BJR

HIP

Ottawa classification for symptomatic acetabular dysplasia: assessment of interobserver and intraobserver reliability

K. Bali,
K. Smit,
M. Ibrahim,
S. Poitras,
G. Wilkin,
R. Galmiche,
E. Belzile,
P. E. Beaulé

The Ottawa Hospital, Ottawa, Canada

Aims

The aim of the current study was to assess the reliability of the Ottawa classification for symptomatic acetabular dysplasia.

Methods

In all, 134 consecutive hips that underwent periacetabular osteotomy were categorized using a validated software (Hip2Norm) into four categories of normal, lateral/global, anterior, or posterior. A total of 74 cases were selected for reliability analysis, and these included 44 dysplastic and 30 normal hips. A group of six blinded fellowship-trained raters, provided with the classification system, looked at these radiographs at two separate timepoints to classify the hips using standard radiological measurements. Thereafter, a consensus meeting was held where a modified flow diagram was devised, before a third reading by four raters using a separate set of 74 radiographs took place.

Results

Intrarater results per surgeon between Time 1 and Time 2 showed substantial to almost perfect agreement among the raters (κappa = 0.416 to 0.873). With respect to inter-rater reliability, at Time 1 and Time 2 there was substantial agreement overall between all surgeons (Time 1 κappa = 0.619; Time 2 κappa = 0.623). Posterior and anterior rating categories had moderate and fair agreement at Time 1 (posterior κappa = 0.557; anterior κappa = 0.438) and Time 2 (posterior κappa = 0.506; anterior κappa = 0.250), respectively. At Time 3, overall reliability (κappa = 0.687) and posterior and anterior reliability (posterior κappa = 0.579; anterior κappa = 0.521) improved from Time 1 and Time 2.

Conclusion

The Ottawa classification system provides a reliable way to identify three categories of acetabular dysplasia that are well-aligned with surgical management. The term 'borderline dysplasia' should no longer be used.

Cite this article: *Bone Joint Res.* 2020;9(5):242–249.

Keywords: Hip dysplasia, Lateral centre edge angle, Retroversion, Classification, Periacetabular osteotomy

Article focus

- Acetabular dysplasia is a complex multidirectional deformity that cannot be appropriately classified with just lateral centre edge angle (LCEA).
- Ottawa classification is a simple yet comprehensive classification of symptomatic acetabular dysplasia that not only identifies patterns of acetabular dysplasia that can be easily missed by traditional classification systems based on just LCEA, but also helps in making treatment decisions and provides guidance for surgical correction.
- The current study focuses on assessment of interobserver and intraobserver reliability of the Ottawa classification of symptomatic hip dysplasia.

Key messages

- Ottawa classification system has substantial overall inter-rater and intrarater reliability.

Correspondence should be sent to P. E. Beaulé;
Email: pbeaule@toh.ca

doi: 10.1302/2046-3758.95.BJR-2019-0155.R1

Bone Joint Res 2020;9:242–249.

- Ottawa classification system identified a large number of dysplastic patients who had no lateral acetabular deficiency.
- A simple anteroposterior (AP) radiograph is needed to classify the hips.

Strengths and limitations

- Strengths include numbers, methodology (power calculation and utilization of multiple blinded raters for evaluation of interobserver and intraobserver reliability), and the study's attempt to refine and simplify the classification system using iterative process.
- Limitations include cases from single surgeon practice, nonutilization of CT scans, and inability to confirm if intraoperative surgical correction correlated with type of dysplasia based on the classification system.

Introduction

Acetabular dysplasia results from maldevelopment in size, shape, and/or orientation of the acetabulum leading to misalignment between acetabulum and femoral head, and is a leading cause of hip instability.[1] If left untreated, hip dysplasia is a known risk factor for osteoarthritis.[2] The risk of progression of arthritis in the asymptomatic hip was first defined by Wiberg[3] using the lateral centre edge angle (LCEA) measurement, with values less than 20° defined as dysplastic and more than 25° defined as normal while values between 20° and 25° are defined as borderline. Consequently, the LCEA has been a key radiological measurement in surgical decision-making, as well as evaluating the adequacy of corrective hip osteotomies.[4]

Some authors have moved away from a quantitative measurement, such as the LCEA, to more of a qualitative assessment, i.e. severe, moderate, and borderline dysplasia.[5] However, a recent meta-analysis showed a lot of heterogeneity in LCEA values used to define dysplasia itself (LCEA 16° to 27°) and to differentiate between 'borderline dysplasia' (LCEA 18° to 28°) and moderate dysplasia.[6] Furthermore, when one looks at the origin of the term 'borderline dysplasia', Wiberg[3] described it in regard to the risk of arthritis progression in hips that were asymptomatic. Thus, the term 'borderline' is not appropriate for symptomatic hips in which intra-articular pathology is the main predictor of outcome.[7]

More importantly, acetabular dysplasia is a complex multidirectional deformity, where the LCEA assesses the femoral head coverage at the most cranial aspect of the acetabular roof.[8] Consequently, basing coverage only on the LCEA could result in missing anterior or posterior under-coverage, which is often present in patients with acetabular dysplasia.[9] In order to address these shortcomings, the authors developed a classification of symptomatic acetabular dysplasia encompassing three discrete prototypical patterns of hip instability: lateral (global), anterior, and posterior.[10] These three patterns encompass the full spectrum of acetabular dysplasia as well as providing valuable guidance for surgical correction of underlying acetabular deficiency, similar to the treatment of hip instability after total hip arthroplasty (THA) secondary to a malorientated acetabular component.

The primary purpose of the current study was to assess the interobserver and intraobserver reliability of the Ottawa classification of acetabular dysplasia. Our hypotheses were that this classification would permit a distribution of dysplastic hips into three categories and would have moderate-to-substantial interobserver and intraobserver reliability among all the evaluators with varied experience in young adult hip preservation.

Methods

After institutional ethics board approval, a retrospective chart review of an institutional database for hip preservation procedures was performed on all patients who underwent a periacetabular osteotomy (PAO) by the senior author (PB) over an 11-year period (2006 to 2016). Patients without digital preoperative supine anteroposterior (AP) radiographs, and the patients who had had prior hip surgery, neuromuscular disorders, and were skeletally immature were excluded from the analysis.

A total of 134 symptomatic hips that had undergone PAO were included in the study. Using standardized supine digital AP radiographs, two blinded observers (KB and MI) independently used a validated software (Hip2Norm; TrollTech, Oslo, Norway)[11] to measure: LCEA; acetabular index (AI); anterior and posterior acetabular coverage indices (anterior wall index (AWI), percentage anterior coverage (AC), posterior wall index (PWI), and percentage posterior coverage (PC)); and retroversion signs (ischial spine sign or crossover sign more than 1 cm from acetabular roof).[12] Using those measurements, all hips were classified based on the Ottawa classification for symptomatic acetabular dysplasia: normal; globally/laterally deficient; anteriorly deficient; and posteriorly deficient. Table I summarizes the prevalence of specific types of dysplasia based upon classification of the 134 symptomatic hips: global dysplasia in 59.0% (79 hips); anterior deficient in 37.3% (50 hips); and posterior deficient in 3.7% (five hips).

To assess reliability of the classification system, we used KappaSize R package version 1.1 (R Foundation for Statistical Computing, Vienna, Austria) for power analysis to calculate the minimum sample size of the number of hips that needed to be reviewed by raters. To test a kappa of 0.5 (moderate agreement) for an outcome with four categories, a precision of 0.1 on each side, an alpha level of 0.05, and six raters, a minimal sample size of 74 was needed.[13] Also, this sample of 74 hips required a distribution of all three types of acetabular dysplasia, and some normal hip radiographs that roughly matched the prevalence of each specific type of acetabular dysplasia in symptomatic hips. Based on statistical analysis and prevalence as determined on Hip2Norm software output,

Table I. Classification of symptomatic hips using Hip2Norm software (TrollTech, Oslo, Norway)

Classification	LCEA	AI	% anterior coverage	% posterior coverage	Prevalence
Overall	18.9	11.4	10.2	40.2	134 hips
Normal (N)	N/A	N/A	N/A	N/A	0.0% (no hips)
Lateral/global dysplasia (L): LCEA < 20°; LCEA 20° to 25° with AI > 10°	13.1	15.9	9.2	36.3	59.0% (79 hips)
Posterior dysplasia (P): LCEA > 25°; percentage posterior coverage < 36%; posterior wall index < 0.80; posterior wall sign; ischial spine sign; crossover sign > 1 cm from acetabular roof	29.1	2	22	29.1	3.7% (five hips)
Anterior dysplasia (A): LCEA > 25°; percentage anterior coverage < 15%; anterior wall index < 0.30.	26.9	5.3	10.6	47.3	37.3% (50 hips)

AI, acetabular index; LCEA, lateral centre edge angle; N/A, not applicable.

44 dysplastic and 30 asymptomatic normal hips were randomly chosen. The 44 dysplastic hips were chosen from the study group as classified based on Hip2Norm analysis, and included 24 hips with global/lateral dysplasia, 15 with anterior dysplasia, and five with posterior dysplasia. The 30 normal hips were chosen from a subset of patients who had been assessed to have symptomatic radiological femoroacetabular impingement in one hip, and an asymptomatic hip on the contralateral side classified as normal on radiological Hip2Norm analysis.

An iterative process was used to refine the classification, where three sequential steps/readings were used. Raters were only shown kappa results between steps for the discussion to identify problems and refine the classification (Figures 1 to 3),[10] and images were randomly sequenced for each step. Steps/iterations were stopped when kappa results demonstrated the appropriateness of the classification.

A group of six blinded raters, which included four surgeons (PB, KS, GW, EB) with an adult hip preserving practice and two fellows (RG and Mansour Abolghasemian (see Acknowledgements)) with career interest in adult hip surgery, were first provided with the Ottawa classification system[10] to understand the classification system (Figure 1), with an accompanying flowchart (Figure 2) used to analyze and classify the 74 radiographs at two separate timepoints (minimum two weeks apart) using standard picture archiving and communication system (PACS) measurements. The second evaluation period images were organized in a different order from the first. After the second reading, the four surgeons (with adult hip preservation practice) met to resolve discrepancies and designed a modified and simplified version of the flowchart (Figure 3) to classify the hips. This flowchart had to be used by these four raters for the third reading. The third reading included a new set of 74 radiographs with a similar dysplasia type distribution.

Figures 4 to 7 show radiological case examples of normal, globally/laterally deficient, posteriorly deficient, and anteriorly deficient hips based on the Ottawa classification system.

Statistical analysis. The kappa statistic was used for assessment of agreement between the raters and within raters (measurement scale - categorical). Intrarater reliability for each surgeon between Time 1 and Time 2 was assessed using Cohen's kappa. Reliability across all raters at each timepoint (separately) was assessed using the Fleiss kappa (adaptation of Cohen's kappa for three or more raters). Kappa results were interpreted as follows: values of ≤ 0 as indicating no agreement and 0.01 to 0.20 as

Class	Clinical findings	Radiological findings	Optimal surgical management
Anterior (A)	Anterior hip pain Exacerbated by: - extension + external rotation - late stance phase - positive PART	- Normal LCEA (> 25°) - Percentage anterior coverage < 15% - Anterior wall index < 0.30 - Excessive posterior wall coverage	PAO to increase anterior coverage while maintaining lateral coverage
Posterior (P)	Anterior + posterior hip pain ± prior diagnosis of SI joint pathology or Piriformis syndrome ± neurological symptoms in sciatic nerve distribution Exacerbated by: - flexion + internal rotation ± axial load - ascending stairs/slope	- Normal LCEA (> 25°) - Percentage posterior coverage < 36% - Posterior wall index < 0.80 - Posterior wall sign - Ischial spine sign - Crossover sign > 1 cm from the acetabular roof	PAO to increase acetabular anteversion ± resection of cam deformity
Lateral (L) • Anterolateral • Posterolateral • Global	Diffuse activity-related pain Abductor fatigue Symptoms of static overload	- LCEA (< 20°). or - LCEA 20° to 25° and AI > 10 - ± features of anterior or posterior uncoverage depending on acetabular version	PAO to increase lateral coverage ± anterior or posterior coverage

Fig. 1

Ottawa classification for symptomatic acetabular dysplasia. AI, acetabular index; LCEA, lateral centre edge angle; PAO, periacetabular osteotomy; PART, prone apprehension/relocation test; SI, sacroiliac. Adapted from and reproduced with permission from **Wilkin GP, Ibrahim MM, Smit KM, Beaulé PE**. A Contemporary Definition of Hip Dysplasia and Structural Instability: Toward a Comprehensive Classification for Acetabular Dysplasia. The Journal of Arthroplasty. 2017;32(9S):S20–S27.[10]

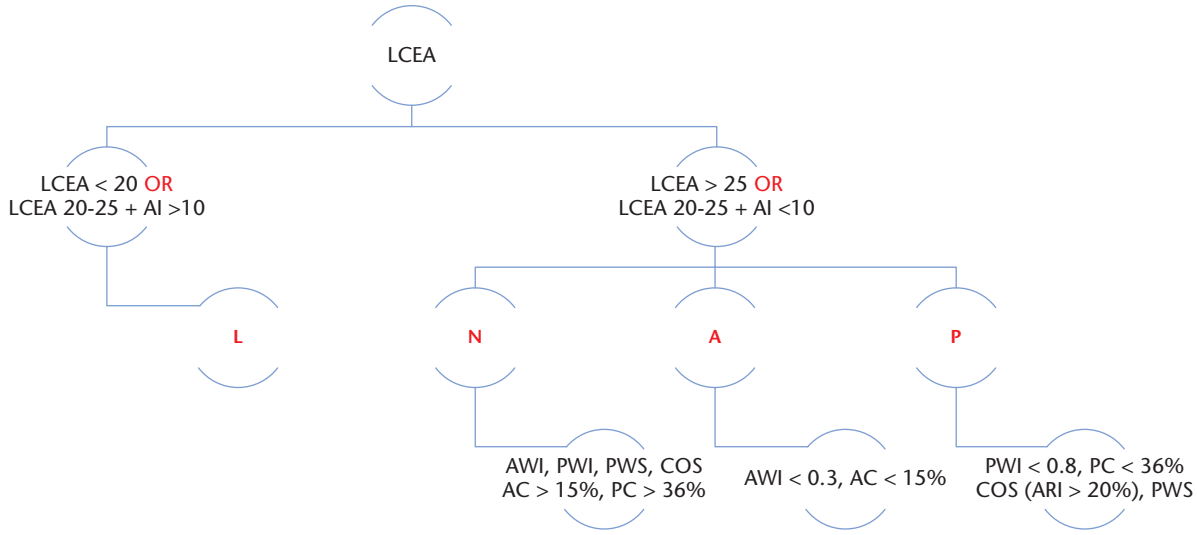

Fig. 2

Flowchart used for first and second readings. A, anterior dysplasia; AC, anterior coverage; AI, acetabular index; ARI, acetabular retroversion index; AWI, anterior wall index; COS, crossover sign; L, lateral/global dysplasia; LCEA, lateral centre edge angle; N, normal/no dysplasia; P, posterior dysplasia; PC, percentage posterior coverage; PWI, posterior wall index; PWS, posterior wall sign.

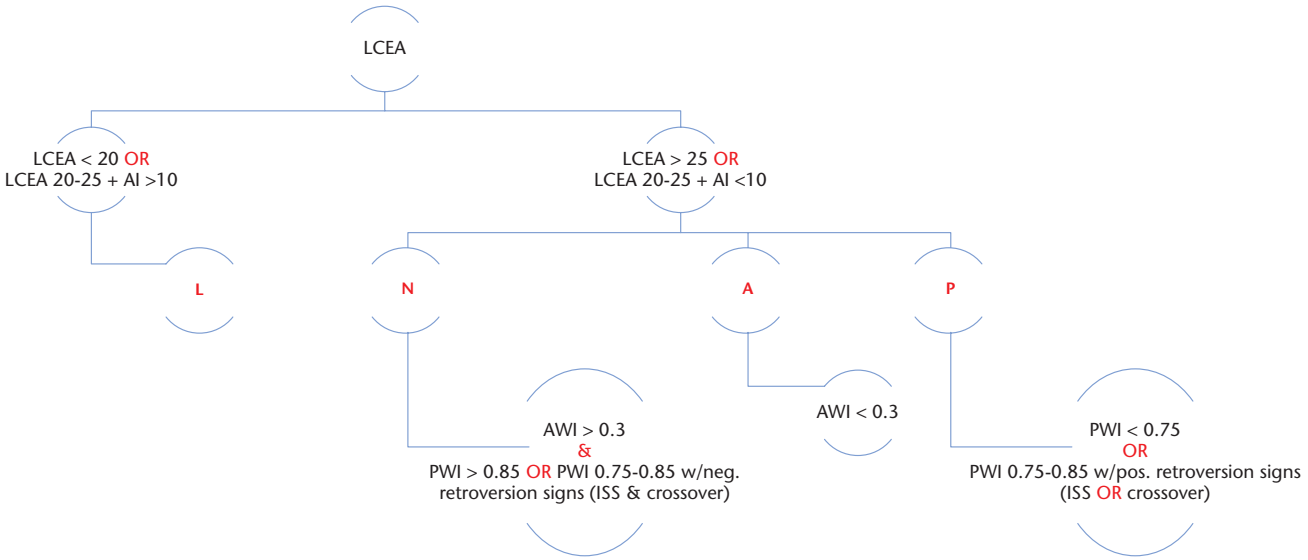

Fig. 3

Modified and simplified flowchart used for third reading. A, anterior dysplasia; AI, acetabular index; AWI, anterior wall index; ISS, ischial spine sign; L, lateral/global dysplasia; LCEA, lateral centre edge angle; N, normal/no dysplasia; neg., negative; P, posterior dysplasia; pos, positive; PWI, posterior wall index.

none to slight, 0.21 to 0.40 as fair, 0.41 to 0.60 as moderate, 0.61 to 0.80 as substantial, and 0.81 to 1.00 as almost perfect agreement.

Results

Intrarater reliability. Table II summarizes the intrarater results per rater between Time 1 and Time 2 using Cohen's kappa. The overall intrarater reliability among the raters ranged from moderate ($\kappa = 0.416$) to almost perfect ($\kappa = 0.873$). Barring one rater, there was substantial to nearly perfect agreement on the classification among individual raters. There was no correlation with respect to intrarater reliability of different raters based on their years of experience in hip preservation surgery.

Inter-rater reliability Time 1 and Time 2. Table III summarizes the reliability across all raters for Time 1 and Time 2, and among the four raters who completed Time 3, using Fleiss kappa.

At Time 1 and Time 2, there was substantial agreement overall between all surgeons (Time 1: $\kappa = 0.619$;

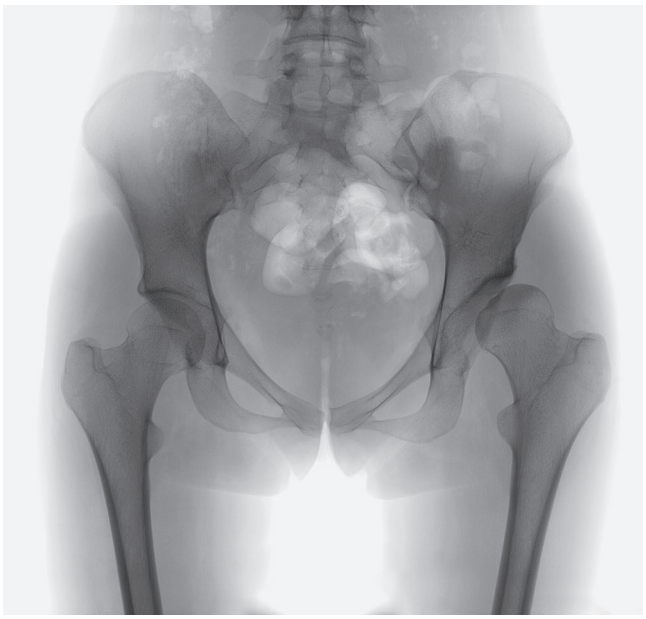

Fig. 4

Preoperative anteroposterior radiographs of a 22-year-old female with a normal right hip with lateral centre edge angle (LCEA) of approximately 28° and acetabular index (AI) of approximately 2.

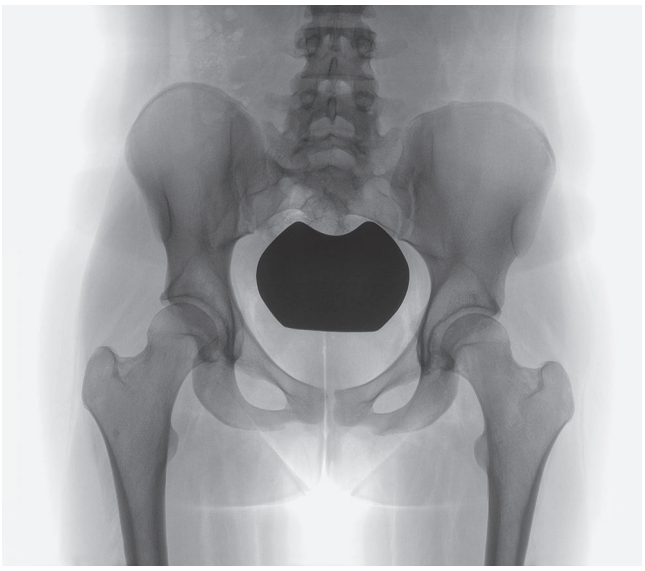

Fig. 6

Preoperative anteroposterior radiographs of a 21-year-old female with a posteriorly deficient left hip with a lateral centre edge angle (LCEA) of approximately 30° and an acetabular index (AI) of approximately -2.

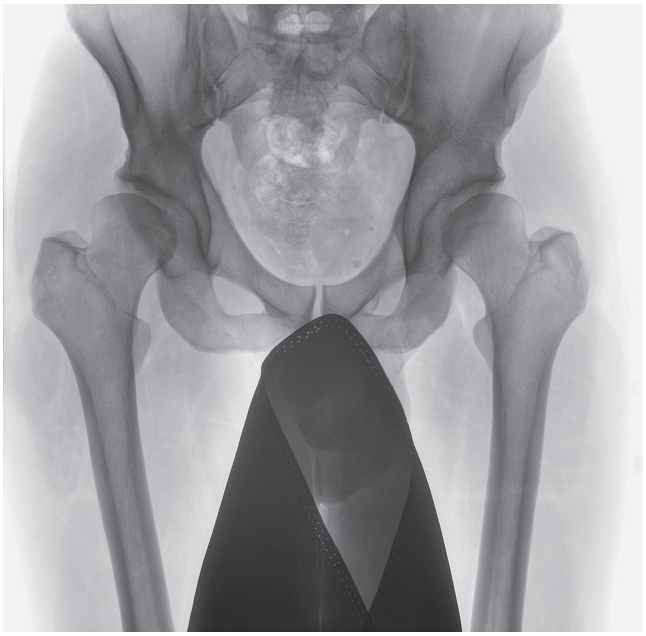

Fig. 5

Preoperative anteroposterior radiographs of a 24-year-old male with a globally/laterally deficient right hip with a lateral centre edge angle (LCEA) of approximately 5° and an acetabular index (AI) of approximately 20.

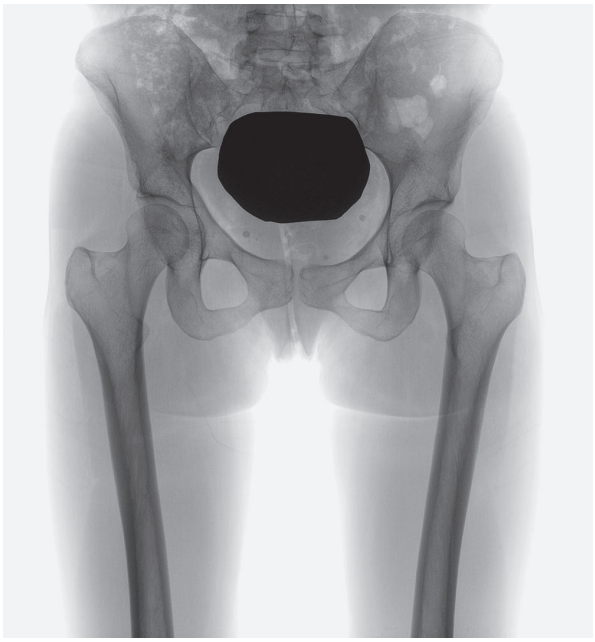

Fig. 7

Preoperative anteroposterior radiographs of a 19-year-old female with an anteriorly deficient right hip with a lateral centre edge angle of approximately 22° and an acetabular index of approximately 7.

Time 2: $\kappa = 0.623$). Agreement for normal rating was also substantial (Time 1: $\kappa = 0.759$; Time 2: $\kappa = 0.785$), while the agreement for lateral rating category was almost perfect (Time 1: $\kappa = 0.847$; Time 2: $\kappa = 0.862$).

At Time 1, both posterior and anterior rating categories had moderate agreement (posterior: $\kappa = 0.557$; anterior: $\kappa = 0.438$). At Time 2, posterior and anterior rating categories had moderate ($\kappa = 0.506$) and fair ($\kappa = 0.250$) agreement, respectively.

Inter-rater reliability Time 3. Following the consensus meeting and implementation of the modified flowchart, the overall inter-rater reliability remained substantial ($\kappa = 0.687$). Agreement for normal and lateral ratings was also substantial (normal: $\kappa = 0.683$; lateral: $\kappa = 0.767$).

Table II. Intrarater results per surgeon between Time 1 and Time 2 using Cohen's kappa

Surgeon	κ (95% CI)	p-value (agreement)
A	0.711 (0.578 to 0.844)	< 0.001 (substantial)
B	0.416 (0.240 to 0.592)	< 0.001 (moderate)
C	0.644 (0.494 to 0.794)	< 0.001 (substantial)
D	0.873 (0.831 to 0.915)	< 0.001 (almost perfect)
E	0.842 (0.795 to 0.889)	< 0.001 (almost perfect)
F	0.665 (0.616 to 0.714)	< 0.001 (substantial)

CI, confidence interval.

Table III. Reliability across surgeons for Time 1, Time 2, and Time 3 using Fleiss kappa

Rating category	κ (95% CI)	p-value (agreement)
Time 1		
Overall	0.619 (0.568 to 0.669)	< 0.001 (substantial)
N	0.759 (0.538 to 0.682)	< 0.001 (substantial)
L	0.847 (0.645 to 0.789)	< 0.001 (almost perfect)
P	0.557 (0.439 to 0.583)	< 0.001 (moderate)
A	0.438 (0.326 to 0.471)	< 0.001 (moderate)
Time 2		
Overall	0.623 (0.556 to 0.689)	< 0.001 (substantial)
N	0.785 (0.550 to 0.736)	< 0.001 (substantial)
L	0.862 (0.657 to 0.843)	< 0.001 (almost perfect)
P	0.506 (0.359 to 0.545)	< 0.001 (moderate)
A	0.250 (0.114 to 0.300)	< 0.001 (fair)
Time 3		
Overall	0.687 (0.621 to 0.752)	< 0.001 (substantial)
N	0.683 (0.589 to 0.776)	< 0.001 (substantial)
L	0.767 (0.673 to 0.861)	< 0.001 (substantial)
P	0.579 (0.485 to 0.673)	< 0.001 (moderate)
A	0.521 (0.427 to 0.615)	< 0.001 (moderate)

A, anterior dysplasia; L, lateral/global dysplasia; N, normal; P, posterior dysplasia.

Reliability across posterior and anterior ratings showed an increase in absolute value of kappa (posterior: $\kappa = 0.579$; anterior: $\kappa = 0.521$) with improvement in anterior ratings, from fair in Time 2 to moderate agreement in Time 3.

Discussion

Hip dysplasia is a 3D deformity and recognizing the underlying pattern of acetabular dysplasia beyond the coronal plane can help better guide treatment and our understanding of causes of failure.[10,14,15] In this study, using well-established radiological parameters, which defined direction of instability as either global/lateral, anterior, or posterior (analogous to instability after THA),[16] we were able to classify dysplastic hips into those three broad categories, with global/lateral being most common (around 59%; 79 hips) followed by anterior (37.3%; 50 hips) and posterior (3.7%; five hips). More importantly, we were able to show a good level of inter-rater and intrarater reliability for this classification system, which is key in establishing its clinical utility.

There are a few limitations of the study. Firstly, the cases selected from a single surgeon practice may not provide a full representation of the various hip pathologies. However, the senior surgeon does also perform hip arthroscopy and surgical dislocations, providing some confidence in the general applicability of this classification system. Secondly, the AP radiographs available to be reviewed were all supine images. Studies have shown that the functional acetabular orientation varies between supine and standing radiographs.[17] Specifically, there is a decrease in the amount of pelvic tilt on standing AP pelvis radiographs, resulting in a decrease in the incidence and amount of crossover sign and ischial spine sign, and a small increase in inclination.[18] As such, supine measurements potentially overestimate posterior dysplasia and may not be the gold standard for evaluating acetabular morphology and orientation. However, in the context of the current study the issue of pelvic tilt is largely irrelevant since the primary goal of this study was to assess the reliability of the classification system. Another issue may be a lack of availability of CT scans, which may be the most accurate way of evaluating acetabular dysplasia.[8,19] However, CT scans do not always improve reliability of a classification system, as shown by Beaulé et al[20] during reliability assessment of the Letournel classification system for acetabular fractures. However, plain radiographs are usually the first line of diagnostic imaging and unlike CT scans, radiographs are readily available and have lower radiation exposure than CT scans. However, CT scans have been used by some authors and Jacobsen et al[21], who looked at a cohort of patients with hip dysplasia with a LCEA > 20° using cross-sectional CT, found that a high proportion of them had deficient anterior acetabular coverage compared to the normal population. In other words, hips that were considered 'borderline' were actually deficient in AC. This finding is in line with our findings, where a high proportion of hips had isolated anterior under-coverage, and supports the measurement of the AWI.

Strengths of the current study included power calculation and utilization of multiple blinded raters for evaluation of interobserver and intraobserver reliability. We used the validated Hip2Norm software[14] for calculating the prevalence of different types of dysplasia in symptomatic patients who underwent periacetabular osteotomy (PAO). The interobserver and intraobserver reliability of computer-assisted measurements have been previously shown to be substantial to excellent.[22] For classifying dysplasia, the raters used PACS measurements that are commonly utilized in a busy outpatient setting, and not the Hip2Norm software[14], which is not widely available. Also, the addition of a third reading with a modified flowchart further improved the reliability of anterior and posterior dysplasia categories and optimized the clinical applicability of the study results.

Although we are not the first to categorize the 3D nature of acetabular dysplasia,[8,23-26] our current study is the first to categorize dysplasia into three distinct patterns of instability based on objective radiological

parameters, as well as the first to include hips with LCEA > 20°. In addition, by involving multiple blinded raters of various training levels, it provides a good sense of its ease of use which is further advanced with the addition of a simplified flowchart.

Stability of the hip is a complex interplay of bone (roof inclination, acetabular and femoral version, and neck shaft angle)[27] and soft tissue structures (muscle, capsule, and ligaments).[28] One key component to this classification system is including acetabular retroversion as a form of dysplasia with associated posterior under-coverage. This inclusion is supported by clinical reports on the treatment of acetabular retroversion in which isolated anterior rim resection was associated with an inferior outcome compared to addressing the posterior deficiency with anteverting PAO.[29,30] It is likely that our current treatment strategies for the young adult with hip pain will continue to evolve, and improving our diagnostic reliability is a critical component of treatment.[31]

In the current study, the inter-rater agreement for anterior or posterior dysplasia was not found to be substantial. In particular, there was only fair-to-moderate agreement among raters for the anterior dysplasia category in the first two readings. This may in part be attributed to challenges in outlining the anterior and posterior wall margins by the raters on AP radiographs. Another reason could be utilization of only AP images and the AWI measurements to calculate the AC. Use of low-dose 3D CT could enhance the diagnosis of anterior wall deficiency in difficult cases. Some have proposed the use of the false profile view to evaluate anterior coverage but recent radiological and CT studies have shown that the anterior centre edge angle does not assess true AC.[32]

Finally, a single AP radiograph for this comprehensive classification adds to its simplicity and likelihood of widespread application. Multiple reliability studies conducted over the years for various classification systems suggest that classification systems that are more comprehensive and detailed[20,33] tend to have better inter-rater and intrarater reliability than more simple classification systems[34,35] that may have substantial overlap between classification subtypes. However, this is not always true, and reliability studies for some classifications with detailed descriptions, such as the Neer classification for proximal humerus fracture, have not shown good inter-rater and intrarater reliability.[36] Such classifications tend to lack sufficient clarity and involve subjective interpretation to some extent, which diminishes the reliability. Although the comprehensive classification of symptomatic acetabular dysplasia is more detailed than other classifications of dysplasia previously described, the classification is based on objective radiological criteria, which is probably why it shows substantial overall inter-rater and intrarater reliability.

In conclusion, the current study shows that the Ottawa classification system has good overall inter-rater and intrarater reliability. Use of a simplified flowchart and understanding of hip dysplasia can help improve the reliability even further. There are a significant number of dysplastic hips without the traditional lateral dysplasia, which can potentially be missed by relying on LCEA measurements alone, hence the term 'borderline dysplasia' is ambiguous and should not be used. Future studies are needed to better understand how this classification can predict clinical outcome in the treatment of symptomatic hip dysplasia and guide effective management.

References

1. **Gala L, Clohisy JC, Beaulé PE.** Hip Dysplasia in the Young Adult. *J Bone Joint Surg Am.* 2016;98-A(1):63-73.
2. **Cooperman D.** What is the evidence to support acetabular dysplasia as a cause of osteoarthritis? *J Pediatr Orthop.* 2013;33 Suppl 1:S2-S7.
3. **Wiberg G.** Studies on dysplastic acetabula and congenital subluxation of the hip joint: with special reference to the complication of osteoarthritis. *Acta Chir Scand.* 1939;83:53-68.
4. **Clohisy JC, Ackerman J, Baca G, et al.** Patient-Reported Outcomes of Periacetabular Osteotomy from the Prospective ANCHOR Cohort Study. *J Bone Joint Surg Am.* 2017;99-A(1):33-41.
5. **Ding Z, Sun Y, Liu S, Chen J.** Hip Arthroscopic Surgery in Borderline Developmental Dysplastic Hips: A Systematic Review. *Am J Sports Med.* 2019;47(10):2494-2500.
6. **Yeung M, Kowalczuk M, Simunovic N, Ayeni OR.** Hip arthroscopy in the setting of hip dysplasia: A systematic review. *Bone Joint Res.* 2016;5(6):225-231.
7. **Dwyer MK, Lee JA, McCarthy JC.** Cartilage Status at Time of Arthroscopy Predicts Failure in Patients With Hip Dysplasia. *J Arthroplasty.* 2015;30(9 Suppl):121-124.
8. **Nepple JJ, Wells J, Ross JR, Bedi A, Schoenecker PL, Clohisy JC.** Three Patterns of Acetabular Deficiency Are Common in Young Adult Patients With Acetabular Dysplasia. *Clin Orthop Relat Res.* 2017;475(4):1037-1044.
9. **Tannast M, Hanke MS, Zheng G, Steppacher SD, Siebenrock KA.** What are the radiographic reference values for acetabular under- and overcoverage? *Clin Orthop Relat Res.* 2015;473(4):1234-1246.
10. **Wilkin GP, Ibrahim MM, Smit KM, Beaulé PE.** A Contemporary Definition of Hip Dysplasia and Structural Instability: Toward a Comprehensive Classification for Acetabular Dysplasia. *J Arthroplasty.* 2017;32(9S):S20-S27.
11. **Tannast M, Mistry S, Steppacher SD, et al.** Radiographic analysis of femoroacetabular impingement with Hip2Norm-reliable and validated. *J Orthop Res.* 2008;26(9):1199-1205.
12. **Beaulé PE, Zaragoza E.** Femoroacetabular Impingement: Diagnosis and Treatment Options. In: Beaulé PE, ed. *AAOS Monograph: The Young Adult with Hip Pain.* Rosemont, Illinois: American Academy of Orthopaedic Surgeons. 2007:63-74.
13. **Rotondi MA, Donner A.** A confidence interval approach to sample size estimation for interobserver agreement studies with multiple raters and outcomes. *J Clin Epidemiol.* 2012;65(7):778-784.
14. **Maradit Kremers H, Schilz SR, Van Houten HK, et al.** Trends in Utilization and Outcomes of Hip Arthroscopy in the United States Between 2005 and 2013. *J Arthroplasty.* 2017;32(3):750-755.
15. **Bogunovic L, Gottlieb M, Pashos G, Baca G, Clohisy JC.** Why do hip arthroscopy procedures fail? *Clin Orthop Relat Res.* 2013;471(8):2523-2529.
16. **Dorr LD, Wolf AW, Chandler R, Conaty JP.** Classification and treatment of dislocations of total hip arthroplasty. *Clin Orthop Relat Res.* 1983;(173):151-158.
17. **Ross JR, Tannenbaum EP, Nepple JJ, Kelly BT, Larson CM, Bedi A.** Functional acetabular orientation varies between supine and standing radiographs: implications for treatment of femoroacetabular impingement. *Clin Orthop Relat Res.* 2015;473(4):1267-1273.
18. **Jackson TJ, Estess AA, Adamson GJ.** Supine and Standing AP Pelvis Radiographs in the Evaluation of Pincer Femoroacetabular Impingement. *Clin Orthop Relat Res.* 2016;474(7):1692-1696.
19. **Larson CM, Moreau-Gaudry A, Kelly BT, et al.** Are normal hips being labeled as pathologic? A CT-based method for defining normal acetabular coverage. *Clin Orthop Relat Res.* 2015;473(4):1247-1254.

20. **Beaulé PE, Dorey FJ, Matta JM.** Letournel classification for acetabular fractures. Assessment of interobserver and intraobserver reliability. *J Bone Joint Surg Am.* 2003;85-A(9):1704-1709.
21. **Jacobsen S, Rømer L, Søballe K.** The other hip in unilateral hip dysplasia. *Clin Orthop Relat Res.* 2006;446(446):239-246.
22. **Nepple JJ, Martell JM, Kim YJ, et al; ANCHOR Study Group.** Interobserver and intraobserver reliability of the radiographic analysis of femoroacetabular impingement and dysplasia using computer-assisted measurements. *Am J Sports Med.* 2014;42(10):2393-2401.
23. **Mast JW, Brunner RL, Zebrack J.** Recognizing acetabular version in the radiographic presentation of hip dysplasia. *Clin Orthop Relat Res.* 2004;(418):48-53.
24. **Li PL, Ganz R.** Morphologic features of congenital acetabular dysplasia: one in six is retroverted. *Clin Orthop Relat Res.* 2003;(416):245-253.
25. **Siebenrock KA, Kistler L, Schwab JM, Büchler L, Tannast M.** The acetabular wall index for assessing anteroposterior femoral head coverage in symptomatic patients. *Clin Orthop Relat Res.* 2012;470(12):3355-3360.
26. **Dandachli W, Islam SU, Liu M, Richards R, Hall-Craggs M, Witt J.** Three-dimensional CT analysis to determine acetabular retroversion and the implications for the management of femoro-acetabular impingement. *J Bone Joint Surg Br.* 2009;91-B(8):1031-1036.
27. **Wyatt M, Weidner J, Pfluger D, Beck M.** The Femoro-Epiphyseal Acetabular Roof (FEAR) Index: A New Measurement Associated With Instability in Borderline Hip Dysplasia? *Clin Orthop Relat Res.* 2017;475(3):861-869.
28. **Le Bouthillier A, Rakhra KS, Belzile EL, Foster RCB, Beaulé PE.** Soft Tissue Structures Differ in Patients With Prearthritic Hip Disease. *J Orthop Trauma.* 2018;32 Suppl 1:S30-S34.
29. **Zurmühle CA, Anwander H, Albers CE, et al.** Periacetabular Osteotomy Provides Higher Survivorship Than Rim Trimming for Acetabular Retroversion. *Clin Orthop Relat Res.* 2017;475(4):1138-1150.
30. **Ibrahim MM, Poitras S, Bunting AC, Sandoval E, Beaulé PE.** Does acetabular coverage influence the clinical outcome of arthroscopically treated cam-type femoroacetabular impingement (FAI)? *Bone Joint J.* 2018;100-B(7):831-838.
31. **Beaulé PE, Allen DJ, Clohisy JC, Schoenecker P, Leunig M.** The young adult with hip impingement: deciding on the optimal intervention. *J Bone Joint Surg Am.* 2009;91-A(1):210-221.
32. **Sakai T, Nishii T, Sugamoto K, Yoshikawa H, Sugano N.** Is vertical-center-anterior angle equivalent to anterior coverage of the hip? *Clin Orthop Relat Res.* 2009;467(11):2865-2871.
33. **Brady OH, Garbuz DS, Masri BA, Duncan CP.** The reliability and validity of the Vancouver classification of femoral fractures after hip replacement. *J Arthroplasty.* 2000;15(1):59-62.
34. **Thomsen NO, Overgaard S, Olsen LH, Hansen H, Nielsen ST.** Observer variation in the radiographic classification of ankle fractures. *J Bone Joint Surg Br.* 1991;73-B(4):676-678.
35. **Frandsen PA, Andersen E, Madsen F, Skjødt T.** Garden's classification of femoral neck fractures. An assessment of inter-observer variation. *J Bone Joint Surg Br.* 1988;70-B(4):588-590.
36. **Sidor ML, Zuckerman JD, Lyon T, Koval K, Cuomo F, Schoenberg N.** The Neer classification system for proximal humeral fractures. An assessment of interobserver reliability and intraobserver reproducibility. *J Bone Joint Surg Am.* 1993;75-A(12):1745-1750.

Author information
- K. Bali, MBBS, MS Orthopaedics, Orthopaedic Surgery Resident,
- M. Ibrahim, MD, PhD, Fellow Orthopaedic Surgery,
- G. Wilkin, MD, FRCSC, Staff Orthopaedic Surgeon,
- R. Galmiche, MD, Fellow Orthopaedic Surgery, Division of Orthopaedic Surgery, Faculty of Medicine, University of Ottawa, The Ottawa Hospital, Ottawa, Canada.
- K. Smit, MD, FRCSC, Staff Orthopaedic Surgeon, Division of Orthopaedic Surgery, Children's Hospital of Eastern Ontario (CHEO), Ottawa, Canada.
- S. Poitras, PT, PhD, Associate Professor and Director, Physiotherapy Program, School of Rehabilitation, Faculty of Health Sciences, University of Ottawa, Ottawa, Canada.
- E. Belzile, MD, FRCSC, Staff Orthopaedic Surgeon, Division of Orthopaedic Surgery, University of Laval, Québec, Canada.
- P. E. Beaulé, MD, FRCSC, Professor of Surgery, University of Ottawa, Ottawa, Canada; Head, Division of Orthopaedic Surgery, The Ottawa Hospital, Ottawa, Canada.

Author contributions
- K. Bali: Designed the study, Drafted the manuscript, Acquired, analyzed, and interpreted the data.
- K. Smit: Designed the study, Drafted the manuscript, Acquired, analyzed, and interpreted the data.
- M. Ibrahim: Designed the study, Drafted the manuscript, Acquired, analyzed, and interpreted the data.
- S. Poitras: Designed the study, Drafted the manuscript, Acquired, analyzed, and interpreted the data.
- G. Wilkin: Designed the study, Drafted the manuscript, Acquired, analyzed, and interpreted the data.
- R. Galmiche: Designed the study, Drafted the manuscript, Acquired, analyzed, and interpreted the data.
- E. Belzile: Designed the study, Drafted the manuscript, Acquired, analyzed, and interpreted the data.
- P. E. Beaulé: Designed the study, Drafted the manuscript, Acquired, analyzed, and interpreted the data.

Funding statement
- No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

Conflict of interest statement
- E. Belzile reports an institutional grant from CIHR (Canadian Institute of Health Research) related to this study.
- S. Poitras reports a personal employment fee from the University of Ottawa, not related to this study.

Acknowledgements
- The authors would like to acknowledge Mansour Abolghasemian for his work on Time 1 and Time 2 evaluation.

Ethical review statement
- This study did not require ethical approval.