Improving Femoral Torsion Evaluation in Infants Through Analysis of Variability and Reliability in 2D and 3D Imaging Measurements of the Femoral Neck Axis

Tamara T Chambers
Department of Mechanical Engineering
Embry-Riddle Aeronautical University
1 Aerospace Boulevard
Daytona Beach, Florida, 32114
chambet2@my.erau.edu

Victoria Melendez
Department of Mechanical Engineering
Embry-Riddle Aeronautical University
1 Aerospace Boulevard
Daytona Beach, Florida, 32114
melendv3@my.erau.edu

Andrew E Anderson
Department of Orthopaedics
University of Utah
590 Wakara Way
Salt Lake City, Utah, 84108
andrew.anderson@hsc.utah.edu

Victor A Huayamave*
Department of Mechanical Engineering
Embry-Riddle Aeronautical University
1 Aerospace Boulevard
Daytona Beach, Florida, 32114
huayamav@erau.edu

Manuscript word count: 3044 words

ABSTRACT

The femoral neck axis serves as a critical parameter in evaluating hip joint health, particularly in the pediatric population. Commonly used metrics for evaluating femoral torsion, such as the femoral neck-shaft and femoral anteversion angles, rely heavily on precise definitions of the position and orientation of the femoral neck axis. Current measurement methods employing radiographs and performing two-dimensional (2D) measurements on computed tomography (CT) scans are susceptible to errors due to their reliance on reader experience and the inherent limitations in 2D measurements. We hypothesized that utilizing volumetric data would mitigate these errors and enable more accurate and reproducible measurements of the femoral neck axis using the femoral anteversion and femoral neck-shaft angles. To test this hypothesis, we analyzed a historical collection of post-mortem infant femoral and pelvic bones (28 hips) aged 0 to 6.5 months, with an average estimated age of 4.68 months ± 1.80 months. Our findings revealed an average neck-shaft angle of

128.00° ± 4.92° and femoral anteversion angle of 35.56° ± 11.68° across all femurs, consistent with literature values. These measurements obtained from volumetric image data were found to be repeatable and reliable compared to conventional methods. Our study suggests that the proposed methodology offers a standardized approach for obtaining repeatable and reproducible measurements, thus potentially enhancing diagnostic accuracy and clinical decision-making in assessing hip developmental conditions in pediatric patients.

Abstract word count: 224 words

1 INTRODUCTION

Quantifying the femoral neck axis in pediatric populations is crucial for diagnosing pathological conditions in the hip joint [1], such as developmental dysplasia of the hip (DDH) in infants or cerebral palsy in older children [2,3]. The femoral neck axis is used as a reference for several measurements used to quantify malformations in the proximal femur, which, if left untreated, can be detrimental to growth and development [4]. Two important morphological features that require referencing the femoral neck axis are the femoral neck anteversion and femoral neck-shaft angles. Femoral anteversion (or femoral neck anteversion) refers to the rotation of the femoral neck about the femur's longitudinal axis and describes cases of increased femoral torsion or increased femoral version [5,6]. The femoral anteversion angle (FAV) averages approximately 30-35 degrees at birth in otherwise healthy infants and decreases to about 10-20 degrees in adulthood [7–10]. However, the reported FAVs measured in pediatric populations with DDH are inconsistent. Some found increased FAVs in infants [11] and children [7] with dysplastic hips, with an average FAV of 55 degrees at year one for severe dislocations [9], while others found no significant differences in FAV between healthy and dysplastic hips [12–14]. Similarly, FAVs in children with cerebral palsy are about 10 degrees higher than in healthy children [5]. The femoral neck-shaft angle (NSA) averages 130-135 degrees at birth in healthy infants, peaks around 145 in years 1-3, and decreases significantly in adolescence and adulthood [4,9].

There are no well-established, gold-standard methods for measuring NSA [15] or the FAV [3,16]. However, both of these metrics are generally obtained using two-dimensional (2D) measurements from radiographs or computed tomography (CT) of one or more transverse image slices for the FAV [6,16] or a single coronal slice for the NSA [4]. These conventional 2D methods may cause some variation in results depending on the patient's positioning during imaging [4]. Due to the lack of established methods, there are

variations in the literature on how the femoral neck axis is defined for measuring the NSA and FAV. Zhang et al. [17] proposed a method for describing the three-dimensional (3D) femoral neck axis in adult femurs. The authors defined the femoral neck axis as the line that connects the geometric center of the femoral head and the midpoint of a tangent line to a concentric sphere at the center of the femoral head. They concluded that their methods provided an accurate and repeatable process for determining the femoral neck axis when measuring the femoral neck torsion angle, which the authors emphasize was different from the femoral neck anteversion angle. Souza et al. [1] explored multiple ways of defining the femoral neck axis for measuring the FAV on adult femurs. They found that using the axis oblique section on the 2D CT provided the best approximation of FAVs. Schmaranzer et al. [18] considered four automated 3D methods for measuring the FAV based on the center of the femoral head and various landmark points on the proximal femur. The authors reported results comparable to those of conventional 2D methods but could not make claims regarding which 3D-based method was superior in performance.

While some authors are investigating using 3D data to define the femoral neck axis, most reported FAVs and NSAs are from 2D methods. There is a lack of consensus on an established method for determining the femoral neck axis, which may lead to variability in the reported values. Some authors suggest that 3D methods are more reliable than 2D methods [3], while others find comparable results [18]. Even in healthy populations, the FAV depended on the landmarks identified and the imaging technique used [5]. The FAV and NSA are important anatomical features in the femur for diagnostic purposes and for understanding healthy femoral growth and development [4]. Therefore, this study aimed to investigate the variability of the femoral neck axis in the infant femur from medical images through two measurements: the NSA and the FAV. We hypothesized that volumetric (or 3D) measurements would minimize the errors associated with two-dimensional measurements and allow for more detailed, reproducible measurements.

2 MATERIALS AND METHODS

The decedents used in this study were from a small historical collection, denoted here as the Ortolani collection. It is comprised of post-mortem infant femoral and pelvic bones with estimated ages ranging from 0 to 6.5 months, with an average estimated age of 4.68 months ± 1.80 months. The decedents from the Ortolani collection included healthy hips and hips with varying degrees of dysplasia and were obtained from

the University of Padua, Italy [19]. The decedents are infants who died of infectious diseases, such as influenza and gastroenteritis, that were common in children during the pre-antibiotic era [12]. Of the 14 total decedents (28 hips) in this collection, we obtained CT (slice thickness: 0.45 mm) of 4 decedents (8 hips). Volumetric anatomical models were generated from the CT of the femurs using Synopsys® Simpleware ScanIP, a medical image processing software. All generated 3D models were constructed using a semi-automatic segmentation approach called thresholding, which generates a binary segmentation by separating a greyscale image into two regions based on the selected threshold value. A threshold value of -50 Hounsfield Units was selected for all decedents. All measurements in the current study were performed in Simpleware ScanIP by two observers following the same protocol.

2.1 Femoral neck-shaft angle

As shown in Fig. 1, the femoral neck shaft angle (NSA) was computed as the angle between the femoral neck and femoral shaft axes. The femoral neck axis was defined by a line connecting the centers of the femoral head (FHC) and the center of the femoral neck (FNC) [20]. The center of the femoral head was defined in 3D using the center of a best-fit sphere, denoting the geometric center of the femoral head. For the femoral neck, a region of interest (ROI) was selected using paint tools in Simpleware ScanIP by "painting" all visible regions of the femoral neck. The geometric center was determined from the center of the sphere completely enclosed by the selected ROI. The ROI was selected as two sections of the femoral shaft, one at a proximal (FS1) point and another at a distal point (FS2). The inner best-fit spheres were computed for each section such that the sphere fit inside each ROI, and the line connecting the center of the two spheres was defined as the femoral shaft axis.

2.2 Femoral anteversion angle

The femoral anteversion angle was computed as the angle between the femoral neck axis and the posterior condylar axis of the knee (Fig. 3). As described in Fig. 2, the posterior condylar axis was defined by two points denoting the most posterior points on the medial (ME) and lateral condyles (LE) [2,16]. Defining the distal femoral axis using the posterior condylar line was reportedly the least-location dependent and most repeatable method [5]. The femoral neck axis reference line created for the NSA measurement was maintained for the FAV measurements, and only the reference line for the posterior condylar axis was

modified. One decedent (2 femurs) was excluded from the FAV measurement because the distal femoral condyles were not intact, which are crucial anatomical features for identifying the posterior condylar axis and consequently measuring the FAV.

2.3 Statistical analysis

The descriptive statistics of the data are presented as the mean value ± standard deviation. The measurements are presented as the mean value calculated by two examiners and as the raw values computed by two examiners. Interclass correlation (ICC) was conducted to determine inter-rater and intra-rater reliability. These were conducted to determine how consistent the two observers were in taking the measurements compared to one another and individually. The inter-rater reliability for NSA and FAV was computed using ICC with a 95% confidence interval (CI) using IBM SPSS Statistics 27 [21].

3 RESULTS AND DISCUSSION

In this study, 8 femurs (4 left, 4 right) were used to test the variability in the femoral neck axis. The average estimated age of the decedents was 3.88 ± 2.84 months (range: 0 to 6 months). A total of 10 measurements were taken for each femur (20 per decedent) for the NSA and the FAV, except for one decedent that was missing crucial anatomy needed for the FAV. The purpose was to investigate the variability of the femoral neck axis in the infant femur through the NSA and the FAV. The hypothesis was that 3D measurements would allow for repeatable and more reproducible measurements since more information can be utilized in 3D measurements.

The NSA ranged from 120.50 to 139.66 degrees for the right femurs, averaging 128.55 ± 5.00 degrees. The NSA ranged from 118.83 to 139.39 degrees for the left femurs, with an average of 127.45 ± 4.84 degrees. The average NSA for all femurs was 128.00 ± 4.92 degrees. A detailed description of the average FAV and NSA values are shown in Table 1. The mean NSA for all decedents was higher for observer 1, while the mean FAV was near exact for both observers. The FAV ranged from 20.35 to 47.80 degrees for the right femurs, with an average of 32.89 ± 10.39 degrees. The FAV ranged from 22.00 to 55.87 degrees for the left femurs, with an average of 38.23 ± 12.44 degrees. The average FAV for all femurs was 35.56 ± 11.68 degrees.

The overall inter-rater reliability for all femurs was 0.74 (CI: 0.53-0.85) for NSA and 0.98 (CI: 0.96-0.99) for the FAV. A breakdown of the inter-rater reliability by femur side is provided in Table 2. The overall

NSA intra-rater reliability for observer 1 and observer 2 was 0.80 (CI: 0.58-0.95) and 0.70 (CI: 0.42-0.92). However, the FAV intra-rater reliability scores for both observers were high, with 0.98 (CI: 0.92-1.00) for observer 1 and 0.98 (CI: 0.94-1.00) for observer 2.

Based on the results, the FAV 3D methods were not largely affected by the observer. The mean values were consistent between observers, with a mean difference of 0.90 ± 2.30 degrees for all measurements. The mean difference between the left and right FAV measurements between observers was 0.45 ± 3.06 degrees and 0.63 ± 1.00 degrees. This demonstrates that the chosen 3D-based measurement methods for FAV allow for consistent, repeatable results between decedents and observers. However, for the NSA, the differences between observers were higher. The mean differences were 1.38 ± 3.41 degrees for all measurements, with differences of 2.14 ± 3.62 degrees and 0.61 ± 3.08 degrees between observers on the left and right sides. The NSA findings further demonstrate the difficulty in defining the femoral shaft axis. This reference axis is known to be challenging to define due to the anterior curvature of the femoral shaft [5].

The inter- and intra-rater reliability for FAV was excellent and was not influenced by the number of measurements per decedent for each observer, as demonstrated by measurements of inter-rater (Fig. 5) and intra-rater (Fig. 6) ICCs. However, the NSA inter- and intra-rater reliability seemed slightly affected by the number of measurements performed by each observer (Fig. 5 and Fig. 6, respectively). Using individual measures of all NSA measurements had the lowest inter-rater reliability (0.66, IC: 0.29-0.85), whereas using the mean of two or more measurements had a reliability of 0.76 or higher. The intra-rater reliability, however, demonstrated some inconsistency in both observers' measurements for the NSA regardless of the number of measurements taken for each decedent, as shown in Fig. 4, which displays the variation in the global location of the FNC and its relationship to both the observer and the 3D measurements analyzed herein. The findings in Fig. 4 suggest that the 3D landmarks for the FAV were less sensitive to the global location of the FNC compared to the NSA.

Overall, the average NSA and FAV values were 128.26 ± 4.87 degrees (range: 118.83-139.66 degrees) and 35.56 ± 11.68 degrees (range: 20.35-55.87 degrees). The 3D NSAs were consistent with 2D-based measurements found in the literature, which were between 115 and 140 degrees [1,22] for second- and third-trimester fetuses. The 3D NSAs also agree with the average 2D-based NSA at birth (130-135 degrees) in healthy infants reported by [9]. Park et al. [2] reported average 2D-based NSA and FAV of 133.7 ± 5.4

degrees and 35.1 ± 12.1 in children with cerebral palsy. Similarly, [15] reported large ranges of NSA in male and female populations, albeit in adult patients. The authors reported NSA of 119.5 to 145 degrees (mean: 132.3 ± 4.4 degrees) in males and 116.5 to 145.5 degrees (mean: 130.1 ± 4.8 degrees) in females. The average 2D-based FAV in the Ortolani collection reported by [12] was 35.8 ± 6.5 degrees. These measurements were described as manual measurements taken using a goniometer and the photographic method. The methods used in the current study mimic the 2D-based photographic method in the definition of the femoral neck axis and distal femoral axis. The average 3D FAV values were close to those reported in [12], with a mean difference of 0.20 ± 11.35 degrees. Although there were minimal differences between 2D- and 3D-based FAVs in the current study, some discrepancies were reported in literature. Cai et al. [3] found significant differences in FAV measurements between 2D and 3D methods in children with hip disorders. The authors attributed these differences to the differences in accuracy and reliability between 2D and 3D methods. Furthermore, the authors found the 2D-based method (2D CT) less accurate and reliable than the 3D-based method. Although Van fraeyenhove et al. [6] did not comment on the accuracy of 2D methods over 3D methods, the authors found that 2D measurements (Murphy's method [23]) underestimated 'true' femoral torsion.

Other studies have performed 3D measurements on the femur from segmented volumetric models. Davis et al. [24] performed 3D measurements of the FAV on adult femurs and reported a mean FAV of 12.7 ± 9.1 degrees when using the distal posterior condylar axis as the distal reference line. Schmaranzer et al. [18] reported similar femoral version angles when comparing automated 3D-based and proximal/distal 2D-based methods. After comparing multiple 3D-based methods, the authors found no significant results supporting using one 3D-based method over another. However, the authors did find comparable results for the neck method and the head-shaft method. The neck method defined the femoral neck axis as a best-fit line between the medial and lateral femoral neck. In contrast, the head-neck method was defined as a line connecting the femoral head center to the femoral shaft at the greater trochanter [18], similar to 2D-based method developed by Murphy in 1987 [23].

The variability in the femoral neck axis may be affected by the age of the decedents. Boniforti et al. [25] investigated the variability in pelvic measurements of infants 3 to 36 months and found that the measurements varied the most during early infancy. This is when the entire femoral head and most of the acetabulum consist

of cartilage, which is difficult to see on pelvic radiographs. The variability they found was of particular concern when diagnosing and managing developmental dysplasia of the hip (DDH) in infants. The authors determined that errors associated with various pelvic measurements varied with the age of the infants, suggesting that age-appropriate indicators of DDH should be used during diagnosis. The results of their study suggest that how the measurements are reported may also influence the variability found in literature, in addition to the variability caused by landmarks and imaging techniques. Based on the FAV method employed, measurements can lead to mean differences of up to 10 degrees or higher in individuals with increased anteversion angles as a result of the imaging modality or the landmarks used [5]. Due to the variability found in literature, some suggest employing threshold values for the various 2D and 3D FAV measurement approaches [6,26].

The study is limited by the sample size and the absence of a healthy control group. The small sample size is precluded by the small number of available high-resolution CT scans of infant femurs. Therefore, the results of the study should not be used for clinical assessment. However, despite the small sample size, the measurements were consistent with those found in the literature. Additionally, there is a possibility that the NSA and FAV measurements were influenced by how the decedents were 1) dissected and 2) positioned during the imaging. Although [15] studied adult patients, the authors found a significant difference in NSA measurements between radiographs acquired with patients in the upright and supine positions. However, since the NSA in the current study was performed only in the 3D axis, it may alleviate some of the variability of this measurement that is influenced by the positioning of the patient. Mayr et al. [27] found that while 2D-based methods (CT) required neutral positional of the legs to get accurate measurements, 3D-based methods allowed for femoral torsion measurements independent of positioning. Regardless, more research is needed on the effect of positioning on 2D and 3D NSA measurements in infants to address variability from positioning properly. Lastly, it is not possible to conclude on the number of measurements that should be taken for each patient or the number of observers that are needed as this is dependent on several factors, such as the landmarks, imaging techniques, the skill of the observer, and the morphology of the patient. Therefore, the findings should not be used for clinical decision making.

4 CONCLUSION

Our study aimed to evaluate infant femoral torsion by exploring the variability of the femoral neck axis through the femoral neck-shaft and anteversion angles using three-dimensional measurements. These angles serve as vital metrics for evaluating hip joint health. By employing a 3D approach, we sought to mitigate some of the variability in femoral neck axis measurements induced by patient positioning. Our findings align with existing literature values and demonstrate the accuracy and reliability of our 3D methodology compared to other conventional methods. Importantly, our approach offers the potential to achieve repeatable and reproducible measurements across different observers. Further research and validation studies are warranted to fully establish the clinical utility and applicability of our approach in diverse patient populations and clinical settings. In summary, our study contributes to advancing the understanding and measurement of femoral neck axis parameters, offering a valuable tool for clinicians and researchers in the fields of orthopedics and pediatric hip health.

ACKNOWLEDGMENT

None.

FUNDING

This study was supported in part by the International Hip Dysplasia Institute (IHDI) and by the US National Science Foundation CAREER award CMMI-2238859. The opinions, findings, conclusions, or recommendations expressed are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

272 **NOMENCLATURE**
273

2D	Two-dimensional
3D	Three-dimensional
CT	Computed tomography
FAV	Femoral anteversion angle
FHC	Center of femoral head
FNC	Center of femoral neck
ICC	Interclass correlation
LC	Lateral condyle
MC	Medial condyle
NSA	Femoral neck-shaft angle
ROI	Region of interest

274
275

REFERENCES

[1] Souza A. D., Ankolekar V., Padmashali S., Das A., Souza A., and Hosapatna M., 2015, "Femoral Neck Anteversion and Neck Shaft Angles: Determination and their Clinical Implications in Fetuses of Different Gestational Ages," Malaysian Orthop. J., **9**(2), pp. 33–6. DOI: 10.5704/moj.1507.009

[2] Park N., Lee J., Sung K. H., Park M. S., and Koo S., 2014, "Design and Validation of Automated Femoral Bone Morphology Measurements in Cerebral Palsy," J. Digit. Imaging., **27**(2), pp. 262–9. DOI: 10.1007/s10278-013-9643-2

[3] Cai Z., Piao C., Zhang T., Li L., and Xiang L., 2021, "Accuracy of CT For Measuring Femoral Neck Anteversion in Children with Developmental Dislocation of the Hip Verified Using 3D Printing Technology," J. Orthop. Surg. Res., **16**(1), pp. 256. DOI: 10.1186/s13018-021-02400-x

[4] Darwich A., Geiselhardt C., Bdeir M., Janssen S., Schoenberg S. O., Gravius S., and Jawhar A., 2021, "Anthropometry of the Proximal Femur and Femoral Head in Children/Adolescents Using Three-Dimensional Computed Tomography-Based Measurements," Surg. Radiol. Anat., **43**(12), pp. 2009–23. DOI: 10.1007/s00276-021-02841-3

[5] Scorcelletti M., Reeves N. D., Rittweger J., and Ireland A., 2020, " Femoral Anteversion: Significance and Measurement," J. Anat., **237**(5), pp. 811–26. DOI: 10.1111/joa.13249

[6] Van fraeyenhove B., Verhaegen J. C. F., Grammens J., Mestach G., Audenaert E., Van Haver A., and Verdonk P., 2023, "The Quest for Optimal Femoral Torsion Angle Measurements: A Comparative Advanced 3D Study Defining the Femoral Neck Axis," J. Exp. Orthop., **10**(1). DOI: 10.1186/s40634-023-00679-9

[7] Fabry G., Macewen G. D., And Shands A. R., 1973, "Torsion of the Femur," J. Bone Jt. Surg., **55**(8), pp. 1726-38. DOI: 10.2106/00004623-197355080-00017

[8] Kingsley P. C., and Olmsted K. L., 1948, " A Study to Determine the Angle of Anteversion of the Neck of the Femur," J. Bone Joint Surg. Am., **30 A**(3), pp. 745–51. DOI: 10.2106/00004623-194830030-00021

[9] Hensinger R. N., 1986, Standards in pediatric orthopedics: tables, charts and Graphs Illustrating Growth, Raven Press, United States, ISBN: 978-0881671834

[10] Kainz H., Mindler G. T., and Kranzl A., 2023, "Influence of Femoral Anteversion Angle and Neck-Shaft Angle on Muscle Forces and Joint Loading During Walking," PLoS One., **18**(10 October). DOI: 10.1371/journal.pone.0291458

[11] McKibbin B., 1970, "Anatomical Factors in the Stability of the Hip Joint in the Newborn," J. Bone Joint Surg. Br., **52**(1), pp. 148–59. DOI: 10.1302/0301-620x.52b1.148

[12] Huayamave V., Chambers T., Fantoni I., Stecco C., De Caro R., and Price C. T., 2023, " Femoral Morphology in Ortolani's Anatomical Collection of Developmental Dysplasia of the Hip: Anteversion is Unrelated to Severity of Infantile Dysplasia," J. Child. Orthop., **17**(2), pp. 1–8. DOI: 10.1177/18632521231152282

[13] Mootha A. K., Saini R., Singh Dhillon M., Aggarwal S., Kumar V., and Tripathy S. K., 2010, "MRI Evaluation of Femoral and Acetabular Anteversion in Developmental Dysplasia of the Hip A Study in An Early Walking Age Group," Acta Orthop. Belg., **76**(2), pp. 174–80.

[14] Li H., Wang Y., Oni J. K., Qu X., Li T., Zeng Y., Liu F., et al., 2014, "The Role of Femoral Neck Anteversion in the Development of Osteoarthritis in Dysplastic Hips," Bone Jt. J., **96B**(12), pp. 1586–93. DOI: 10.1302/0301-620X.96B12.33983

[15] Haddad B., Hamdan M., Al Nawaiseh M., Aldowekat O., Alshrouf M. A., Karam A. M., Azzam M. I., et al., 2022, "Femoral Neck Shaft Angle Measurement on Plain Radiography: Is Standing or Supine

Radiograph A Reliable Template For the Contralateral Femur?," BMC Musculoskelet. Disord., **23**(1), pp. 1092. DOI: 10.1186/s12891-022-06071-5

[16] Song K. S., Yon C. J., Heo Y. R., Lee J. H., Lee S. B., Ko Y. K., Lee K. J., et al., 2022, "Using the Axial Oblique View of Computed Tomography (CT) in Evaluating Femoral Anteversion: A Comparative Cadaveric Study," Diagnostics., **12**(8), pp. 1820. DOI: 10.3390/diagnostics12081820

[17] Zhang R. Y., Su X. Y., Zhao J. X., Li J. T., Zhang L. C., and Tang P. F., 2020, "Three-Dimensional Morphological Analysis of the Femoral Neck Torsion Angle—An Anatomical Study," J. Orthop. Surg. Res., **15**(1), pp. 192. DOI: 10.1186/s13018-020-01712-8

[18] Schmaranzer F., Movahhedi M., Singh M., Kallini J. R., Nanavati A. K., Steppacher S. D., Heimann A. F., et al., 2024," Computed Tomography-Based Automated 3D Measurement of Femoral Version: Validation Against Standard 2D Measurements in Symptomatic Patients," J. Orthop. Res. DOI: 10.1002/jor.25865

[19] Stecco C., Porzionato A., Macchi V., Fantoni I., Ortolani L., and de Caro R., 2014, "Marino Ortolani: "Does That Baby's Hip Go Click?"," Perspect. Biol. Med., **57**(4), pp. 538–46. DOI: 10.1353/pbm.2014.0035

[20] Fischer C. S., Kühn J. P., Völzke H., Ittermann T., Gümbel D., Kasch R., Haralambiev L., et al., 2020, " The Neck–Shaft Angle: An Update on Reference Values and Associated Factors," Acta Orthop., **91**(1), pp. 53–7. DOI: 10.1080/17453674.2019.1690873

[21] IBM Corp, 2020, "IBM SPSS Statistics for Windows," Armonk, IBM Corp.

[22] Maslon A., Sibinski M., Topol M., Krajewski K., and Grzegorzewski A., 2013, "Development of Human Hip Joint in the Second and the Third Trimester of Pregnancy: A Cadaveric Study," BMC Dev. Biol., **13**. DOI: 10.1186/1471-213X-13-19

[23] Murphy S. B., Simon S. R., Kijewski P. K., Wilkinson R. H., and Griscom N. T., 1987, "Femoral Anteversion.," JBJS., **69**(8)

[24] Davis E., LeBrun D. G., McCarthy T., and Westrich G. H., 2024, "Femoral Neck Anteversion: Which Distal Femur Landmark Matters?," Arthroplast. Today., **26**, pp. 101318. DOI: 10.1016/j.artd.2024.101318

[25] Boniforti F. G., Fujii G., Angliss R. D., and Benson M. K. D., 1997, "The Reliability of Measurements of Pelvic Radiographs in Infants," J. Bone Joint Surg. Br., **79-B**(4), pp. 570–5. DOI: 10.1302/0301-620X.79B4.0790570

[26] Kaiser P., Attal R., Kammerer M., Thauerer M., Hamberger L., Mayr R., and Schmoelz W., 2016, "Significant Differences in Femoral Torsion Values Depending on the CT Measurement Technique," Arch. Orthop. Trauma Surg., **136**(9), pp. 1259–64. DOI: 10.1007/s00402-016-2536-3

[27] Mayr H. O., Schmidt J. P., Haasters F., Bernstein A., Schmal H., and Prall W. C., 2021, "Anteversion Angle Measurement in Suspected Torsional Malalignment of the Femur in 3-Dimensional EOS vs Computed Tomography—A Validation Study," J. Arthroplasty., **36**(1), pp. 379–86. DOI: https://doi.org/10.1016/j.arth.2020.07.058

380 **Figure Captions List**
381

Fig. 1 The femoral neck-shaft angle (α) measurement. (a) projection of the femur onto the coronal plane, (b) projection onto the sagittal plane, and (c) 3D generated model. The femoral neck axis was defined as the line connecting the femoral head center (FHC – red) with the center of the femoral neck (FNC – blue). The femoral shaft axis was defined as the line connecting a center point on the proximal femoral shaft (FS1) with a center point on the distal femoral shaft (FS2).

Fig. 2 The landmarks used to define the distal femoral axis on the 3D generated models, which was defined using the most posterior distal medial (MC) and lateral (LC) condyles. The 3D models are represented in the (a) posterior isometric view, (b) medial side view (sagittal plane), (c) top-down view of the anterior femur, and (d) inferior axial view (axial plane).

Fig. 3 The femoral anteversion angle (β) showing (a) a head-on view of the femur (axial plane), (b) the 3D generated model, and (c) an isometric axial view. The femoral neck axis was defined as the line connecting the center of the femoral head with the center of the femoral neck. The posterior condylar axis was defined as the line connecting most posterior distal medial (MC) and lateral (LC) condyles.

Fig. 4 The position of the center of the femoral neck (FN) in the global x- (medial-lateral direction), y (anterior-posterior), and z-positions (inferior-superior) and how it related to the measured FAV and NSA for both observers.

Fig. 5 Inter-rater reliability by computing ICC of the mean of a varying number of measurements for all decedents by each observer. upper bound (UB) and lower bound (LB) on a 95% confidence interval.

Fig. 6 Intra-rater reliability for both observers by computing ICC of a varying number of measurements for all decedents. upper bound (UB) and lower bound (LB) on a 95% confidence interval.

382
383

384 **Table Caption List**
385

Table 1 Descriptive statistics for FAV and NSA for both observers, separated by left and right sides.

Table 2 Inter-rater reliability using ICC by using all five measurements per subject by each observer.

386

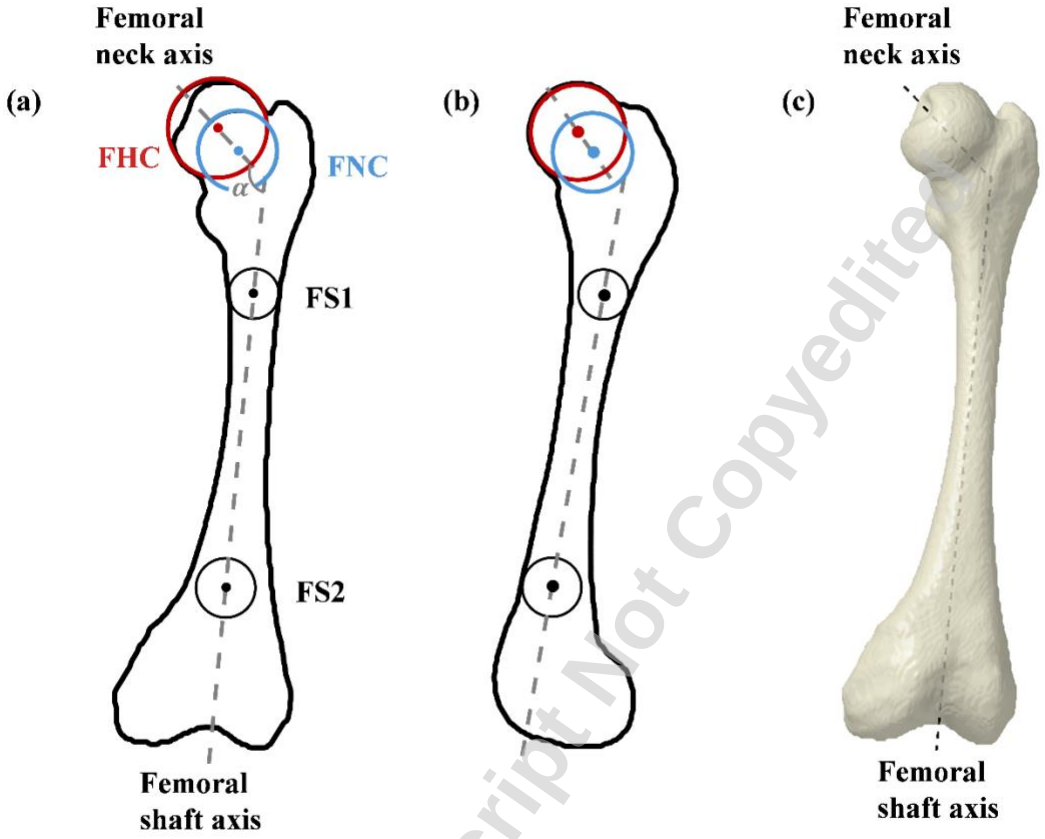


Fig. 1. The femoral neck-shaft angle (α) measurement. (a) projection of the femur onto the coronal plane, (b) projection onto the sagittal plane, and (c) 3D generated model. The femoral neck axis was defined as the line connecting the femoral head center (FHC – red) with the center of the femoral neck (FNC – blue). The femoral shaft axis was defined as the line connecting a center point on the proximal femoral shaft (FS1) with a center point on the distal femoral shaft (FS2).

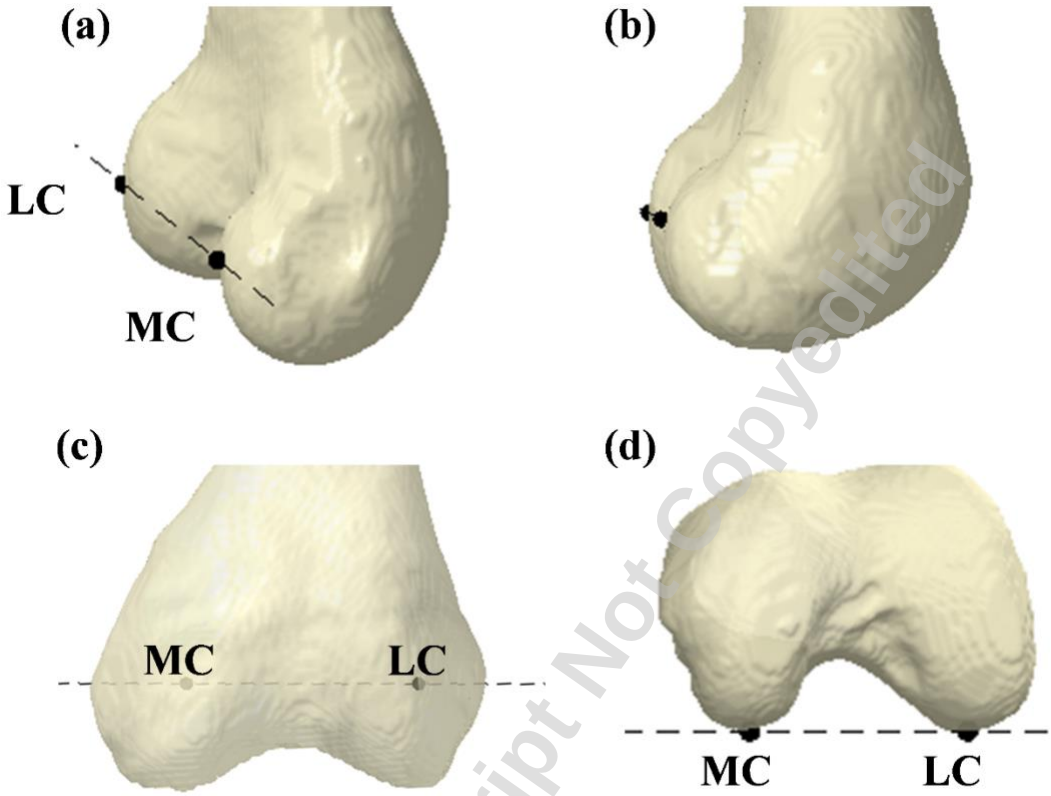


Fig. 2. The landmarks used to define the distal femoral axis on the 3D generated models, which was defined using the most posterior distal medial (MC) and lateral (LC) condyles. The 3D models are represented in the (a) posterior isometric view, (b) medial side view (sagittal plane), (c) top-down view of the anterior femur, and (d) inferior axial view (axial plane).

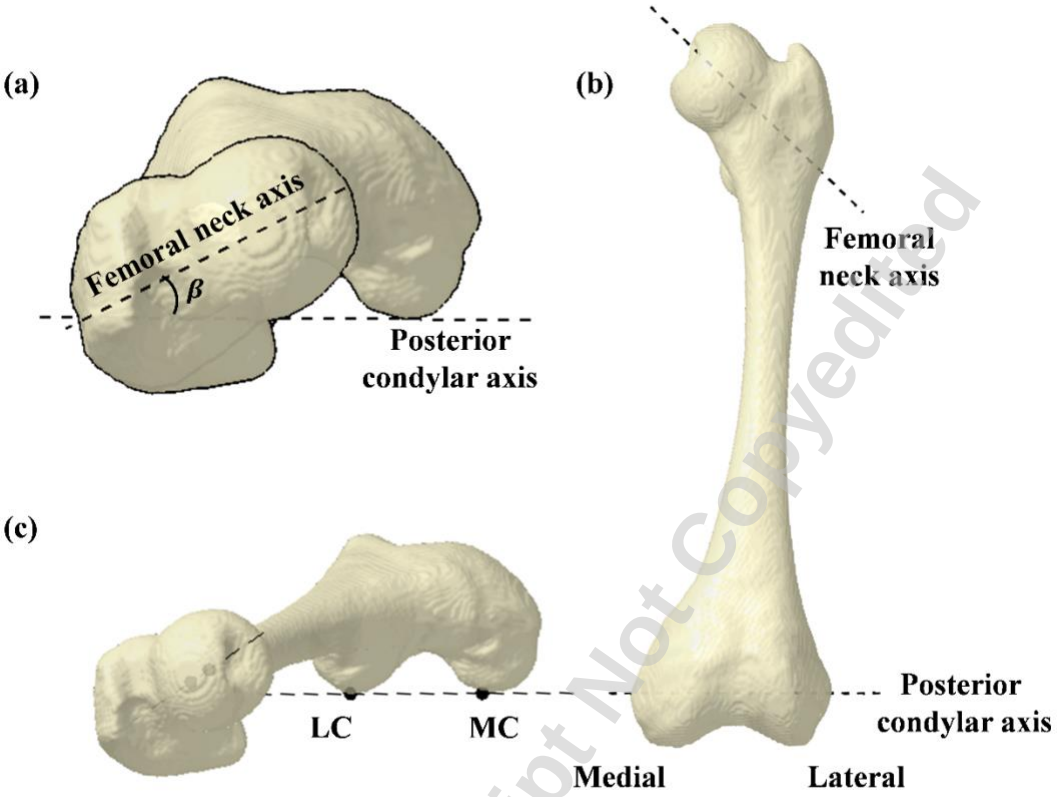


Fig. 3. The femoral anteversion angle (β) showing (a) a head-on view of the femur (axial plane), (b) the 3D generated model, and (c) an isometric axial view. The femoral neck axis was defined as the line connecting the center of the femoral head with the center of the femoral neck. The posterior condylar axis was defined as the line connecting most posterior distal medial (MC) and lateral (LC) condyles.

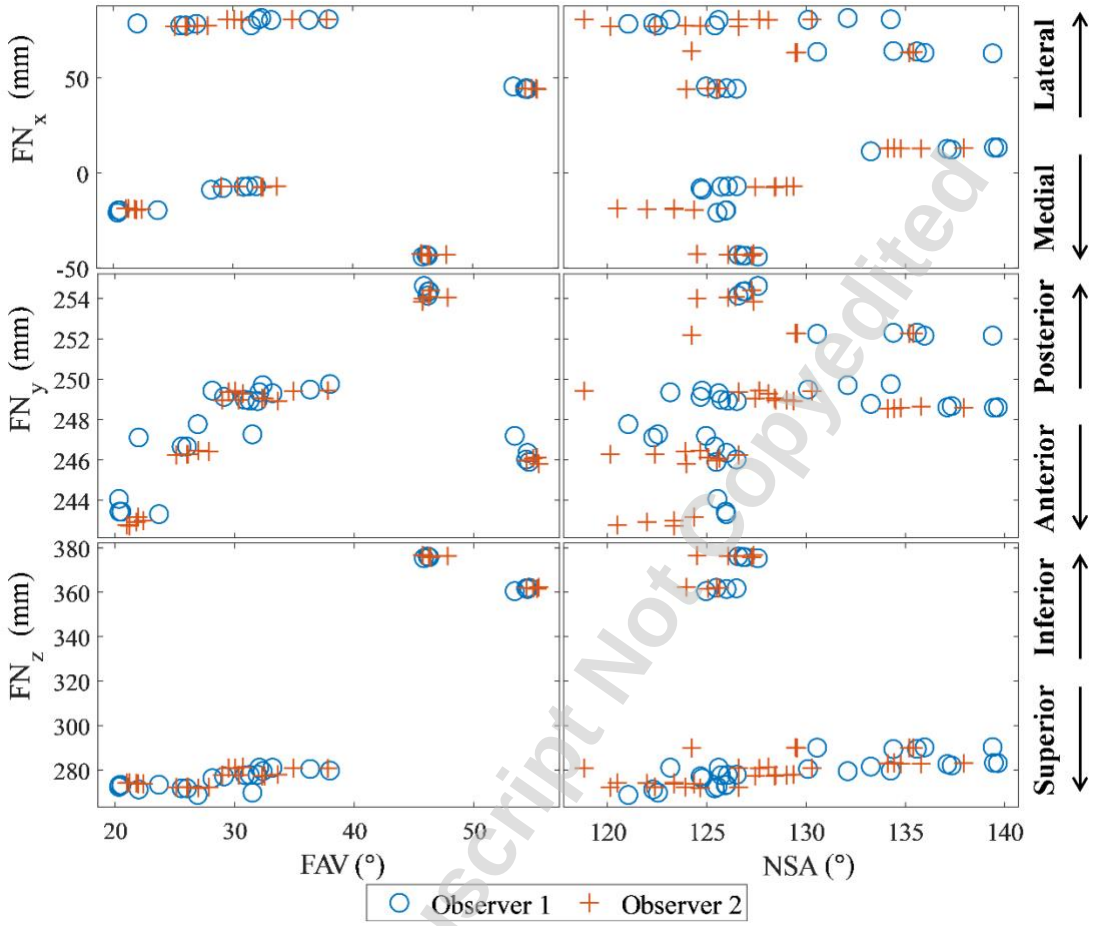


Fig. 4. The position of the center of the femoral neck (FN) in the global x- (medial-lateral direction), y (anterior-posterior), and z-positions (inferior-superior) and how it related to the measured FAV and NSA for both observers

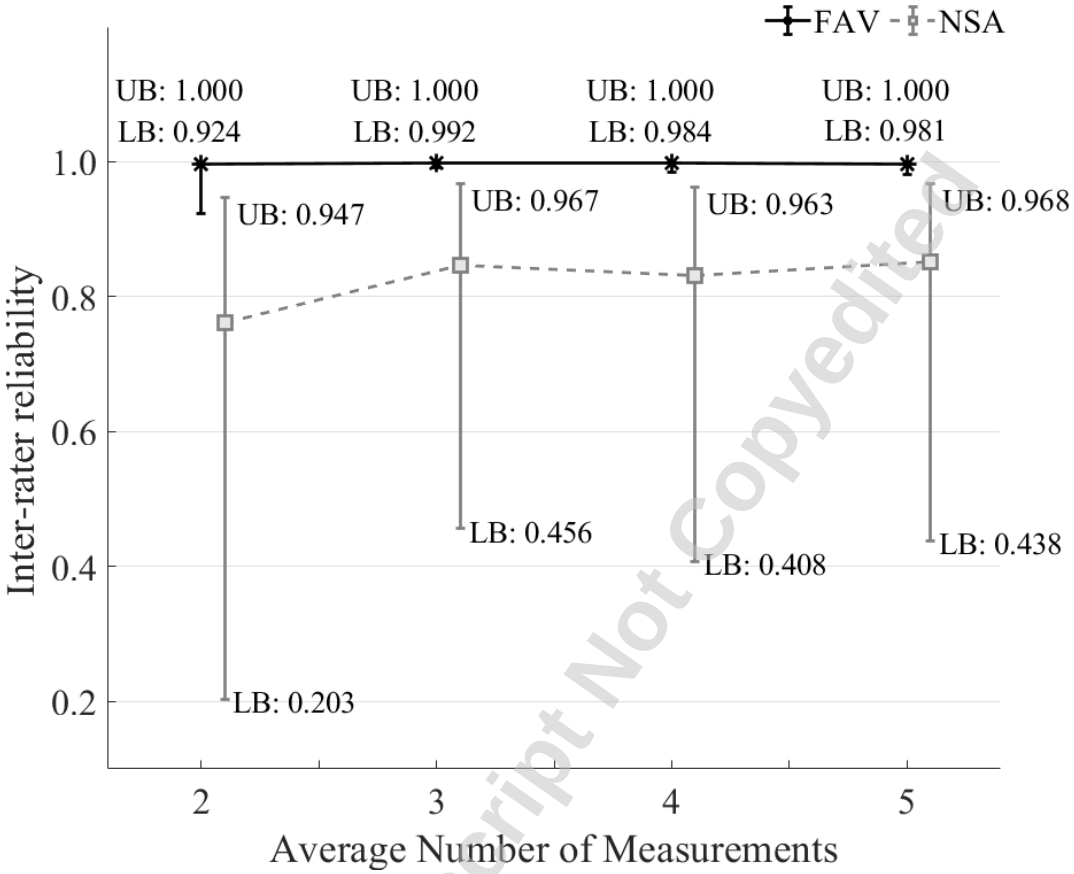
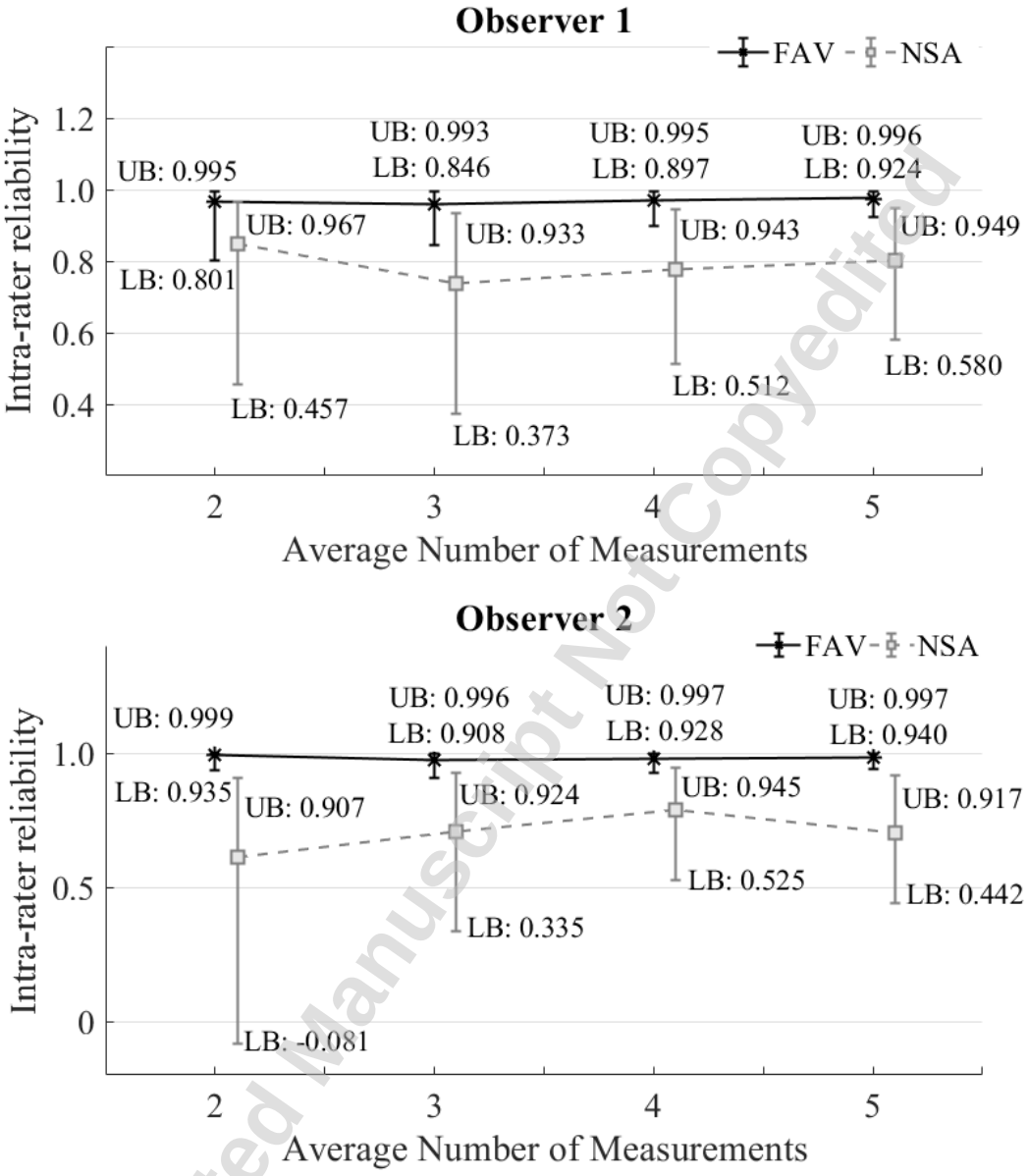


Fig. 5. Inter-rater reliability by computing ICC of the mean of a varying number of measurements for all decedents by each observer. upper bound UB) and lower bound (LB) on a 95% confidence interval



411
412 Fig. 6. Intra-rater reliability for both observers by computing ICC of a varying number of measurements for
413 all decedents. upper bound (UB) and lower bound (LB) on a 95% confidence interval.
414

415 Table 1. Descriptive statistics for FAV and NSA for both observers, separated by left and right sides.

	Mean ± SD (Right)	**Mean ± SD (Left)**	**Mean ± SD (All)**
Observer 1			
NSA	128.86 ± 5.22	128.52 ± 5.29	128.69 ± 5.19
FAV	32.57± 10.57	38.45 ± 12.48	35.51 ± 11.75
Observer 2			
NSA	128.24 ± 4.89	126.38 ± 4.21	127.31 ± 4.60
FAV	33.21 ± 10.57	38.00 ± 12.85	35.60 ± 11.81
ALL			
NSA	128.55 ± 5.00	127.45 ± 4.84	128.00 ± 4.92
FAV	32.89 ± 10.39	38.23 ± 12.44	35.56 ± 11.68

416

417 Table 2. Inter-rater reliability using ICC by using all five measurements per subject by each observer.

	Inter-rater score	**95% Confidence Interval**	
		Lower	*Upper*
NSA			
Left	0.66	0.27	0.85
Right	0.82	0.60	0.92
Overall	0.74	0.53	0.85
FAV			
Left	0.97	0.92	0.99
Right	0.99	0.98	1.00
Overall	0.98	0.96	0.99

418