

An index for diagnosing infant hip dysplasia using 3-D ultrasound: the acetabular contact angle

Myles G. Mabee¹ · Abhilash Rakkunedeth Hareendranathan¹ · Richard B. Thompson² · Sukhdeep Dulai³ · Jacob L. Jaremko^{1,2}

Received: 2 July 2015 / Revised: 3 November 2015 / Accepted: 21 January 2016 / Published online: 11 February 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract

Background Developmental dysplasia of the hip (DDH) is a common condition that is highly treatable in infancy but can lead to the lifelong morbidity of premature osteoarthritis if left untreated. Current diagnostic methods lack reliability, which may be improved by using 3-D ultrasound.

Objective Conventional 2-D US assessment of DDH has limitations, including high inter-scan variability. We quantified DDH on 3-D US using the acetabular contact angle (ACA), a property of the 3-D acetabular shape. We assessed ACA reliability and diagnostic utility.

Materials and methods We prospectively collected data from January 2013 to December 2014, including 114 hips in 85 children divided into three clinical diagnostic groups: (1) normal, (2) initially borderline but ultimately normal without treatment and (3) dysplastic requiring treatment. Using custom software, two observers each traced acetabula twice on two 3-D US scans of each hip, enabling automated generation of 3-D surface models and ACA calculation. We computed inter-observer and inter-scan variability of repeatability coefficients and generated receiver operating characteristic (ROC) curves.

Results The 3-D US acetabular contact angle was reproduced 95% of the time within 6° in the same scan and within 9° in different scans of the same hip, vs. 9° and 14° for the 2-D US alpha angle ($P < 0.001$). Areas under ROC curves for diagnosis of developmental dysplasia of the hip were 0.954 for ACA and 0.927 for alpha angle.

Conclusion The 3-D US ACA was significantly more reliable than 2-D US alpha angle, and the 3-D US measurement predicted the presence of DDH with slightly higher accuracy. The ACA therefore shows promising initial diagnostic utility. Our findings call for further study of 3-D US in the diagnosis and longer-term follow-up of infant hip dysplasia.

Keywords Developmental dysplasia of the hip · Hip · Infants · Quantitative indices · Three-dimensional ultrasound · Ultrasound

Introduction

Developmental dysplasia of the hip (DDH) is common, with a prevalence of 1.6–28.5/1,000 infants [1, 2]. Two-dimensional ultrasound (2-D US) is the preferred imaging modality for evaluation of DDH in infants younger than 6 months [1]. Imaging is essential because clinical diagnostic tests, such as the Barlow and Ortolani maneuvers, lack sensitivity after the neonatal period and in cases of mild disease [3]. The most common technique for diagnosing DDH on 2-D US, in combination with Harcke dynamic testing, is the Graf method, which requires a coronal plane image containing several landmarks: flat horizontal iliac wing, labrum, acetabular roof, os ischium, and a round femoral head. Imaging with visualization of these landmarks is intended to standardize the plane in which the acetabulum is imaged; however this method is still susceptible to high inter-observer [4–6] and inter-scan [7, 8]

✉ Myles G. Mabee
mmabee@ualberta.ca

¹ Department of Radiology and Diagnostic Imaging, University of Alberta, 2A2.42 Walter Mackenzie Health Sciences Centre, Edmonton, Alberta T6G 2B7, Canada

² Department of Biomedical Engineering, University of Alberta, Edmonton, Alberta, Canada

³ Department of Surgery, University of Alberta, Edmonton, Alberta, Canada

variability. Because diagnosis of DDH relies heavily upon the indices defined by the Graf method, especially the alpha angle, the diagnosis can change as a result of these sources of variability. If the diagnostic category (normal, borderline, moderate or severe dysplasia) is based solely on the alpha angle, diagnosis could change in half of hips in infants and up to three-quarters in neonates simply by changing 2-D US probe orientation [8]. Given these flaws, we seek a more reliable approach. Three-dimensional US captures the entire acetabular shape in a single scan and therefore should be less susceptible to inter-scan variability, with the potential to provide a more effective method for DDH diagnosis. Such diagnosis might ultimately rely on a combination of visual qualitative assessment and quantitative indices analogous to the alpha angle. To our knowledge, no quantitative indices have been developed for 3-D US in DDH.

Bony remodeling of the acetabulum is based on forces and loading factors imposed by the femoral head in the ball and socket joint [9]; consequently it may be meaningful to develop indices that define the geometry of the portion of acetabulum that contacts the femoral head. Differences in the contact zone might differentiate between normal and dysplastic hips. We have developed an index for measuring infant hip dysplasia on 3-D US, called the acetabular contact angle (ACA). The ACA is a property of the 3-D acetabular shape and indicates the direction of the forces the acetabular cup would experience from standing body weight. This paper focuses on evaluating the diagnostic utility and reliability of the ACA index.

Materials and methods

Patients

Our institutional health research ethics board approved this prospective study. Imaging was performed at a tertiary pediatric hospital from January 2013 to December 2014. At the first routine clinical 2-D US examination of each hip we obtained written informed consent from a guardian to perform 3-D US as part of the examination. Scan indications included clinical suspicion of DDH (hip laxity or asymmetrical skin creases) or risk factors such as positive family history. Because dysplasia can be unilateral or bilateral, we included each hip separately, and in cases of normal hips we only included one hip per subject in the study. We had 114 hips in 85 infants (60/85, or 71%, female). Infants underwent US at the mean age of 56.1 days (range 4–183 days). We observed clinical care for at least 3 months after initial scan (average 9.4 months; maximum 26 months) to classify each imaged hip based on orthopedic surgeon expert opinion at the first orthopedic assessment. Hips were categorized as normal (category 0; 40 hips and 40 infants; follow-up mean 157 days, range 94–619 days); questionably abnormal initially but with

findings that resolved spontaneously at follow-up imaging and clinical examination (category 1; 34 hips in 23 infants; follow-up mean 464 days, range 94–801 days); or dysplastic and proceeding to treatment by Pavlik harness or surgery (category 2; 40 hips in 22 infants; follow-up mean 314 days, range 94–733 days).

Imaging

We performed conventional 2-D US in both hips using a 12-megahertz (MHz) linear transducer (L12-5; Philips Healthcare, Andover, MA), in the coronal Graf standard plane by usual clinical protocol per American College of Radiology recommendations [10]. In addition, two study team members, including a radiologist, technologist or medical or graduate student trained by the study radiologists, used a high-resolution 13-MHz 3-D linear transducer (13VL5; Philips Healthcare, Andover, MA) to obtain coronal 3-D US images of each hip. The 3-D probe was held in a similar manner to the 2-D ultrasound protocol, with the transducer head resting near the greater trochanter of the infant. The user first aligned the 3-D probe such that the central slice of the 3-D image (displayed on screen) approximated the Graf standard plane (Fig. 1). As the 3-D sweep ran, the sonographer would note the presence of the os ischium as well as the posterior and anterior ends of the acetabulum, while minimizing motion artifacts. The 3-D transducer was used to perform a 3.2-second automated sweep through $\pm 15^\circ$ range to generate a 3-D dataset of 256 US sections, each 0.13-mm thick and containing 411×192 pixels measuring 0.11×0.20 mm. We had 114 hips to evaluate, with two 2-D US images and two 3-D US scans of each hip obtained by different team members to assess inter-scan variability.

Image processing

Three-dimensional US

Three-dimensional US images were analyzed off-line using custom software built with Python (Anaconda Distribution 2.1.0; Continuum Analytics Inc., Austin, TX). Images of each

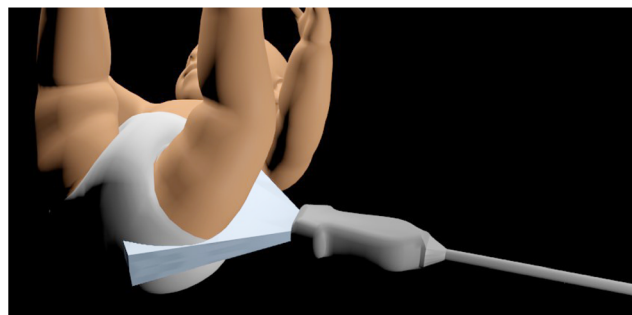


Fig. 1 Diagram of a typical sonographic examination of an infant hip using a 3-D US probe (fan beam dimensions are $0.11 \times 0.20 \times 0.13$ mm)

hip were reviewed by two observers (observer 1, M.G.M., a graduate student in biomedical engineering/radiology; and observer 2, J.L.J., the lead radiologist, with fellowship training in pediatric and musculoskeletal radiology and 11 years of experience).

For each hip, the acetabulum was traced on selected slices between two landmark points on opposite ends of the acetabulum (Fig. 2) using a customized semi-automated interactive interface in Python, including functions from the Visualization Toolkit (Kitware Inc., Clifton Park, NY). Contours traced on 5–7 selected sections were automatically interpolated to intervening sections, then reviewed and corrected by manual nudging of contours if needed. These tracings were used to create a surface model of the acetabulum. The process and tool are similar to those used to generate ventricle contours in cardiac MRI. Tracing of the acetabulum bone for surface models took 30–90 s, depending on reader experience and acetabular complexity.

The apex point on each slice was the point farthest from the line joining end points of ilium and acetabulum (Fig. 3). For each point on the acetabular surface, we calculated a normal vector, which is an engineering term for the line angled 90° from that portion of the surface and uniquely defines the orientation of that surface (Fig. 4). We then added all the normal vectors for points on the ilium above the apex line, considering each point to be equally important in our sum. The summed normal vector gives a net direction representing the average orientation of the iliac wall. We similarly added all the normal vectors for points on the acetabular roof below the apex line, to get an acetabular summed normal vector. The angle between these two lines is the ACA (Fig. 4).

Reading exercise

Readers performed bone tracings after consensus training on randomly selected data outside the study cohort.

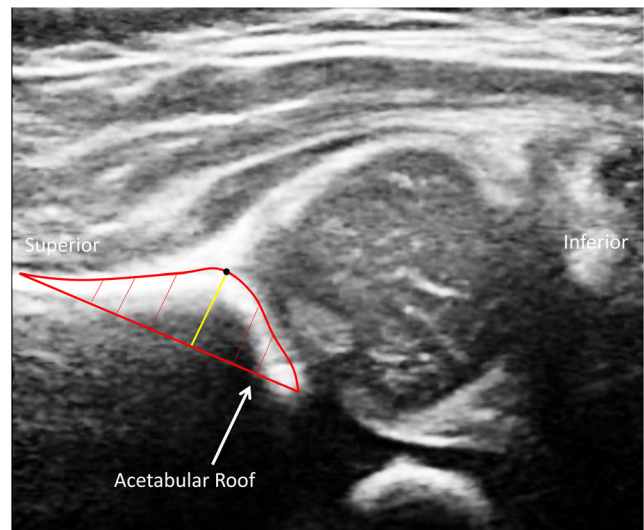


Fig. 3 Apex point. Typical 2-D US tracing of acetabulum is shown with its apex definition point on a coronal image of the left hip of a normal 23-day-old boy. The black dot overlaid on acetabulum tracing represents the apex point as defined by the largest perpendicular distance between the acetabulum tracing and the straight line connecting the start and end points of the tracings. The apex point is calculated using the convexity of adjacent points

Observers 1 and 2 (M.G.M. and J.L.J.) evaluated both scans of each hip in the cohort ($n=114$ scan 1, $n=113$ scan 2). To assess intra-observer reliability, each user repeated all measurements on one set of hips ($n=114$) >1 week after the initial readings.

Two-dimensional US

The alpha angle was also measured from the 2-D US image of each hip most closely meeting Graf standards by the usual clinical technique, by both users (observer 1/observer 2) ($n=114$ hips). Both users also re-measured on the same images 1 week later for intra-

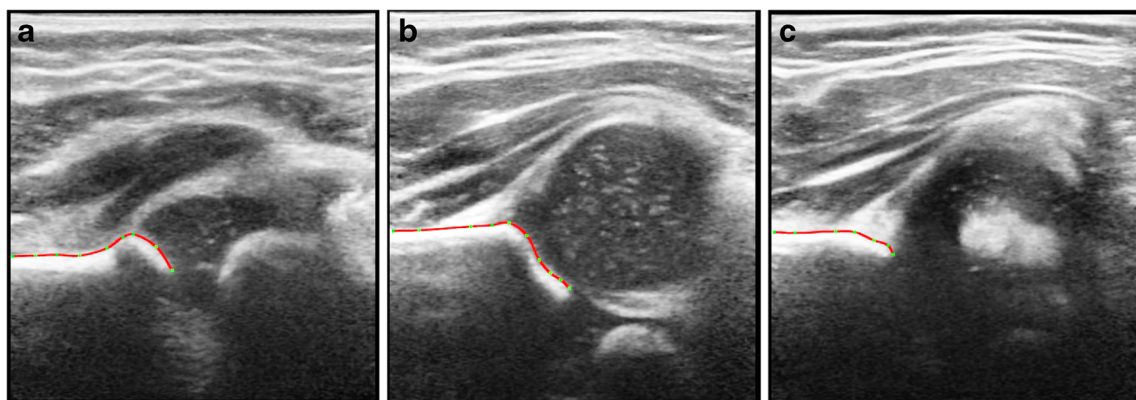


Fig. 2 Acetabulum landmarks and tracing techniques. Coronal images show the normal left hip of a 48-day-old girl from posterior (a) to anterior (c). **a, c** The acetabulum landmarks at the posterior (a) and anterior (c) edges of the acetabulum; slices more anterior or posterior no longer

showed acetabulum edges. **b** Typical slice between landmarks with associated acetabulum (red). The green points on each figure indicate user defined points defining the acetabulum

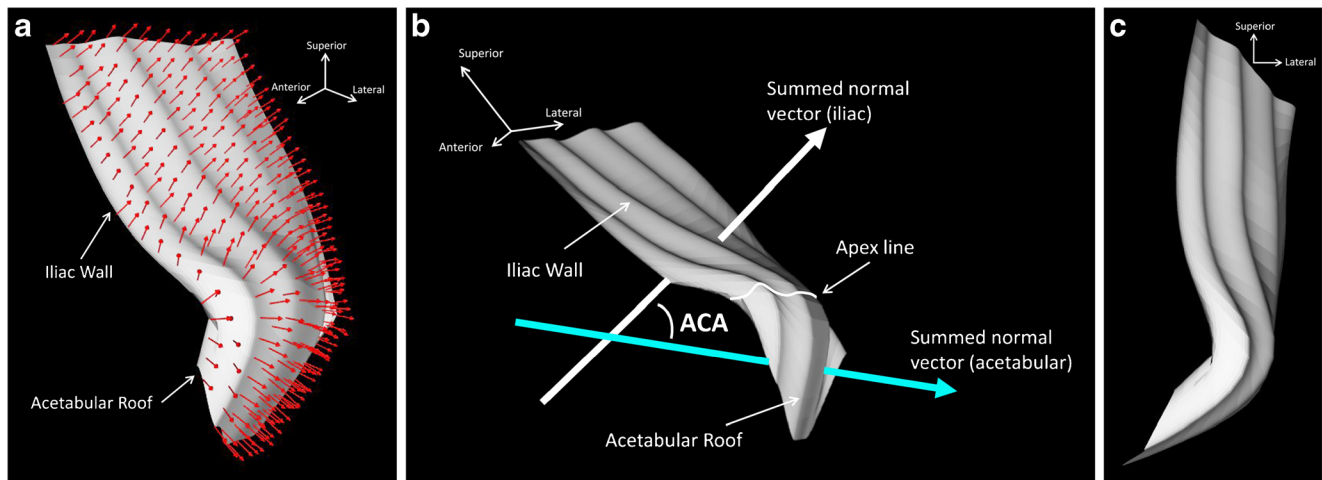


Fig. 4 Surface normals, apex point and acetabular contact angle (ACA). **a** Three-dimensional US image shows a limited number of the surface normal vectors (*red arrows*) used in ACA calculation. These are defined as the lines angled at 90° to each tiny portion of the bony surface. Mathematically, the surface normal uniquely defines the orientation of that surface. **b** ACA measurement. The irregular line traversing the acetabulum is the apex line. The thick white line represents the summed, or net, iliac wall normal vector calculated by adding the

normal vectors from each point on the iliac surface above the apex line. The thick blue line represents the acetabular summed normal vector, calculated by adding all normal vectors from the acetabular surface below the apex line. The ACA is the angle between these two lines. A smaller ACA indicates less difference between the orientation of the iliac wall and acetabular roof, i.e. a more dysplastic hip. **c** The true anterior view. Three-dimensional images created from raw data in a 48-day-old girl

observer reliability, and also measured alpha angle on 2-D US images of a subset ($n=83$) of these hips, which had been scanned at the same initial visit by a different user, to assess inter-scan variability.

Statistical analysis

Statistics were calculated on SPSS software v. 20 (IBM, Armonk, NY). Descriptive statistics were recorded as mean \pm standard deviation (SD). For repeated measurements we calculated the mean difference in values and characterized limits of agreement by the repeatability coefficient, i.e. the range of values within which there is 95% limit of agreement that a second observation will fall from the first [11], and the coefficient of variation. The intraclass correlation coefficient (ICC 3,1) was also calculated to assess conformity between users. For ACA and alpha angle we calculated intra- and inter-observer variability for evaluation of a single 3-D US or 2-D US scan of each hip, and the inter-scan variability for each observer between the first and second 3-D US or 2-D US scans of each hip. We compared inter-scan reliability of alpha angle and ACA indices by paired Student's *t*-test and also by nonparametric Wilcoxon signed rank test. To assess the utility of ACA and alpha angle as diagnostic tests for DDH requiring treatment, we computed the area under the corresponding receiver operating characteristic (ROC) curves.

Results

Reliability of alpha angle

Alpha angle reliability is summarized in Table 1. In 114 hips, intra-observer variability of the 2-D US alpha angle was $1.3 \pm 3.5^\circ$ (median 1.1) for observer 1, and $2.6 \pm 3.9^\circ$ (median 1.9) for observer 2, with repeatability coefficient= 7.1° and 7.6° , ICC=0.96 and 0.95, and coefficient of variation=4.4% and 5.3%, respectively. Inter-observer reliability of the alpha angle was $0.9 \pm 4.5^\circ$ (median 0.5) between users (observer 1 and observer 2), repeatability coefficient= 8.8° , ICC=0.95, coefficient of variation=5.4%. In the 83 hips with two 2-D US scans, inter-scan reliability of the alpha angle was $1.6 \pm 6.5^\circ$ (median 2.0) for observer 1 and $0.8 \pm 7.1^\circ$ (median 0.4) for observer 2, repeatability coefficient= 12.3° and 13.8° , ICC=0.88 and 0.88, coefficient of variation 7.5% and 8.2%, respectively. Inter-scan reliability was significantly poorer for alpha angle than ACA by paired *t*-test ($P=0.001$; Fig. 5).

Reliability of acetabular contact angle

Reliability information for ACA is also summarized in Table 1. In 114 hips, the intra-observer variability of the ACA was $0.3 \pm 2.8^\circ$ (median 0) for observer 1, and $1.5 \pm 3.0^\circ$ (median 1.6) for observer 2, with repeatability coefficients= 5.5° and 5.7° and ICC=0.97 and 0.96, respectively. The coefficient of variation was 4.3% for observer 2 and

Table 1 Reliability analysis of alpha angle and acetabular contact angle by mean and standard deviation (SD), repeatability coefficients (RC) and coefficient of variation (CV)

		Alpha angle (2-D US)				Acetabular contact angle (3-D US)			
		Mean ± SD (°)	Median (°)	RC (°)	CV (%)	Mean ± SD (°)	Median (°)	RC (°)	CV (%)
Intra-observer	Observer 2	2.6 ± 3.9	1.9	7.6	5.3	1.5 ± 3.0	1.6	5.7	4.3
	Observer 1	1.3 ± 3.5	1.1	7.1	4.4	0.3 ± 2.8	0	5.5	3.9
Inter-observer	Observer 2 vs. Observer 1	0.9 ± 4.5	0.5	8.8	5.4	2.6 ± 3.0	2.5	5.6	6.0
Inter-scan	Observer 2	0.8 ± 7.1	0.4	13.8	8.2	0.1 ± 4.4	1.6	8.6	6.5
	Observer 1	1.6 ± 6.5	2.0	12.3	7.5	0.9 ± 4.1	1.2	8.1	5.4

Note: Bolded values indicate the largest RC/CV value in each row

3.9% for observer 1. Inter-observer reliability of the ACA on the same 3-D US scan was 2.6 ± 3.0° (median 2.5) between readers (observer 1 and observer 2), repeatability coefficient=5.6°, ICC=0.91, coefficient of variation=6.0%. In 113 hips, the inter-scan reliability of the ACA was 0.9 ± 4.1° (median 1.2, observer 1.) and 0.1 ± 4.4° (median 1.6, observer 2), repeatability coefficient=8.1° and 8.6°, ICC=0.93 and 0.92, and coefficient of variation 5.4% and 6.5%, respectively.

Diagnostic utility: normal vs. dysplastic hips

Clinical categories of dysplasia (0=normal, 1=borderline initially but ultimately normal, 2=dysplastic and treated) were well separated into distinct groups by the 2-D alpha angle and by the ACA (Table 2). The 2-D US alpha angle was 63.3 ± 4.7° (mean ± standard deviation [SD]), with a 95% confidence interval [CI] 62.8°, 65.8° for category 0; 56.1 ± 4.5°, with a 95% CI 54.6°, 57.7° for category 1; and 45.5 ± 8.1° with a 95% CI 42.8°, 45.9° for category 2. The 3-D US ACA was 53.6 ± 3.7° (95% CI 52.4°, 54.7°) for category 0;

45.7 ± 4.5° (95% CI 44.2°, 47.2°) for category 1; and 36.8 ± 5.5° (95% CI 35.1°, 38.7°) for category 2. The mean values found within the 2-D US and 3-D US approaches, within each of the three diagnostic categories, were found to be statistically different (P<0.001). The Pearson correlation coefficient between observer 1’s ACA and alpha angle values (n=114) was 0.759, P<0.001 (Fig. 6). The raw alpha angle and ACA readings were normally distributed.

Case-by-case frequency distribution shown on pyramid graphs (Fig. 7) demonstrates optimal separation of normal vs. dysplastic hips at a threshold of 60.0° for 2-D US alpha angle (normal above 60°, dysplastic below 60°), and 48.0° for 3-D US ACA (normal above 48°, dysplastic below 48°). Receiver operating characteristic (ROC) curves were generated (Fig. 8) for a diagnostic test using either the ACA or alpha angle imaging index to detect DDH requiring treatment, compared to the clinical gold standard diagnosis by an orthopedic surgeon using all available clinical and imaging information. The area under the curve for 2-D US alpha angle was 0.927 and for 3-D US ACA was 0.954. There was only one case in which an infant was clinically considered normal but had an ACA in the abnormal range, compared to six overlapping cases between normal and dysplastic for the 2-D US alpha angle (15% of the dysplastic patients). On review, the single overlapping case with normal ACA had an acetabulum with an appropriately steep edge but which was quite shallow with subluxed femoral head, and alpha angle was also in the normal range in this hip.

This study included 34 hips that were considered borderline clinically (initially concerning enough for dysplasia to require follow-up but normalizing later without treatment). Of these, 25 had ACA below the 48° normal threshold, 27 had alpha angle below the 60° normal threshold, and 22 had both alpha angle and ACA indices that were abnormal. Of the 25 borderline cases with ACA below 48°, seven had ACA only slightly low (45–48°), and 19 (76%) were in infants <6 weeks of age.

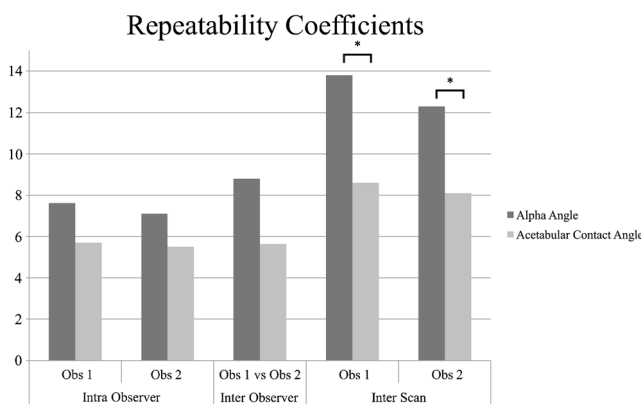


Fig. 5 Repeatability. Bar graph illustrates the difference between alpha angle and acetabular contact angle’s repeatability coefficients. A smaller repeatability coefficient is desirable because it indicates a narrower range of variability. Note the asterisk on inter-scan columns denotes a significant difference (P<0.001)

Table 2 Alpha and acetabular contact angle by diagnostic category (Dx)

Dx	Alpha angle			Acetabular contact angle		
	Mean ± SD ^{o*}	Median ^o	Confidence interval ^o	Mean ± SD ^{o*}	Median ^o	Confidence interval ^o
0	63.3 ± 4.7	65.0	[62.8, 65.8]	53.6 ± 3.7	53.7	[52.4, 54.7]
1	56.1 ± 4.5	55.9	[54.6, 57.7]	45.7 ± 4.5	45.2	[44.2, 47.2]
2	45.5 ± 8.1	45.9	[42.8, 45.9]	36.8 ± 5.5	37.7	[35.1, 38.7]

*Mean values found within the 2-D US and 3-D US approaches, within each of the three diagnostic categories, were found to be statistically significant ($P < 0.001$)

Qualitative review of 3-D US images showed the dysplastic hips receiving treatment to have shallower, simplified shapes of 3-D acetabular sockets compared to the deeper shapes of normal hips.

Alpha angle and ACA had similar Pearson correlation coefficients when compared with age: ACA vs. age was 0.551 and alpha angle vs. age was 0.456.

The intra-scan variability between angle measurements was compared for both users (observer 1 and observer 2) using the normalized difference values within users. For both users, the intra-observer intra-scan differences in ACA values (i.e. read on successive sessions by the same user from the same scan data) was significantly less than the intra-observer intra-scan difference in alpha angle ($P < 0.05$), whether using data normalized to the largest angle in the dataset or not.

Discussion

This study introduced the acetabular contact angle (ACA), an index generated from 3-D US to quantify developmental dysplasia of the hip. We tested the measurement’s reliability and

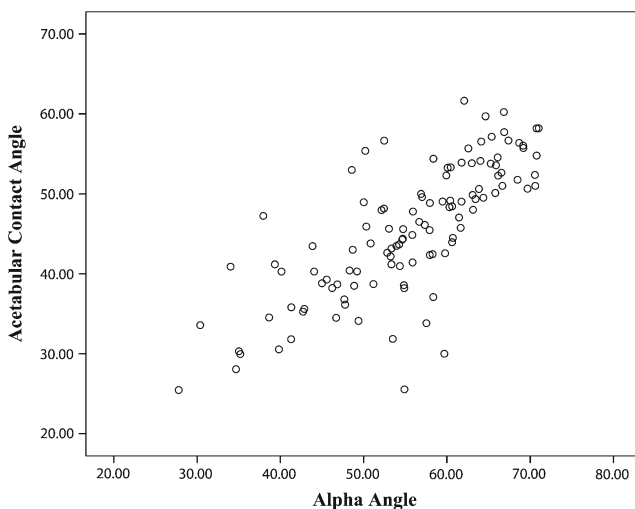


Fig. 6 Graph shows correlation between alpha angle and acetabular contact angle; $r = 0.759$

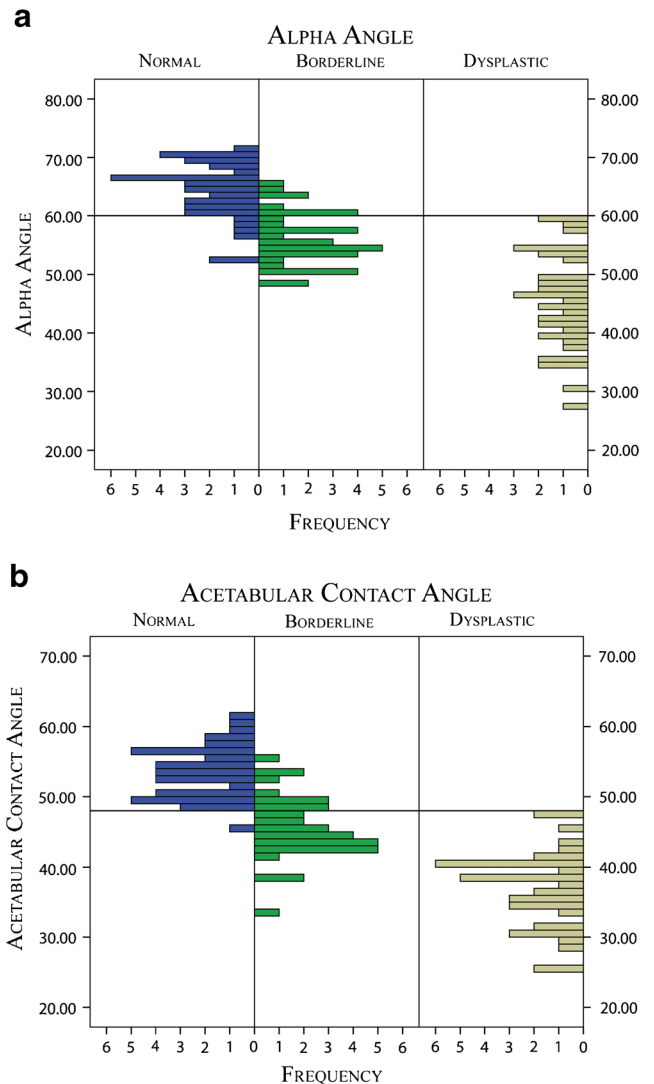


Fig. 7 Case-by-case frequency distribution pyramid graphs for alpha and acetabular contact angles (ACA) by diagnostic category (0=normal, 1=borderline, 2=dysplastic requiring treatment). For the alpha angle (a), category 0 hips are best separated from category 2 at a threshold of 60° (horizontal line), with no missed cases of dysplasia (upper right portion of graph) and six normal hips with abnormal alpha angle values (lower left portion of graph). For the ACA (b), the optimal threshold was 48°, giving no missed dysplastic hips and just one normal hip with an abnormal ACA. Most clinically borderline cases were in the dysplastic range by alpha angle or ACA

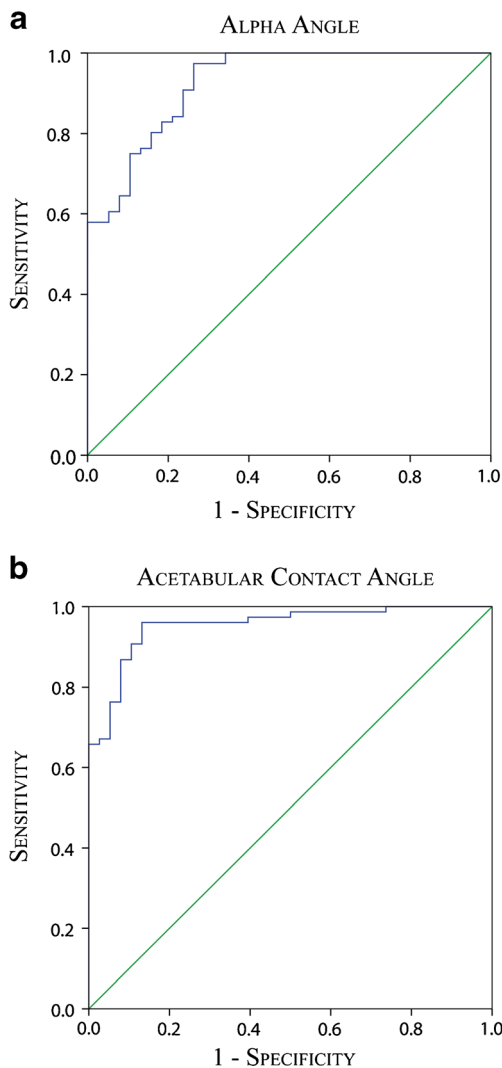


Fig. 8 Receiver operating characteristic curve for alpha angle (a) and acetabular contact angle (b). For the alpha angle the area under the ROC curve was 0.927. For the acetabular contact angle the area under the ROC curve was 0.954

diagnostic utility. We compared ACA reliability to that of the 2-D US alpha angle, which is the index most widely used in DDH diagnosis and which has the lowest variability of any 2-D US index, especially in comparison to beta angle [4, 5, 12, 13] or femoral head coverage [5, 7].

The intra-observer variability of the ACA was similar to that reported for the 2-D US alpha angle ($SD=3.5^\circ$ in this study, and 3° [12], 3.2° [13], 2.8° [6] and 3.1° [14] in the literature). The inter-observer reliability of ACA (interpreting the same images) is also comparable to that of the 2-D US alpha angle in other studies ($SD=2^\circ$ [4] to 3.5° [6]), and better than repeatability coefficient= 17° reported in one study performed by orthopedic surgeons [5]. In our study the 2-D US alpha angle inter-observer variability was relatively high. The inter-scan variability of the ACA (i.e. from different scans of the same hip) was somewhat higher than inter-observer

variability (when two observers analyze the same 3-D US scan), but this was significantly and substantially less than inter-scan variability of the 2-D US alpha angle ($P < 0.05$ parametric and nonparametric tests). Overall the ACA was measured on 3-D US with reliability equivalent to that of the 2-D US alpha angle when observers each reviewed the same 3-D US or 2-D US images, but the ACA was substantially more reliable than the alpha angle on different scans of the hip. This difference is likely a result of the intrinsic benefit of using 3-D US to capture the full acetabulum, compared to 2-D US, in which there is inevitably variation in which plane the sonographer captures a representative image of the hip. Because the ACA is calculated from the entire acetabular shape rather than a selected slice of it, ACA therefore ought to be more reliable than alpha angle, which is what we have demonstrated.

The reported repeatability coefficients for users observer 2 and observer 1, respectively, for the 2-D US findings are 7.6° and 7.1° and for the 3-D US are 5.7° and 5.5° . These values measure the 95% limits of agreement based on $mean \pm 1.96 SD$. By definition, each of these values provides an interval within which 95% of test-retest measurement differences lie — in this case, for observer 2 the values were $\pm 7.6^\circ$ for 2-D US and $\pm 5.7^\circ$ for 3-D US approaches. This implies that the probability of detecting a change in 3-D US ACA greater than 5.7° (i.e. z greater than $+1.96$) in the test population is only 2.5%. Further, if the repeatability coefficient value of 5.7 is halved to 2.85 ($mean + 1SD$, $z = +1$), 84% of subjects would be expected to exhibit test-retest changes in ACA less than 2.85° . The probability of such a change being from measurement noise is 16%, and for a change of 5.7° , just 2.5%. In other words, on the basis of probability, such differences would more than likely represent a genuine difference in 3-D US ACA rather than measurement noise. The significantly lower inter-scan repeatability coefficient for ACA than for the alpha angle implies that the 3-D US ACA is less subject to measurement noise than the 2-D US alpha angle.

The receiver operating characteristic (ROC) curves of the ACA and alpha angle provide a useful insight into the diagnostic utility of these indices. Although both indices were highly effective in diagnosis of hip dysplasia in this dataset, the ACA has greater area under its curve, suggesting that the ACA better measures the condition. The area under both ROC curves is exceptionally high, likely because these scans were performed in a rigorous research-protocol environment by experienced scanners in a dedicated multidisciplinary clinic. In everyday clinical practice the area under the ROC curve would likely be less than this for both indices, as seen in a recent study where this area was 0.836 for the alpha angle [15]. Both ACA and alpha angle showed high diagnostic utility on all categories of hips in this study (i.e. normal, borderline and dysplastic), but ACA over-called only one case of dysplasia requiring treatment in a hip considered clinically normal, compared to six cases for alpha angle.

The moderate correlation of $r=0.759$ between acetabular contact angle and alpha angle indicates that just over half of the variation ($r^2=0.57$) in one index accounts for variation in the other. This implies that the ACA measures different aspects of dysplasia from the alpha angle. The ACA is intended to represent the line of action of the net force that can act from the acetabulum on the femoral head. If the ACA were simply a 3-D version of the alpha angle, the correlation between the two would be expected to be much higher.

It is generally recommended to scan infants at about 6 weeks because before this the hip is considered immature [1, 16, 17]. Although delayed scanning is typically an accepted practice for 2-D US, this pilot study suggests that 3-D US might best be performed after some delay as well, because three-quarters of the clinically borderline hips with ACA overlapping in the dysplastic range were <6 weeks of age.

The 48° threshold we determined between normal and abnormal for the ACA was that which visually best distinguished normal from dysplastic hips in this particular dataset, with the least false-positive readings for DDH, but the optimal threshold value could only be set after extensive further validation with larger study and longer-term follow-up.

The use of 3-D US gives a more complete view of the acetabulum than is available from 2-D US. Given the high reliability and discriminatory ability demonstrated here, 3-D US is likely to be a useful diagnostic tool. Indices such as the ACA can be used quantitatively to screen for hip dysplasia and categorize its severity, analogous to the use of the 2-D alpha and beta angles to establish Graf classifications. Even without the use of a quantitative index such as the ACA, review of 3-D US images and surface models might provide subjective and overall impressions of a child's physiological development of value beyond a numeric index of dysplasia.

This study has limitations. We are the first clinic to systematically add 3-D US for infant hip dysplasia and this gave us a limited sample size to draw from; however, we were still able to include a full range of hips in each diagnostic category ranging from normal to severely dysplastic. More important, the clinical diagnoses and treatment decisions were made by clinical orthopedic surgeons as per usual practice, and were not validated by any external gold standard. Of note, because neither the 3-D US images nor the ACA values were available to the orthopedic surgeons, the diagnoses and treatment decisions were also made entirely independent of 3-D US. Still, our results regarding diagnostic accuracy should be interpreted cautiously because the ultimate diagnosis might vary in individual children after longer-term follow-up. It is especially important to note that the gold standard clinical diagnosis of DDH, although it includes assessment of hip stability and clinical risk factors as well as acetabular morphology, does rely to some extent on the 2-D US alpha angle, so it is unsurprising that the alpha angle correlates highly to the diagnosis. It is difficult to avoid this element of circular

reasoning in study of DDH diagnostic accuracy without multi-year follow-up, because whether a hip is subtly dysplastic or not might not be determined until a child is much older, sometimes not until the third or fourth decade when premature osteoarthritis sometimes develops in these children. Therefore, the true clinical utility of the ACA is uncertain, and extensive further testing against long-term clinical outcomes is needed to confirm its validity as a diagnostic tool.

Conclusion

The acetabular contact angle (ACA) generated from 3-D US data of infant hips is measured more reliably than the traditional alpha angle and can be reproduced 95% of the time within 6° in the same scan, and within 9° in different scans of the same hip. The ACA predicted the presence of hip dysplasia with higher accuracy than the alpha angle, despite the alpha angle forming a key component of the diagnosis of hip dysplasia in current practice. The ACA therefore shows promising initial diagnostic utility. Our findings call for further study into the clinical use of 3-D US examinations for infant hip dysplasia, and these studies should include longer-term follow-up.

Acknowledgments The authors are grateful for the Radiology Society of North America (RSNA) Research Seed Grant #RSD1425 and the Canadian Institutes of Health Research (CIHR) Institute of Human Development Child and Youth Health (IHDCYH) Grant NI15-004, which provided research funding for this work.

Compliance with ethical standards

Conflicts of interest None

References

1. Storer SK, Skaggs DL (2006) Developmental dysplasia of the hip. *Am Fam Physician* 74:1310–1316
2. Shorter D, Hong T, Osborn DA (2013) Cochrane review: screening programmes for developmental dysplasia of the hip in newborn infants. *Evid Based Child Health* 8:11–54
3. Dezateux C, Rosendahl K (2007) Developmental dysplasia of the hip. *Lancet* 369:1541–1552
4. Zieger M (1986) Ultrasound of the infant hip. Part 2. Validity of the method. *Pediatr Radiol* 16:488–492
5. Gwynne Jones DP, Vane AG, Coulter G et al (2006) Ultrasound measurements in the management of unstable hips treated with the Pavlik harness: reliability and correlation with outcome. *J Pediatr Orthop* 26:818–822
6. Morin C, Zouaoui S, Delvalle-Fayada A et al (1999) Ultrasound assessment of the acetabulum in the infant hip. *Acta Orthop Belg* 65:261–265

7. Falliner A, Schwinzer D, Hahne HJ et al (2006) Comparing ultrasound measurements of neonatal hips using the methods of Graf and Terjesen. *J Bone Joint Surg Br* 88:104–106
8. Jaremko JL, Mabee M, Swami VG et al (2014) Potential for change in US diagnosis of hip dysplasia solely caused by changes in probe orientation: patterns of alpha-angle variation revealed by using three-dimensional US. *Radiology* 273:870–878
9. Ozonoff MB (1992) *Pediatric orthopedic radiology*, 2nd edn. Saunders, Philadelphia
10. American Institute of Ultrasound in Medicine (2013) AIUM practice guideline for the performance of an ultrasound examination for detection and assessment of developmental dysplasia of the hip. *J Ultrasound Med* 32:1307–1317
11. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307–310
12. Simon EA, Saur F, Buerge M et al (2004) Inter-observer agreement of ultrasonographic measurement of alpha and beta angles and the final type classification based on the Graf method. *Swiss Med Wkly* 134:671–677
13. Roovers EA, Boere-Boonekamp MM, Geertsma TS et al (2003) Ultrasonographic screening for developmental dysplasia of the hip in infants. Reproducibility of assessments made by radiographers. *J Bone Joint Surg Br* 85:726–730
14. Mabee M, Dulai S, Thompson RB et al (2014) Reproducibility of acetabular landmarks and a standardized coordinate system obtained from 3D hip ultrasound. *Ultrason Imaging* 37:267–276
15. Cheng E, Mabee M, Swami VG et al (2014) Ultrasound quantification of acetabular rounding in hip dysplasia: reliability and correlation to treatment decisions in a retrospective study. *Ultrasound Med Biol* 1:56–63
16. Bialik V, Bialik GM, Blazer S et al (1999) Developmental dysplasia of the hip: a new approach to incidence. *Pediatrics* 103:93–99
17. Roposch A, Liu LQ, Hefti F et al (2011) Standardized diagnostic criteria for developmental dysplasia of the hip in early infancy. *Clin Orthop Relat Res* 469:3451–3461