SCIENTIFIC ARTICLE

Reliability of the assessment of lower limb torsion using computed tomography: analysis of five different techniques

**Emmanouil Liodakis · Iosifina Doxastaki ·
Kongfai Chu · Christian Krettek · Ralph Gaulke ·
Musa Citak · Mohamed Kenawey**

Abstract

Objective Various methods have been described to define the femoral neck and distal tibial axes based on a single CT image. The most popular are the Hernandez and Weiner methods for defining the femoral neck axis and the Jend, Ulm, and bimalleolar methods for defining the distal tibial axis. The purpose of this study was to calculate the intra- and interobserver reliability of the above methods and to determine intermethod differences.

Methods Three physicians separately measured the rotational profile of 44 patients using CT examinations on two different occasions. The average age of patients was 36.3 ± 14.4 years, and there were 25 male and 19 female patients. After completing the first two sessions of measurements, one observer chose certain cuts at the levels of the femoral neck, femoral condylar area, tibial plateau, and distal tibia. The three physicians then repeated all measurements using these CT cuts.

Results The greatest interclass correlation coefficients were achieved with the Hernandez (0.99 intra- and 0.93 interobserver correlations) and bimalleolar methods (0.99 intra- and 0.92 interobserver correlations) for measuring the femoral neck and distal tibia axes, respectively. A statistically significant decrease in the interobserver median absolute differences could be achieved through the use of predefined CT scans only for measurements of the femoral

condylar axis and the distal tibial axis using the Ulm method. The bimalleolar axis method underestimated the tibial torsion angle by an average of 4.8° and 13° compared to the Ulm and Jend techniques, respectively.

Conclusions The methods with the greatest inter- and intra-observer reliabilities were the Hernandez and bimalleolar methods for measuring femoral anteversion and tibial torsion, respectively. The high intermethod differences make it difficult to compare measurements made with different methods.

Introduction

The accuracy of a method can be defined as how close a measured value is to a true value, for example, how close a CT measurement of femoral anteversion is to the real anteversion measured in a cadaver. A highly accurate method must be one that is reliable when used by different observers (interobserver reliability) or even for the same observer when measurements are repeated one or more times (intraobserver reliability).

The steps and angles used for the analysis of the frontal plane alignment of the lower extremities are well described with a relatively narrow range of normal values and good intraobserver and interobserver reliabilities [1–5]. Contrary to that, major concerns can be easily raised for the assessment of the rotational profile of the lower limbs using CT examination, and this can even be deemed to be a completely unreliable tool [6]. Two major difficulties face every physician who tries to assess rotation using axial CT

E. Liodakis (✉) · I. Doxastaki · K. Chu · C. Krettek · R. Gaulke ·
M. Citak
Trauma Department, Hannover Medical School,
Carl-Neuberg-Str. 1,
30625 Hannover, Germany
e-mail: manoliodakis@yahoo.gr

M. Kenawey
Orthopaedic Surgery Department, Sohag University Hospital,
82524 Sohag, Egypt

Table 1 The methods used for the assessment of the femoral neck axis and the distal tibial axis

Area	Method	Definition
Femoral neck axis	Hernandez et al. (1981)	The femoral neck axis is created by a line passing through the center of the femoral head and the midpoint of the femoral neck in a CT cut where the femoral head, isthmus of the femoral neck, and the superior border of the greater trochanter are evident.
	Weiner et al. (1978)	A line passing through the middle of the neck parallel to the ventral and dorsal cortices is considered to be the femoral neck axis. The CT image needed for this measurement is obtained distal to the femoral head, where the ventral and dorsal cortices are approximately parallel to each other.
Distal tibial axis	Ulm method (1992)	The distal tibial axis is created by a line passing through the middle of the ellipse formed from the incisura fibularis tibiae and the middle of an ellipse formed from the surface of the medial malleolus.
	Jend method (1980)	The line connecting the middle of the line connecting the end points of the incisura fibularis tibiae with the center of a circle created from the junction of tibial pilon und incisura fibularis tibiae is defined as the distal tibial axis.
	Bimalleolar axis (1989)	The distal tibial axis is created by a line connecting the centers of the medial and lateral malleolus.

cuts: the first is defining the femoral neck axis and the second is defining the distal tibial or malleolar axis. Several methods have been proposed for each location with conflicting reports regarding their accuracy compared to anatomic references or three-dimensional CT modeling [7–10]. One other important issue is choosing the level of the cut at which you make your measurements [7]. Moreover, at the present time, no sufficient data are available regarding the intraobserver, interobserver, and intermethod reliability for the famous techniques used in these measurements.

The rotational alignment of the lower extremity can be determined by defining four main axes: the femoral neck axis, the distal femoral condylar axis, and the proximal and distal tibial axes. The objective of the current study was to evaluate the intraobserver and interobserver reliabilities of measurements made at these four different axes. Two popular methods were used to define the femoral neck axis (the Hernandez et al. and Weiner et al. methods), and three methods were used for the distal tibial axis (the Ulm, Jend, and bimalleolar axis

methods) (Table 1 and Figs. 1, 2, and 3) [11–17]. In addition, we evaluated the following questions: (1) Can the interobserver differences be decreased by using predefined CT cuts for all observers? (2) How significant are the intermethod differences (e.g., differences between the measurements made by different methods for either the femoral neck axis or distal tibial axis)?

Patients and methods

Patients

We retrospectively analyzed 44 consecutive torsion difference CTs performed between 2008 and 2010 at our institution. Three physicians had separately measured the lower limb torsion of one healthy extremity. The mean age of these patients at the time of the examination was 36.3± 14.4 years. There were 25 male and 19 female patients. The

Fig. 1 Defining the femoral neck axis. **a** The Hernandez et al. method was used within a cut in the area of the femoral head with the isthmus of the femoral neck and the greater trochanter. The femoral neck axis is the line between the center of the femoral head and that of the isthmus of the neck. **b** The Weiner et al. method was used in a cut with the ventral and dorsal cortices of the femoral neck parallel to each other. The femoral neck axis is the midline between the ventral and dorsal cortices
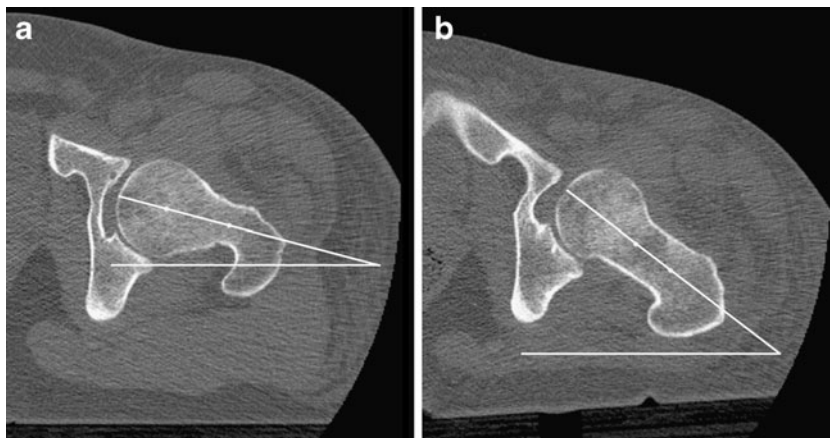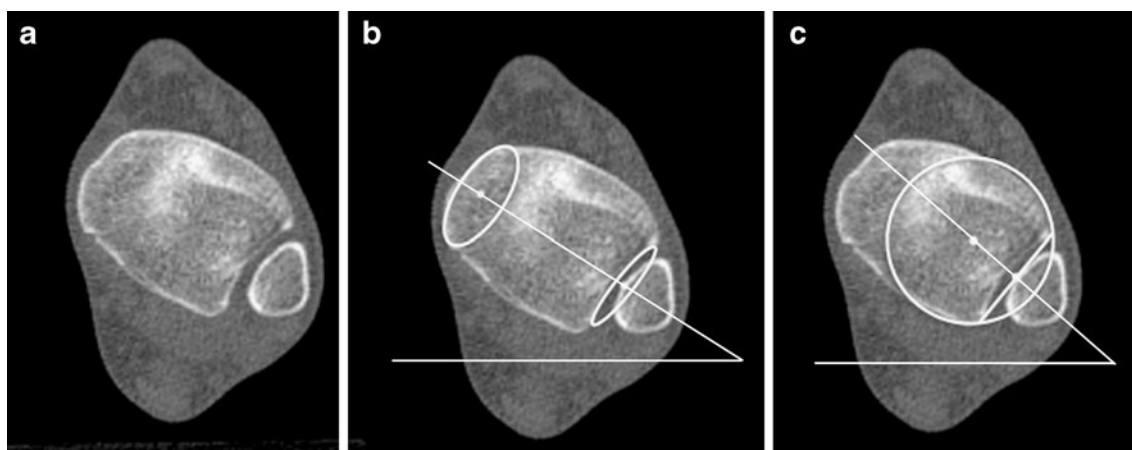
Fig. 2 **a** The distal tibial axis can be defined in a cut in the distal tibia with the fibula articulating in the incisura fibularis. **b** In the Ulm method, the distal tibial axis is drawn between the centers of an ellipse from the surface of the medial malleolus and another ellipse formed by the incisura fibularis. **c** According to the Jend method, the distal tibial axis is defined as the line intersecting the middle of the line connecting the end points of the incisura fibularis of the tibia with the center of a circle created from the junction of the tibial pilon and incisura fibularis

contralateral limb had fractures of the femur and/or tibia and the CT scans were performed on average 3 days after reduction and fixation of the fracture (range 1–15 days). The angles to be measured were the angle of the femoral neck axis, the femoral condylar axis (dorsal tangent to the femoral condyles), the tibial plateau (dorsal tangent to the tibial plateau), and the distal tibial axis. The femoral neck axis was evaluated with two different methods: the methods of Hernandez et al. [11] and Weiner et al. [12]. The distal tibial axis was evaluated with three different methods: the Ulm, Jend, and bimalleolar methods (Table 1 and Figs. 1, 2, and 3) [13, 14, 17, 18].

One physician was an attending surgeon experienced in the area of deformity correction and reconstructive surgery (observer 1), the second was a chief resident of orthopedic

and trauma surgery (observer 2), and the third was a senior house officer (observer 3). The three physicians had a meeting at the start of the study to clarify the methods and objectives of this research and then they worked separately for the measurements. Each CT study was evaluated on two separate occasions by each physician, and there was at least 2 weeks between each set of measurements to avoid memorization of the results. After all observers completed the first two sessions of measurements, one physician was asked to individually review all 44 CTs and to choose certain cuts at the levels of the femoral neck, distal femur, tibial plateau, and distal tibia. These cuts were as follows: one cut based on the Hernandez method and one based on the Weiner method in the area of the femoral neck; one cut in the area of the femoral condyles, one at the level of the
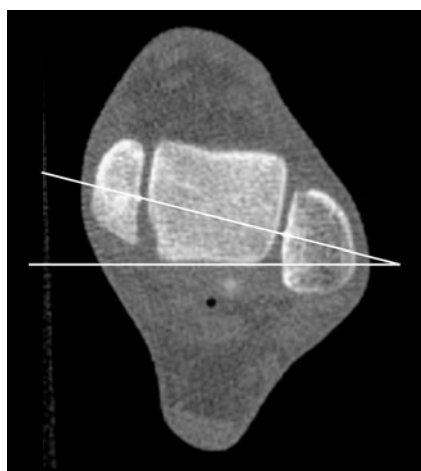

Fig. 3 The bimalleolar axis is drawn in a cut just below the tibial pilon's articular surface with the medial and lateral malleoli and talar dome evident between the centers of the dense surfaces of the malleoli

Table 2 Intraclass correlation coefficient (ICC) scores for intraobserver and interobserver measurements in the first two sessions

	Intraobserver			Interobserver
	Observer 1	Observer 2	Observer 3	
Femoral neck axis (Hernandez)	0.997	0.99	0.986	0.931
Femoral neck axis (Weiner)	0.961	0.949	0.88	0.884
Femoral condylar axis	0.995	0.993	0.991	0.864
Proximal tibial axis	0.997	0.998	0.988	0.95
Distal tibial axis (Ulm)	0.993	0.978	0.983	0.918
Distal tibial axis (Jend)	0.969	0.959	0.942	0.916
Distal tibial axis (bimalleolar axis)	0.998	0.999	0.996	0.92

Table 3 Absolute differences (median and range) for the intraobserver measurements in the first two sessions

	Intraobserver		
	Observer 1	Observer 2	Observer 3
Femoral neck axis (Hernandez)	0.5 (0–2.6)	0.9 (0–5.7)	1.2 (0.1–6.9)
Femoral neck axis (Weiner)	2.4 (0.2–9)	3.1 (0.1–8)	4 (0.1–15)
Femoral condylar axis	0.6 (0–4.6)	0.4 (0–6.6)	0.6 (0–6.6)
Proximal tibial axis	0.7 (0–2.5)	0.5 (0–2.5)	1.5 (0.1–4.7)
Distal tibial axis (Ulm)	0.9 (0.1–6.2)	0.9 (0–7.9)	1.3 (0.1–9.5)
Distal tibial axis (Jend)	1.8 (0–9.4)	3.4 (0.1–10.7)	2.9 (0.1–12)
Distal tibial axis (bimalleolar axis)	0.4 (0–2.2)	0.6 (0–1.4)	0.8 (0–3.6)

tibial plateau, and one at the level of the distal tibia based on both the Ulm and Jend methods; and one cut based on the bimalleolar axis method. The three physicians then repeated all measurements individually in a third session using these defined CT cuts.

This study was performed according to the international guidelines of the Declaration of Helsinki for clinical research.

CT examination

Scans were obtained with LightSpeed QX/i CT equipment (GE Healthcare, USA). The limbs were extended during examination and fixed to a foot rest to stabilize the position during scans. Sections of 1.25 mm thickness were taken through the hip, knee, and ankle joints with both limbs in the same positions. Internal torsion is assigned a minus (−) sign and external torsion a positive (+) sign. All data analysis was done digitally on a computer with the use of FDA-approved medical planning software (MediCAD version 2.0, Hectec, Altfraunhofen, Germany).

Statistical analysis

Intraobserver and interobserver reliability was evaluated using the intraclass correlation coefficient (ICC). The scoring system of Fleiss et al. [19] was utilized in the analysis of our results (good >0.75, fair 0.4–0.75, poor <0.4). For the intraobserver
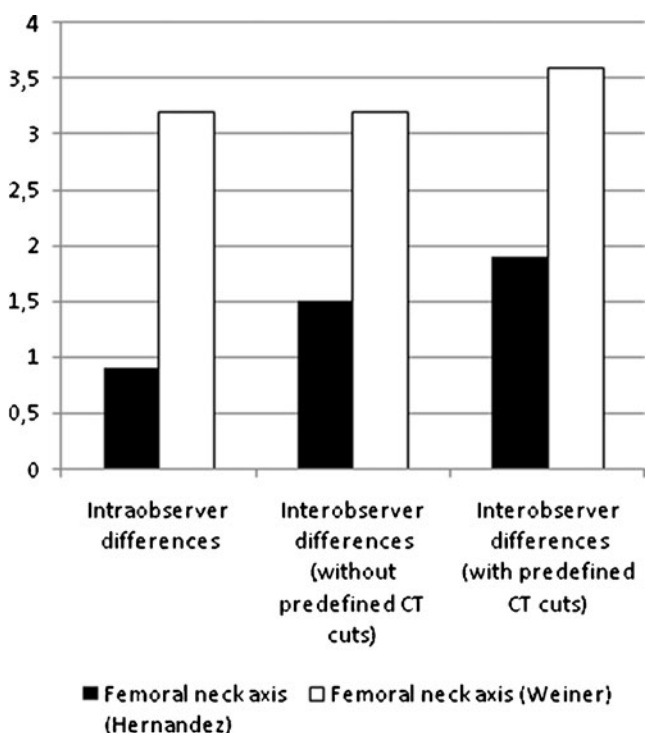

Fig. 4 Measurements of the femoral neck axis with the Hernandez and Weiner methods. Intra- and interobserver differences, measured in degrees, show better results for the Hernandez method
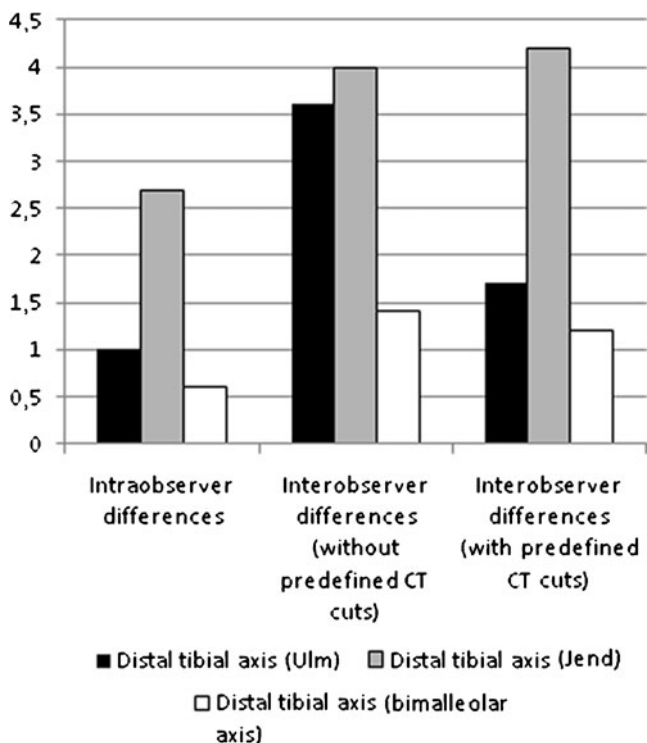

Fig. 5 Measurements of the distal tibial axis (degrees) with three different methods. Intra- and interobserver differences show better results for the bimalleolar axis method

Table 4 Average interobserver differences (median and range) with or without defined CT cuts

	Without predefined CT cuts	With predefined CT cuts	P value
Femoral neck axis (Hernandez)	1.5 (0.3–8.5)	1.9 (0–12.1)	0.418
Femoral neck axis (Weiner)	3.2 (0.8–15.9)	3.6 (0–10.9)	0.518
Femoral condylar axis	0.9 (0.3–6.4)	0.4 (0–7.1)	<0.0001
Proximal tibial axis	1.5 (0.2–5.2)	1.6 (0.1–6.5)	0.429
Distal tibial axis (Ulm)	3.6 (0.7–10.6)	1.7 (0.5–11.6)	0.008
Distal tibial axis (Jend)	4 (0.9–9.3)	4.2 (0.4–11.1)	0.564
Distal tibial axis (bimalleolar axis)	1.4 (0–11.2)	1.2 (0.1–12.4)	0.567

data, we calculated absolute differences between measurements of the first and second sessions for each observer, and the outcome is presented in the form of median value and range for each angle. Interobserver absolute differences were also calculated for the first two sets of measurements without predefined CT cuts for the three pairs of observers and then recalculated for the third session separately. The Wilcoxon test was used to detect the presence of statistically significant changes in the magnitude of the interobserver median absolute differences in measurements made without and with predefined CT cuts.

Differences between the first and the second method for the evaluation of the angle of the femoral neck axis (Hernandez versus Weiner et al. methods) or among the three methods in the distal tibial axis (Ulm versus Jend methods, Ulm versus bimalleolar axis, and Jend versus bimalleolar axis) were quantified using the measurements from the first two sessions for each observer (intermethod analysis). The measurements of the first two sessions were used to avoid the change in accuracy after defining certain CT cuts.

All statistical analyses were performed using the SPSS program (SPSS 15.0, SPSS, Chicago, IL, USA).

Results

All measured intra- and interobserver interclass correlations were greater than 0.75 (good). However, the best scores were achieved with the Hernandez (0.99 intra- and 0.93 interobserver correlations) and bimalleolar methods (0.99 intra- and 0.92 interobserver correlations, Table 2). For all observers, the intraobserver median absolute differences were lower for the Hernandez method of evaluating the femoral neck axis when compared to the Weiner et al. method. This was also found for the bimalleolar axis technique in the distal tibia compared to the Ulm and Jend methods (Table 3). Intraobserver variability was lowest for observer 1 followed by observer 2 and the highest variability was for observer 3 in almost all measured angles (Figs. 4 and 5).

Interobserver differences were in agreement with the intraobserver data (Figs. 4 and 5). Moreover, there was a statistically significant decrease in the interobserver median absolute differences for measurements of the femoral condylar axis and the distal tibial axis using the Ulm method with predefined CT cuts compared to the first two sessions (P<0.05, Table 4).

It was evident from the intermethod analysis that the highest values for the distal tibial axis were recorded with the Jend method followed by the Ulm method, and the lowest were recorded using the bimalleolar axis. The Jend method overestimated the distal tibial axis angle by an average of 8° compared to the Ulm method, whereas the bimalleolar axis underestimated it by an average of 5°. In the area of the femoral neck, there were fluctuating results among different observers (Table 5).

Discussion

The best intraobserver and interobserver ICC scores were found with the Hernandez method and the bimalleolar axis method in the area of the femoral neck and distal tibia, respectively. A statistically significant decrease in the interobserver median absolute differences could be achieved by using predefined CT scans for measurements of the femoral condylar axis and the distal tibial axis. Finally, as far as the intermethod differences are concerned, the bimalleolar axis method underestimated the tibial torsion angle by an average of 4.8° and 13° compared to the Ulm and Jend techniques, respectively.

Computed tomography is believed to be the most accurate method for assessment of the rotational profile of the lower

Table 5 Average intermethod differences for different techniques used in measuring the angle of the femoral neck and the distal tibial axes

	Observer 1	Observer 2	Observer 3	Mean
Neck axis	−2.4±4.6	−0.2±2	0.5±7.3	−0.7±3.6
Ulm vs. Jend	−8.9±3.3	−8.5±3.3	−7.3±3	−8.3±2.5
Ulm vs. bimalleolar	6.2±2.2	2.6±2.8	5.6±2.3	4.8±1.9
Jend vs. bimalleolar	15.1±3.4	11.1±3.6	12.9±3.1	13±2.6

extremities [18, 20]. The incidence of rotational malalignment following closed nailing of femoral and tibial diaphyseal fractures may reach up to 28% and 22%, respectively [21, 22]. Rotational differences of more than 15°, compared to the healthy side, are considered true deformities and can cause long-term complaints [21, 23]. However, most studies that put a limit on the accepted range of differences between the two lower limbs evaluated the rotational profile using one observer and one method [20, 21, 24].

On the other hand, few reports have analyzed the reliability of CT measurements of rotation in the lower extremities, and most focused on the accuracy of different methods versus anatomic references (intermethod reliability) [6–9]. In cases of posttraumatic rotational deformities, we are not interested in the accuracy of the method used to measure rotational angles as there is no gold standard for comparison, only the contralateral healthy side, which is used to determine side to side differences. With a limit of 15° for side to side differences, above which some authors recommend a corrective procedure, the reliability of the method used is a more important issue [9, 21, 23].

According to our results, defining the femoral neck axis using the Hernandez et al. method was associated with lower intraobserver and interobserver variability, associations also found for the bimalleolar axis in the distal tibia. This variability was not further improved by using predefined CT cuts. To define the femoral neck axis using the Weiner et al. method, we first chose a CT cut where the ventral and dorsal cortices of the femoral neck are parallel to each other (Fig. 1) [12]. However, it is not common to find this axial cut; in most patients, we find cuts where the neck is curved or oblong shaped, and therefore its axis cannot be easily drawn (Fig. 6). Sugano et al. [8] also came to the conclusion that the method of Weiner et al. is not accurate for routine clinical use.

In the distal tibia, measurements using the Ulm and Jend methods depend on CT cuts in the distal tibial pilon with the fibula articulating in the incisura fibularis. Both techniques had lower reliability compared to the bimalleo-lar axis, and this is mainly due to the difficulty in choosing the level at which we can make measurements. This was also evident by the statistically significant improvement in the interobserver variability when using the Ulm method with predefined CT cuts; however, this was not the case for the Jend method. In contrast, to draw the bimalleolar axis, we choose the first axial CT cut just below the tibial pilon's articular surface in which the medial and lateral malleoli and the talar dome are present, which is an easier task [17].

ICC scores for the femoral condylar axis were lower for the interobserver data compared to the intraobserver scores. Furthermore, the interobserver variability was significantly lowered by using predefined CT cuts. This is due to the size of the femoral condyles, as they are large and it may be difficult to choose the cut of their maximum prominence. A final observation related to the intraobserver variability was the effect of observer experience on the results. The attending surgeon had the lowest variability followed by the chief resident and the senior house officer.

Kuo et al. [9] reported an intraobserver difference up to 14.2° when using the Hernandez et al. method for the measurement of femoral anteversion. In another study that assessed femoral torsion using CT, Jaarsma et al. [6] reported intraobserver differences up to 10.8° and interobserver differences up to 15.6° in 95% of the measurements. An earlier study by the same authors recommended a corrective procedure for side to side femoral anteversion differences of more than 15° following closed nailing procedures for diaphyseal femoral fractures; however, with the reported magnitude of intraobserver and interobserver differences, this limit seems unreasonable [21]. For the measurement of tibial torsion, an interobserver variability of ±3° was reported [10].

When using the Hernandez or Weiner et al. method to assess the rotational angle of a given anatomic area, the results fluctuated for the femoral neck axis. The bimalleolar axis method underestimated the tibial torsion angle by an average of 4.8° and 13° compared to the Ulm and Jend techniques, respectively. Furthermore, compared to the Ulm
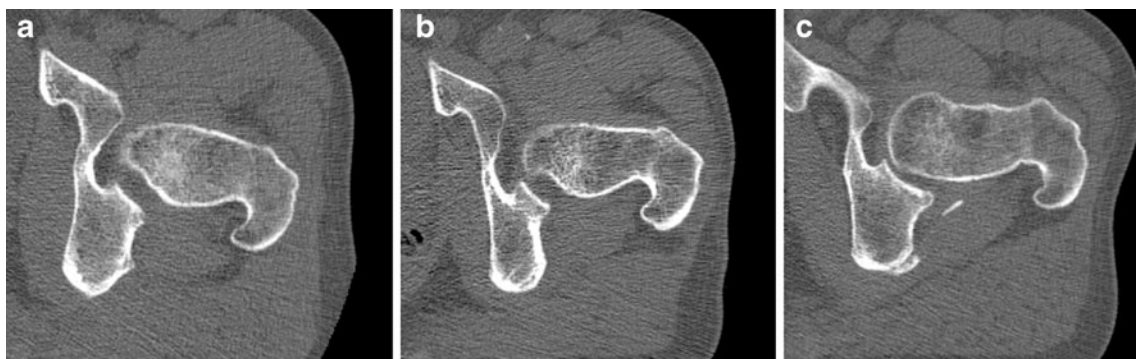

Fig. 6 In most patients, the ventral and dorsal cortices of the femoral neck are not parallel to each other; instead, we find an oblong shaped femoral neck (**a** and **b**) or a curved femoral neck (**c**)

technique, the Jend method overestimated the results by an average of 8.3°. This represents another difficulty when interpreting the results of different studies assessing lower limb torsion with the use of inconsistent methods to define different axes and the impact of choosing one method over the others.

In conclusion, the most reliable methods are the Hernandez and bimalleolar methods for measuring femoral and tibial torsion, respectively. With the exception of the Ulm method, the reliability of the other methods could not be improved by using predefined CT cuts. Measurements of lower limb torsion should only be compared between different physicians if they are using the same methods to define different axes.

Conflict of interest statement The authors certify that they have no commercial associations (e.g., consultancies, stock ownership, equity interest, etc.) that might pose a conflict of interest in connection with the submitted article.

References

1. Paley D, Herzenberg JE, Tetsworth K, McKie J, Bhave A. Deformity planning for frontal and sagittal plane corrective osteotomies. Orthop Clin North Am. 1994;25:425–65.
2. Paley D, Tetsworth K. Mechanical axis deviation of the lower limbs. Preoperative planning of multiapical frontal plane angular and bowing deformities of the femur and tibia. Clin Orthop Relat Res 1992:65–71.
3. Paley D, Tetsworth K. Mechanical axis deviation of the lower limbs. Preoperative planning of uniapical angular deformities of the tibia or femur. Clin Orthop Relat Res 1992:48–64.
4. Feldman DS, Henderson ER, Levine HB, et al. Interobserver and intraobserver reliability in lower-limb deformity correction measurements. J Pediatr Orthop. 2007;27:204–8.
5. Noyes FR, Goebel SX, West J. Opening wedge tibial osteotomy: the 3-triangle method to correct axial alignment and tibial slope. Am J Sports Med. 2005;33:378–87.
6. Jaarsma RL, Bruggeman AW, Pakvis DF, et al. Computed tomography determined femoral torsion is not accurate. Arch Orthop Trauma Surg. 2004;124:552–4.
7. Murphy SB, Simon SR, Kijewski PK, Wilkinson RH, Griscom NT. Femoral anteversion. J Bone Joint Surg Am. 1987;69:1169–76.
8. Sugano N, Noble PC, Kamaric E. A comparison of alternative methods of measuring femoral anteversion. J Comput Assist Tomogr. 1998;22:610–4.
9. Kuo TY, Skedros JG, Bloebaum RD. Measurement of femoral anteversion by biplane radiography and computed tomography imaging: comparison with an anatomic reference. Invest Radiol. 2003;38:221–9.
10. Bouchard R, Meeder PJ, Krug F, Libicher M. Evaluation of tibial torsion—comparison of clinical methods and computed tomography. Rofo. 2004;176:1278–84.
11. Hernandez RJ, Tachdjian MO, Poznanski AK, Dias LS. CT determination of femoral torsion. AJR Am J Roentgenol. 1981;137:97–101.
12. Weiner DS, Cook AJ, Hoyt Jr WA, Oravec CE. Computed tomography in the measurement of femoral anteversion. Orthopedics. 1978;1:299–306.
13. Jend HH, Heller M, Dallek M, Schoettle H. Measurement of tibial torsion by computer tomography. Acta Radiol Diagn (Stockh). 1981;22:271–6.
14. Jend HH, Heller M, Schontag H, Schoettle H. A computer tomographic method for the determination of tibial torsion (author's transl). Rofo. 1980;133:22–5.
15. Reikeras O, Hoiseth A. Torsion of the leg determined by computed tomography. Acta Orthop Scand. 1989;60:330–3.
16. Yagi T, Sasaki T. Tibial torsion in patients with medial-type osteoarthritic knee. Clin Orthop Relat Res. 1986:177–182.
17. 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;88:2439–47.
18. Waidelich HA, Strecker W, Schneider E. Computed tomographic torsion-angle and length measurement of the lower extremity. The methods, normal values and radiation load. Rofo. 1992;157:245–51.
19. Fleiss J. The design and analysis of clinical experiments. New York: John Wiley; 1986.
20. Strecker W, Keppler P, Gebhard F, Kinzl L. Length and torsion of the lower limb. J Bone Joint Surg Br. 1997;79:1019–23.
21. Jaarsma RL, Pakvis DF, Verdonschot N, Biert J, van Kampen A. Rotational malalignment after intramedullary nailing of femoral fractures. J Orthop Trauma. 2004;18:403–9.
22. Puloski S, Romano C, Buckley R, Powell J. Rotational malalignment of the tibia following reamed intramedullary nail fixation. J Orthop Trauma. 2004;18:397–402.
23. Jaarsma RL, Ongkiehong BF, Gruneberg C, et al. Compensation for rotational malalignment after intramedullary nailing for femoral shaft fractures. An analysis by plantar pressure measurements during gait. Injury. 2004;35:1270–8.
24. Tornetta 3rd P, Ritz G, Kantor A. Femoral torsion after interlocked nailing of unstable femoral fractures. J Trauma. 1995;38:213–9.