

Developmental Hip Dysplasia Diagnosis at Three-dimensional US: A Multicenter Study¹

Dornoosh Zonoobi, MEng, PhD
 Abhilash Hareendranathan, PhD
 Emanuel Mostofi, BSc
 Myles Mabee, MSc
 Saba Pasha, PhD
 Dana Cobzas, PhD
 Padma Rao, MBBS, MRCP, FRCR, FRANZCR
 Sukhdeep K. Dulai, MD, BSc, MHSc, FRCSC
 Jeevesh Kapur, MD, MMed, FRCR
 Jacob L. Jaremko, MD, PhD, FRCPC

Purpose:

To validate accuracy of diagnosis of developmental dysplasia of the hip (DDH) from geometric properties of acetabular shape extracted from three-dimensional (3D) ultrasonography (US).

Materials and Methods:

In this retrospective multi-institutional study, 3D US was added to conventional two-dimensional (2D) US of 1728 infants (mean age, 67 days; age range, 3–238 days) evaluated for DDH from January 2013 to December 2016. Clinical diagnosis after more than 6 months follow-up was normal ($n = 1347$), borderline (Graf IIa, later normalizing spontaneously; $n = 140$) or dysplastic (Graf IIb or higher, $n = 241$). Custom software accessible through the institution's research portal automatically calculated indexes including 3D posterior and anterior alpha angle and osculating circle radius from hip surface models generated with less than 1 minute of user input. Logistic regression predicted clinical diagnosis (normal = 0, dysplastic = 1) from 3D indexes (ie, age and sex). Output represented probability of hip dysplasia from 0 to 1 (output: >0.9 , dysplastic; 0.11 – 0.89 , borderline; <0.1 , normal). Software can be accessed through the research portal.

Results:

Area under the receiver operating characteristic curve was equivalently high for 3D US indexes and 2D US alpha angle (0.996 vs 0.987). Three-dimensional US helped to correctly categorize 97.5% (235 of 241) dysplastic and 99.4% (1339 of 1347) normal hips. No dysplastic hips were categorized as normal. Correct diagnosis was provided at initial 3D US scan in 69.3% (97 of 140) of the studies diagnosed as borderline at initial 2D US scans.

Conclusion:

Automatically calculated 3D indexes of acetabular shape performed equivalently to high-quality 2D US scans at tertiary medical centers to help diagnose DDH. Three-dimensional US reduced the number of borderline studies requiring follow-up imaging by over two-thirds.

© RSNA, 2018

¹From the Department of Radiology and Diagnostic Imaging, WC Mackenzie Health Sciences Centre, University of Alberta, 8440-112 St, 2A2.41 Edmonton, AB, Canada T6G 2B7 (D.Z., A.H., E.M., M.M., S.K.D., J.L.J.); Division of Orthopedic Surgery, The Children's Hospital of Philadelphia, Philadelphia, Pa (S.P.); Department of Computing Science, MacEwan University, Edmonton, Canada (D.C.); Department of Medical Imaging, Royal Children's Hospital, Melbourne, Australia (P.R.); and Department of Diagnostic Imaging, National University Hospital, Singapore (J.K.). Received November 5, 2017; revision requested January 5, 2018; revision received March 13; accepted March 14. **Address correspondence to** J.L.J. (e-mail: jjaremko@ualberta.ca).

Study supported by Capital Health Endowed Chair in Diagnostic Imaging; the Institute of Human Development, Child and Youth Health (New Investigator Research Grant); the Canadian Medical Association (Joule Innovation Grant); the Radiological Society of North America (Research Seed Grant); and Canadian Institutes of Health Research – Hospital for Sick Children (Young Investigator Grant). J.J. is supported by Capital Health Endowed Chair in Radiology and Diagnostic Imaging. D.Z. is supported by Alberta Innovates – Health Solutions (Postdoctoral Fellowship Award).

© RSNA, 2018

Developmental dysplasia of the hip (DDH), found in 1.6–28.5 of 1000 infants (1), is associated with premature osteoarthritis and is responsible for nearly 30% of hip arthroplasties in patients younger than 60 years (2). Because treatment by Pavlik harness is highly effective, prompt and accurate diagnosis of DDH in infants could substantially reduce the long-term morbidity of DDH. Diagnosis by clinical examination alone is difficult because Barlow test and Ortolani maneuvers lack sensitivity after the neonatal period and for mild disease (3), especially by nonexpert examiners. Ultrasonography (US) is preferred for infant DDH imaging because it is portable, avoids ionizing radiation, and can view the nonossified cartilaginous infant femoral head (3). Two-dimensional (2D) US examinations usually include dynamic assessment for stability and measurements on a standardized coronal plane static image by Graf method (4,5). Drawbacks of the Graf method include

relatively high interobserver (6–8) and interscan (9,10) variability, potentially altering the final diagnosis in 50%–75% of infants if scanned by a nonexpert (8). In practice, over diagnosis of DDH leads to increased health care costs and unnecessary treatment (2). There are frequent recalls for follow-up imaging of borderline examinations (usually Graf IIa), which limit cost-effectiveness of 2D US in screening for DDH (2).

Three-dimensional (3D) US may overcome some of the limitations of 2D US by providing a more complete view of hip geometry than 2D US, which shows only a partial view of the complex 3D acetabular shape (10). Three-dimensional US for DDH was first proposed and studied in the 1990s (11–15), but probes used slow manual sweeps to produce scans and cumbersome postprocessing was necessary, limiting real-world utility of 3D US. More recently, probes with increases in processing power that perform automatic rapid 3D sweeps underpin a renewed interest in 3D US for DDH (16–22). Three-dimensional US surface models have high fidelity to 3D magnetic resonance imaging reconstructions (21). In small pilot studies (17,18,23), indexes from these surface models helped to predict DDH at least as accurately and reliably as the conventional 2D US alpha angle. However, these findings have not been validated in a large patient group, and it remains unclear to what extent 3D US could reduce borderline scans requiring repeated assessment.

We hypothesized that 3D US scans could provide diagnostic accuracy equivalent to that obtained at expert 2D US scans and that the use of 3D US scans would reduce the need for follow-up in borderline examinations. The purpose of our study was to validate accuracy of diagnosis of DDH from geometric properties of acetabular shape extracted from 3D US.

Materials and Methods

Patients

We undertook a comprehensive study involving a 3-year, four-center

collaboration between the University of Alberta hospital Edmonton (Canada), Royal Children's Hospital Melbourne (Australia), Children's Hospital of Philadelphia (Pa), and National University of Singapore Hospital. Our retrospective study was approved by the health research ethics boards of the participating centers. With written parental consent, 3D US was added to the first conventional 2D US scan of infants who presented to radiology departments for initial evaluation for DDH. Images in our study were collected from the main center from January 2013 to December 2016, and more recently from the other centers as they began obtaining hip 3D US images. The indication for imaging was clinical suspicion of DDH, usually because of breech presentation, laxity at examination, asymmetric hip or thigh creases, and/or other risk factors such as a family history positive for DDH. Because dysplasia can be unilateral or bilateral, we included each hip separately. By December 31, 2016, we had 1884 consecutively scanned hips, performed at mean age 67 days (age range, 3–238 days; Fig 1). In some of our early work (16–18)

Implications for Patient Care

- Three-dimensional (3D) US is at least equivalent to two-dimensional (2D) US in diagnostic accuracy for developmental hip dysplasia, with advantages including a four-fold reduction in the number of patients with inconclusive borderline results who require follow-up imaging.
- Three-dimensional US has potential advantages in feasibility in a screening setting for hip dysplasia because the 3D indexes of dysplasia are calculated automatically from surface models generated with minimal user input, or potentially completely automatically calculated by using deep-learning tools.
- Qualitative review of 3D US images and surface models may help experts to provide subjective impressions of hip development, providing value beyond a numeric index of dysplasia and aiding in individualized treatment planning.

<https://doi.org/10.1148/radiol.2018172592>

Content codes: **PD** **MK** **US**

Radiology 2018; 287:1003–1015

Abbreviations:

DDH = developmental dysplasia of the hip
 OCR = osculating circle radius
 3D = three-dimensional
 $3D\alpha_{Post}$ = 3D alpha angle posterior
 $3D\alpha_{Ant}$ = 3D alpha angle anterior
 2D = two-dimensional

Author contributions:

Guarantors of integrity of entire study, D.Z., A.H., J.L.J.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, D.Z., E.M., D.C., P.R., J.K.; clinical studies, D.Z., M.M., S.P., P.R., S.K.D., J.L.J.; experimental studies, D.Z., A.H., P.R.; statistical analysis, D.Z., A.H., E.M., P.R., J.L.J.; and manuscript editing, all authors

Conflicts of interest are listed at the end of this article.

See also the editorial by Jaramillo in this issue.

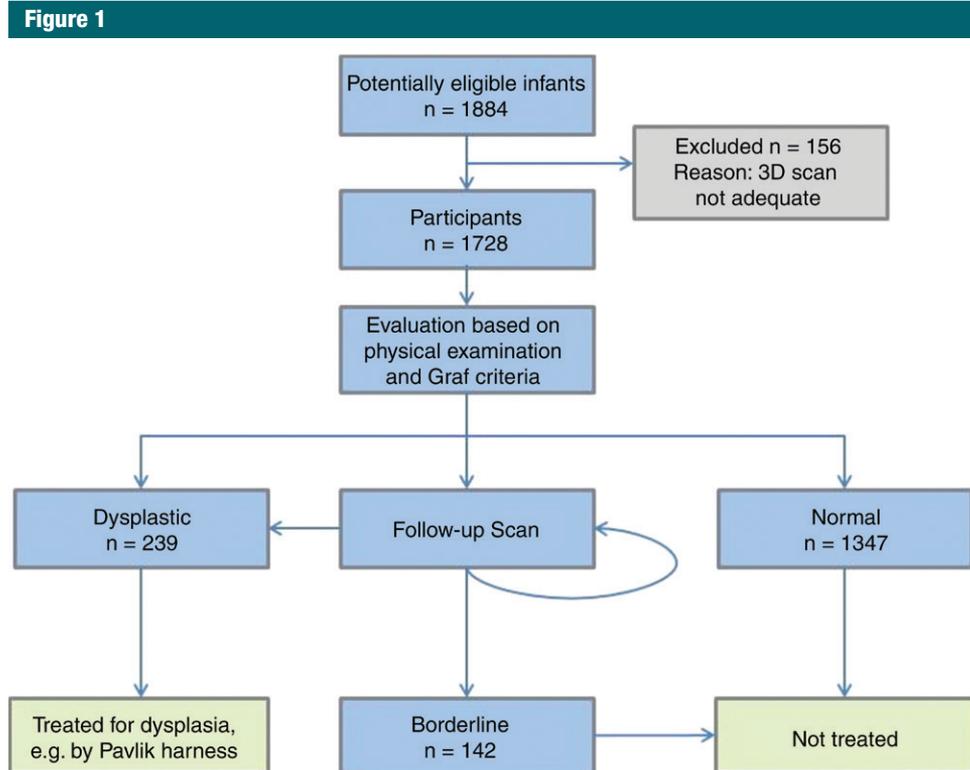


Figure 1: Flowchart shows the initial patient population, reason for exclusion from study, and the three diagnostic categories.

developing the scan protocol, not all of the Graf landmarks (acetabular roof, os ischium, iliac wing, femoral head, and labrum) were adequately visible. We excluded 156 hips (8.3%; 156 of 1884) from our study for this reason, and 41.0% (64 of 156) if these excluded hip scans were from 2013. We found that 95.6% (576 of 602) of our scans from 2016 were adequate.

To assess outcomes, we observed results of routine clinical care for at least 6 months after the initial US scan (average, 8 months; maximum, 23 months) by orthopedic surgeons at each center who were blinded to 3D US images and findings. The reference standard diagnosis was provided by the responsible clinician on the basis of usual clinical care, and it was performed by synthesizing all available information (except 3D US), including 2D US (initial and any follow-up within 6 months), risk factors, and clinical examination. Each imaged hip was classified as normal at initial

assessment (category 0; $n = 1347$), borderline (ie, questionably abnormal initially [generally Graf IIa] but with findings that resolved spontaneously at follow-up imaging and what clinical examination referred to as borderline) (category 1; $n = 140$), or dysplastic and undergoing treatment by Pavlik harness and/or surgery (category 2; $n = 241$). Clinicians were blinded to 3D US reconstructions. Three-dimensional US images were read blindly as anonymized data sets, with no other data about each patient made available to the readers.

US Imaging

Imaging was performed on Philips (iU22, Philips Healthcare, Andover, MA, USA) or Toshiba (Aplio 500, now manufactured by Canon Medical Systems, Otawara, Japan) platforms. We performed conventional 2D US for both hips by using a linear transducer of approximately 12 MHz (Philips 12 L5, Philips Healthcare, Andover, Mass;

Toshiba 14LV7), including static coronal imaging in the Graf standard plane and axial dynamic imaging to assess for hip stability by usual clinical protocol conforming to American College of Radiology recommendations (24). Two-dimensional US images were interpreted by a pediatric radiologist at each center, and results and images were made available to referring clinicians. In addition, one of the clinicians used a high-resolution (11–14 MHz) 3D linear transducer (Philips VL13–5; Toshiba 14LV7, Toshiba Medical Systems, Otawara, Japan) in the coronal orientation to obtain 3D US data sets of each hip (10). At the main study center, scans were performed by either a radiologist (J.J., 15 years of experience), technologist (D.Z. and A.H., with 3 and 4 years of experience, respectively), or a medical or graduate student (M.M. and E.M., with 5 and 1 years of experience, respectively) trained by our radiologist. Teams at other centers were similarly structured. The transducers

Figure 2

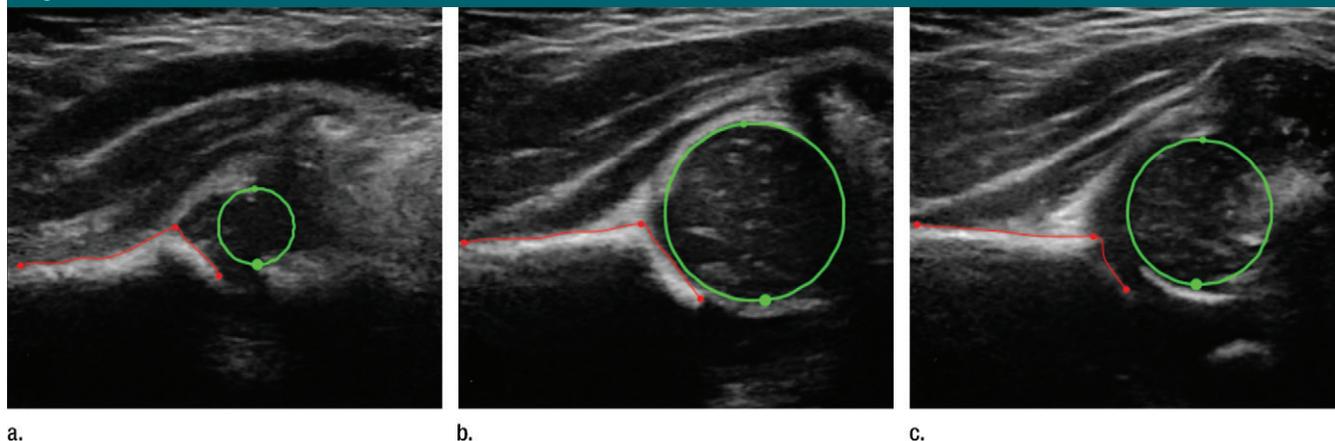


Figure 2: US images of acetabulum and femoral head landmarks and tracing. Coronal images show a normal left hip from (a) anterior, (b) middle, to (c) posterior; sections that are anterior or posterior no longer showed acetabulum edges. There was a typical section between landmarks with associated acetabulum (red) and femoral head (green). The red and green points on each figure indicate user-defined points.

specifications are typical: 3.2-second automated sweep through a range of $\pm 15^\circ$ to generate a 3D data set of 256 US sections that were 0.13 mm thick, each containing 411×192 pixels measuring 0.11×0.20 mm. For reliability analysis, there was a subset of hips for which two of our study members each performed two scans. Three-dimensional US was not released for use in clinical treatment. We selected the highest quality 3D US image for each hip on which the Graf landmarks (acetabular roof, os ischium, iliac wing, femoral head, and labrum) were observable.

Image Processing

The clinically acquired 2D US image of each hip that best met the Graf standard plane requirements was chosen and the alpha angle and acetabular coverage were remeasured centrally as per Graf methods (4) by one of the authors (M.M., with 5 years of experience in DDH imaging) trained by the lead pediatric musculoskeletal radiologist (J.J.).

We performed semiautomated acetabulum and femoral head surface extraction. For each hip, the acetabulum was traced by selecting seed points: three at the acetabular contour on each of three sections (anterior, middle, and posterior) and an additional two points at the midfemoral head (total of

$3 \times (3 + 2) = \leq 15$ user-placed points) (Fig 2). User point selection occurred for 30–75 seconds depending on shape complexity and user experience. Automated interpolation along acetabular and femoral surfaces in all other image sections with these seed points then

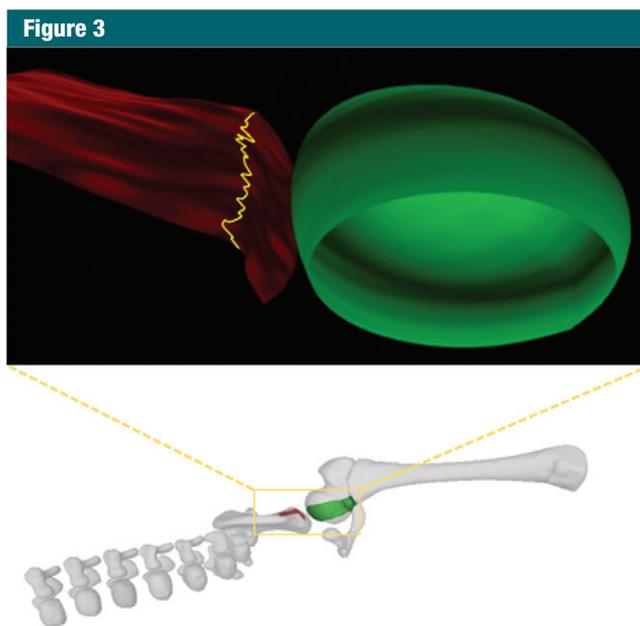


Figure 3: (Top) Illustration of the extracted acetabulum bone and femoral head surfaces from the seed points. The apex line is marked in yellow. (Bottom) Picture of the region segmented from three-dimensional US overlaid on a model of the hip segmented by using computed tomography (CT). Note that model of the hip generated at CT is from a different participant, and is used here to illustrate the anatomic context of the segmented region.

generated a surface model of both the acetabulum and femoral head (Fig 3). The software used in this study was developed in-house by using Python programming language (version 2.7; Python Software Foundation, Beaverton, Ore) and the Visualization Toolkit

Figure 4

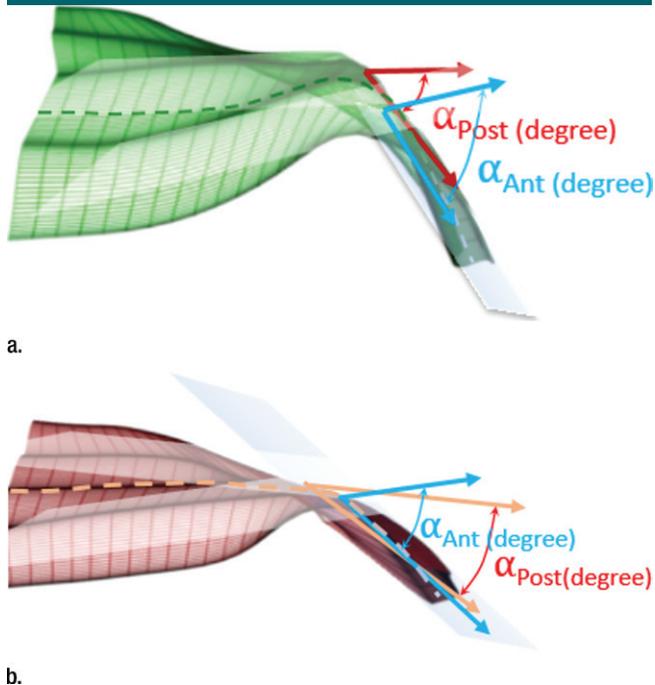


Figure 4: (a) The figure in green shows the surface model of normal hip ($3D\alpha_{Post} = 65^\circ$, $3D\alpha_{Ant} = 57^\circ$, where $3D\alpha_{Post}$ and $3D\alpha_{Ant}$ are three-dimensional [3D] alpha angle posterior and anterior, respectively) and (b) the figure in red shows a dysplastic hip ($3D\alpha_{Post} = 41^\circ$, $3D\alpha_{Ant} = 36^\circ$).

Figure 5

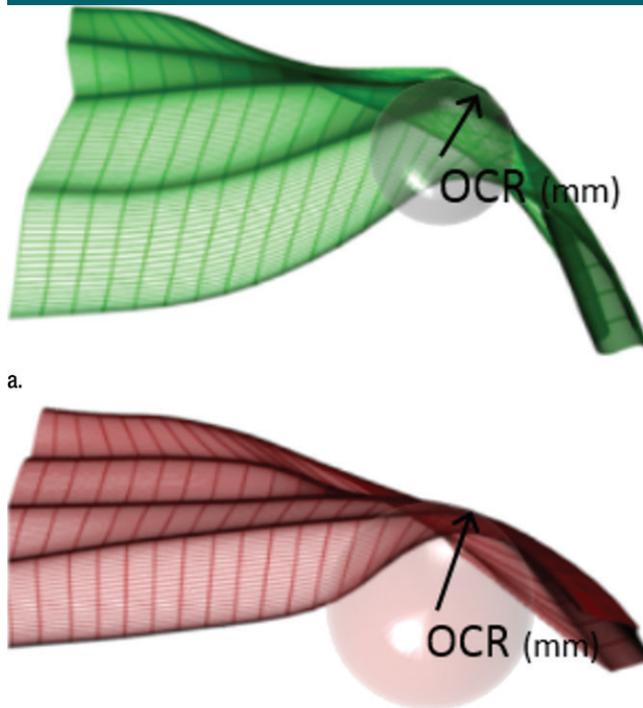


Figure 5: Illustration of the osculating circle radius (OCR) for (a) a normal hip (OCR, 6.2 mm) and (b) a dysplastic hip (OCR, 15.3 mm).

(version 5.10; Kitware, Clifton Park, NY).

We incorporated the software into our website where it can be used as an online tool (<http://niduscanada.com/contact>).

After 3D US surface model generation, we automatically generated the apex line along the lateral acetabular roof, which marks the separation of the ilium and the acetabulum (yellow line in Fig 3), then we calculated three indexes of acetabular shape: 3D alpha angle posterior and anterior ($3D\alpha_{Post}$ and $3D\alpha_{Ant}$, respectively), and the osculating circle radius (OCR) (Figs 4, 5).

We separated the acetabulum into anterior and posterior segments by using the middle section tracing (Fig 2b) as the dividing landmark between the anterior (Fig 2a) and the posterior (Fig 2c) acetabulum. The alpha angle for each segment was calculated as the angle between planes best fitting the 3D surface at the lateral iliac wall and acetabular

roof, expressed as the angle between normal vector of the acetabular plane (N_a) and the normal vector of the iliac plane (N_i), where \cos is cosine:

$$\alpha = \cos^{-1} \left(\frac{N_a \cdot N_i}{|N_a| |N_i|} \right)$$

The OCR is an index that represents the radius of the largest sphere fit under the curve of the lateral acetabulum along the apex line; the larger the OCR, the more rounded the acetabular roof margin. Mathematically, the OCR is the inverse of the first principal curvature K_1 of the 3D acetabular surface, calculated from the relationship between the Gaussian curvature K_G (an intrinsic property of the surface) and the mean curvature K_M (which depends on the surface's orientation in space). We first parametrize the extracted surface as $s(x, y) = [x, y, h(x, y)]$, where $h(x, y)$ is

the height of surface at (x, y) . The first

and second derivatives are $s_x = (1, 0, h_x)$, $s_y = (0, 1, h_y)$, $s_{xx} = (0, 0, h_{xx})$, $s_{xy} = (0, 0, h_{xy})$, $s_{yy} = (0, 0, h_{yy})$, and the

normal vector at point (x, y) is given by

$$N(x, y) = \frac{s_x \times s_y}{|s_x \times s_y|} = \frac{(-h_x, -h_y, 1)}{\sqrt{1 + h_x^2 + h_y^2}}$$

$$\text{At point } (x, y), K_G = \frac{h_{xx}h_{yy} - h_{xy}^2}{(1 + h_x^2 + h_y^2)^2}$$

$$K_M = \frac{h_{xx}(1 + h_y^2) + h_{yy}(1 + h_x^2) - 2h_xh_yh_{xy}}{2(1 + h_x^2 + h_y^2)^{3/2}}$$

Because $K_G = K_1K_2$, and

$$K_M = (K_1 + K_2)/2, K_1 \text{ is computed}$$

from K_G and K_M , and then $OCR = \frac{1}{K_1}$.

A fourth index, 3D femoral head coverage, was also calculated for our study but was not found to be useful in improving the performance of our classifier and was not assessed further in our study.

Reliability of Indexes

To assess index reliability, two sonographers performed two 3D US scans of 60 hips in 60 patients. We invited seven readers (two radiologists, three medical and engineering postdoctoral fellows, and two graduate students) to trace acetabular contours on both images as in Figure 2c, twice on two separate days. We then automatically generated 3D acetabular surface models and the three indexes above from each tracing. We computed root mean square error between the generated 3D surfaces, and we computed intraobserver, interobserver, and interscan variability of the calculated indexes as intraclass correlation coefficients.

Automatic Diagnosis

We performed simple logistic regression to predict clinical diagnosis (normal = 0, dysplastic = 1) from the 3D indexes and basic demographics (age and sex). The regression model output was interpreted as a probability of hip dysplasia from 0 to 1. We considered regression outputs greater than 0.9 to be dysplastic; 0.11–0.89, borderline; and less than 0.1, normal. Statistics were calculated with software (SPSS 22; SPSS, Chicago, Ill).

Results

Surface model reliability expressed as root mean square distance between surfaces derived from pairs of tracings by different readers was mean 0.27 mm \pm 0.22 (standard deviation). Intraclass correlation coefficients (two-way mixed effects, consistency, single rater/measurement 3,1) for $3D\alpha_{Post}$, $3D\alpha_{Ant}$, and OCR indexes averaged across all seven observers were as follows, respectively: at the same scan, intraobserver variability was 0.91, 0.87, and 0.73; interobserver variability was 0.89, 0.83, and

Table 1

Mean Results for the Two-dimensional US Alpha Angle and Three-dimensional Indexes

Index	Normal Result	Borderline Result	Dysplastic Result
$2D\alpha$ ($^{\circ}$)	67.2 \pm 4.9 (67.0, 67.6)	60.5 \pm 5.9 (59.6, 61.5)	49.1 \pm 8.1 (47.6, 49.7)
$3D\alpha_{Post}$ ($^{\circ}$)	61.1 \pm 8.4 (60.6, 61.5)	56.3 \pm 9.7 (54.6, 57.9)	40.1 \pm 8.4 (39.0, 41.1)
$3D\alpha_{Ant}$ ($^{\circ}$)	54.5 \pm 9.2 (53.9, 54.9)	48.9 \pm 11.2 (47.0, 50.7)	25.8 \pm 8.4 (24.8, 26.9)
OCR (mm)	10.8 \pm 2.7 (10.6, 10.9)	11.6 \pm 2.6 (11.1, 12)	15.4 \pm 2.5 (15.0, 15.7)

Note.—Data are mean \pm standard deviation; data in parentheses are 95% confidence intervals. OCR = osculating circle radius, $2D\alpha$ = two-dimensional alpha angle, $3D\alpha_{Post}$ = posterior three-dimensional alpha angle, $3D\alpha_{Ant}$ = anterior three-dimensional alpha angle.

0.71; and on different scans obtained the same day, interscan variability was 0.68, 0.62, and 0.50.

The clinical categories of dysplasia (scores of 0, 1, and 2) were well separated into distinct groups by the 2D US alpha angle and by each 3D US index, with statistically significant differences in mean values for each index (Table 1) ($P < .001$). Case-by-case frequency distributions shown on pyramid graphs show optimal separation of normal versus dysplastic hips at a threshold of 60° for 2D US alpha angle (Fig 6), 45° for $3D\alpha_{Post}$, and 39° for $3D\alpha_{Ant}$ (Fig 7).

The moderate correlation ($r = 0.6570$) between $3D\alpha_{Post}$ and $3D\alpha_{Ant}$ angles implied that just over half of the variation ($r^2 = 0.43$) in one index accounted for variation in the other (ie, these two indexes measured different aspects of dysplasia) (Fig 8).

Receiver operating characteristic curves were generated (Fig 9) for a diagnostic test by using 2D US or 3D US indexes from the first scan to help detect DDH that required treatment, compared with clinical reference standard diagnosis. The area under the receiver operating characteristic curve was slightly higher for the combination of $3D\alpha_{Post}$ and $3D\alpha_{Ant}$ compared with the 2D alpha angle (0.994 and 0.987, respectively). The highest area under the receiver operating characteristic curve was 0.996, achieved through inclusion of the OCR (Fig 9).

Table 2 shows the confusion matrix for the regression model that classifies each hip from 3D indexes versus clinical diagnosis categories.

The 3D US model accurately categorized 97.5% (235 of 241) and 99.4% (1339 of 1347) of dysplastic and normal hips, respectively. No dysplastic hips were categorized as normal and no normal hips were categorized as dysplastic. It also gave the correct diagnosis directly at initial 3D US scan in 69.3% (97 of 140) of the hips that were borderline at initial 2D US (requiring at least one follow-up 2D US to establish a diagnosis). Three-dimensional US resulted in 39 borderline diagnoses compared with the 140 generated by 2D US, a reduction of 72.1% (101 of 140).

The clinical implications of the use of 3D US to categorize hips as normal, borderline, or dysplastic are shown in Figure 10. Figure 11 shows that nearly all hips were confidently determined as either normal or dysplastic by the 3D US model.

Discussion

In our large multicenter study of the diagnostic utility of 3D US to quantify and categorize infant hip dysplasia, we found that performance of automatically calculated 3D indexes of acetabular shape (angularity and roundness) was at least equivalent to high-quality 2D US scans performed at tertiary medical centers to diagnose DDH, and it reduced the number of indeterminate borderline scans by over two-thirds.

The specific 3D indexes used in the classification algorithm were selected after preliminary testing in which acetabular coverage and the

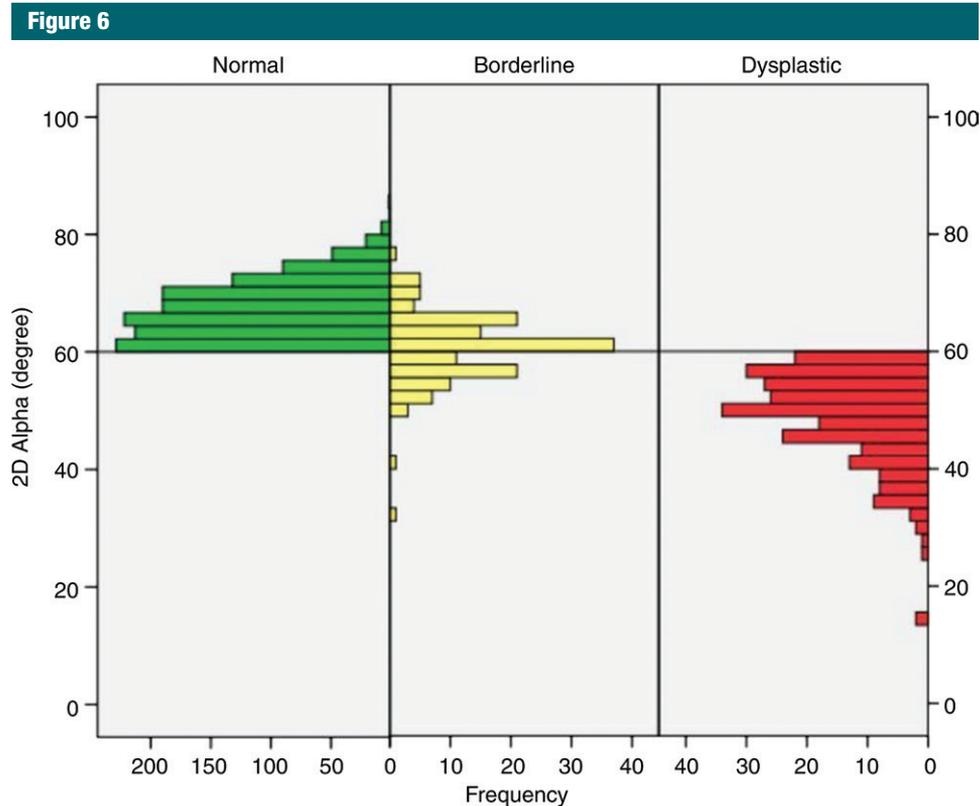


Figure 6: Case-by-case frequency distribution pyramid graphs for the two-dimensional alpha angle by diagnostic category (normal, borderline, and dysplastic requiring treatment). Normal hips are best separated from dysplastic hips at an angle of 60° (horizontal line).

acetabular contact angle were found to be unhelpful. The acetabular contact angle, another 3D index of acetabular angularity, was previously found to be highly accurate for diagnosis of DDH (18). In that study, the acetabular contact angle was calculated on smooth, manually segmented acetabular surface models. However, the surfaces in our study were generated semiautomatically and were more prone to minor irregularities, which we found to affect acetabular contact angle calculation. The indexes we proposed are robust and well suited to use on irregular surfaces.

Reliability of 3D alpha angle indexes was higher interobserver than interscan, which was expected because of reduction (but not elimination) of the effect of varying the transducer position. We did not focus on formally comparing reliability of 2D US versus 3D US indexes in this study, but we

previously demonstrated 3D US indexes to be more reliable than concurrently obtained 2D US alpha angle (17). Just as other 2D US indexes (such as the beta angle) are less reliable than the 2D US alpha angle, the OCR was less reliable than 3D alpha angles. Lower reliability of OCR was likely because of higher sensitivity of OCR to small variations in acetabular curvature. Our data validated that 3D acetabular surface models and indexes automatically calculated from them can be rapidly and reliably generated from 3D US by users with a wide variety of imaging experience. Users with minimal training could generate indexes to predict DDH from 3D US in approximately 1 minute of processing time. Because the indexes are generated from 3D data, they provide additional information regarding the 3D shape of the acetabular surface and are more reliable than the single-plane image of 2D US.

The receiver operating characteristic curves of the 3D indexes and 2D alpha angle provide useful insight into the diagnostic utility of these indexes. Although all indexes were highly effective in diagnosis of hip dysplasia in this dataset (Fig 9), the combination of the 3D indexes produced an area under the receiver operating characteristic curve of 0.996 versus 0.987 for the 2D US alpha angle (ie, the 3D US indexes slightly outperformed high quality 2D US) despite the fact that DDH is a condition in which the current definition relies heavily on findings at 2D US. This difference, however, did not meet statistical significance ($P = .0636$). The area under the receiver operating characteristic curve is exceptionally high for all indexes, likely because the scans were performed in a rigorous research environment by experienced radiologists in dedicated multidisciplinary clinics,

Figure 7

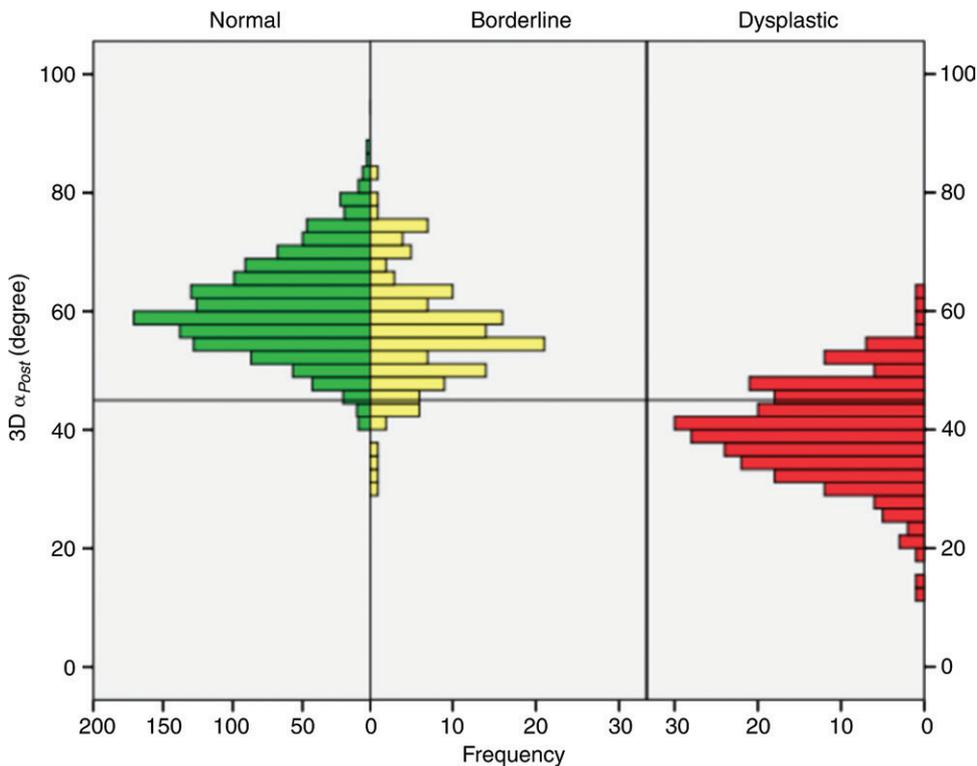
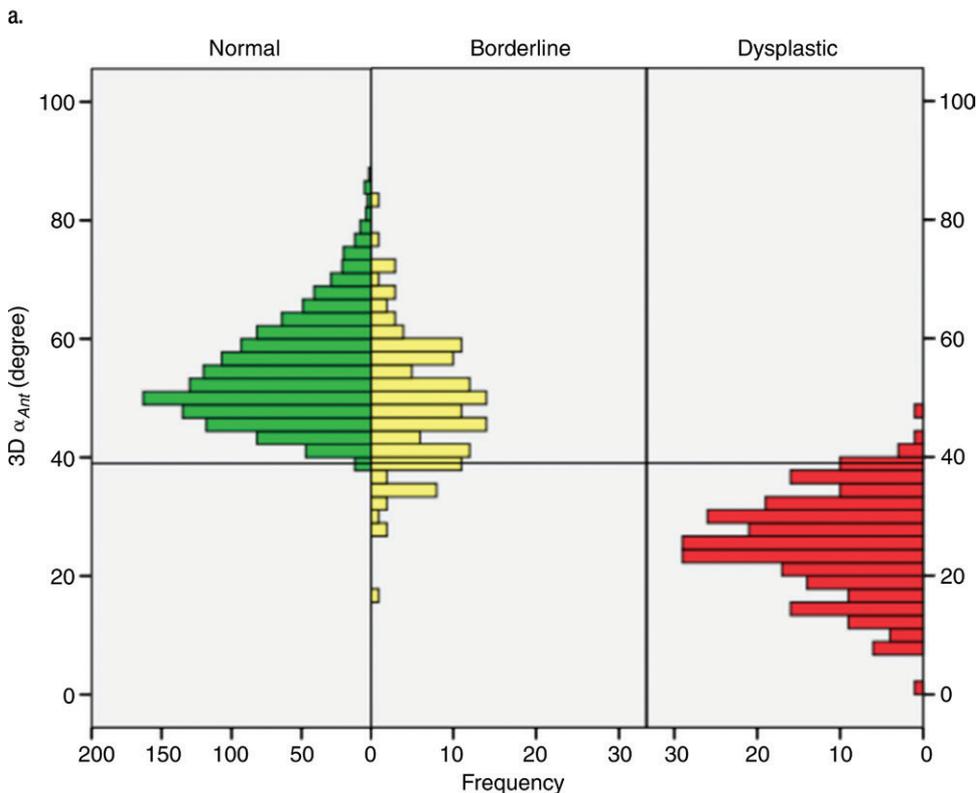
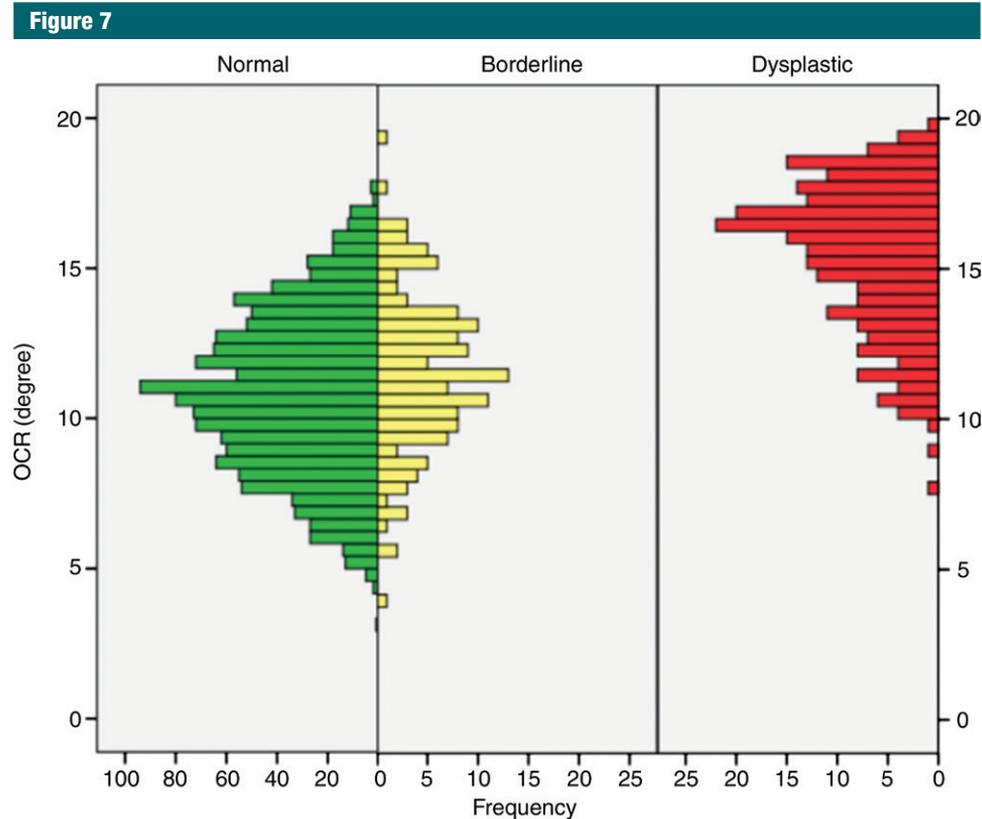


Figure 7: (a-c) Case-by-case frequency distribution pyramid graphs for three-dimensional (3D) (a) posterior and (b) anterior alpha angles (Fig 7 continues).



a.

b.



c.

Figure 7 (continued): (c) Osculating circle radius (OCR), in millimeters, by diagnostic category (normal, borderline, and dysplastic requiring treatment). Normal hips are best separated from the dysplastic hips at an angle of 45° for the three-dimensional (3D) alpha angle posterior ($3D\alpha_{Post}$, **a**) and 39° for the 3D alpha angle anterior ($3D\alpha_{Ant}$, **b**).

representing an idealized situation. In everyday clinical practice at centers with less experienced providers, diagnostic accuracy would decline, as in a recent study (17) where area under the receiver operating characteristic curve was 0.836 for the 2D US alpha angle. Because of the ease of use of 3D US, accuracy may decline less for 3D US than for 2D US in the hands of novice radiologists; this will be tested in future studies.

The crucial borderline category, in which hips initially indeterminate for dysplasia at 2D US normalize at follow-up examination, usually represents Graf IIa hips (alpha angle, 50° – 59° ; age, <3 months). Hips in this category eventually transition to normal Graf I (alpha angle, $>60^\circ$) 97% of the time (5). This category is important because a high rate of

infants who require follow-up imaging limits cost effectiveness of screening for DDH. In our study, the initial 3D US was able to confidently and correctly help define as normal to be borderline at the initial 2D US (Fig 10). Thus, 3D US screening could help to substantially save costs, reduce the number of required follow-up scans, ease parental anxiety by eliminating the 6-week waiting periods for follow-up scanning, and prevent unnecessary treatment of patients whose hips might be placed in a Pavlik harness as a precaution during the follow-up period.

Our classification model on the basis of 3D indexes misclassified six dysplastic hips (category 2) as borderline (category 1) (Table 2). However, manual consensus inspection of these

six hips revealed that all six could have reasonably been categorized as Graf IIa at initial 2D US; all were at most mildly dysplastic. This highlights the limitations of the current reference standard for clinical diagnosis of DDH: the difference between Graf I (normal) and Graf IIa (borderline) can be a tiny subjective variation in alpha angle measurement. Future use of 3D US in longitudinal studies may provide further insight into which acetabular shapes are unlikely to normalize spontaneously over time, helping redefine hip dysplasia in a more prognostically useful way. Eventually 3D US analysis might prevent overtreatment of mild hip dysplasia.

The rate of Graf IIa (borderline) classification in our study (8.1% [140 of 1728]), is lower than in most published studies (10%–15% of hips)

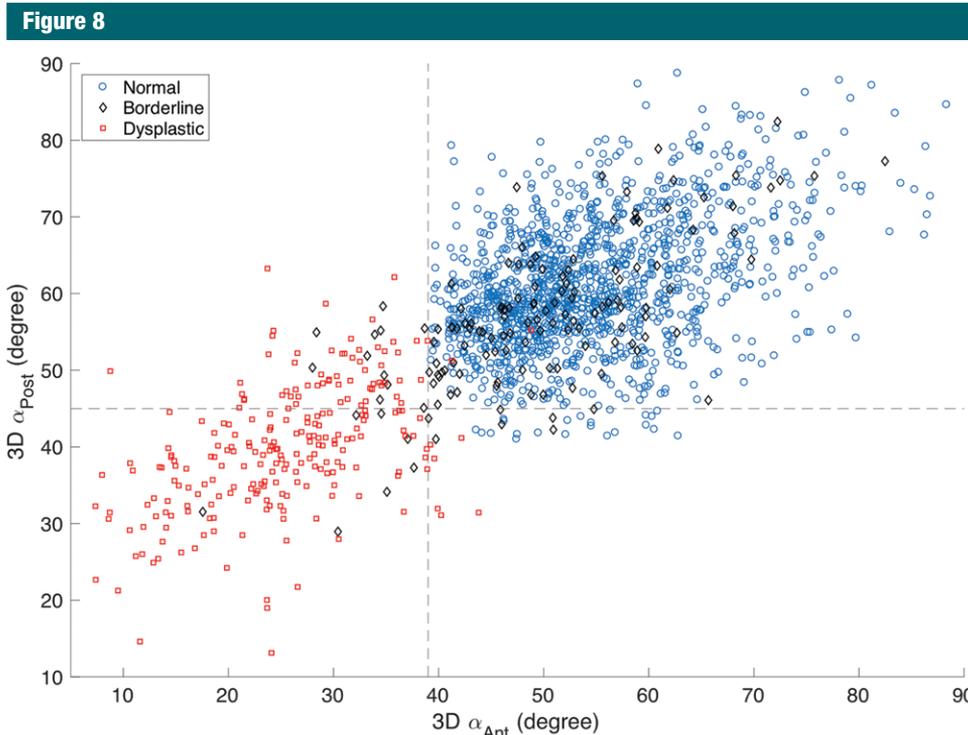


Figure 8: Scatterplot shows the correlation between three-dimensional (3D) alpha angle posterior ($3D\alpha_{Post}$) and 3D alpha angle anterior ($3D\alpha_{Ant}$) for different categories ($r = 0.6570$; $P < .0001$).

(25,26). The area under the receiver operating characteristic curve is likely because of the controlled environment of our study in tertiary centers with experienced radiologist. The magnitude of reduction in borderline hips could potentially be even greater if 3D US is extended to general clinical practice.

We observed during our study that the posterior and anterior acetabular curvature at hip 3D US have different meanings. The mean $3D\alpha_{Post}$ is higher than the $3D\alpha_{Ant}$ in both normal and dysplastic hips (Fig 7), with the difference between these two angles higher in dysplastic than normal hips. In dysplastic hips the anterior acetabulum became more flattened than the relatively well-preserved posterior acetabular lip, so that a smaller $3D\alpha_{Ant}$ may be a better indicator of dysplasia than its posterior counterpart. This compares well with other work (27) that demonstrates a primary anterior

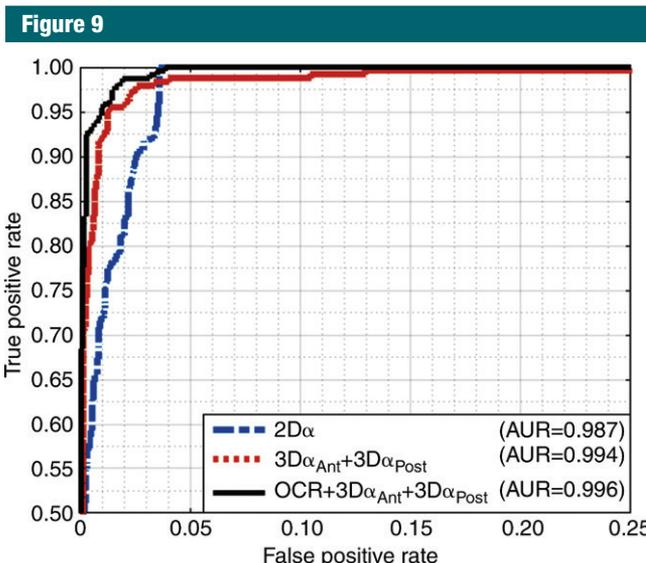


Figure 9: Receiver operating characteristic curve for two-dimensional (2D) alpha angle ($2D\alpha$) and 3D indexes. Note that the area under the curve obtained from the combination of anterior 3D alpha angle ($3D\alpha_{Ant}$), posterior 3D alpha angle ($3D\alpha_{Post}$), and osculating circle radius (OCR) gave a marginally higher accuracy than 2D alpha angle (but this difference is small and was not statistically significant; $P = .0636$).

Figure 10

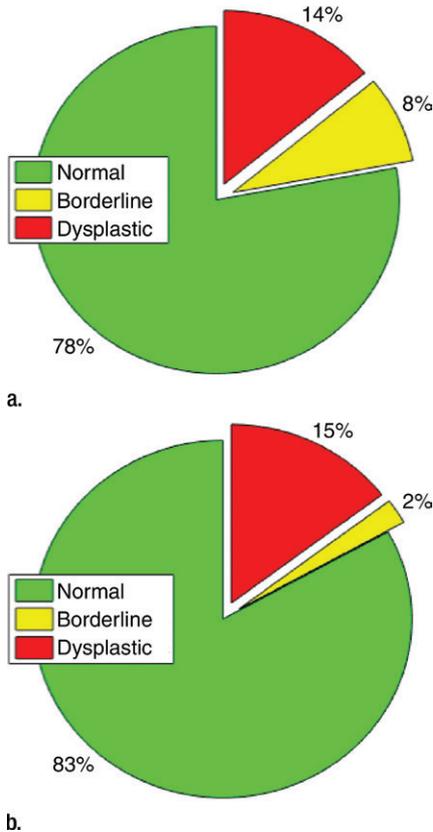


Figure 10: Portion of different categories on the basis of the (a) clinical diagnosis and (b) prediction from three-dimensional (3D) US indexes. In clinical assessment of over 1700 hips, 8% were considered borderline. At concurrent but blinded assessment, only 2% of these hips were categorized as borderline and requiring follow-up at 3D US. If 3D US had been in use for screening in this population, the number of follow-up examinations required would have been reduced substantially, reducing system costs. The 3D US classification was also accurate relative to clinical diagnosis, with no dysplastic hip classified as normal (Table 2).

deficiency of acetabular coverage in patients with DDH.

Regarding acetabular rounding, we found that despite measurement variability a larger OCR signified increased acetabular rounding and an increased risk of dysplasia. In our study population, normal hips could have a wide range of OCR (ie, sharp or rounded curvature), but

Table 2

Confusion Matrix: Patients in Clinical Diagnostic Categories versus Predicted Categories at Three-dimensional US

Clinical Diagnostic Categories	Predicted Categories		
	Normal	Borderline	Dysplastic
Normal	1339	8	0
Borderline	97	25	18
Dysplastic	0	6	235

Note.—Data are number of patients.

Figure 11

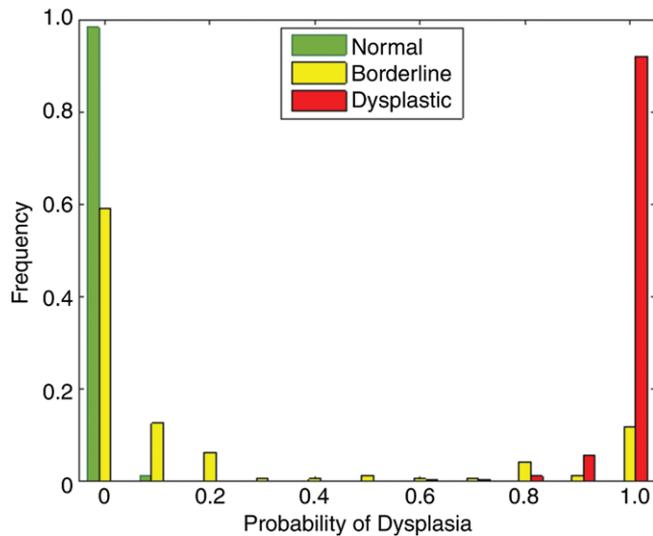


Figure 11: Normalized histogram of probability values showing the occurrence of each diagnostic category versus probability of dysplasia (as computed by the regression classifier). This histogram helps to understand the meaning of the probability of dysplasia generated from the three-dimensional (3D) US analysis package. The 3D US was decisive (ie, nearly all hips had a probability of dysplasia near 0 or 1) and it was accurate (ie, a probability near 0 was never associated with dysplasia by clinical reference standard and probability near 1 was never associated with a normal hip).

nearly all dysplastic hips had OCR of 10 or greater (rounded; Fig 7). Any hip in our study with OCR less than 10 was normal regardless of the value of the alpha angle; this rule may help assess borderline hips. While the OCR index is not itself diagnostic for hip dysplasia, it may be a useful supplementary index for DDH classification.

This study had limitations. First, it was conducted by using thin-section 3D US transducers that, although readily available from various manufacturers, are specialized

transducers not often found in clinical US departments. We have not yet tested whether surface models generated from the more widely available lower-frequency 3D US transducers used in obstetrical imaging (eg, 7 MHz) would be equivalent in accuracy. However it has been shown in earlier works (20) that the segmentation algorithm was robust to moderate levels of noise and the scale of these variations would be less than 2 mm. Second, our assessment of accuracy and reliability of 2D and 3D US

in this study was on the basis of examinations performed by experienced users at high-volume tertiary referral centers. Two-dimensional US has an inherent limitation: it relies on a 2D plane to analyze complex 3D anatomy. This leads to high interscan variability (10). Interscan variability is relatively small in 3D US, but still greater than interreader intrascan variability (21). Our study does not test the so-called real-world performance of 2D US versus 3D US in lower-volume or peripheral centers with less experienced users. Diagnostic accuracy (ie, area under the receiver operating characteristic curve) would be expected to be lower in those settings. A third limitation, intrinsic to any evaluation of DDH, is the lack of an external pathologic tissue reference standard. Diagnosis and treatment decisions made by experienced orthopedic surgeons are necessarily subject to individual variability. Because of the reliance of the reference standard clinical diagnosis of DDH on the 2D US alpha angle, it is unsurprising that this index is associated with DDH diagnosis. This element of circular reasoning is difficult to avoid in a study of DDH diagnostic accuracy. A multiyear follow-up assessment for 3D acetabular shape with teenage patients, and for development of premature osteoarthritis with young adult patients, might eventually allow us to refine our definition of hip dysplasia. Three-dimensional US shows promising diagnostic utility because it is equivalent or superior to 2D US versus the existing reference standard, but further testing against long-term clinical outcomes is necessary to confirm its validity as a diagnostic tool.

Disclosures of Conflicts of Interest: **D.Z.** disclosed no relevant relationships. **A.H.** Activities related to the present article: disclosed patent on file for the segmentation algorithm that was used for this study. Activities not related to the present article: disclosed no relevant relationships. Other relationships: disclosed no relevant relationships. **E.M.** disclosed no relevant relationships. **M.M.** Activities related to the present article: disclosed patent on file for the segmentation algorithm that was used for this study. Activities not related to

the present article: disclosed no relevant relationships. Other relationships: disclosed no relevant relationships. **S.P.** disclosed no relevant relationships. **D.C.** disclosed no relevant relationships. **P.R.** disclosed no relevant relationships. **S.K.D.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: disclosed travel support paid to author for board membership from Canadian Orthopedic Association; travel support paid to author from St Justine Pediatric Orthopedic Review Course. Other relationships: disclosed no relevant relationships. **J.K.** disclosed no relevant relationships. **J.L.J.** disclosed no relevant relationships.

References

1. Storer SK, Skaggs DL. Developmental dysplasia of the hip. *Am Fam Physician* 2006;74(8):1310–1316.
2. Shorter D, Hong T, Osborn DA. Cochrane Review: Screening programmes for developmental dysplasia of the hip in newborn infants. *Evid Based Child Health* 2013;8(1):11–54.
3. Dezateux C, Rosendahl K. Developmental dysplasia of the hip. *Lancet* 2007;369(9572):1541–1552.
4. Graf R. The diagnosis of congenital hip-joint dislocation by the ultrasonic Compound treatment. *Arch Orthop Trauma Surg* 1980;97(2):117–133.
5. Graf R, Mohajer M, Plattner F. Hip sonography update. Quality-management, catastrophes - tips and tricks. *Med Ultrason* 2013;15(4):299–303.
6. Zieger M. Ultrasound of the infant hip. Part 2. Validity of the method. *Pediatr Radiol* 1986;16(6):488–492.
7. Gwynne Jones DP, Vane AG, Coulter G, Herbison P, Dumbar JD. Ultrasound measurements in the management of unstable hips treated with the pavlik harness: reliability and correlation with outcome. *J Pediatr Orthop* 2006;26(6):818–822.
8. Morin C, Zouaoui S, Delvalle-Fayada A, Delforge PM, Leclat H. Ultrasound assessment of the acetabulum in the infant hip. *Acta Orthop Belg* 1999;65(3):261–265.
9. Falliner A, Schwinzer D, Hahne HJ, Hedderich J, Hassenpflug J. Comparing ultrasound measurements of neonatal hips using the methods of Graf and Terjesen. *J Bone Joint Surg Br* 2006;88(1):104–106.
10. Jaremko JL, Mabee M, Swami VG, Jamieson L, Chow K, Thompson RB. 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* 2014;273(3):870–878.
11. Gerscovich EO, Greenspan A, Cronan MS, Karol LA, McGahan JP. Three-dimensional sonographic evaluation of developmental dysplasia of the hip: preliminary findings. *Radiology* 1994;190(2):407–410.
12. Graf R, Lercher K. Experiences with a 3-D ultrasound system in infant hip joints [in German]. *Ultraschall Med* 1996;17(5):218–224.
13. Sohn C, Lenz GP, Thies M. 3-dimensional ultrasound image of the infant hip [in German]. *Ultraschall Med* 1990;11(6):302–305.
14. Böhm K, Niethard FU. Three-dimensional ultrasound image of the infant hip [in German]. *Bildgebung* 1994;61(2):126–129.
15. Ozonoff MB. *Pediatric Orthopedic Radiology*. RD 732.3 C48 O99. University of Alberta. JW Scott Health Sciences, 1992.
16. Mabee M, Dulai S, Thompson RB, Jaremko JL. Reproducibility of acetabular landmarks and a standardized coordinate system obtained from 3D hip ultrasound. *Ultrasound Imaging* 2015;37(4):267–276.
17. Mabee MG, Hareendranathan AR, Thompson RB, Dulai S, Jaremko JL. An index for diagnosing infant hip dysplasia using 3-D ultrasound: the acetabular contact angle. *Pediatr Radiol* 2016;46(7):1023–1031.
18. Hareendranathan AR, Mabee M, Punithakumar K, Noga M, Jaremko JL. Toward automated classification of acetabular shape in ultrasound for diagnosis of DDH: Contour alpha angle and the rounding index. *Comput Methods Programs Biomed* 2016;129:89–98.
19. Quader N, Hodgson A, Mulpuri K, Cooper A, Abugharbieh R. Towards reliable automatic characterization of neonatal hip dysplasia from 3d ultrasound images. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, October 17, 2016; 602–609.
20. Hareendranathan AR, Mabee M, Punithakumar K, Noga M, Jaremko JL. A technique for semiautomatic segmentation of echogenic structures in 3D ultrasound, applied to infant hip dysplasia. *Int J CARS* 2016;11(1):31–42.
21. Diederichs C, Heath A, Hareendranathan AR, et al. Cross-modality validation of acetabular surface models using 3-D ultrasound versus magnetic resonance imaging in normal and dysplastic infant hips. *Ultrasound Med Biol* 2016;42(9):2308–2314.

22. Hareendranathan AR, Zonoobi D, Mabee M, et al. Semiautomatic classification of acetabular shape from three-dimensional ultrasound for diagnosis of infant hip dysplasia using geometric features. *Int J CARS* 2017;12(3):439–447.
23. Hareendranathan AR, Zonoobi D, Mabee M, et al. Toward automatic diagnosis of hip dysplasia from 2D ultrasound. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), April 18, 2017; 982–985.
24. American Institute of Ultrasound in Medicine. AIUM practice guideline for the performance of an ultrasound examination for detection and assessment of developmental dysplasia of the hip. *J Ultrasound Med* 2013;32(7):1307–1317.
25. Dogruel H, Atalar H, Yavuz OY, Sayli U. Clinical examination versus ultrasonography in detecting developmental dysplasia of the hip. *Int Orthop* 2008;32(3):415–419.
26. Kosar P, Ergun E, Gökharman FD, Turgut AT, Kosar U. Follow-up sonographic results for Graf type 2A hips: association with risk factors for developmental dysplasia of the hip and instability. *J Ultrasound Med* 2011;30(5):677–683.
27. Fujii M, Nakashima Y, Sato T, Akiyama M, Iwamoto Y. Pelvic deformity influences acetabular version and coverage in hip dysplasia. *Clin Orthop Relat Res* 2011;469(6):1735–1742.

- McDonald RJ, McDonald JS, Kallmes DF, et al. Intracranial gadolinium deposition after contrast-enhanced MR imaging. *Radiology* 2015;275(3):772–782.

Response

From

Ashkan Heshmatzadeh Behzadi, MD,* and Martin R. Prince, MD, PhD*†‡

Departments of Radiology* and Healthcare Policy and Research,† Weill Cornell Medical Center, 416 E 55th St, New York, NY 10022

e-mail: map2008@med.cornell.edu

Department of Radiology, Columbia College of Physicians and Surgeons, New York, NY‡

We thank Drs Adin and Yousem for their thought-provoking comments on our article, “Dentate Nucleus Signal Intensity Decrease on T1-weighted MR Images after Switching from Gadopentetate Dimeglumine to Gadobutrol” (1). With regard to the patients who were excluded because they had only three or fewer gadobutrol-enhanced follow-up examinations, no changes were apparent during their short follow-up interval. Regarding the reference to McDonald et al (2), who did not observe washout in autopsy specimens, it is interesting that Smith et al (3) more recently reported observing washout of elemental gadolinium accumulated in rat brains from 20 gadodiamide doses. Use of inductively coupled plasma mass spectroscopy to quantify elemental gadolinium content means that actual washout of the gadolinium was occurring. More data are needed, especially in humans, to more completely understand this phenomenon.

Disclosures of Conflicts of Interest: A.H.B. disclosed no relevant relationships. M.R.P. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: is a paid consultant for GE Healthcare and Bracco; receives payment for lectures including service on speak-

ers bureau from Bayer; has patent agreements with Bayer, Bracco, GE Healthcare, Guerbet-Mallinckrodt, and Lantheus. Other relationships: disclosed no relevant relationships.

References

- Behzadi AH, Farooq Z, Zhao Y, Shih G, Prince MR. Dentate nucleus signal intensity decrease on T1-weighted MR images after switching from gadopentetate dimeglumine to gadobutrol. *Radiology* 2018;287(3):816–823.
- McDonald RJ, McDonald JS, Kallmes DF, et al. Intracranial gadolinium deposition after contrast-enhanced MR imaging. *Radiology* 2015;275(3):772–782.
- Smith AP, Marino M, Roberts J, et al. Clearance of gadolinium from the brain with no pathologic effect after repeated administration of gadodiamide in healthy rats: an analytical and histologic study. *Radiology* 2017;282(3):743–751.

Erratum

Originally published in:

Radiology 2018;287(3):1003–1015

DOI: 10.1148/radiol.2018172592

Developmental Hip Dysplasia Diagnosis at Three-dimensional US: A Multicenter Study

Dornoosh Zonoobi, Abhilash Hareendranathan, Emanuel Mostofi, Myles Mabee, Saba Pasha, Dana Cobzas, Padma Rao, Sukhdeep K. Dulai, Jeevesh Kapur, Jacob L. Jaremko

Erratum in:

Radiology 2018;288(3):912

DOI:10.1148/radiol.2018184016

In Figure 1, the number of dysplastic and borderline patients is incorrect; these numbers should be **241** and **140**, respectively, as stated in the text of the Materials and Methods, not 239 and 142 as shown in the Figure.