CT-Based Automated 3D Measurement of Femoral Version: Validation against Standard 2D Measurements in Patients with Hip Pain

Mohammadreza Movahhedi[1], Florian Schmaranzer[2], Mallika Singh[1], Andreas Nanavati[3], Simon Steppacher[3], Klaus Siebenrock[3]
Eduardo N. Novais[1], Ata M. Kiapour[1]
[1]Department of Orthopaedic Surgery, Boston Children's Hospital, Harvard Medical School, Boston MA
[2]Department of Diagnostic, Interventional and Pediatric Radiology, Inselspital, Bern University, Hospital, University of Bern, Bern, Switzerland
[3]Department of Orthopaedic Surgery and Traumatology, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland
Mohammadreza.Movahhedi@childredns.harvard.edu

Disclosures: M. Movahhedi: None. F. Schmaranzer: None. M. Singh: None. A. Nanavati: None. S. Steppache: None. K. Siebenrock: Nne. E.N. Novais: None. A.M. Kiapour: 3B; MIACH Orthopaedics: 8; BMC Musculoskeletal Disorders and American Journal of Sports Medicine.

INTRODUCTION: Calculation of an accurate femoral version is vital for surgical planning ahead of joint-preserving hip surgery. Yet, the current 2D methods for measuring the femoral version can have extensive variability due to landmark selection or patient positioning, limiting their reproducibility. 3D methods for the femoral version measurement have been proposed, however, no study to date has validated two separate 3D methods based on proximal and distal references. Here we have developed and validated a fully automated technique to quantify the femoral version in 3D from CT scans. We hypothesized that the automated 3D measurements are in strong agreement with manual 2D measurements.

METHODS: Between May 2017 and June 2018, this retrospective study evaluated 81 patients with hip pain who underwent low-dose CT of the pelvis to the knee. 36 patients were excluded due to possessing only external CT examinations, leaving 45 patients (57 hips) in our final cohort, with a mean age of 18.7 ± 5.1 years. The 2D femoral version was assessed with four measurement methods differing by the landmark levels for the proximal femoral reference axis and included measurements at the level of the greater trochanter (Lee method), femoral neck (Reikeras method), the base of the femoral neck (Tomczak method), and level of the lesser trochanter (Murphy method). A custom program (VirtualHip, Boston Children's Hospital) was used to automatically segment the femur, identify the proximal and distal femoral landmarks and then measure the femoral version based on the femoral neck-shaft and femoral head-shaft approaches (Figure 1A). Mean femoral versions with 95% confidence intervals were calculated, along with intraclass correlation coefficients (ICCs) and Bland-Altman analysis to assess bias between measurement methods. The study was approved by Boston Children's Hospital IRB.

RESULTS: We found our automatic 3D segmentations to be highly accurate, with mean dice coefficients of 0.98 ± 0.03 for the femur and landmarks. The mean difference between the automatic 3D head-shaft- (27.4 ± 16.6º) and 3D neck methods (12.9 ± 13.7º) was 14.5 ± 10.7º ($p<0.001$). Further comparison of the 3D methods with the 2D methods demonstrated that the neck method was closer to the more proximal Lee (2 ± 5.6º, 95% CI 0.2 to 3.8º, p=0.03) and Reikeras (-2.4 ± 5.9º, -4.4 to 0.5º, p=0.009) methods, with the majority of measurements (>50%) falling within the range of interobserver variation. On the other hand, the head-shaft method showed mean differences closest to the more distal Tomczak (-1.3 ± 7.5º, 95% CI -3.8 to 1.1º, p=0.57) and Murphy (1.5 ± 5.4º, -0.3 to 3.3º, p=0.12) methods. Of these two, the Murphy method yielded differences that were mostly (>50%) within the range of intra-observer measurement variation (Figure 1B).

DISCUSSION: To the best of our knowledge, this is the first study to clinically apply a fully automated deep learning approach for the comparison of 3D methods based on proximal and distal approaches with the most commonly used 2D measurement methods. We found that our automatic 3D neck-based- and head-shaft methods yielded femoral version angles comparable to the proximal and distal 2D-based methods when applying fully-automated segmentation using a convolutional neural network. Having a consistent methodology for measuring femoral version is imperative, given the previously reported variation of up to 20º depending on landmark selection which was confirmed in the current study. This difference further increases at the extremes of femoral anteversion and valgus deformity. Our neck-based method closely approximated the two most proximal 2D methods of Lee and Reikeras. This is sensible, given that the 3D femoral neck method uses the center of the femoral neck axis (between the medial and lateral cortices) as the reference point, similar to the approximation of the femoral neck axis when applying the methods described by Reikeras. By contrast, 3D head-shaft-based femoral version angles were comparable to the distal 2D methods of Tomczak and Murphy. This too, is a reflection of the landmarks used.

CLINICAL RELEVANCE: This automated 3D femoral version measurement approach holds promise for establishing a more streamlined and standardized technique which is independent from measurement bias and patient positioning in the scanner for assessing the femoral version and aiding in treatment planning.

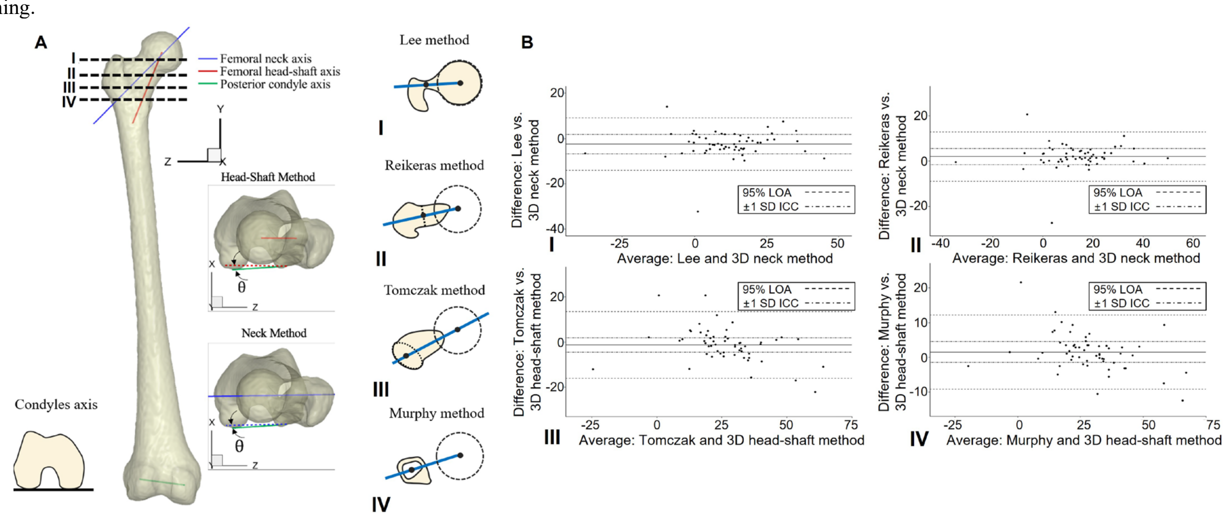

Figure 1. (A) Automated 3D measurement of the femoral version (neck method and head-shaft method) and the four different 2D measurements of the femoral version. (B) Bland Altman plots to evaluate the systematic bias between 2D and 3D proximal and distal measurement methods. The solid black line represents the mean difference between the respective 2D and 3D methods. The dark, dashed black line represents the 95% limits of agreement. The lighter dashed line represents 1 standard deviation of the corresponding intra-rater reliability of CT measurements.