

Original Research

The Reliability of Measurements for Tibial Torsion: A Comparison of CT, MRI, Biplanar Radiography, and 3D Reconstructions With and Without Standardized Measurement Training

Eric W. Edmonds, MD^{1,2}; Kevin C. Parvaresh, MD¹; Mason J. Price, BS¹; Christine L. Farnsworth, MS²; James D. Bomar, MPH²; Jessica L. Hughes, MD²; Vidyadhar V. Upasani, MD^{1,2}

¹University of California San Diego Medical Center, San Diego, CA; ²Rady Children's Hospital-San Diego, San Diego, CA

Correspondence: Vidyadhar V. Upasani, MD, Department of Orthopedic Surgery, Rady Children's Hospital-San Diego, 3020 Children's Way, MC5062, San Diego, CA 92123. E-mail: vupasani@rchsd.org

Received: February 3, 2023; Accepted: April 21, 2023; Published: August 1, 2023

DOI: 10.55275/JPOSNA-2023-661

Abstract

Background: Accurate and reliable assessment of tibial torsion is important for the identification and treatment of tibial rotation malalignment; however, the ideal rotational measurement modality and technique are controversial. This study compares rotational measurements between computed tomography (CT), magnetic resonance imaging (MRI), biplanar erect radiograph (BER) reconstructions, and three-dimensional (3D) reconstructed CT before and after standardized training of observers to evaluate the reliability of each assessment method.

Methods: Eight adult cadaveric specimens underwent CT, MRI, and BER imaging. Tibial torsion was measured by three independent observers (one resident and two experienced orthopaedic surgeons) both before and after standardized measurement instruction. Reliability for inter-observer measurement was evaluated using the Intraclass Correlation Coefficient (ICC). Measurement values for CT, MRI, and BER reconstructions were compared to 3D CT reconstructions analyzed using custom software to align and measure tibial torsion (used as the reference standard).

Results: Before training, there was poor inter-observer reliability for CT (ICC=0.492, p=0.014) and moderate inter-observer reliability for MRI (ICC=0.633, p=0.002). There was no inter-method reliability between 3D CT and MRI for two of the three surgeons and moderate to good reliability between 3D CT and CT. After training, the inter-observer reliability for CT improved to 0.536 and the inter-observer reliability for MRI improved to 0.701. The BER measurements (no observer involvement) had moderate reliability compared to the 3D CT reconstructions (ICC=0.69,

$p=0.026$). Measurement error was 4 degrees for CT pre-training and 7 degrees post-training, and 7 degrees for MRI pre-training and 8 degrees post-training.

Conclusions: A standardized training regimen for MRI measurements improved both inter-observer and intra-observer reliability. Inter-method reliability between CT, MRI, and BER compared to reference 3D CT reconstructions demonstrated that all imaging modalities are a valid means to measure tibial rotation but differ in reliability from moderate to good. When assessing tibial torsional deformities, it is important to consider these variations from true rotation and feel comfortable using them for pre-operative planning purposes.

Level of Evidence: N/A for a cadaveric imaging study

Key Concepts

- A standardized training regimen for MRI measurements of tibial torsion significantly improved inter-observer and intra-observer reliability.
- CT, MRI, and BER measurements are valid imaging modalities compared to 3D CT reference standards for measuring tibial rotation.
- Established manual measurements of tibial torsion from MRI may be improved with adequate training.
- Measurement error of up to 8 degrees can be seen from true tibial torsion during the imaging assessment.

Introduction

Rotational alignment is an important factor influencing lower extremity function and pathology. Excessive tibial torsion can affect growth and development of multiple lower extremity disorders.¹ In cases where torsion becomes symptomatic and interferes with function, surgical management typically involves tibial derotational osteotomy.^{2,3} Precise measurement of rotational deformity and correction is therefore important yet can be a challenge to reproduce pre-, intra-, and post-operatively.

Accurate measurement of tibial torsion is important for diagnosis and management of many lower extremity deformity conditions, including but not limited to tibial torsion disorders and Blount's disease. In the era of shifting imaging protocols from CT scan to MRI due to a desire to minimize the risk of radiation to children, it is important to identify whether there is the ability to measure tibial torsion accurately. Tibial rotation, or torsional profile, may be measured by direct

visualization, clinical exam using thigh-foot axis, foot progression angle assessed by gait analysis, ultrasound, computed tomography (CT), magnetic resonance imaging (MRI), or even from biplanar erect radiograph (BER) imaging using computer-generated three-dimensional (3D) reconstructions and torsional profile measurement software (sterEOS®, EOS Imaging, Paris, France).

Previous studies have always used another measurement modality to compare the tested modality in question, but none have known the true measure of tibial torsion. This investigation aims to determine that true reference measure of rotation using custom software (MATLAB, MathWorks, Natick, MA) to identify the true torsional profile of cadaver tibia and then compare the reliability of various techniques of measurement against that true value. Our hypothesis is that all radiographic modalities (CT, MRI, and BER) will have variations from the true value but have reliable inter-observer measurements that

reflect past publications. A secondary aim of this study is to evaluate the efficacy of a standardized measurement training program for evaluating tibial torsion based on established manual measurements from MRI and CT exams.

Materials and Methods

Eight skeletally mature adult cadaveric lower extremities were obtained after institutional approval. None of the cadaveric lower extremities had prior extremity surgery, fracture, infection, osteoporosis, or known systemic disorder affecting musculoskeletal development. Each specimen underwent CT (120 kV, 190-240 mA, 0.35s rotation speed helical imaging, 0.625 mm slice thickness; GE VCT 64-Slice LightSpeed CT Scanner, Piscataway, NJ), and MRI (1.5 T, used pulse sequence 3D gradient recalled echo MERGE T2 weighted image with no fat saturation, slice thickness of 1.0 mm, field-of-view 14%, repetition time 66.1 ms, echo time min/Full; GE Discovery MR450, Milwaukee, WI).

Tibial torsion measurements were performed manually using the imaging tools on the PACS software (Amicas Inc., Brighton, MA). Initial rotation measurements were performed by three evaluators (A—orthopaedic surgery resident, B—fellowship-trained pediatric orthopaedic surgeon over a decade in practice, and C—fellowship-trained pediatric orthopaedic surgeon 5 years into practice), without standardized agreement or instruction in order to mimic a routine clinical setting. Following these measures, the same evaluators received training consisting of reviewing a presentation of four slides. Each slide contained an example axial image with specific reference points defined and angular measurement guidelines (slide 1—axial CT image of proximal tibia [Figure 1 left], 2—axial CT image of distal tibia [Figure 1 right], 3—axial MRI image of proximal tibia [Figure 2 left], 4—axial MRI image of distal tibia [Figure 2 right]). The CT measurement was defined via the Goutallier method (Figure 1),⁴ and the MRI measurement was defined by the Rosskopf modification of the Goutallier method (Figure 2).⁵

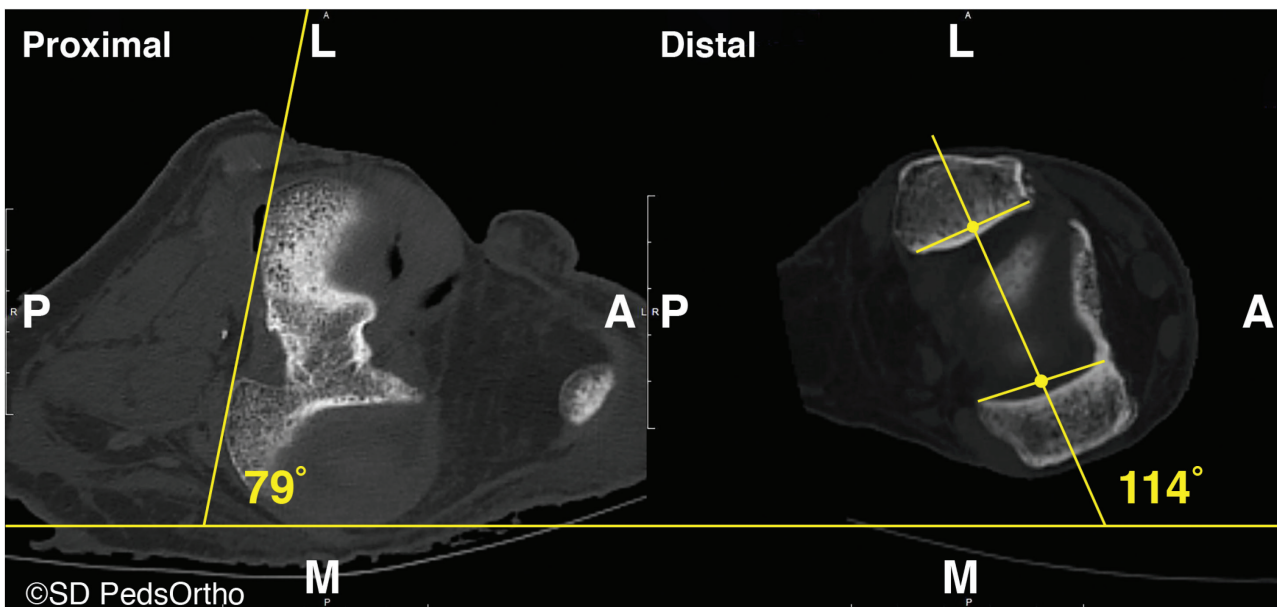


Figure 1. The Goutallier method⁴ of tibial torsion calculation from axial CT images is demonstrated. Proximally, the CT section immediately distal to the femorotibial space was used to draw the proximal tibial line through the most prominent part of the tibial plateau and an angle was measured between this line and a reference horizontal line. Distally, the first section distal to the tibial pilon on which the posterior malleolus was not visible was used to determine a bimalleolar line, and an angle was measured between this line and a reference horizontal line. Here, tibial torsion is calculated as the difference between the distal and proximal angles: 114 degrees minus 79 degrees = 35 degrees.

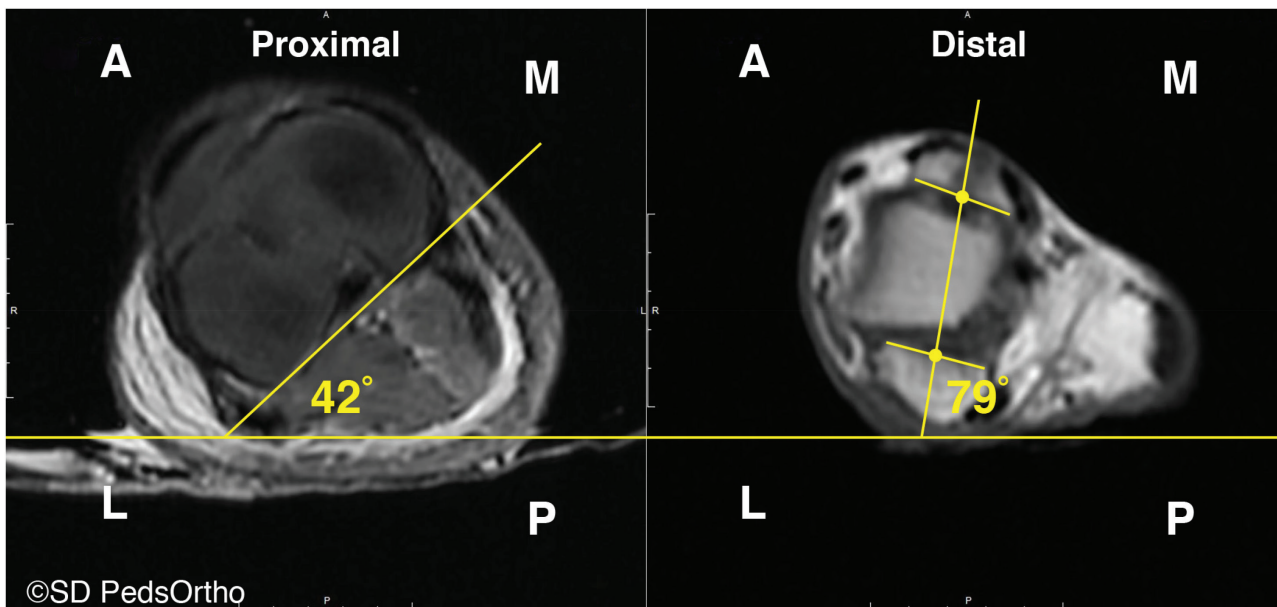


Figure 2. The Rosskopf modification⁵ of the Goutallier method of tibial torsion calculation from axial MRI images is demonstrated. Proximally, a line along the dorsal contour of the tibial plateau and a reference horizontal line determined the proximal angle. Distally, a line between the mid-points (yellow circles) of the lines drawn along the cartilage-covered joint surface of the medial and lateral malleolus determined the intermalleolar axis, and the angle between this and the reference horizontal line was the distal angle. Here, tibial torsion is calculated as the difference between the distal and proximal angles: 79 degrees minus 42 degrees = 37 degrees.

The proximal angle was measured on the axial slice through the tibial plateau just proximal to appearance of the fibular head, defined by the horizontal line parallel to the bottom of the image and the bone line along the most prominent part of the posterior aspect of the medial and lateral plateaus, excluding any osteophytes. The distal angle was measured on the first axial slice distal to the tibiotalar joint including the superior talar dome, defined by the horizontal line parallel to the bottom of the image and the bone line between the center of the medial and lateral malleolar articular surfaces (the intermalleolar axis). Torsion was defined as the overall difference between the angles of the proximal and distal measurements relative to the bone orientation, with external torsion defined as positive and internal torsion defined as negative. The three evaluators performed the independent CT and MRI measurements again after training.

For BER imaging, the lower extremities were suspended upright with weight-bearing on the foot and patella

oriented toward the AP/PA scanner, with the scan proceeding from the hip distally to the foot. BER DICOM files were then provided to the system developer “in-house” (EOS Imaging, Paris, France) so that they could perform their best-fit 3D reconstructions using their sterEOS® software. This process utilizes software that produces a generic digital tibia model which is overlaid onto biplanar radiographs. A technician uses predefined landmarks to create a best-fit, “patient-specific” model by stepwise manipulations of the generic model to conform them to the specific landmarks of each tibia. A full 3D model of the tibia is generated from the modified generic overlay from which the software calculates tibial torsion, amongst other metrics.

A 3D CT measurement was performed utilizing the same images obtained for the aforementioned measurements to have a “reference standard” of true torsion (torsion was calculated from uniformly aligned 3D reconstructed models so was void of user measurement errors). DICOMs of axial CT images of each tibia were

imported into Mimics software (Materialise NV, Leuven, Belgium), and a 3D stereolithography (STL) file of the tibia was rendered. STL data were imported into custom MATLAB (MathWorks, Natick, MA) software. The posterior aspects of the proximal tibial condyles and the anterior and posterior edges of the medial malleolus were identified from the axial view of 3D images, which could be manipulated by the software user. The code aligned the tibia to its mechanical axis. In 3D space, code identified lines between the two proximal condyles and between the midpoint of the medial and lateral malleoli. These lines were projected onto the transverse plane orthogonal to the mechanical axis of the tibia, and tibial torsion was calculated as the angle between these lines (Figure 3).

Basic descriptive statistics are reported. Intraclass Correlation Coefficient (ICC) values were compared to evaluate inter-observer reliability for authors pre- and post-instruction between CT and MRI measurements. A p-value of <0.05 was set for significance with 95% confidence intervals reported. ICC values with a

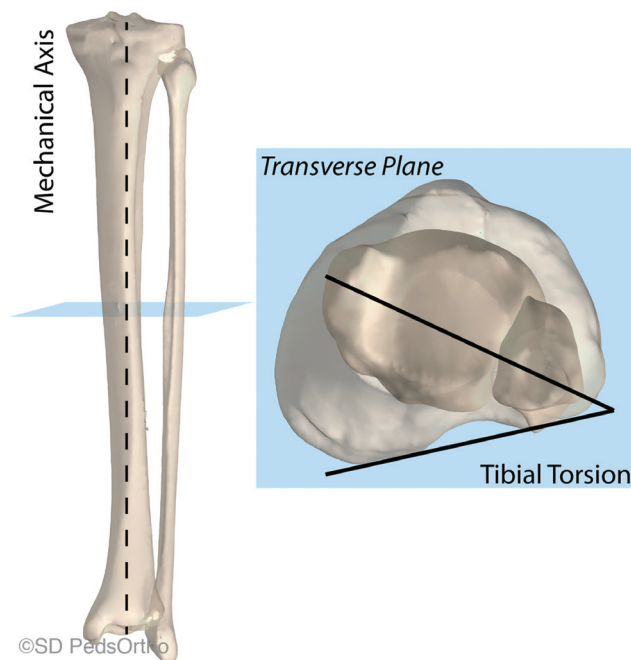


Figure 3. The 3D CT method using custom software aligning the tibia in the mechanical axis. The transverse plane to the mechanical axis is shown in blue and the 3D tibial torsion angle projected onto the plane on the right.

significance of $p>0.05$ are reported as “no agreement.” Agreement strength was assessed based on tiers outlined by Koo and Li (<0.5 poor, 0.5 to 0.75 moderate, 0.75 to 0.9 good, ≥ 0.9 excellent).⁶ ICC values were then compared to evaluate intra-observer reliability for authors that completed multiple measures both pre- and post-instruction for both CT and MRI measurements. ICC values were then used to compare measures between BER best-fit calculated measurement values and the reference standard 3D CT measurements. Measurement error was calculated by determining the difference between each observer’s measurements and the measurement derived from the 3D CT, which was used as the reference standard. The standard error of the difference between observer measurement and the reference standard was determined. The standard error was then multiplied by 1.96, and the product was added to the mean difference between observer measurements and the reference standard. Statistical analysis was performed using SPSS (version 26, IBM, Armonk, NY).

Results

Eight lower extremities were used in the study: five left and three right extremities were from four female and four male donors, 72.9 ± 16.8 (range 53-90) years old. CT images confirmed no bony defects. Raw measurement data are available in the Appendix. The average absolute difference in tibial torsion between the 3D CT reference standard and the BER was lower than for CT and MRI modalities (Table 1). ICC values for pre- and post-training inter-observer agreement between CT and MRI are listed in Table 2.

Before training, there was poor inter-observer agreement for CT measures (ICC=0.492, $p=0.014$), with moderate inter-observer agreement for MRI (ICC=0.633, $p=0.002$). There was no agreement between CT and MRI for Surgeons A and B ($p>0.05$) but good agreement for Surgeon C. After training, the agreement improved for all measures; Inter-observer agreement was improved to moderate for CT measures (ICC=0.536, $p=0.008$) and improved for MRI measures (ICC=0.701, $p<0.001$). Significant agreement between CT and MRI after

Table 1. Absolute Difference in Measurements from Our 3D CT Reference Standard

	Mean±SD	Minimum	Maximum
BER	3.5±2.8	0	8
Surgeon A – CT	3.6±3.7	0	10
Surgeon B – CT	8.5±8.4	2	27
Surgeon C – CT	4.4±4.1	0	11
Surgeon A – MRI	5.6±3.4	1	11
Surgeon B – MRI	9.6±6.1	1	21
Surgeon C – MRI	4.0±2.4	0	7

training was present for all three surgeons, with Surgeon A (ICC=0.760, $p=0.009$) and Surgeon B (ICC=0.792, $p=0.005$) improving from no agreement to good agreement.

ICC values for pre- and post-training agreement between CT and MRI utilizing 3D CT measurements as the reference standard are listed in Table 3. Before training, there was good agreement between CT and 3D reconstruction for two of the three surgeons (Surgeon A (ICC=0.795, $p=0.005$), Surgeon B (ICC=0.769,

$p=0.008$), with Surgeon C having moderate agreement (ICC=0.716, $p=0.015$). There was no agreement between measurements from MRI and the reference standard for two of the surgeons prior to training and moderate agreement for the other surgeon. Post-training agreement between CT and the reference standard ranged from no agreement to moderate. Post-training agreement between MRI and the reference standard improved for Surgeons A and C but not for Surgeon B.

The ICC intra-observer values for pre- and post-training agreement between CT and MRI were completed for only Surgeon B, who did at least three different measurements on each modality both pre- and post-training. The intra-observer measurement agreement improved for CT from moderate to good agreement and for MRI from poor to good agreement. These ICC values are listed in Table 4.

There was moderate agreement between 3D CT measurements and BER measurements provided by EOS (ICC=0.690, $p=0.026$, Table 3). The training did not improve ICC for CT measurements (0.891 and 0.636) but resulted in improved ICC for MRI measurements (0.606 and 0.690). All ICC values relative to the standard reference of MatLab 3D CT are provided in Table 5.

Table 2. Inter-Observer and Inter-Modality Reliability between CT and MRI Before and After Training

			ICC	Lower	Upper	Sig.
Pre-training	Surgeon A vs B vs C	CT	0.492	0.050	0.850	$p = 0.014$
		MRI	0.633	0.216	0.901	$p = 0.002$
	CT vs MRI	Surgeon A	No agreement			$p = 0.398$
		Surgeon B				$p = 0.105$
		Surgeon C	0.864	0.466	0.971	$p = 0.001$
Post-training	Surgeon A vs B vs C	CT	0.536	0.097	0.867	$p = 0.008$
		MRI	0.701	0.315	0.923	$p < 0.001$
	CT vs MRI	Surgeon A	0.760	0.189	0.947	$p = 0.009$
		Surgeon B	0.792	0.266	0.955	$p = 0.005$
		Surgeon C	0.764	0.198	0.948	$p = 0.008$

No agreement occurred when $p > 0.05$.

Table 3. Measurement Reliability between CT and MRI versus 3D CT Measurements as the Reference Standard

			ICC	Lower	Upper	Sig.
Pre-training	3D CT vs	Surgeon A – CT	0.795	0.272	0.955	p = 0.005
		Surgeon B – CT	0.769	0.209	0.949	p = 0.008
		Surgeon C – CT	0.716	0.095	0.936	p = 0.015
		Surgeon A – MRI	No agreement			p = 0.098
		Surgeon B – MRI	No agreement			p = 0.092
		Surgeon C – MRI	0.651	0.027	0.919	p = 0.029
Post-training	3D CT vs	Surgeon A – CT	0.728	0.121	0.939	p = 0.013
		Surgeon B – CT	No agreement			p = 0.216
		Surgeon C – CT	0.690	0.043	0.929	p = 0.020
		Surgeon A – MRI	0.754	0.176	0.945	p = 0.009
		Surgeon B – MRI	No agreement			p = 0.269
		Surgeon C – MRI	0.786	0.251	0.953	p = 0.006

No agreement occurred when $p > 0.05$.

Table 4. Measurement Reliability between Pre- and Post-Training for Both CT and MRI

		ICC	Lower	Upper	Sig.
CT	Pre	0.692	0.301	0.920	p < 0.001
	Post	0.883	0.667	0.973	p < 0.001
MRI	Pre	0.452	0.009	0.833	p = 0.023
	Post	0.785	0.459	0.947	p < 0.001

Measurement error was calculated to be 4 degrees for CT pre- and 7 degrees post-training. Pre-training measurement error for MRI was 7 degrees and post-training was 8 degrees. These differences were not statistically significant ($p=0.177$).

Discussion

Many means of measuring tibial torsion have been independently measured but until the present study,

Table 5. Measurement Reliability Using Two-Way Analysis Across All Measurement Types (All Surgeons Averaged) Against the Reference Standard of the 3D CT Reconstruction Method

	ICC	Lower	Upper	Sig.
3D CT vs BER	0.690	0.011	0.931	p = 0.026
3D CT vs Pre-training CT	0.891	0.553	0.977	p < 0.001
3D CT vs Post-training CT	0.636	0.053	0.915	p = 0.033
3D CT vs Pre-training MRI	0.606	0.102	0.906	p = 0.042
3D CT vs Post-training MRI	0.690	0.044	0.929	p = 0.020

have not compared across modality types or assessed for measurement error against a 3D standard. This study found a marginal improvement in measurement reliability for MRI images after implementation of a standardized measurement training method across experience levels but had a notable decrease in reliability for post-training CT measurements. The BER measurements (using best-fit modeling) had moderate agreement with the 3D CT reconstructions (our reference standard).

Direct visualization by clinical exam or gait analysis, although easy and efficient, has not demonstrated accurate results utilizing the bimalleolar axis or the thigh-foot axis.⁷⁻⁹ Ultrasound can reproducibly provide measures of tibial torsion with no radiation exposure but is highly dependent on the skill of the ultrasonographer.^{10,11} CT has been shown to provide precise torsional measurements¹² but involves higher radiation that may have long-term consequences in a pediatric population.¹³ Studies have also verified accurate torsional measurements via MRI;^{14,15} however, cost and availability hinder its universal application as well as the ability for young children or those with spasticity to be immobile for an adequate length of time without adding the risk of needing sedation. Recently, reconstructed best-fit 3D BER imaging software (sterEOS, EOS Imaging, Paris, France) has been used to assess the rotational structures of the lower limb but with less-than-optimal results when performed at the time of imaging.^{16,17} Furthermore, the intra-operative utility of CT and MRI is very limited and logistically challenging, whereas biplanar imaging is presently impossible in the operating room given patient positioning restrictions.

When assessing CT and MRI for external validity, moderate to good reliability was found for both CT and MRI in the present study. Clinical applicability of 3D lower extremity morphology may extend beyond congenital tibial torsion. Lower extremity fracture malalignment may be more common than traditionally reported, with studies ranging from 38 to 61%.^{18,19} Tibial rotation has also been identified as a risk factor for patellar instability, and surgical techniques involving

tibial tubercle and derotational osteotomy have been described.^{20,21} Patient-specific cutting guidance and instrumentation are at the forefront of advancing surgical technology and currently rely on validated three-dimensional measurements from CT or MRI.^{22,23} Accurate and reproducible measurement of tibial torsion both pre- and post-operatively may therefore allow for improved diagnostic accuracy for more precise torsion correction.

This is the first study to examine the relationship of CT and MRI measurement reliability following implementation of a standard measurement protocol. Prior studies have validated both CT and MRI measurements of tibial torsion individually along with validating CT compared to MRI measures with good correlation. However, none of these studies examined the effect of training discrepancies between resident and attending surgeons. Standardizing measurements is important not only for diagnostic value but also for surgical management and standardized outcomes research. As there is high variability in measurement options described in the literature, it is critical to clearly outline measurement protocols for future research and analyses.

The present results comparing CT and MRI are similar to a prior study by Muhamad et al. evaluating tibial torsional measurement.²⁴ When measured by two fellowship-trained orthopaedic surgeons, they found a significant inter-observer ICC of 0.930 for CT and 0.878 for MRI, which are slightly higher compared to the post-training values seen in this study of 0.536 and 0.701, respectively. The differences may be attributable to measurement technique or training level. Together, these results support both CT and MRI as internally validated means of assessing tibial torsion. In this study, we found a post-training measurement error of 7 degrees for CT and 8 degrees for MRI.

The 4 to 8 degrees of error is on the order of what is accepted as appropriate for most musculoskeletal evaluations.²⁵ Part of this accepted measurement error exists in the fact that the ability to obtain similarly precise measurements of tibial torsion intra-operatively without the same CT or MRI assessments is considered

less achievable. Moreover, despite the varied methods used for assessing tibial torsion, there may be a range of acceptable intra-operative changes that allow good functional and clinical outcomes after corrective osteotomy (particularly considering the potential for muscle memory of the child to affect post-osteotomy gait measures, regardless of intraoperative measurements). Thus, the ultimate significance of small variations in torsional measurements is unclear. Therefore, considering these unknowns, the current practice in our institution has changed to obtain MRI torsional profiles to reduce radiation exposure in our patient population even though it may have a higher error in measurement reliability.

There are a number of limitations in this study. A standard measurement protocol was used that is easy to learn and to apply in a clinical setting with demonstrable reproducibility. These methods were chosen in order for the interpreting physician to obtain precise and accurate values regardless of the MRI or CT protocol. Unfortunately, the inter-rater correlation was only marginally improved with training despite a significant improvement in the intra-rater correlations observed. This suggests that the training on good methodology at least reduced the error (improved precision) in measurement for an observer but did not actually improve the accuracy of the measurements obtained. Moreover, true interpretation of the patient's torsional profile is gained from multiple facets of the clinical visit, including physical exam, gait analysis, plus imaging. Therefore, it is imperative that providers use all their potential tools before making clinical decisions.

Although this study includes eight specimens for data collection, with significant agreement for a number of measurements, the smaller sample size accounts for the large confidence intervals identified. It is additionally possible that measures without significance could have demonstrated type two errors due to the low number of specimens. The use of adult specimens may not mimic the exact situation of pediatric or adolescent specimens who are more likely to have thicker periosteum and abundant cartilage present potentially confounding

the measurements in that younger group; however, procurement of younger specimens is extremely difficult. In addition, torsion measurements were compared amongst observers, so the exact amount of each measurement is not as crucial, as it relates to a child's torsion versus an adult's tibial torsion. A more recent study has suggested that 3D models from low-dose biplanar radiographs have similar reliability for measuring torsion versus standard CT scan.²⁶ Our results are no different regarding the reliability of this imaging modality and even suggest that the best-fit reconstructions provided by sterEOS are just as reliable (if not more so) than the physician-derived measures of CT and MRI. For the purposes of attaining ALARA (As Low As Reasonably Achievable) radiation safety principles, we would then advocate the lower radiation exposure of BER, if available.

In conclusion, standardized CT and MRI torsional measurement protocols provide both an accurate and reliable means of assessing tibial torsion across multiple levels of training (within an 8-degree measurement error). Moreover, the sterEOS best-fit 3D reconstructions may also provide reliable measures with a time-consuming process that cannot be performed real-time in clinic but can be performed prior to surgical intervention. The presented comparison across all radiographic means of obtaining a rotational measure of tibial torsion provides insight into the reliability of each measure and begs the question of necessary accuracy in those measures. Yet, it also indicates that a provider may choose CT, MRI, or BER methodologies with equal assurance to the measurements obtained relative to the true value of tibial torsion. Additional research will need to be directed toward assessing reproducibility of torsional correction utilizing intraoperative imaging guidance to improve clinical outcomes, and how these pre-operative measures relate to those outcomes.

Acknowledgements

The authors would like to thank Megan Jeffords, MS, for developing software methods for 3D measurements used in this manuscript.

Disclaimer

Research support was obtained from the Orthopedic Division of Children's Specialists of San Diego, and a tissue grant was obtained from the University of California, San Diego, School of Medicine, Anatomical Materials program. The authors report no conflicts of interest related to this manuscript.

References

1. Parikh S, Noyes FR. Patellofemoral disorders: role of computed tomography and magnetic resonance imaging in defining abnormal rotational lower limb alignment. *Sports Health*. 2011;3:158-169.
2. Davids JR, Davis RB, Jameson LC, et al. Surgical management of persistent intoeing gait due to increased internal tibial torsion in children. *J Pediatr Orthop*. 2014;34:467-473.
3. Krenkel WF, Staheli LT. Tibial rotational osteotomy for idiopathic torsion. A comparison of the proximal and distal osteotomy levels. *Clin Orthop*. 1992;283:285-289.
4. Goutallier D, Van Driessche S, Manicom O, et al. Influence of lower-limb torsion on long-term outcomes of tibial valgus osteotomy for medial compartment knee osteoarthritis. *J Bone Joint Surg Am*. 2006;88:2439-2447.
5. Roskopf AB, Buck FM, Pfirrmann CWA, et al. Femoral and tibial torsion measurements in children and adolescents: comparison of MRI and 3D models based on low-dose biplanar radiographs. *Skeletal Radiol*. 2017;46:469-476.
6. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155-163.
7. Carriero A, Zavatsky A, Stebbins J, et al. Influence of altered gait patterns on the hip joint contact forces. *Comput Methods Biomech Biomed Engin*. 2014;17:352-359.
8. Radler C, Kranzl A, Manner HM, et al. Torsional profile versus gait analysis: consistency between the anatomic torsion and the resulting gait pattern in patients with rotational malalignment of the lower extremity. *Gait Posture*. 2010;32:405-410.
9. Tamari K, Tinley P, Briffa K, et al. Validity and reliability of existing and modified clinical methods of measuring femoral and tibiofibular torsion in healthy subjects: use of different reference axes may improve reliability. *Clin Anat*. 2005;18:46-55.
10. Joseph B, Carver RA, Bell MJ, et al. Measurement of tibial torsion by ultrasound. *J Pediatr Orthop*. 1987;7:317-323.
11. Krishna M, Evans R, Sprigg A, et al. Tibial torsion measured by ultrasound in children with talipes equinovarus. *J Bone Joint Surg Br*. 1991;73:207-210.
12. Sayli U, Bölükbaşı S, Atik OS, et al. Determination of tibial torsion by computed tomography. *J Foot Ankle Surg*. 1994;33:144-147.
13. Brody AS, Frush DP, Huda W, et al. Radiation risk to children from computed tomography. *Pediatrics*. 2007;120:677-682.
14. Koenig JK, Pring ME, Dwek JR. MR evaluation of femoral neck version and tibial torsion. *Pediatr Radiol*. 2012;42:113-115.
15. Schneider B, Laubenberger J, Jemlich S, et al. Measurement of femoral antetorsion and tibial torsion by magnetic resonance imaging. *Br J Radiol*. 1997;70:575-579.
16. Cho BW, Lee TH, Kim S, et al. Evaluation of the reliability of lower extremity alignment measurements using EOS imaging system while standing in an even weight-bearing posture. *Sci Rep*. 2021;11(1):22039.
17. Westberry DE, Carpenter AM. 3D modeling of lower extremities with biplanar radiographs: reliability of measures on subsequent examinations. *J Pediatr Orthop*. 2019;39(10):521-526.
18. Buckley R, Mohanty K, Malish D. Lower limb malrotation following MIPO technique of distal femoral and proximal tibial fractures. *Injury*. 2011;42:194-199.
19. Phan VC, Wroten E, Yngve DA. Foot progression angle after distal tibial physeal fractures. *J Pediatr Orthop*. 2002;22:31-35.
20. Drexler M, Dwyer T, Dolkart O, et al. Tibial rotational osteotomy and distal tuberosity transfer for patella subluxation secondary to excessive external tibial torsion: surgical technique and clinical outcome. *Knee Surg Sports Traumatol Arthrosc*. 2014;22:2682-2689.
21. Teitge R. Osteotomy in the treatment of patellofemoral instability. *Tech Knee Surg*. 2006;5:2-18.
22. Munier M, Donnez M, Ollivier M, et al. Can three-dimensional patient-specific cutting guides be used to achieve optimal correction for high tibial osteotomy? Pilot study. *Orthop Traumatol Surg Res*. 2017;103:245-250.
23. Victor J, Premanathan A. Virtual 3D planning and patient specific surgical guides for osteotomies around the knee: a feasibility and proof-of-concept study. *Bone Joint J*. 2013;95-B:153-158.
24. Muhamad AR, Freitas JM, Bomar JD, et al. CT and MRI lower extremity torsional profile studies: measurement reproducibility. *J Child Orthop*. 2012;6:391-396.
25. Morrissy RT, Goldsmith GS, Hall EC, et al. Measurement of the Cobb angle on radiographs of patients who have scoliosis. Evaluation of intrinsic error. *J Bone Joint Surg Am*. 1990;72:320-327.
26. Buck FM, Guggenberger R, Koch PP, et al. Femoral and tibial torsion measurements with 3D models based on low-dose biplanar radiographs in comparison with standard CT measurements. *AJR Am J Roentgenol*. 2012;199:W607-W612.

Appendix

Raw Data for Tibial Torsion

					CT (°)			MRI(°)		
	Specimen	3D CT Standard(°)*	BER(°)*		Surgeon A	Surgeon B	Surgeon C	Surgeon A	Surgeon B	Surgeon C
Pre-training	1	32.5	34.5	25	32		38	29	36	36
	2	34.1	35.8	32	26		40	26	45	33
	3	36.6	31.4	29	31		32	21	32	30
	4	23.1	22.7	21	21		21	10	16	21
	5	30.2	32.1	19	34		35	30	38	33
	6	41.6	33.9	28	41		40	35	42	38
Post-training	7	29.4	31.1	23	30		30	34	23	34
	8	21.2	28.2	15	26		31	26	37	30
	1	32.5	34.5	33	30		33	32	27	37
	2	34.1	35.8	37	22		42	29	20	39
	3	36.6	31.4	27	10		36	28	16	32
	4	23.1	22.7	21	20		21	13	14	17
	5	30.2	32.1	30	33		38	28	31	30
	6	41.6	33.9	33	35		41	36	30	41
	7	29.4	31.1	24	18		25	23	20	33
	8	21.2	28.2	22	19		32	25	26	28

All Values Are Degrees

*Note: *3D CT Standard and BER values are not changed with training*