Evaluation of Limb Length Discrepancies Using Manual Measurements of Mechanical Axis Films

Ryan Jaworski[1], Luke Bornstein[1], Zuzanna Janusz[1], Swathi Venkatesan[1], Lisa Tamburini[2], Adam Weaver[3], Donna M. Pacicca[3],John F. Drazan[1]
[1]Fairfield University, Fairfield CT, [2] University of Connecticut Medical Center, Farmington CT, [3]Connecticut Children's Hospital, Hartford CT
jdrazan@fairfield.edu

Disclosures: Ryan Jaworski (N), Luke Bornstein (N), Zuzanna Janusz (N), Swathi Venkatesan (N), Lisa Tamburini (N), Adam Weaver (N), Donna Pacicca (N), John Drazan (N)

INTRODUCTION: Limb length discrepancy (LLD) is a common orthopedic condition in pediatrics, characterized by having legs of unequal length. LLDs greater than 2 cm are considered clinically significant and impact long term health, estimated to occur 1 in 1000 in the general population. LLD is associated with back and joint pain, gait imbalances, and chronic musculoskeletal conditions later in life. Mechanical Axis Films (MAF) are the present standard of care for assessment of suspected LLDs, however, the use of ionizing radiation generates concern from parents and precludes the collection of images outside of 6-month time intervals. Early and accurate diagnosis of LLD in pediatric patients experiencing rapid growth is especially important in allowing orthopedic specialists to determine treatment methods and prevent the condition from worsening. As such, there is significant interest in developing alternative approaches for the measurement of bone length in the lower limbs, without the need for ionizing radiation. To the best of our knowledge, there have been few studies evaluating the agreement of MAF measurements across raters and days, making the development of alternative modalities a challenge without comparative benchmarks. Therefore, the purpose of this study is to assess the inter and intra-rater reliability across 5 raters of varying experience for the measurement of lower legs bone length to assess LLDs. Additionally, human measurements were compared with automatically calculated measurements from Picture Archiving and Communication System (PACS) system. We hypothesized that there would be concordance of measurements between raters and the PACs system.

METHODS: Our study investigates the reproducibility and reliability of adolescent MAF measurements across five raters based on historical imaging from 49 patients treated at a single clinical center (33/24 male/female). After IRB approval, we identified 150 potential patient charts for inclusion in the study based on the presence of 3 MAF films in their chart with spacing of approximately 6 months between each image. After reviewing each patient chart, a total of 93 patients were excluded from further analysis due to these criteria: 1) no ruler on film, 2) age out of range. Of the remaining 57 patients, we randomly selected 8 patients data to be used in a training protocol for MAF measurement among the raters, leaving a total of 49 patient charts with 3 MAF images each. For each patient, we noted the following information from their charts: age, weight, height, and time in between studies for further analysis. The attending physician gave a training session on how to measure MAFs to the other raters. Each leg on each MAF was measured at the top of the femur, bottom of the femur, top of the tibia, and bottom of the tibia using specified landmarks the attending physician familiarized everyone with. There were a total of two rounds of measurements a week apart for each image, using ImageJ which mimics the DICOM software utilized in the clinic (EUnity, Mach7, Burlington, VT). A MATLAB script was written to randomize the data, giving each rater a unique sequence of blinded images to measure. The measurements were then taken across at least 7 days and put into a spreadsheet. The data from the spreadsheet was then uploaded into another MATLAB script with a specific key for each rater, generated from the first script, that derandomized the data.

RESULTS: 49 pediatric patients with mechanical axis films were analyzed. Inter-rater reliability, which assesses consistency of measurements across different raters, was excellent across all anatomical sites, with intraclass correlation coefficients (ICCs) ranging from 0.962 to 0.996 at both time points (95% CIs consistently narrow). Intra-rater reliability, which evaluates the repeatability of measurements by the same rater over time, was similarly high for all raters, with ICCs generally above 0.94. The attending physician demonstrated excellent repeatability for the femur (ICCs 0.94–0.99) and good repeatability for the right tibia (ICC 0.92). The medical resident and two undergraduate raters achieved excellent intra-rater reliability across all sites (ICCs >0.97). Agreement between the attending physician and automated measurements from Computer was also strong, with ICCs of 0.993 (L. femur), 0.992 (L. tibia), 0.944 (R. femur), and 0.916 (R. tibia), (Figure 1 and 2).

DISCUSSION: This study demonstrated excellent inter- and intra-rater reliability of ruler-based mechanical axis film measurements across all raters, including those with limited prior experience, and strong agreement between human and automated computer measurements. Outliers in the right tibia and femur were suspected to be single instance keystroke errors. These consistent results indicate that ruler-based measurements are not only accurate and reproducible but also straightforward to teach, suggesting that tasks could be reliably delegated to trained personnel such as technicians to improve efficiency in pediatric imaging workflows. The high concordance between computer-generated and attending physician measurements further supports the integration of automated tools as a time-saving alternative without loss of accuracy. While advanced modalities such as EOS may reduce radiation exposure, their broader adoption is limited by cost, availability, risk of equipment breakdown, and potential image stitching errors. Overall, our findings highlight both the practicality of ruler-based training and the promise of computer-assisted methods for reliable, efficient assessment of limb length discrepancy in children. Additionally, it provides benchmarking for the development of novel LLD monitoring approaches.

SIGNIFICANCE/CLINICAL RELEVANCE: The findings have significant implications for reducing cumulative radiation exposure in pediatric patients while maintaining diagnostic accuracy, potentially informing future diagnostic protocols and technological innovations in limb length assessment.

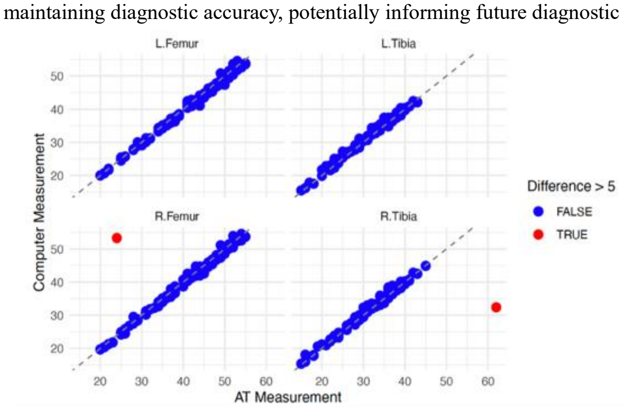

Figure 1: Comparison of Attending Physician (AT) and automated analysis (Computer).

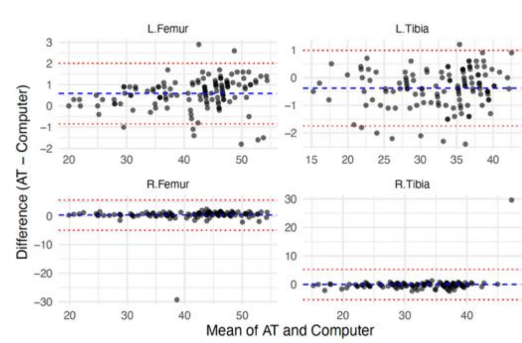

Figure 2: Comparison of Attending Physician (AT) and automated analysis (Computer)