minimum of two-year follow-up. Although radiographic selection added a component of variability, we believe this aspect of the study's methodology is a strength and better simulates the clinical practice of determining mLPC by an individual surgeon. When assigning mLPC in the clinical setting, the correct radiograph is not pre-selected. Our study found that although radiograph selection varied across raters, the particular radiograph selected was not a critical factor affecting rater agreement. In the absence of a clear reason or reasons for the reduced performance in these previous studies, the ability to select a radiograph from a series may have allowed each rater to gain multiple qualitative views of a hip and allowed them to choose the radiograph that best fit their perception of the appropriate mLPC.

Examining specific cases with poor inter-observer agreement revealed that these cases fell on the border of Group B and Group C hips. More severe group B hips and milder group C hips appeared to show poor inter-observer agreement along with B/C border hips, in general, which only had 42% agreement. These results suggest that the classification system needs additional clarifications and refinement to help define hips that fall within the B/C border range of severity. Akgun et al emphasized the importance of actual measurement of lateral pillar height versus visual estimation, especially in borderline cases.¹⁸ Since most radiographs are now reviewed digitally, attention should be focused on digital measurement tools in order to quantitatively assess the lateral pillar and better discern borderline cases more reliably.

Due to continued difficulties in classifying hips on the border of B and C, as well as B/C border, we also question whether the modification of the LPC is useful or if the B/C border group should be withdrawn. In Herring et al's study of the radiographic outcome of children with Perthes disease, it should be noted that B/C border hips had intermediate outcomes, between those of Groups B and C.¹⁹ With respect to Stulberg I and II outcomes, they reported that Lateral Pillar Group B did relatively well at all ages (6 years to 11 years), and Lateral Pillar Group C did relatively poorly at all ages. In contrast, the outcome for B/C border hips showed greater correlation with age at onset: there was a sharp decrease in Stulberg I or II results with increasing age at diagnosis (e.g. 54% Stulberg I and II for 6 year olds to 6.9 year olds versus 19% in 8 year olds to 8.9 year olds). Additionally, in children with disease onset greater than eight years, children with B/C border hips seemed to benefit from osteotomy, while children with Group C hips did not.¹⁹ Until new data on the prognostic value of B/C border hips based on age becomes available, currently available

evidence supports the notion that the B/C border group has prognostic value. We do recognize that the B/C border hips pose a classification challenge, however, simply classifying a B/C border hip as typical Group B or C could risk an incorrect assessment of potential outcome. The authors do not advocate for withdrawal of the B/C border group at this time. Nevertheless, we strongly recognize the need to refine the B/C border group further and are committed to engaging in next steps to help better define it.

This study has several strengths. To our knowledge, it is the most comprehensive assessment of inter-rater and intra-rater reliability for mLPC to date, using a large number of surgeons from multiple centres in the United States and four other countries with varied clinical experience. We believe that this study design adds to a wide applicability of the results. Additionally, this is the only study of the mLPC that mimics the clinical practice of surgeons independently selecting the appropriate radiograph for mLPC assessment.

Our study does have limitations. Patients with the onset of Perthes disease prior to age six years were not included, so these findings may not be applicable to younger patients. Limitations also include the lack of Group A hips included in the study. This likely represents the low prevalence of these mild cases, and reflects the exclusion of patients with < 50% femoral head involvement on perfusion MRI. The absence of Group A hips may also indirectly reflect the inclusion of frog lateral images of the hip. Podeszwa et al reported that when the lateral radiograph was taken into account, LPC assessments tended to be more severe.⁹ If the inclusion of these lateral images in the present study biased the mLPC assessments towards being more involved, this may have affected the lack of Group A rating. One could argue that the lateral image is typically available in the clinical setting, and that the results of this study would likely reflect the assignment of the respective classifications in the clinic. Lastly, while experience level was assessed, all of the raters were paediatric orthopaedic surgeons with a dedicated interest in Perthes disease. Trainees or physicians within other fields were not included in the study.

The results of this study indicate that the mLPC has low good inter-observer reliability among paediatric orthopaedic surgeons with a specific interest in Perthes disease from a large number of centres in the USA and four other countries. Despite the intention of bringing greater clarity by adding the B/C border group in the modified classification, we found ambiguity regarding the cases between the severe B, B/C border, and mild C groups. The mLPC will benefit from further refinement for these B/C border cases. The results of this study support continued use of the mLPC in the clinical setting.

Received 28 March 2020; accepted after revision 29 September 2020

COMPLIANCE WITH ETHICAL STANDARDS

FUNDING STATEMENT

This study was funded by three grants from the Pediatric Orthopaedic Society of North America, as well the Texas Scottish Rite Hospital for Children and the Gillette Children's Foundation. No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the content of this article.

OA LICENCE TEXT

This article is distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 International (CC BY-NC 4.0) licence (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed.

ETHICAL STATEMENT

Ethical approval: All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: This study was reviewed and approved by our institutional review board and informed consent was obtained from all individual participants included in this study.

ICMJE CONFLICT OF INTEREST STATEMENT

JCL reports the International Perthes Study Group's receipt of three awards from Pediatric Orthopaedic Society of North America (POSNA). This work was also partially funded by Texas Scottish Rite Hospital and the Gillette Children's Foundation. JCL also reports being a member of the Board of Directors of the Legg Calvé Perthes Foundation (this is a volunteer position with no financial compensation), and receiving grant funding from Gillette Children's Foundation and POSNA to develop a new animal model of Perthes disease, outside the submitted work.

The other authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

JCL: Study design, Data acquisition, Data analysis and interpretation, Actively involved in the drafting and critical revision of the manuscript, Provided final approval of the submitted manuscript.

SAN: Data acquisition, Data analysis and interpretation, Actively involved in the drafting and critical revision of the manuscript, Provided final approval of the submitted manuscript.

SH: Data acquisition, Data analysis and Interpretation, Actively involved in the drafting and critical revision of the manuscript, Provided final approval of the submitted manuscript.

AJR: Data analysis and interpretation, Actively involved in the drafting and critical revision of the manuscript, Provided final approval of the submitted manuscript.

JET: Study design, Data acquisition, Actively involved in the drafting and critical revision of the manuscript, Provided final approval of the submitted manuscript.

WNS: Study design, Data acquisition, Data analysis and interpretation, Actively involved in the drafting and critical revision of the manuscript, Provided final approval of the submitted manuscript.

CHJ: Data analysis and interpretation, Actively involved in the drafting and critical revision of the manuscript, Provided final approval of the submitted manuscript.

HKWK: Study design, Data acquisition, Data analysis and interpretation, Actively involved in the drafting and critical revision of the manuscript, Provided final approval of the submitted manuscript.

APPENDIX

International Perthes Study Group: Benjamin D. Martin, Benjamin Joseph, Charles T. Mehlman, Courtney M. Selberg, Derek M. Kelly, Eric D. Fornari, Fábio Ferri-de-Barros, Hitesh Shah, Jeffrey I. Kessler, Joseph A. Janicki, Junichi Tamai, Leah Cobb, Mihir M. Thacker, Patricia Moreno Grangeiro, Rachel Y. Goldstein, Ralf D. Stuecker, Roberto Guarnieiro, Scott Yang, Shawn R. Gilbert, Tim Schrader, John A. Herring, Virginia F. Casey, Yasmin D. Hailer, Phillip W. Mack, Vidyadhar V. Upasani, Theresa A. Hennessey, Joshua E. Hyman, Zhongli Zhang, Crystal A. Perkins, Scott B. Rosenfeld, David H. Godfried, Stephen B. Sundberg, Rebecca J. Dieckmann, Molly F. McGuire.

REFERENCES

- Catterall A.** The natural history of Perthes' disease. *J Bone Joint Surg [Br]* 1971;53:37-53.
- Huhnstock S, Svenningsen S, Merckoll E, Catterall A, Terjesen T, Wiig O.** Radiographic classifications in Perthes disease. *Acta Orthop* 2017;88:522-529.
- 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.
- 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.
- Christensen F, Søballe K, Ejsted R, Luxhøj T.** The Catterall classification of Perthes' disease: an assessment of reliability. *J Bone Joint Surg [Br]* 1986;68:614-615.
- 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;66:479-489.
- Herring JA, Neustadt JB, Williams JJ, Early JS, Browne RH.** The lateral pillar classification of Legg-Calvé-Perthes disease. *J Pediatr Orthop* 1992;12:143-150.
- Ritterbusch JF, Shantharam SS, Gelinas C.** Comparison of lateral pillar classification and Catterall classification of Legg-Calvé-Perthes' disease. *J Pediatr Orthop* 1993;13:200-202.
- 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.
- Nathan Sambandam S, Gul A, Shankar R, Goni V.** Reliability of radiological classifications used in Legg-Calvé-Perthes disease. *J Pediatr Orthop B* 2006 Jul;15:267-270.
- 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 Oct;86:2103-2120.
- 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;33:120-123.
- Joseph B, Varghese G, Mulpuri K, Narasimha Rao K, Nair NS.** Natural evolution of Perthes disease: a study of 610 children under 12 years of age at disease onset. *J Pediatr Orthop* 2003;23:590-600.

14. **Hyman JE, Trupia EP, Wright ML, et al; International Perthes Study Group Members.** Interobserver and intraobserver reliability of the modified Waldenström classification system for staging of Legg-Calvé-Perthes disease. *J Bone Joint Surg [Am]* 2015;97:643-650.
15. **Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG.** Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377-381.
16. **Harris PA, Taylor R, Minor BL, et al; REDCap Consortium.** The REDCap consortium: building an international community of software platform partners. *J Biomed Inform* 2019;95:103-208.
17. **Landis JR, Koch GG.** The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174.
18. **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-Calvé-Perthes disease. *J Pediatr Orthop* 2004;24:651-653.
19. **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 Oct;86:2121-2134.