ORIGINAL CLINICAL ARTICLE

Reliability of plain radiographic parameters for developmental dysplasia of the hip in children

Vidyadhar V. Upasani · James D. Bomar ·
Gaurav Parikh · Harish Hosalkar

Abstract

Introduction Few studies have evaluated the reliability and reproducibility of the femoral neck-shaft angle (NSA), center-edge angle (CEA), and acetabular index (AI) in young children with developmental dysplasia of the hip (DDH). We wanted to determine whether these parameters could be used reliably by practitioners.

Methods Fifty radiographs from 21 children with DDH were reviewed. Analysis was performed by three observers, at two time periods. The intra- and inter-observer reliability for each measure was assessed.

Results At time period one, we noted a "high" level of agreement between observers when measuring the NSA, a "low" level when measuring the CEA, and a "moderate" level when measuring the AI. At time period two, we noted a "very high" level of agreement between observers when measuring the NSA and a "high" level when measuring the CEA and AI. When comparing the measurements of observer 1 at the two different time periods, we noted nearly "very high" agreement when measuring the NSA, a "moderate" agreement when measuring the CEA, and a "high" agreement for the AI. In comparing the measurements of observer 2, we noted "very high" agreement for the NSA and "high" agreement for the CEA and AI. In comparing the measurements for observer 3, we noted nearly "very high" agreement for the NSA, nearly "high" agreement for the CEA, and "high" agreement for the AI.

Conclusion It is difficult to reliably measure three-dimensional pelvic morphology on a frontal plane radiograph, especially when important pelvic landmarks have yet to ossify.

V. V. Upasani
Department of Orthopedics, University of California,
200 West Arbor Drive, San Diego, CA 92103, USA

J. D. Bomar · H. Hosalkar (✉)
Department of Orthopedics, Rady Children's Hospital,
3030 Children's Way, Suite 410, San Diego, CA 92123, USA
e-mail: hhosalkar@rchsd.org

G. Parikh
Department of Orthopedics, AMC MET Medical College,
Maninagar, Ahmedabad 380008, India

Introduction

Developmental dysplasia of the hip (DDH) is a common disorder in the pediatric population, with an overall incidence of approximately 3–4 per 1,000 live births. Risk factors for DDH include female gender, first born, and breech position. Although clinical examination remains the gold standard for diagnosing DDH in early infancy, ultrasonography has gained popularity worldwide as a screening tool. However, as the femoral head ossifies, ultrasonography is less accurate at evaluating the development of the acetabulum, and biplanar anteroposterior (AP) pelvic radiographs are used [1, 2].

Several radiographic parameters have been described to define hip dysplasia. Some of the most commonly used include the femoral neck-shaft angle (Fig. 1a), the center-edge angle of Wiberg (Fig. 1b) [3], and the acetabular index angle of Hilgenreiner (Fig. 1c) [4]. The center-edge
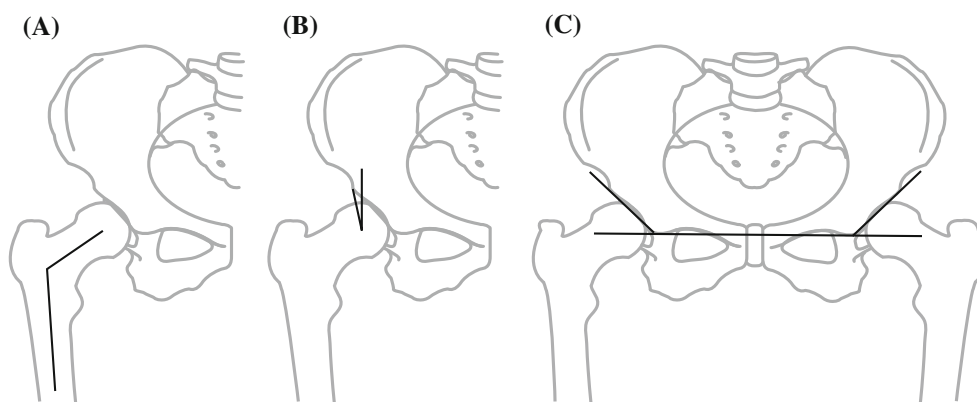
(A) **(B)** **(C)**

Fig. 1 Illustration of measurements. **a** Femoral neck-shaft angle, **b** center-edge angle of Wiberg, **c** acetabular index angle of Hilgenreiner

angle of Wiberg evaluates the degree of lateral femoral head coverage by the acetabulum in the frontal plane. The acetabular index angle of Hilgenreiner measures the slope of the acetabulum in the frontal plane.

Clinically, orthopedic surgeons commonly measure only one or two of these indices to make the radiographic diagnosis of DDH, to decide on treatment recommendations, and to assess the effect of the treatment. Although normal values for these indices have been described, they are often based on single observations performed by one observer. Few studies have evaluated the reliability and reproducibility of these radiographic measurements in the pediatric population. The purpose of this study was to assess the inter- and intra-observer reliability for three of the common radiographic measures of hip dysplasia (femoral neck-shaft angle, center-edge angle, and acetabular index). We did not attempt to evaluate the accuracy or validity of these particular indices; however, we did want to determine whether they could be used reliably by orthopedic surgeons and trainees. We hypothesized that these three measures would have significant reproducibility and that they could be used clinically to identify DDH.

Methods

Fifty AP pelvis radiographs from 21 patients with DDH (mean age 10 months, range 4–33 months) were reviewed. Most of the patients in this study had multiple radiographs evaluated. Seven patients in this study had one radiographic evaluation included, one patient had two, 11 patients had three, and two patients had four. These patients were consecutively evaluated in our pediatric hip specialty clinic and were referred to this clinic by their primary pediatrician to be evaluated for hip dysplasia. None of the patients had previous hip surgery.

The radiographs were analyzed by three independent observers who were blinded to the measurements of each other and to the identity of the patients. The observers included: (1) an attending fellowship trained pediatric orthopedic surgeon with a special interest in pediatric hip surgery, (2) a current pediatric orthopedic fellow who had completed the hip rotation, and (3) an orthopedic chief resident who had completed 6 months on the pediatric orthopedic rotation. All radiographs were printed on paper for analysis. Any patient-identifying information was removed from all radiographs.

Potential errors in radiographic measurements include identification of landmarks (i.e., drawing the lines), as well as actually measuring the angles. As such, the exact measurement technique for each radiographic parameter was agreed upon by the three observers. The measurements were based on the classic monographs of these indices from the literature. Once the observers completed drawing the lines necessary to measure these parameters, a fourth participant used a protractor to measure and record each angle. Using multiple examiners to identify the landmarks and a single examiner for goniometric measurements presumably minimizes the latter source of error and focuses on the former. One week later, the 50 AP pelvis radiographs were put in a random order and printed out again, and the exercise outlined above was repeated.

Statistical analysis

Statistical analyses were performed utilizing SPSS software (version 12, IBM, Armonk, NY, USA). Intra- and inter-observer reliability for each measure was assessed using intraclass correlation coefficient (ICC) analysis. The ICC generally reports a value between 0.0 and 1.0. Values closer to 1.0 represent stronger agreement. Although there are no definitive values that clearly differentiate between acceptable and unacceptable correlation [5], for the purposes of this study, we have adopted Munro's correlation strength categories (0.9–1.0 = "very high"; 0.7–0.89 = "high"; 0.5–0.69 = "moderate"; 0.26–0.49 = "low"; 0.0–0.25 = "little, if any") [6]. The two-way mixed model (absolute

agreement) was utilized as the observers in this study were not randomly selected and were measuring identical radiographs.

Results

In total, six individual measures were evaluated for each of the 50 radiographs. The right center-edge angle of Wiberg, however, was not measured on one radiograph because the hip was dislocated.

Inter-observer reliability

At time period one, we noted a "high" level of agreement between observers when measuring the neck-shaft angle [ICC = 0.868, 95 % confidence interval (CI) = 0.812–0.908], a "low" level of agreement between observers when measuring the center-edge angle (ICC = 0.473, 95 % CI = 0.190–0.666), and a "moderate" level of agreement between observers when measuring the acetabular index (ICC = 0.604, 95 % CI = 0.498–0.698). At time period two, we noted a "very high" level of agreement between observers when measuring the neck-shaft angle (ICC = 0.900, 95 % CI = 0.864–0.928), a "high" level of agreement between observers when measuring the center-edge angle (ICC = 0.742, 95 % CI = 0.613–0.828), and a "high" level of agreement between observers when measuring the acetabular index (ICC = 0.775, 95 % CI = 0.703–0.836) (see Table 1).

Intra-observer reliability

When comparing the measurements of observer 1 at the two different time periods, we noted nearly "very high" agreement when measuring the neck-shaft angle (ICC = 0.898, 95 % CI = 0.853–0.930), a "moderate" agreement when measuring the center-edge angle (ICC = 0.657, 95 % CI = 0.529–0.756), and a "high" agreement for the acetabular index (ICC = 0.721, 95 % CI = 0.609–0.805). In comparing the measurements of

observer 2, we noted "very high" agreement for the neck-shaft angle (ICC = 0.931, 95 % CI = 0.895–0.955), "high" agreement for the center-edge angle (ICC = 0.730, 95 % CI = 0.618–0.813), and "high" agreement for the acetabular index (ICC = 0.702, 95 % CI = 0.587–0.789). In comparing the measurements for observer 3, we noted nearly "very high" agreement for the neck-shaft angle (ICC = 0.893, 95 % CI = 0.826–0.932), nearly "high" agreement for the center-edge angle (ICC = 0.699, 95 % CI = 0.274–0.855), and "high" agreement for the acetabular index (ICC = 0.804, 95 % CI = 0.721–0.863) (see Table 2).

Discussion

Frontal plane pelvis radiographs are currently used to diagnose and assess DDH in children over 6 months of age and are almost a standard protocol in all pediatric centers across the world. Although various radiographic parameters have been described to quantify the deformity in acetabular and proximal femoral development, the reliability and reproducibility of some of the most common radiographic indices used today has not been defined within trainees and consultants. In this study, we demonstrated "moderate" to "high" inter- and intra-observer reliability for measurements of the center-edge angle and the acetabular index, performed by an orthopedic surgeon, a pediatric orthopedic fellow, and an orthopedic senior resident. The neck-shaft angle measurements, however, were found to be consistent between observers as well as between time periods for each observer.

To our knowledge, only three previous studies have evaluated the reliability of radiographic indices for hip dysplasia in the adult and pediatric population. Most recently, in 1999, Nelitz et al. [7] reviewed 100 radiographs in skeletally mature adults between 16 and 32 years of age. They found a high correlation for inter- and intra-observer reliability for the center-edge angle of Wiberg (0.87), acetabular index (0.85), and neck-shaft angle (0.79). They described no difficulties in identifying the important landmarks, such as the center of the femoral head, lateral acetabular edge, and the lateral margin of the teardrop, on skeletally mature radiographs and recommended using any one of the above parameters to plan treatment and follow outcomes.

In contrast, the other two studies were performed in a pediatric cohort ranging in age from 3 months to 15 years. Overall, both these studies demonstrated a wide range in values recorded by different observers and by one observer on two occasions. Broughton et al. [8] found good reliability of the center-edge angle of Wiberg and neck-shaft angle over the age of 6 years. Boniforti et al. [9] calculated

Table 1 Inter-observer reliability

	Measures	ICC	95 % CI		Significance
			Lower bound	Upper bound	
Time period 1	NSA	0.868	0.812	0.908	$p < 0.001$
	CE	0.473	0.190	0.666	$p < 0.001$
	AI	0.604	0.498	0.698	$p < 0.001$
Time period 2	NSA	0.900	0.864	0.928	$p < 0.001$
	CE	0.742	0.613	0.828	$p < 0.001$
	AI	0.775	0.703	0.836	$p < 0.001$

Table 2 Intra-observer reliability

	Measures	ICC	95 % CI		Significance
			Lower bound	Upper bound	
Observer 1	NSA	0.898	0.853	0.930	$p < 0.001$
	CE	0.657	0.529	0.756	$p < 0.001$
	AI	0.721	0.609	0.805	$p < 0.001$
Observer 2	NSA	0.931	0.895	0.955	$p < 0.001$
	CE	0.730	0.618	0.813	$p < 0.001$
	AI	0.702	0.587	0.789	$p < 0.001$
Observer 3	NSA	0.893	0.826	0.932	$p < 0.001$
	CE	0.699	0.274	0.855	$p < 0.001$
	AI	0.804	0.721	0.863	$p < 0.001$

the error in measurement between observers and between two measurements of a single observer and found large variances, especially when the acetabular margin was notched and the child was under 7 months old.

The current study supports the conclusions of the two previous evaluations of pediatric hip dysplasia. It is difficult to accurately measure three-dimensional pelvic abnormalities on a frontal plane radiograph, especially when important pelvic landmarks have yet to ossify. Radiographs evaluated in our study were from patients between 4 and 33 months of age. In our experience, the greatest variability occurred in identifying the center of the minimally ossified femoral head, as well as the lateral margin of the acetabulum. Measurement of the neck-shaft angle, however, was consistent in this cohort, likely because the femoral metaphysis and diaphysis could be well defined radiographically.

The results of this study contradicted our hypothesis and we were surprised to find that the acetabular index was not more reproducible. In our clinic, most of the junior faculty rely on only the acetabular index to guide their clinical decision-making, while the more senior faculty base their decisions on a gestalt of the AP pelvis radiograph, without making any objective measurements. Overall, our findings question the reliability and reproducibility of the common radiographic parameters for diagnosing DDH in children less than 3 years of age, primarily, the acetabular index.

If significant variability exists between observers, as well as between time periods for a single observer, it is difficult to create clinical pathways to treat these patients and determine the impact of a certain treatment method over time. We are all aware that clinical experience is invaluable in the management of this condition. However, in the era of advanced imaging, including the ability to

three dimensionally assess the bony and soft-tissue morphology with ultrasound technology, it is likely that there will be options available in the future to move from traditional biplanar radiographs (that obviously have radiation exposure) to non-radiation, non-anesthesia-based dynamic three-dimensional assessments, which could allow for a more reproducible assessment of hip dysplasia.

References

1. Tönnis D (1976) Normal values of the hip joint for the evaluation of X-rays in children and adults. Clin Orthop Relat Res 119:39–47
2. Catterall A (1994) The early diagnosis of congenital dislocation of the hip. J Bone Joint Surg Br 76(4):515–516
3. Wiberg G (1939) Studies on dysplastic acetabula and congenital subluxation of the hip joint: with special reference to the complication of osteoarthritis. Acta Chir Scand Suppl 58:7–38
4. Hilgenreiner H (1925) Early diagnosis and early treatment of congenital dislocation of the hip. Med Klin 21:1385–1388
5. Portney LG, Watkins MP (2000) Statistical measures of reliability. In: Foundations of clinical research: applications to practice. Prentice Hall, Upper Saddle River, pp 557–588
6. Munro BH (2005) Statistical methods for health care research. Lippincott Williams & Wilkins, Philadelphia
7. Nelitz M, Guenther KP, Gunkel S, Puhl W (1999) Reliability of radiological measurements in the assessment of hip dysplasia in adults. Br J Radiol 72(856):331–334
8. Broughton NS, Brougham DI, Cole WG, Menelaus MB (1989) Reliability of radiological measurements in the assessment of the child's hip. J Bone Joint Surg Br 71(1):6–8
9. Boniforti FG, Fujii G, Angliss RD, Benson MK (1997) The reliability of measurements of pelvic radiographs in infants. J Bone Joint Surg Br 79(4):570–575