

# Differences in Femoral Torsion Among Various Measurement Methods Increase in Hips With Excessive Femoral Torsion

Florian Schmaranzer MD, Till D. Lerch MD, Klaus A. Siebenrock MD, Moritz Tannast MD, Simon D. Steppacher MD

Received: 14 June 2018 / Accepted: 26 November 2018 / Published online: 0, xxxx  
Copyright © 2019 by the Association of Bone and Joint Surgeons®

## Abstract

**Background** Correct quantification of femoral torsion is crucial to diagnose torsional deformities, make an indication for surgical treatment, or plan the amount of correction. However, no clear evaluation of different femoral torsion measurement methods for hips with excessive torsion has been performed to date.

**Questions/purposes** (1) How does CT-based measurement of femoral torsion differ among five commonly used measurement methods? (2) Do differences in femoral

torsion among measurement methods increase in hips with excessive femoral torsion? (3) What is the reliability and reproducibility of each of the five torsion measurement methods?

**Methods** Between March and August 2016, we saw 86 new patients (95 hips) with hip pain and physical findings suggestive for femoroacetabular impingement at our outpatient tertiary clinic. Of those, 56 patients (62 hips) had a pelvic CT scan including the distal femur for measurement of femoral torsion. We excluded seven patients (seven hips) with previous hip surgery, two patients (two hips) with sequelae of Legg-Calvé-Perthes disease, and one patient (one hip) with a posttraumatic deformity. This resulted in 46 patients (52 hips) in the final study group with a mean age of  $28 \pm 9$  years (range, 17–51 years) and 27 female patients (59%). Torsion was compared among five commonly used assessment measures, those of Lee et al., Reikerås et al., Jarrett et al., Tomczak et al., and Murphy et al. They differed regarding the level of the anatomic landmark for the proximal femoral neck axis; the method of Lee had the most proximal definition followed by the methods of Reikerås, Jarrett, and Tomczak at the base of the femoral neck and the method of Murphy with the most distal definition at the level of the lesser trochanter. The definition of the femoral head center and of the distal reference was consistent for all five measurement methods. We used the method described by Murphy et al. as our baseline measurement method for femoral torsion because it reportedly most closely reflects true anatomic femoral torsion. With this method we found a mean femoral torsion of  $28 \pm 13^\circ$ . Mean values of femoral torsion were compared among the five methods using multivariate analysis of variance. All differences between two of the measurement methods were plotted over the entire range of femoral

The institution of one or more of the authors (FS, KAS, MT) has received, during the study period, funding from the Swiss National Science Foundation, outside the submitted work.

Each author certifies that neither he, nor any member of his immediate family, has funding or commercial associations (consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that might pose a conflict of interest in connection with the submitted article.

*Clinical Orthopaedics and Related Research*® neither advocates nor endorses the use of any treatment, drug, or device. Readers are encouraged to always seek additional information, including FDA approval status, of any drug or device before clinical use.

Each author certifies that his institution approved the human protocol for this investigation and that all investigations were conducted in conformity with ethical principles of research. Drs Schmaranzer and Lerch share first co-authorship.

Department of Orthopaedic Surgery, Inselspital Bern, University of Bern, Bern, Switzerland

F. Schmaranzer (✉), Department of Orthopaedic Surgery, Inselspital Bern, University of Bern, Switzerland, Freiburgstrasse, 3010 Bern, Switzerland, email: florian.schmaranzer@insel.ch

All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research*® editors and board members are on file with the publication and can be viewed on request.

torsion to evaluate a possible increase in hips with excessive femoral torsion. All measurements were performed by two blinded orthopaedic residents (FS, TDL) at two different occasions to measure intraobserver reproducibility and interobserver reliability using intraclass correlation coefficients (ICCs).

**Results** We found increasing values for femoral torsion using measurement methods with a more distal definition of the proximal femoral neck axis: Lee et al. (most proximal definition:  $11^\circ \pm 11^\circ$ ), Reikerås et al. ( $15^\circ \pm 11^\circ$ ), Jarrett et al. ( $19^\circ \pm 11^\circ$ ), Tomczak et al. ( $25^\circ \pm 12^\circ$ ), and Murphy et al. (most distal definition:  $28^\circ \pm 13^\circ$ ). The most pronounced difference was found for the comparison between the methods of Lee et al. and Murphy et al. with a mean difference of  $17^\circ \pm 5^\circ$  (95% confidence interval,  $16^\circ$ – $19^\circ$ ;  $p < 0.001$ ). For six of 10 possible pairwise comparisons, the difference between two methods increased with increasing femoral torsion and decreased with decreasing femoral torsion. We observed a fair-to-strong linear correlation (R range, 0.306–0.622; all  $p$  values  $< 0.05$ ) for any method compared with the Murphy method and for the Reikerås and Jarrett methods when compared with the Tomczak method. For example, a hip with  $10^\circ$  of femoral antetorsion according to Murphy had a torsion of  $1^\circ$  according to Reikerås, which corresponds to a difference of  $9^\circ$ . This difference increased to  $20^\circ$  in hips with excessive torsion; for example, a hip with  $60^\circ$  of torsion according to Murphy had  $40^\circ$  of torsion according to Reikerås. All five methods for measuring femoral torsion showed excellent agreement for both intraobserver reproducibility (ICC, 0.905–0.973) and interobserver reliability (ICC, 0.938–0.969).

**Conclusions** Because the quantification of femoral torsion in hips with excessive femoral torsion differs considerably among measurement methods, it is crucial to state the applied methods when reporting femoral torsion and to be consistent regarding the used measurement method. These differences have to be considered for surgical decision-making and planning the degree of correction. Neglecting the differences among measurement methods to quantify femoral torsion can potentially lead to misdiagnosis and surgical planning errors.

**Level of Evidence** Level IV, diagnostic study.

## Introduction

Torsional deformities of the femur have increasingly been recognized as a cause of hip pain in young adults [15, 17, 29] and femoral torsion abnormalities recently have been integrated into the concept of femoroacetabular impingement (FAI) [3, 17, 24] separate from cam and pincer deformities. Femoral torsion considerably affects impingement-free hip ROM and can thereby aggravate or compensate cam- or

pincer-type impingement [5, 15]. More specifically, decreased femoral torsion can predispose patients to anterior intraarticular FAI, whereas excessively high femoral torsion can lead to posterior extraarticular impingement [24]. Severe torsional deformities have been reported in one of six symptomatic patients eligible for joint-preserving surgery [17]. Femoral osteotomies to correct these deformities have increasingly been performed with encouraging results [4, 8, 13, 25].

The correct quantification of femoral torsion is of utmost importance for diagnosing femoral torsional deformities, for making an indication for surgical correction osteotomy, and for planning the amount of correction needed [7, 17]. Differences among the various methods described to measure femoral torsion have been shown in cadavers; however, these studies have neither included symptomatic hips with excessively high femoral torsion nor did they analyze the effect of high femoral torsion on the resulting differences among different measurement methods [1, 12]. In practice, we have seen that patients with excessive femoral torsion and posterior ischiofemoral FAI appear to have differing degrees of femoral torsion depending on the method we have used to measure femoral torsion. More specifically, we have found that the method of Murphy et al. [19] typically produces the highest values for femoral torsion, but this differs from previously published results [12].

Therefore, we asked: (1) How does CT-based measurement of femoral torsion differ among five commonly used measurement methods? (2) Do differences in femoral torsion among measurement methods increase in hips with excessive femoral torsion? (3) What is the reliability and reproducibility of each of the five torsion measurement methods?

## Patients and Methods

After institutional review board approval, we performed a retrospective diagnostic study to compare five commonly used assessment measures to measure femoral torsion on a CT scan, those of Lee et al. [16], Reikerås et al. [22], Jarrett et al. [11], Tomczak et al. [28], and Murphy et al. [19], in a series of symptomatic patients. Between March and August 2016, we saw 86 new patients (95 hips) with hip pain and physical findings suggestive of FAI at our outpatient tertiary clinic. These physical findings include a positive anterior and/or posterior impingement test [24] and restricted ROM of the hip (typically restricted internal rotation with anterior impingement and restricted external rotation with posterior impingement). Furthermore an in-/out-toeing gait can be found in hips with femoral malrotation [31]. Thirty patients (33 hips) without a CT scan were excluded.

Fifty-six patients (62 hips) with suspected complex osseous deformities had a pelvic CT including the distal femur for measurement of femoral torsion and underwent virtual impingement analysis using collision detection software [20, 21, 27]. Seven patients (seven hips) who had undergone previous surgery, two patients (two hips) with a sequelae of Legg-Calvé-Perthes disease, and one patient (one hip) with a posttraumatic hip deformity were excluded. This left 46 patients (52 hips) in the final study group. The mean age in the patient series was  $28 \pm 9$  years (range, 17–51 years) with 27 female patients (59%).

Three-dimensional CT was performed (Somatom Definition Flash; Siemens Medical Solutions, Erlangen, Germany). The scanned volume included the entire pelvis and the distal femoral condyles using the following parameters: voltage 100 kVp; intensity 160 mAs; pitch 0.8; field of view 39 cm; voxel size  $1 \text{ mm}^3$ ; and reconstruction kernels I31f and I70f. Patients were positioned supine, with both feet taped together, resulting in  $10^\circ$  to  $15^\circ$  of internal hip rotation.

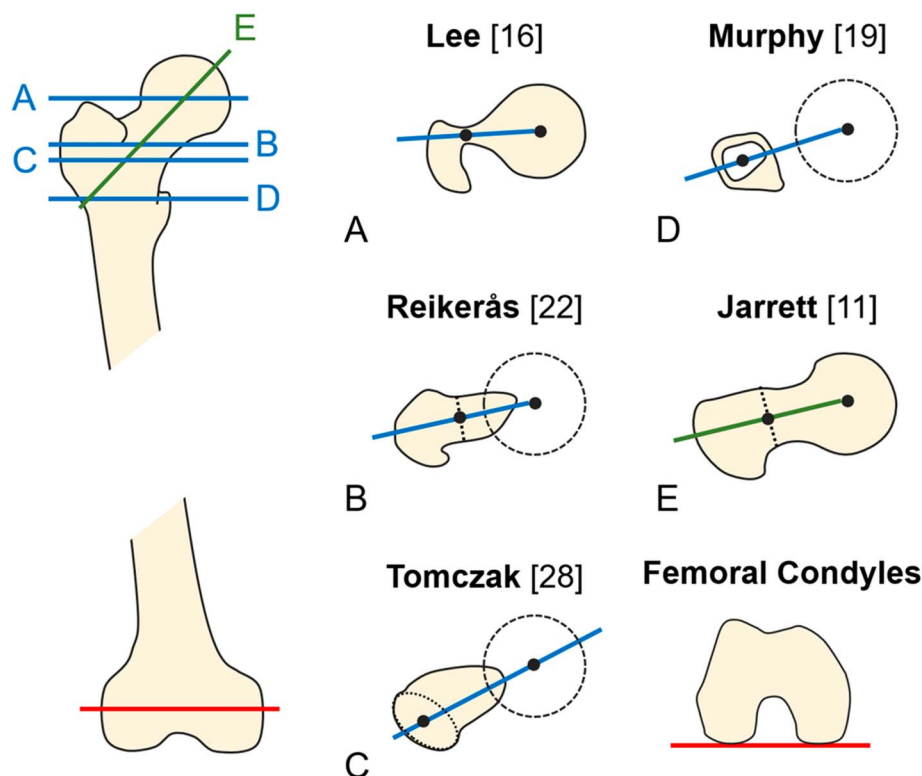
To determine how CT-based measurement of femoral torsion differs among the five measurement methods, one of the authors (FS), an orthopaedic resident with 5 years of experience in diagnostic imaging of the hip and who was not involved in the surgical treatment of the patients, measured femoral torsion in all 46 patients (52 hips) using five methods (Table 1). All measurements were performed

on 1-mm thick slices of the pelvis and the femoral condyles. Similar to our clinical practice, we selected the method described by Murphy et al. as our baseline for measurement because it reportedly most closely reflects true femoral torsion [1, 19] and found a relatively high mean femoral torsion of  $28^\circ \pm 13^\circ$ . Using this measurement method, the lowest value of femoral torsion was  $5^\circ$  and the highest value was  $63^\circ$ . Four of those measurement methods [16, 19, 22, 28] are based on true axial images, whereas the method according to Jarrett et al. [11] is based on reformatted axial-oblique images that are aligned to the femoral neck axis (Fig. 1). The femoral head center is used as a proximal landmark for all methods. The definition of the proximal femoral neck axis differs among the five methods. According to the methods of Lee et al. [16] (axial slices) or Jarrett et al. [11] (axial-oblique slices), the proximal femoral neck axis is defined on the same slice as the femoral head center (Table 1). The other three methods use two superimposed and more distal axial slices to define the proximal femoral neck axis (Fig. 1). The method of Reikerås et al. [22] defines this landmark at the level of the femoral neck where the anterior and posterior cortices run parallel. Tomczak et al. [28] uses the center of the greater trochanter at the base of the femoral neck. Murphy et al. [19] has defined the proximal femoral neck axis at the level just superior to the lesser trochanter (Fig. 1). The distal femoral reference was a line connecting the medial and

**Table 1.** Definitions of the five methods for CT-based measurement of femoral torsion are summarized

Methods	Slice orientation	Single/superimposed image(s)	Definition of landmarks for proximal femoral reference line
Lee et al. [16]	Axial	Single	A line is drawn on the first image on which the femoral head center can be connected with the most cephalic junction of the greater trochanter and the femoral neck
Reikerås et al. [22]	Axial	Two superimposed images	A line connecting the femoral head center with the femoral neck center is drawn on an image where the anterior and posterior cortices run parallel to each other
Jarrett et al. [11]	Axial-oblique	Single	A line is drawn on a single image that runs from the femoral head center through the center of the femoral neck
Tomczak et al. [28]	Axial	Two superimposed images	The femoral head center is connected with the center of the greater trochanter at the base of the femoral neck
Murphy et al. [19]	Axial	Two superimposed images	The femoral head center is connected with the center of the base of the femoral neck directly superior to the lesser trochanter

Distal femoral reference was a line connecting the medial and lateral posterior condyles in all five definitions of torsion measurement; see also Fig. 1 for schematic illustration of each method.



**Fig. 1** Schematic drawings show the five different methods that were used for CT-based measurement of femoral torsion. The blue lines indicate the methods that are based on true axial slices (A–D); the green line indicates the method (E) that is based on axial-oblique images. All five methods used the femoral head center and the posterior distal femoral condyles as landmarks. They differed regarding the definition of the proximal femoral reference axis, which gradually moves from proximal (A) to distal (D). For a detailed definition of each measurement method, see also Table 1.

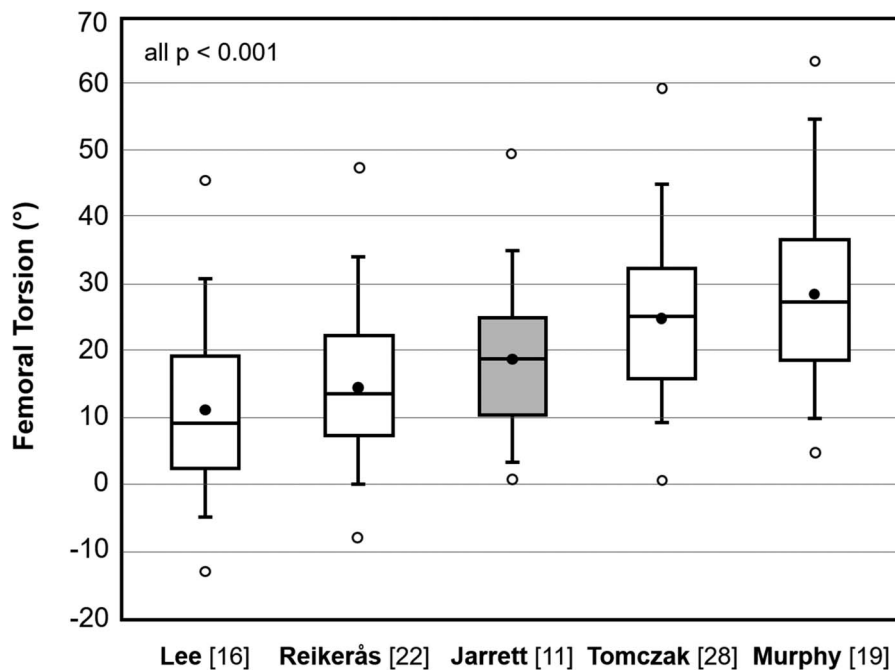
lateral posterior condyles in all five methods (Fig. 1). The means of femoral torsion were compared among all five measurement methods.

To determine whether the difference in femoral torsion among measurement methods increases in hips with excessive femoral torsion, the difference between two measurement methods was plotted over the entire range of femoral antetorsion from 5° to 63°, which had been found for the baseline measurement performed according to Murphy et al. [19]. We evaluated whether the difference in femoral torsion methods increases with increasing femoral torsion. To convert one of the five methods of femoral torsion measurement into another, we provided the linear regression equations.

To assess the intraobserver reliability and interobserver reproducibility of each of the five torsion measurement methods, all the measurements were performed twice by two observers (FS, TDL), neither of whom were involved in the surgical treatment of the patients, with at least 5 years of experience in diagnostic hip assessment. Image analysis was performed independently from each other at two

different timepoints at a 3-month interval and blinded to patients' names and identifying features to minimize recall bias.

We confirmed normal distribution with the Kolmogorov-Smirnov test. To assess differences between measurement methods, the five definitions to measure femoral torsion were compared using multivariate analysis of variance. For pairwise comparison of two measurement methods, we used the pairwise t-test. Bonferroni correction for five groups was used ( $p$  adjusted =  $0.05/5 = 0.01$ ). To assess whether differences between techniques increased with increased torsion, linear regression analysis was performed for pairwise comparison of the measurement methods. The differences between two methods were plotted against the femoral torsion measured using the method of Murphy as a baseline reference [19]. Linear regression analysis was performed to compare the difference of two methods with the method of Murphy et al. [19] or to directly compare two methods of femoral torsion measurement. The Pearson correlation coefficient was graded as:  $R < 0.2$  for very weak; 0.20 to 0.39 for weak;



**Fig. 2** The five methods for CT-based measurement of torsion differed ( $p < 0.001$ ; white box plots for definitions on axial slices; gray box plot for definition of axial-oblique slices). The lowest values of torsion were found for the Lee method, which had the most proximal definition of the landmark for the proximal femoral reference. With the Murphy method, which had the most distal definition of the landmark for the proximal femoral axis, we found the highest values of torsion. All pairwise comparisons showed significant differences (each  $p < 0.001$ ; for detailed results, see also Table 3).

0.40 to 0.59 for moderate; 0.60 to 0.79 for strong; and  $\geq 0.8$  for very strong correlation [6]. To assess intraobserver reliability and interobserver reproducibility, we calculated intraclass correlation coefficients (ICCs) including 95% confidence intervals (CIs). The ICC was graded as ICC  $< 0.20$  for slight; 0.21 to 0.40 for fair; 0.41 to 0.60 for moderate; 0.61 to 0.80 for substantial; and  $> 0.80$  for almost perfect agreement [18]. We used SPSS Version 25.0 (IBM, Armonk, NY, USA) for statistical analysis.

**Table 2.** CT-based measurements of femoral torsion differed among the five definitions for torsion ( $p < 0.001$ )

Methods	Slice orientation	Femoral torsion (°)
Lee et al. [16]	Axial	$11 \pm 11$ (-13 to 45)
Reikerås et al. [22]	Axial	$15 \pm 11$ (-8 to 47)
Jarrett et al. [11]	Axial-oblique	$19 \pm 11$ (1–49)
Tomczak et al. [28]	Axial	$25 \pm 12$ (0–60)
Murphy et al. [19]	Axial	$28 \pm 13$ (5–63)

Notably negative values of femoral torsion (that is, retro-torsion) were found for the Lee et al. [16] and Reikerås et al. [22] methods, whereas no negative values of femoral torsion were measured with the remaining methods; values are expressed as mean  $\pm$  SD with range in parentheses.

## Results

We found increasing values for femoral torsion using methods with a more distal definition of the proximal femoral neck axis (Fig. 2). Comparing the four methods for measurement of torsion on axial slices, the lowest values of femoral torsion were found for the method of Lee et al. [16] with a mean torsion of  $11^\circ \pm 11^\circ$  (Table 2). The highest values of femoral torsion were found for the Murphy et al. method [19] with the most distal definition for the proximal femoral neck axis and a mean torsion of  $28^\circ \pm 13^\circ$  (Table 2). The most pronounced difference between two measurement methods was found for the comparison between the methods of Lee et al. [16] and Murphy et al. [19] with a mean difference of  $17^\circ \pm 5^\circ$  (95% CI,  $16^\circ$ – $19^\circ$ ;  $p < 0.001$ ; Table 3). On axial-oblique slices, the method according to Jarrett et al. [11] showed a mean femoral torsion of  $19^\circ \pm 11^\circ$  (Table 2). Compared with the other methods on axial slices, this method showed an increased torsion compared with Lee et al. [16] ( $11^\circ \pm 11^\circ$ ; difference:  $8^\circ \pm 5^\circ$ , 95% CI,  $6^\circ$ – $9^\circ$ ;  $p < 0.001$ ) or Reikerås et al. [22] ( $15^\circ \pm 11^\circ$ ; difference:  $3^\circ \pm 5^\circ$ , 95% CI,  $2^\circ$ – $5^\circ$ ;  $p < 0.001$ ) and decreased torsion compared with Tomczak et al. [28] ( $25^\circ \pm 12^\circ$ ; difference:  $7^\circ \pm 5^\circ$ , 95% CI,  $5^\circ$ – $8^\circ$ ;  $p <$

**Table 3.** Pairwise comparison of the five measurement methods

Methods	Lee [16]	Reikerås [22]	Jarrett [11]	Tomczak [28]	Murphy [19]
Lee et al. [16]	—	4° ± 3° (4°–5°)	8° ± 5° (6°–9°)	14° ± 3° (13°–15°)	17° ± 5° (16°–19°)
Reikerås et al. [22]	p < 0.001	—	3° ± 5° (2°–5°)	10° ± 3° (9°–11°)	13° ± 5° (12°–14°)
Jarrett et al. [11]	p < 0.001	p < 0.001	—	7° ± 5° (5°–8°)	10° ± 5° (8°–11°)
Tomczak et al. [28]	p < 0.001	p < 0.001	p < 0.001	—	3° ± 4° (2°–4°)
Murphy et al. [19]	p < 0.001	p < 0.001	p < 0.001	p < 0.001	—

In the upper right-hand corner, the values represent the mean difference of femoral torsion ± SD with 95% confidence intervals in parentheses; in the lower left-hand corner, the values represent the level of significance for each comparison; the dashes represent the cells for which no pairwise comparison was possible because they represent the same measurement methods.

0.001) or Murphy et al. [19] (28° ± 13°; difference: 10° ± 5°, 95% CI, 8°–11°; p < 0.001) (Table 3).

For six of 10 possible pairwise comparisons, the difference between two methods increased with increasing femoral torsion and decreased with decreasing femoral torsion (Table 4). A strong correlation was found for the differences between Murphy et al. [19] and Reikerås et al. [22] as well as between Murphy et al. [19] and Jarrett et al. [11] (R of 0.622 and 0.619, respectively; both p < 0.001; Fig. 3). A moderate correlation was found for the differences between Murphy et al. [19] and Lee et al. [16] (R = 0.534; p < 0.001); Murphy et al. [19] and Tomczak et al. [28] (R = 0.493; p < 0.001) as well as Tomczak et al. [28] and Reikerås et al. [22] (R = 0.426; p < 0.001). Correlation was fair for Tomczak et al. [28] and Jarrett et al. [11] (R = 0.306; p = 0.014). The method of Murphy et al. [19] showed an increasing difference compared with any of the other four methods with increasing femoral torsion (Table 4). For example, in a hip with normal femoral torsion of 10°, the difference between the methods of Murphy et al. [19] and Reikerås et al. [22] was 9° (Fig. 3A). The

difference in the same hip increased to 14° when we compared the methods of Murphy et al. [19] and Lee et al. [16] (Fig. 3B). In contrast, in a hip with an excessive femoral torsion of 60°, the difference between methods of Murphy et al. [19] and Reikerås et al. [22] increased to 20° (Fig. 3A). The difference in the same hip increased to 23° when we compared Murphy et al. [19] and Lee et al. [16] (Fig. 3B).

We found excellent agreement (ICC > 0.800 [18]) for both intraobserver reproducibility (ICC range, 0.905–0.973) and interobserver reliability (ICC range, 0.938–0.969) for all five CT-based definitions of femoral torsion (Table 5).

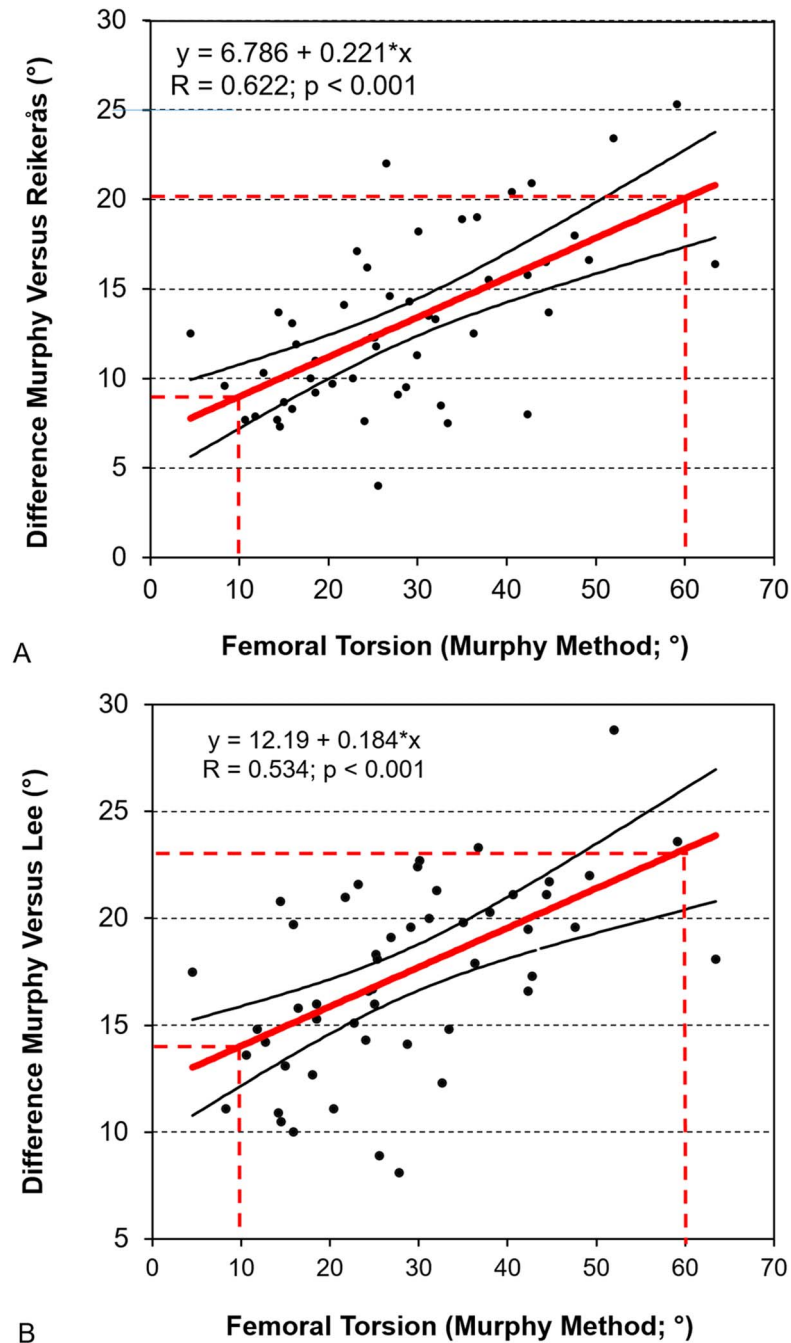
## Discussion

The correct CT-based quantification of femoral torsion is essential for surgical decision-making and planning of corrective osteotomies [29]. However, different measurement methods have to our knowledge not been compared

**Table 4.** Pairwise comparison of the five different methods to measure femoral torsion correlated with femoral torsion using linear regression analysis

Correlation between differences in two methods versus Murphy method	Equation	R	p value
Murphy et al. [19] – Reikerås et al. [22]	y = 6.786 + 0.221*x	0.622	< 0.001
Murphy et al. [19] – Jarrett et al. [11]	y = 2.790 + 0.242*x	0.619	< 0.001
Murphy et al. [19] – Lee et al. [16]	y = 12.191 + 0.184*x	0.534	< 0.001
Murphy et al. [19] – Tomczak et al. [28]	y = -0.664 + 0.134*x	0.493	< 0.001
Tomczak et al. [28] – Reikerås et al. [22]	y = 7.449 + 0.087*x	0.426	< 0.001
Tomczak et al. [28] – Jarrett et al. [11]	y = 3.453 + 0.108*x	0.306	0.014
Tomczak et al. [28] – Lee et al. [16]	y = 12.854 + 0.050*x	0.218	0.060
Reikerås et al. [22] – Lee et al. [16]	y = 5.405 – 0.037*x	-0.180	0.100
Jarrett et al. [11] – Lee et al. [16]	y = 9.401 – 0.058*x	-0.156	0.134
Reikerås et al. [22] – Jarrett et al. [11]	y = 3.996 – 0.021*x	-0.059	0.337

Six of 10 comparisons showed a significant linear correlation and all six showed an increasing difference between the two measurement methods with increasing femoral torsion (positive correlation coefficient); the strongest correlation was found for the difference between the methods of Reikerås et al. and Murphy et al. (see Fig. 3 for the scatterplot); with every degree of femoral torsion, the difference between these two methods increased by 0.22°.



**Fig. 3 A-B** With increasing femoral torsion, the difference between two measurement methods increased for six of 10 possible pairwise comparisons (for detailed information, see Table 4). **(A)** The strongest correlation was found for the difference between the methods of Reikerås et al. [21] and Murphy et al. [18] ( $R = 0.622; p < 0.001$ ). In a hip with normal femoral torsion of  $10^\circ$ , the difference between the Reikerås et al. [21] and the Murphy et al. [18] method was  $9^\circ$ . In a hip with excessive femoral torsion of  $60^\circ$ , the difference between these two measurement methods increased to  $20^\circ$ . **(B)** Moderate correlation was found for the difference between the methods of Lee et al. [14] and Murphy et al. [18] ( $R = 0.534; p < 0.001$ ). In a hip with normal femoral torsion of  $10^\circ$ , the difference between the Lee et al. [14] and the Murphy et al. [18] methods was  $14^\circ$ . In a hip with excessive femoral torsion of  $60^\circ$ , the difference between these two measurement methods increased to  $23^\circ$ .

**Table 5.** Results of reproducibility and reliability for the five measurement methods

Methods	Reproducibility (ICC reader 1)	Reproducibility (ICC reader 2)	Reliability (ICC reader 1 versus reader 2)
Lee et al. [16]	0.970 (0.948–0.983)	0.970(0.948–0.982)	0.969 (0.947–0.982)
Reikerås et al. [22]	0.962 (0.934–0.978)	0.948 (0.912–0.970)	0.939 (0.897–0.965)
Jarrett et al. [11]	0.928 (0.878–0.958)	0.905 (0.839–0.944)	0.938 (0.894–0.964)
Tomczak et al. [28]	0.973 (0.954–0.984)	0.947 (0.910–0.969)	0.946 (0.908–0.969)
Murphy et al. [19]	0.958 (0.928–0.976)	0.962 (0.934–0.978)	0.964 (0.939–0.979)

Values are expressed as mean with 95% confidence interval in parentheses; ICC = intraclass correlation coefficient.

in a group of symptomatic patients who have a wide range of femoral torsion. Consequently, the effect of increased femoral torsion on the resulting differences between measurement methods has not been defined. We therefore compared five commonly used methods to measure CT-based femoral torsion. We found considerable differences between measurement methods with a maximum mean difference up to 17° (Fig. 2). Femoral torsion measured at the most proximal landmark for definition of the neck axis such as with the method of Lee et al. [16] yielded lower values compared with the methods with a more distal definition of the landmark such as the Murphy et al. method [19] (Fig. 1). In addition, we observed that differences between measurement methods rose with increasing femoral torsion; for example, the difference in femoral torsion between the method of Murphy et al. and Lee et al. increased from 14° (in a hip with normal torsion of 10°) to 23° (in a hip with excessive torsion of 60°; Fig. 3B). Thus, especially in hips with excessively high femoral torsion, it is not sufficient to just add differences of mean femoral torsion between two measurement techniques for conversion of torsion values. However, we found a high linear correlation (correlation coefficient R ranging from 0.904 to 0.975; Table 6) among all 10 possible pairwise comparisons for the five evaluated methods for measurement of

femoral torsion (Table 6). Therefore, these linear regression models can be used to convert femoral torsion measurements from one definition to another. Finally, we found that variations among two readers and differences on repeated measurements were negligible (Table 5).

This study has a number of limitations. First, the range of femoral torsion from 5° to 63° (according to the Murphy et al. method [19]) included in the current study does not cover the entire possible range of femoral torsion. A range of 100° of femoral torsion (-16° to 84° measured according to the Murphy et al. method) [19] has been reported [17]. Therefore, we could have missed the most extreme forms of excessive femoral antetorsion or retrotorsion. Regardless, mean femoral torsion as measured with the Murphy method [19] was 28° ± 13°, which is higher than the reported normal range of 10° to 25° [29], and thus we believe it was adequate for answering our study questions. Second, no evaluation of femoral torsion measurement has been performed on MR images. MRI is increasingly used in clinical routine to reduce radiation in the typically young patients who are eligible for joint-preserving surgery [10, 23, 26]. We note that MR-based measurements of femoral torsion showed a high reproducibility and reliability in one study, comparable with CT-based measurements [10]. However, CT is still considered the gold standard for

**Table 6.** Conversion of one method for femoral torsion measurement into another using linear regression analysis

Original method of torsional measurement (X)	Designated method of torsional measurement (Y)	Equation	R	p value
Murphy et al. [19]	Tomczak et al. [28]	$y = 0.664 + 0.866 \cdot x$	0.964	< 0.001
Murphy et al. [19]	Reikerås et al. [22]	$y = -6.786 + 0.779 \cdot x$	0.941	< 0.001
Murphy et al. [19]	Lee et al. [16]	$y = -12.191 + 0.816 \cdot x$	0.942	< 0.001
Murphy et al. [19]	Jarrett et al. [11]	$y = -2.790 + 0.758 \cdot x$	0.927	< 0.001
Tomczak et al. [28]	Reikerås et al. [22]	$y = -7.364 + 0.899 \cdot x$	0.975	< 0.001
Tomczak et al. [28]	Lee et al. [16]	$y = -12.590 + 0.934 \cdot x$	0.967	< 0.001
Tomczak et al. [28]	Jarrett et al. [11]	$y = -2.427 + 0.838 \cdot x$	0.920	< 0.001
Jarrett et al. [11]	Reikerås et al. [22]	$y = -1.906 + 0.920 \cdot x$	0.909	< 0.001
Jarrett et al. [11]	Lee et al. [16]	$y = -6.975 + 0.958 \cdot x$	0.904	< 0.001
Reikerås et al. [22]	Lee et al. [16]	$y = -4.625 + 1.018 \cdot x$	0.972	< 0.001



assessing femoral torsion [2]. The five evaluated methods to measure femoral torsion may be used as well on MRI. It would seem plausible that our finding of increasing differences in femoral torsion values resulting from different measurement methods in hips with excessive femoral torsion apply as well for MRI-based measurements. However, this assumption has to be tested in future studies. Third, not all known methods for measuring femoral torsion have been included for evaluation in the current study. Often, no detailed description was available in the original reports or only small variations of a method have been described [9, 30, 32]. Fourth, we cannot rule out selection bias because our results are based on a predominantly female western European patient population seen at a tertiary hip preservation center and did not include patients with more complex deformities such as sequelae of slipped capital femoral epiphysis or Legg-Calvé-Perthes disease. Although overall differences in femoral torsion related to sex and ethnicity may be negligible according to a recent study [14], our findings should be reproduced first, before being directly applied to patients with pediatric hip deformities.

All five evaluated measurement methods for femoral torsion differed (Fig. 2). We found that femoral torsion was dependent on the definition of the proximal femoral neck axis; the more distal this landmark was defined (Fig. 1), the higher was the measured femoral torsion (Table 1). Therefore, the largest difference was found between the methods of Lee et al. [16] (the most proximal definition of the femoral axis) and Murphy et al. [19] (the most distal definition) with a mean difference of  $17^\circ$  (Table 3). In the original description of Murphy et al. [19], two measurement methods of femoral torsion using a single axial image and two superimposed images were compared. Both methods were evaluated at different levels and compared with the true anatomic femoral torsion in cadaveric femora. At the level of the femoral head center, the method with a single axial image (similar to the method later described by Lee et al. [16]) underestimated torsion by a mean of  $18^\circ$  [19]. By selecting more distal axial cuts at the level of the neck, the underestimation of femoral torsion decreased gradually. The femoral torsion closest to the anatomic value was found for the method with two superimposed images, one at the level of the femoral head center and another 6 mm proximal to the center of the lesser trochanter (axial slice including typically the most proximal part of the lesser trochanter in the dorsal aspect; Fig. 1) [19]. Berryman et al. [1] reported comparable results. They observed the highest values of femoral torsion with the method of Murphy et al. [19] and the lowest values with the method of Lee et al. [16] and reported a mean difference of  $15^\circ$  [1]. By contrast, Kaiser et al. [12] did not report the highest values of femoral torsion using the method according to Murphy et al. [19]. Instead, the method of Waidelich et al. [30] (comparable to the method of Tomczak et al. [28] used in

the current evaluation [Table 1]), in which the proximal femoral neck axis is defined as the center of the greater trochanter, yielded the highest values of femoral torsion. This difference may be related to the fact that the slice for the method of Murphy et al. [19] was chosen more proximal as shown in the published figures [12]. According to Murphy et al. [19], femoral torsion should be measured directly superior to the center of the lesser trochanter, typically with the proximal part of the lesser trochanter in the dorsal aspect of the cross-section to minimize the difference to true anatomic femoral torsion. This was recently confirmed in a study that used a computerized neck-fitting method and showed that the Murphy et al. [19] method most closely reflects anatomic femoral torsion compared with more proximal measurement methods [1].

With increasing femoral torsion, we found increasing differences between the method of Murphy et al. [19] and any other measurement method (Fig. 3). In addition, differences between the Tomczak et al. [28] method, the second most distal method, and the more proximal methods of Reikerås et al. [22] or Jarrett et al. [11] rose with increasing torsion (Table 4). Conversely, these differences declined with decreasing femoral torsion; for example, in a hip with a normal femoral torsion of  $10^\circ$  as measured according to Murphy et al. [19], the resulting difference of femoral torsion compared with the methods of Reikerås et al. [22] or Lee et al. [16] was  $9^\circ$  and  $14^\circ$ , respectively (Fig. 3). Notably, this is less than the observed mean difference between two methods:  $13^\circ$  between Murphy et al. [19] and Reikerås et al. [22] and  $17^\circ$  between Murphy et al. [19] and Lee et al. [16] (Table 3). By simply adding this mean difference to the measurements of Reikerås et al. [22] or Lee et al. [16] would result in an overestimation of  $4^\circ$  and  $3^\circ$ , respectively, in a hip with  $10^\circ$  of femoral torsion according to Murphy et al. [19]. In a hip with excessive femoral torsion of  $60^\circ$ , this difference increased to  $20^\circ$  for the method of Reikerås et al. [22] and  $23^\circ$  for the method of Lee et al. [16] (Fig. 3A-B). In this hip with excessive femoral torsion, these differences exceed the mean difference between two methods ( $13^\circ$  between Murphy et al. [19] and Reikerås et al. [22] and  $17^\circ$  between Murphy et al. [19] and Lee et al. [16]). Therefore, by simply adding the mean difference to the measurements of Reikerås et al. [22] or Lee et al. [16] would result in an underestimation of  $7^\circ$  and  $6^\circ$ , respectively, in a hip with  $60^\circ$  of femoral torsion according to Murphy et al. [19]. This effect has to be considered, especially when planning the degree of torsional correction, which is usually performed in hips with severe torsional deformities [4, 8, 13, 25, 29]. Because a strong linear correlation ( $R > 0.904$ ,  $p < 0.001$ ; Table 6) was observed among all 10 possible pairwise comparisons, conversion of one measurement method to another is possible (Table 6). However, in a clinical setting, we recommend the use of a consistent method for measuring

femoral torsion and use the corresponding reported normal values and the respective pathologic thresholds to which the surgeon refers. Also, one should always state which measurement method was used when reporting femoral torsion. Simply adding a mean difference of torsion between two methods can result in overestimation in a hip with low torsion or underestimation in a hip with high torsion when converting to the Murphy et al. method [19].

Agreement was excellent for each of the five evaluated measurement methods for both reproducibility (ICC range, 0.905–0.973) and reliability (ICC range, 0.938–0.969; Table 5). This is in accordance with previously published results in cadavers that showed excellent reproducibility (ICC range, 0.88–0.99) and reliability (ICC range, 0.91–0.99) for different methods to measure femoral torsion [12] among two trauma surgeons and two radiologists. Thus, errors related to person and time are negligible when measuring femoral torsion.

In summary, we found that femoral torsion differed considerably depending on the measurement method used. We found that the differences depend on the definition of the anatomic landmark for the proximal reference line; the more distal this landmark was defined, the higher the measured femoral torsion. In contrast, reproducibility and reliability were excellent, suggesting that errors related to person and time do not reflect a relevant source of error between measurement techniques. Increased differences between different measurement methods were found in hips with excessive femoral torsion. In these hips, the applied measurement method may play an even more important role. One has to keep in mind that correction among various methods just by adding reported mean differences will result in a discrepancy of femoral torsion. Neglecting the differences among methods for measuring femoral torsion can potentially lead to misdiagnosis and errors in surgical planning. Especially in hips with an excessive femoral torsion potentially undergoing femoral derotation osteotomy, the differences among the methods to quantify femoral torsion are increased. It is crucial to state the used measurement methods when reporting femoral torsion and to be consistent regarding the used measurement method.

## References

- Berryman F, Pynsent P, McBryde C. A semi-automated method for measuring femoral shape to derive version and its comparison with existing methods. *Int J Numer Method Biomed Eng*. 2014; 30:1314–1325.
- Botser IB, Ozoude GC, Martin DE, Siddiqi AJ, Kuppuswami S, Domb BG. Femoral anteversion in the hip: comparison of measurement by computed tomography, magnetic resonance imaging, and physical examination. *Arthroscopy*. 2012;28:619–627.
- Bouma HW, Hogervorst T, Audenaert E, Krekel P, van Kampen PM. Can combining femoral and acetabular morphology parameters improve the characterization of femoroacetabular impingement? *Clin Orthop Relat Res*. 2015;473:1396–1403.
- Buly RL, Sosa BR, Poultsides LA, Caldwell E, Rozbruch SR. Femoral derotation osteotomy in adults for version abnormalities. *J Am Acad Orthop Surg*. 2018;26:e416–e425.
- Chadayammuri V, Garabekyan T, Bedi A, Pascual-Garrido C, Rhodes J, O'Hara J, Mei-Dan O. Passive hip range of motion predicts femoral torsion and acetabular version. *J Bone Joint Surg Am*. 2016;98:127–134.
- Evans J. *Straightforward Statistics for the Behavioral Sciences*. 1<sup>st</sup> ed. Pacific Grove, CA, USA: Brooks/Cole Publishing; 1996.
- Fabricant PD, Fields KG, Taylor SA, Magennis E, Bedi A, Kelly BT. The effect of femoral and acetabular version on clinical outcomes after arthroscopic femoroacetabular impingement surgery. *J Bone Joint Surg Am*. 2015;97:537–543.
- Hartigan DE, Perets I, Walsh JP, Domb BG. Femoral derotation osteotomy technique for excessive femoral anteversion. *Arthrosc Tech*. 2017;6:e1405–e1410.
- Hernandez RJ, Tachdjian MO, Poznanski AK, Dias LS. CT determination of femoral torsion. *AJR Am J Roentgenol*. 1981;137: 97–101.
- Hesham K, Carry PM, Freese K, Kestel L, Stewart JR, Delavan JA, Novais EN. Measurement of femoral version by MRI is as reliable and reproducible as CT in children and adolescents with hip disorders. *J Pediatr Orthop*. 2017;37:557–562.
- Jarrett DY, Oliveira AM, Zou KH, Snyder BD, Kleinman PK. Axial oblique CT to assess femoral anteversion. *AJR Am J Roentgenol*. 2010;194:1230–1233.
- Kaiser P, Attal R, Kammerer M, Thauerer M, Hamberger L, Mayr R, Schmoelz W. Significant differences in femoral torsion values depending on the CT measurement technique. *Arch Orthop Trauma Surg*. 2016;136:1259–1264.
- Kamath AF, Ganz R, Zhang H, Grappiolo G, Leunig M. Subtrochanteric osteotomy for femoral mal-torsion through a surgical dislocation approach. *J Hip Preserv Surg*. 2015;2:65–79.
- Koerner JD, Patel NM, Yoon RS, Sirkin MS, Reilly MC, Liporace FA. Femoral version of the general population: does 'normal' vary by gender or ethnicity? *J Orthop Trauma*. 2013;27: 308–311.
- Kraeutler MJ, Chadayammuri V, Garabekyan T, Mei-Dan O. Femoral version abnormalities significantly outweigh effect of cam impingement on hip internal rotation. *J Bone Joint Surg Am*. 2018;100:205–210.
- Lee YS, Oh SH, Seon JK, Song EK, Yoon TR. 3D femoral neck anteversion measurements based on the posterior femoral plane in ORTHODOC system. *Med Biol Eng Comput*. 2006;44: 895–906.
- Lerch TD, Todorski IAS, Steppacher SD, Schmaranzer F, Werlen SF, Siebenrock KA, Tannast M. Prevalence of femoral and acetabular version abnormalities in patients with symptomatic hip disease: a controlled study of 538 hips. *Am J Sports Med*. 2018;46:122–134.
- Montgomery AA, Graham A, Evans PH, Fahey T. Inter-rater agreement in the scoring of abstracts submitted to a primary care research conference. *BMC Health Serv Res*. 2002;2:8.
- Murphy SB, Simon SR, Kijewski PK, Wilkinson RH, Griscom NT. Femoral anteversion. *J Bone Joint Surg Am*. 1987;69: 1169–1176.
- Puls M, Ecker TM, Steppacher SD, Tannast M, Siebenrock KA, Kowal JH. Automated detection of the osseous acetabular rim using three-dimensional models of the pelvis. *Comput Biol Med*. 2011;41:285–291.
- Puls M, Ecker TM, Tannast M, Steppacher SD, Siebenrock KA, Kowal JH. The equidistant method—a novel hip joint simulation algorithm for detection of femoroacetabular impingement. *Comput Aided Surg*. 2010;15:75–82.

22. Reikerås O, Bjerkreim I, Kolbenstvedt A. Anteversion of the acetabulum and femoral neck in normals and in patients with osteoarthritis of the hip. *Acta Orthop Scand*. 1983;54:18–23.
23. Schmaranzer F, Todorski IAS, Lerch TD, Schwab J, Cullmann-Bastian J, Tannast M. Intra-articular lesions: imaging and surgical correlation. *Semin Musculoskelet Radiol*. 2017;21:487–506.
24. Siebenrock KA, Steppacher SD, Haefeli PC, Schwab JM, Tannast M. Valgus hip with high antetorsion causes pain through posterior extraarticular FAI. *Clin Orthop Relat Res*. 2013;471:3774–3780.
25. Suren C, Burgkart R, Banke IJ, Hertel G, Schauwecker J, von Eisenhart-Rothe R, Gollwitzer H. [Surgical therapy of ischiofemoral impingement by lateralizing intertrochanteric osteotomy] [in German]. *Oper Orthop Traumatol*. 2018;30:98–110.
26. Sutter R, Dietrich TJ, Zingg PO, Pfirmann CWA. Femoral antetorsion: comparing asymptomatic volunteers and patients with femoroacetabular impingement. *Radiology*. 2012;263:475–483.
27. Tannast M, Kubiak-Langer M, Langlotz F, Puls M, Murphy SB, Siebenrock KA. Noninvasive three-dimensional assessment of femoroacetabular impingement. *J Orthop Res*. 2007;25:122–131.
28. Tomczak RJ, Guenther KP, Rieber A, Mergo P, Ros PR, Brambs HJ. MR imaging measurement of the femoral antetorsional angle as a new technique: comparison with CT in children and adults. *AJR Am J Roentgenol*. 1997;168:791–794.
29. Tönnis D, Heinecke A. Acetabular and femoral anteversion: relationship with osteoarthritis of the hip. *J Bone Joint Surg Am*. 1999;81:1747–1770.
30. Waidelich HA, Strecker W, Schneider E. [Computed tomographic torsion-angle and length measurement of the lower extremity. The methods, normal values and radiation load] [in German]. *Rofo*. 1992;157:245–251.
31. Westberry DE, Wack LI, Davis RB, Hardin JW. Femoral anteversion assessment: comparison of physical examination, gait analysis, and EOS biplanar radiography. *Gait Posture*. 2018;62:285–290.
32. Yoshioka Y, Cooke TD. Femoral anteversion: assessment based on function axes. *J Orthop Res*. 1987;5:86–91.