Comparison of an Automatic Technique to Current Practice in Calculating Dysplasia Metrics in 2D Ultrasound Images of the Neonatal Hip

**Purpose:** Ultrasound (US) is the standard imaging modality used to screen for developmental dysplasia of the hip in infants. In current practice, images are deemed either adequate or inadequate for interpretation by radiologists or orthopaedic surgeons and, if adequate, diagnostic dysplasia metrics are determined. Both manual steps introduce significant sources of inter-observer variability which can affect misdiagnosis rates. To eliminate subjectivity in this process, we developed an automatic method to identify adequate US images and extract dysplasia metrics. The purpose of this study was to validate the efficacy of this automatic method by comparing results with observer-determined dysplasia metrics on a set of US images.

**Methods:** A total of 693 US images from scans of 35 infants were analyzed. Images were labeled as adequate or inadequate by trained clinicians at a single institution. The same clinicians measured $\alpha$ and $\beta$ angles on adequate images to diagnose dysplasia. We trained our image classifier on random sets of 415 images and used it to assess the remaining 278 images. On the images judged adequate, we automatically estimated $\alpha$ and $\beta$ angles. We compared the two methods for discrepancies in adequacy determination, metric variability and incidences of missed early diagnosis or over-treatment.

**Results:** Excellent agreement was seen in image adequacy classification between the automatic and manual methods (Kappa coefficient =0.912). On each adequate US image, $\alpha$ and $\beta$ angle measurements were compared, producing mixed levels of agreement between methods. Mean discrepancies of $1.78\pm4.72^\circ$ and $8.91\pm6.437^\circ$ were seen for $\alpha$ and $\beta$ angles, respectively. Standard deviations of the angle measures across multiple images from a single patient scan were significantly reduced by the automatic method for both $\alpha$ ($p<0.05$) and $\beta$ ($p<0.01$) angles. Additionally, the automatic method classified three hips (two patients) as Graf type II and two hips (two patients) as type III, while the manual method classified them as type I and II, respectively. Both cases flagged as type III patients by the automatic system subsequently failed Pavlik harness treatment and were booked for surgery.

**Conclusion:** The automatic method produced excellent agreement with radiologists in scan adequacy classification and significantly reduced measurement variability. Good agreement between methods was found in Graf classification. In instances of disagreement, subsequent clinical findings seemed to support the classification of the automatic method.

**Significance:** This proposed method presents an alternative automatic, near-real-time analysis for US images that may potentially significantly improve dysplasia metric reliability and reduce missed early diagnoses without increasing over-treatment.