Title

A reliability study of radiographic measures of total ankle replacement position: an analysis from the OARS cohort.

Authors

Samantha Low BSc FRCR[1]

Matthew Kim MB BS FRCR[1]

Toby Smith PhD[2]

David T Loveday FRCS (Tr & Orth)[3]

Alex MacGregor PhD FRCP[4]

Andoni P Toms PhD FRCR[1]

Affiliations

1. Department of Radiology, Norfolk & Norwich University Hospital
2. School of Health Sciences, University of East Anglia
3. Department of Trauma & Orthopaedics, Norfolk & Norwich University Hospital
4. Norwich Medical School, University of East Anglia

ABSTRACT

Objective

There is no validated radiographic measurement to diagnose prosthetic complication(s) following total ankle replacements (TAR) although a number of angular and linear measurements, used to define the TAR position on postoperative radiographs, have been recommended to detect prosthetic loosening. The aim of this study was to test the intra and interobserver reliability of these measurements.

Materials and methods

This is a prospective study embedded within a multicentre cohort study. Following sample size calculation, 62 patients were analysed. Six measurements were performed on the first postoperative anteroposterior and lateral ankle radiographs: angles α and β, and length “a” defined the craniocaudal position of the tibial component; while angle γ, lengths “b” and “c” defined the angular position of the talar component. Measurements were recorded by three independent observers. Inter and intraobserver reliability was assessed with: intraclass correlation coefficient (ICC); Bland-Altman plots; and within-subject coefficients of variation (CV).

Results

The intrarater ICC was “almost perfect” (ICC 0.83 – 0.97) for all six measurements. The interrater ICC was “substantial” to “almost perfect” (ICC 0.69 – 0.93). The mean difference in intrarater angular measurements was ≤0.6 degree, and ≤0.8mm for linear measurements; ≤2.2 degrees and ≤2.1mm for interrater measurements. Maximum CV for the interrater linear measurements (≤17.7%) more than doubled that of the angular measurements (≤8.0%). The maximum width of the 95% limits of agreement was 6.5 degrees and 8.4mm for intrarater measures, and 8.9 degrees and 10.6mm for interrater measurements.

Conclusion

Angular measures are more reliable than linear measures and have potential in routine clinical practice for TAR position assessment.

KEYWORDS

Total ankle replacement; Total ankle arthroplasty; Radiograph; Reliability

INTRODUCTION

Since their introduction in the early 1970s, total ankle replacements (TAR) have made a lesser impact in clinical practice when compared to total hip and knee replacements. The highly anticipated first generation cemented two component constrained designed TAR was soon abandoned because of high complications rates which were thought to relate to the complex anatomy of the ankle joint.(1,2) A second generation of uncemented anatomically designed TARs was subsequently developed to address previous flaws, incorporating cementless technology to minimise bone resection and encourage bone ingrowth.(3) The current third generation designs remain uncemented but are three component prostheses with a minimally constrained, independent polyethylene component to represent the mobile-bearing meniscus.(3) These prostheses are the gold standard with studies reporting clinical efficacies comparable to ankle arthrodesis in the treatment of advanced ankle osteoarthritis.(3,4)

Radiographs are used in the postoperative evaluation of TAR. Radiological abnormalities are common in many of the high and intermediate grade complications.(3,5) Glazebrook et al(5) proposed classifying complications based on the likelihood of TAR failure for a given complication; intermediate grade complications lead to TAR failure in <50% of the time while high grade complications accounted for >50%. Radiographically evident intermediate grade complications include implant subsidence, postoperative bone fracture, and medial impingement while high grade complications include aseptic loosening (>2mm), periprosthetic osteolysis, deep infection and implant failure from background systemic disease.(3)

The clinician can identify prosthesis position change by comparing serial postoperative radiographs. The addition of angular and linear measurements of the position of the TAR relative to the surrounding bones affords quantitative comparison between serial radiographic examinations. Quantification of these measurements allows for statistical analysis of the association of radiographic position of the TAR with other outcome measures in large cohort studies.

Several methods have been proposed for establishing radiologic reference points to monitor component migration in TARs. Bestic et al.(6) recommended a number of angular and linear measurements, defining the position of TARs on serial postoperative radiographs for detecting prosthesis loosening. The reproducibility (reliability) and validity of these methods have not been established. Accordingly, the aim of this study was to measure the intra and interrater reliability of the linear and angular measures of radiographic TAR position as proposed by Bestic et al.(6).

MATERIALS AND METHODS

Participants

This investigation is embedded within the Outcomes in Ankle Joint Replacement (OARS) study. The OARS study is a prospective, multisite longitudinal study, aimed to enrich the National Joint Registry for England, Wales, Northern Ireland and the Isle of Man (NJR) dataset on TAR by collecting preoperative measures of clinical and radiological disease severity, postoperative Patient Reported Outcome Measures (PROMS), together with clinical and radiological measures of surgical outcome for 12 months postoperatively. The inclusion criteria for our study were patients who have had the TAR surgery between January 2016 – December 2017, followed by a baseline postoperative ankle radiograph. The TAR systems used in this study comprised of the Zenith (Corin) and INFINITY (Wright Medical Technology).

This study conforms to relevant research ethnical guidelines and was approved by the Health Research Authority (HRA) on 7th July 2016 (IRAS project ID: 146735).

Image acquisition and analysis

Anteroposterior (AP) and lateral views of the ankle were obtained in the standing position to ensure physiologic positioning. If these were not possible, nonweightbearing radiographs in positions, which were as closely matched to weightbearing ankle radiographs as possible, were obtained.

Six angular and linear measurements(6) were performed on AP and lateral ankle radiographs: angles α and β, and length “a” defined the craniocaudal position of the tibial component; while angle γ, lengths “b” and “c” defined the angular position of the talar component (**Figure 1**). Measurements were recorded twice by three independent observers, a musculoskeletal radiologist with 20 years' experience, and two radiology trainees with three- and five-years' experience. All observers were blinded to the clinical data. The two observations were performed at least two weeks apart.

The alpha (α) angle and beta (β) angles are subtended between a line drawn along the long axis of the tibia and the articular surface of the tibial component on the AP and lateral projections respectively (**Figure 1A, 1B**). The gamma (γ) angle is subtended by a line between the most anterior and posterior points of the talar component, on the lateral projection, and a line drawn from the most posterior point of the prosthesis along the middle of the neck of the talus (**Figure 1B**).

Length a is a perpendicular measure from a line drawn laterally from the articular surface of the tibial component to the tip of the lateral malleolus (**Figure 1C**). Talar tilt measures b and c are drawn perpendicular to a line connecting the anterior and posterior margins of the talar component, on the lateral projection, to a line connecting the anterosuperior corner of the head of the talus to the posterosuperior margin of the posterior subtalar facet of the calcaneus (**Figure 1D**).

Statistical analysis

Inter and intraobserver reliability was assessed with: intraclass correlation coefficient (ICC) for consistency of single measures by fixed raters; Bland-Altman plots; and within-subject coefficients of variation.

ICCs will be interpreted using the Landis and Koch (8) method where values < 0 indicate poor agreement, 0-0.2 as slight, 0.21-0.4 as fair, 0.41-0.6 as moderate, 0.61-0.8 as substantial and 0.81-1 as almost perfect agreement.

Bland-Altman plots yielded limits of agreement and coefficient of repeatability (CR) or smallest real difference (SRD) which was calculated as 1.96 multiplied by the standard deviation (SD) of the measurement difference between observers. The within-subject coefficients of variation (CV) was calculated using the root mean square approach.(7) The significance threshold was set at .05. All statistical analyses were completed using R version 3.2.2 with the psych package.(8)

Sample size

A sample size calculation of this embedded reliability study was made after the first 25 sets of measurements. Based on the first 25 mean α and β measurements for Raters one and two, assuming a standard deviation of two degrees, a sample size of 62 was required for a total 95% confidence interval (CI) around the mean of one degree.(9)

RESULTS

Sixty-two patients were included in the study. There were 35 male and 27 females (mean age 63 years; range 24 to 86). The median time from TAR to radiographs was 45 days (Interquartile Range: 42, 60).

The summary statistics for each variable (derived from the mean of six observations) were as follows: mean α angle = 89.0 degrees (SD 2.1, 95% CI: 88.5, 89.5), median β = 89.3 degrees (IQR = 87.2, 90.3), mean γ = 21.6 degrees (SD 2.9; 95% CI: 20.8, 22.3), mean length a = 31.7mm (SD 3.9; 95% CI: 30.7, 32.6), mean talar tilt length b = 9.3mm (SD 2.6; 95% CI: 8.7, 10.0), and mean talar tilt length c = 14.1mm (SD 4.4; 95% CI: 12.9, 15.2). The variable β failed the Shapiro-Wilk test of normality and therefore summary data for this has been presented as median and IQR (**Figure 2**).

The mean difference in intrarater angular measurements (α, β, γ) was $\leq$ 0.6 degree. The width of the 95% limits of agreement was $\leq$ 5.6 degrees for α and β, and $\leq$ 6.2 degrees for γ. The intrarater CV for α and β was less than 1.2% for α and β, and $\leq$ 5.7% for γ which reflects the smaller magnitude of mean γ (21.6 degrees) compared to mean α and β (89 degrees and 89 degrees).

The mean difference in intrarater linear measures (a, b, c) was $\leq$ 0.8mm with the width of the 95% limits of agreement measuring up to 8.2mm. The CV for $a \leq$ 5.1%, for $b \leq$ 13.0% and for $c \leq$ 10.3%. The difference in the magnitude of the mean for a (31.7mm) and for b and c (9.3 and 14.1mm) is again associated with these differences in CV.

The mean difference in interrater angular measurements (α, β, γ) was ≤ 2.2 degrees. The width of the 95% limits of agreement was ≤ 7.9 degrees for α and β, and ≤ 7.7 degrees for γ. The interrater CV was less than 2.2% for α and β, and ≤ 8.0% for γ.

The mean difference in interrater linear measures (*a*, *b*, *c*) was ≤ 2.1mm with the width of the 95% limits of agreement measuring up to 10.6mm. The CV for *a* was ≤ 7.9%, for *b* ≤ 16.9%, and for *c* ≤ 17.7%.

The intrarater ICC was "almost perfect" for all variables. The interrater ICC was "substantial" to "almost perfect" (**Tables 1 and 2**).

DISCUSSION

There are many radiographic findings that may predict clinical outcomes after TAR such as measures of severity of joint disease, hindfoot alignment and initial TAR position. There are also findings such as changes in position of the TAR with time, periprosthetic lucency, cyst formation and periostitis that may indicate a failing or failed prosthesis.

Radiographic loosening by angular and linear measurements was originally described by Carlsson et al.(10) in 1994. An adaptation of Carlsson's angular measurements was used by Wood et al.(11) in 2003 and most recently by Bianchi et al.(4) in 2019. The measurements as recommended by Bestic et al.(6) are a combination of Carlsson's original linear and the later adapted angular measurements by Wood et al. To our knowledge, these measurements have never been validated before. As such, the aim of this study was to focus on these radiographic measures of TAR position and to identify the amount of variability in these measurements between and within observers.

We report the width of the confidence intervals for the mean for each of the six variables was narrow, indicating the sample was clinically representative. The intra and interrater reliability results suggest angular measurements α, β and γ were the most reliable while the linear measurements *a, b* and *c* were the least reliable. The intrarater reliability measurements were similar across all levels of experience.

The 95% limits of agreement are relatively wide, and the coefficients of variance are large for linear measures of talar component positioning. Therefore, the linear measures of TAR position are so variable that the clinical utility of these measures in their current form is questionable.

The lowest ICC for intrarater reliability was 0.83 (**Table 1**), and for interrater reliability the lowest ICC was 0.69 (**Table 2**).

All the other comparisons produced an ICC of 0.7 or more. These ICCs can be interpreted as “substantial” to “almost perfect” and “almost perfect” respectively using the Landis and Koch(12) method. While this method is a commonly cited classification for interpreting ICC results, the categories described are arbitrary descriptions of consistency and do not necessarily mean that a test is reliable enough to use in research or routine clinical practice. Previous authors have suggested an ICC of at least 0.8(13) can be regarded as reliable, others suggesting that this should be measured from the lower of the 95% CIs.(14) In this study, almost all of the lower 95% CIs for intrarater ICCs, and one third of the interrater ICCs, are >0.8. Thirty-four of 36 comparisons demonstrate “good” or “excellent” reliability according to Koo & Li.(14) While ICCs are commonly used to calculate consistency in repeated measurements, there are limitations to their use in reliability studies. Repeated measurements may be consistent but discordant and the magnitude of differences between raters may depend on the size of the observations.(15) These effects can be assessed using Bland-Altman plots (**Figure 3**).

The distribution of data across the Bland-Altman plots demonstrated no evidence of funnelling. Therefore, the magnitude of the measurements does not appear to have an effect on reliability (**Figure 3**).

The mean difference in intrarater angular measurements was <1 degree, and <1 mm for linear measurements. These mean differences increased to a maximum of 2.2 degrees and 2.1mm for interrater angular and linear measurements respectively (**Tables 1 and 2**). While these differences are small, the variation in the differences, described by the 95% limits of agreement, are larger. The maximum width of the 95% limits of agreement was 6.5 degrees and 8.4mm for intrarater measures, and 8.9 degrees and 10.6mm for interrater measurements. While the absolute measure of these 95% confidence intervals was similar across comparisons, they do vary as a proportion of the variable being measured. For α and β angles, most variation was less than 10% of the mean but because γ angles are smaller, the variation constituted some 25% of the mean γ angle measurement.

There are a number of reasons why some of these measurements are more reliable than others. The variation in measurements between observers adversely affects small measures more than large measures (**Tables 3 and 4**). The variation in angular measures for α and γ were similar (~1 degree) but mean α (89 degrees) was much larger than mean γ (22 degrees) and therefore the ratio of variability to magnitude (i.e. the *coefficient of variance*) was greater for γ (~5%) compared to α (~1%) (**Tables 3 and 4**). α and β are probably the most straight-forward to measure and least dependent on patient positioning and implant type.

The landmarks for the γ angle vary between type of implant and are also dependent on obtaining a true lateral projection on the radiograph. The linear measurements are more dependent on anatomical landmarks which can be more difficult to define and which may be obliterated in severe deforming arthropathies. For instance, the calcaneal tubercle used in measuring the talar tilt distances *b* and *c* is not a single point and there are a number of definitions of the calcaneal tubercle. Some refer to this as being on the inferior surface of the calcaneus for the attachment of the long plantar ligament,(16) others refer to this as the trochlear process where it is a raised projection separating the two oblique grooves on the lateral surface of the calcaneus, for the tendons of the peroneal muscles.(17) This could be improved by choosing a more anatomically consistent landmark, such as the posterior margin of the subtalar joint, but this would need to tested.

As mentioned before, the angular measurements α, β, γ were used in a study of 200 post-TAR participants in the immediate postoperative phase by Wood and Deakin.(11) The angular measurement α was also used in Willeger et al.(18) where they analysed different methods to measure α in postoperative ankle arthrodesis. Our angular measurements are comparable to the values as obtained by Wood and Deakin(11) and Willeger et al.,(18) regardless of the measurement method. Neither of these studies reported measures of reliability or reproducibility.

This study presents with two key limitations. Firstly, all patients originated from three hospitals; more than half from one hospital. Therefore, some selection bias may have occurred in the choice of patients, prostheses, and operative and radiographic techniques that might influence reliability. The proposed measurements by Bestic et al.(6) were performed on the Scandinavian Total Ankle Replacement (STAR) device while our study included a variety of TAR devices due to individual surgeon preference, having recruited patients from multiple sites. While there was heterogeneity in the TAR devices, they were all second generation devices(19) with either two or three component designs. The difference between two or three component designs will not affect the angular or linear measurements as the landmark for the measurements involve the radio-opaque metal component and not the radiolucent polyethylene spacer. Although reliability measures might be improved by limiting the study to a single prosthesis and rigidly standardised radiographs the results of this study are more likely to reflect the heterogeneity of practice in the real world. Secondly, while this study describes the reliability of radiographic TAR measurements it does not address test retest reliability where measurements from one radiograph are compared to measurements from a second interval radiograph. This is an important next step in defining the minimum radiographic changes that can detect positional change of a prosthesis on serial radiographs.

The intra and interrater reliability of radiographic measurements of TAR measured by ICC is “good” to “excellent” but this does not reflect the degree of variability in some of the linear measurements. The coefficients of variance for linear measures of the position of the talar component of the TAR are large and therefore, these are probably not clinically useful in their current form. Angular measures of TAR position are reliable across all statistical methods and have potential for use in routine clinical practice.

REFERENCES

1. Bolton-Maggs BG, Sudlow RA, Freeman MA. Total ankle arthroplasty. A long-term review of the London Hospital experience. J Bone Joint Surg Br. 1985 Nov;67(5):785–90.

2. Kitaoka HB, Patzer GL. Clinical results of the Mayo total ankle arthroplasty. J Bone Joint Surg Am. 1996 Nov;78(11):1658–64.

3. Kim D-R, Choi YS, Potter HG, Li AE, Chun K-Y, Jung YY, et al. Total Ankle Arthroplasty: An Imaging Overview. Korean J Radiol. 2016;17(3):413–23.

4. Bianchi A, Martinelli N, Hosseinzadeh M, Flore J, Minoli C, Malerba F, et al. Early clinical and radiological evaluation in patients with total ankle replacement performed by lateral approach and peroneal osteotomy. BMC Musculoskelet Disord. 2019 Dec;20(1):132.

5. Glazebrook MA, Arsenault K, Dunbar M. Evidence-based classification of complications in total ankle arthroplasty. Foot Ankle Int. 2009 Oct;30(10):945–9.

6. Bestic JM, Peterson JJ, DeOrio JK, Bancroft LW, Berquist TH, Kransdorf MJ. Postoperative Evaluation of the Total Ankle Arthroplasty. Am J Roentgenol. 2008 Apr;190(4):1112–23.

7. Bland M, Altman D. How should I calculate a within-subject coefficient of variation? [Internet]. 2006. Available from: https://www-users.york.ac.uk/~mb55/meas/cv.htm

8. R Development Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2008. Available from: http://www.R-project.org

9. Eng J. Sample Size Estimation: How Many Individuals Should Be Studied? Radiology. 2003 May;227(2):309–13.

10. Carlsson ÅS, Nilsson J-Å, Redlund-Johnell I, Henricson A, Linder L. A survival analysis of 52 Bath & Wessex ankle replacements A clinical and radiographic study in patients with rheumatoid arthritis and a critical review of the literature. The Foot. 1994 Mar 1;4(1):34–40.

11. Wood PLR, Deakin S. Total ankle replacement: THE RESULTS IN 200 ANKLES. J Bone Joint Surg Br. 2003 Apr;85-B(3):334–41.

12. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977 Mar;33(1):159–74.

13. Bernstein IH, Nunnally JC. The assessment of reliability. Psychometric Theory. New York: McGraw-Hill Companies; 1994. 248–92 p.

14. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016 Jun;15(2):155–63.

15. Ranganathan P, Pramesh C, Aggarwal R. Common pitfalls in statistical analysis: Measures of agreement. Perspect Clin Res. 2017;8(4):187.

16. Micheau A. Calcaneal tubercle [Internet]. IMAIOS. 2019 [cited 2019 Dec 5]. Available from: https://www.imaios.com/en/e-Anatomy/Anatomical-Parts/Calcaneal-tubercle

17. Thieme Atlas of Anatomy: General Anatomy and Musculoskeletal System. Thieme; 2006. 541 p.

18. Willegger M, Holinka J, Nemecek E, Bock P, Wanivenhaus AH, Windhager R, et al. Reliability of the Radiographic Sagittal and Frontal Tibiotalar Alignment after Ankle Arthrodesis. Tsuchiya H, editor. PLOS ONE. 2016 Apr 28;11(4):e0154224.

19. Bestic JM, Bancroft LW, Peterson JJ, Kransdorf MJ. Postoperative imaging of the total ankle arthroplasty. Radiol Clin North Am. 2008 Nov;46(6):1003–15, v–vi.

TABLES

Table 1 Intrarater reliability for the three raters for each TAR position variable (mean differences, 95% limits of agreement, intraclass correlation coefficients and 95% confidence intervals)

Rater	Alpha	Beta	Gamma	Length *a*	Talar Tilt Length *b*	Talar Tilt Length *c*
1	0.06 (-2.03, 2.16)* 0.90 (0.85, 0.94)†	0 (-2.10, 2.10) 0.92 (0.86, 0.95)	-0.02 (-2.95, 2.92) 0.89 (0.82, 0.93)	0.05 (-2.15, 2.25) 0.96 (0.94, 0.98)	0.08 (-2.19, 2.35) 0.92 (0.86, 0.95)	-0.26 (-4.38, 3.85) 0.89 (0.82, 0.93)
2	0.16 (-1.76, 2.08) 0.91 (0.86, 0.95)	-0.06 (-2.04, 1.91) 0.92 (0.87, 0.95)	-0.31 (-2.65, 2.04) 0.92 (0.87, 0.95)	-0.78 (-4.97, 3.41) 0.87 (0.79, 0.92)	-0.03 (-1.54, 1.49) 0.97 (0.95, 0.98)	-0.43 (-2.67, 1.81) 0.97 (0.95, 0.98)
3	-0.03 (-2.64, 2.57) 0.83 (0.73, 0.89)	-0.34 (-3.13, 2.45) 0.86 (0.78, 0.91)	-0.61 (-3.73, 2.51) 0.88 (0.81, 0.93)	0.4 (-2.01, 2.81) 0.95 (0.93, 0.97)	0.38 (-2.16, 2.91) 0.89 (0.83, 0.93)	0.16 (-2.71, 3.03) 0.95 (0.92, 0.97)

*mean difference (95% limits of agreement)

† Intraclass Correlation Coefficient for consistency of single measures by fixed raters (95% confidence intervals)

Table 2 Interrater reliability for the three raters for each TAR position variable (mean differences, 95% limits of agreement, intraclass correlation coefficients and 95% confidence intervals)

Raters	Alpha	Beta	Gamma	Length *a*	Talar Tilt Length *b*	Talar Tilt Length *c*
1 v 2	-0.13 (-3.58, 3.32)* 0.74 (0.65, 0.81)†	0.82 (-2.52, 4.16) 0.79 (0.71, 0.84)	0.68 (-3.16, 4.51) 0.82 (0.75, 0.87)	-1.57 (-6.51, 3.38) 0.86 (0.81, 0.90)	0.75 (-2.65, 4.15) 0.83 (0.77, 0.88)	1.12 (-2.99, 5.22) 0.86 (0.81, 0.90)
2 v 3	-0.63 (-4.59, 3.33) 0.69 (0.58, 0.77)	-2.23 (-5.39, 0.94) 0.75 (0.66, 0.82)	0.60 (-3.2, 4.39) 0.79 (0.72, 0.85)	2.10 (-3.19, 7.39) 0.80 (0.73, 0.86)	-0.37 (-3.98, 3.24) 0.77 (0.69, 0.84)	-1.15 (-5.64, 3.35) 0.89 (0.84, 0.92)
1 v 3	-0.76 (-4.08, 2.57) 0.71 (0.61, 0.79)	-1.40 (-4.68, 1.88) 0.77 (0.68, 0.83)	1.27 (-2.53, 5.08) 0.78 (0.70, 0.84)	0.54 (-2.12, 3.19) 0.93 (0.89, 0.95)	0.38 (-2.91, 3.66) 0.82 (0.75, 0.87)	-0.03 (-4.30, 4.24) 0.87 (0.82, 0.91)

*mean difference (95% limits of agreement)

† Intraclass Correlation Coefficient for consistency of single measures by fixed raters (95% confidence intervals)

Table 3 Intrarater reliability for the three raters for each TAR position variable (within subject root-mean-square coefficient of variation and 95% confidence intervals)

Rater	Alpha	Beta	Gamma	Length *a*	Talar Tilt Length *b*	Talar Tilt Length *c*
1	0.85 (0.68, 0.99)*	0.85 (0.70, 0.98)	4.67 (3.85, 5.37)	2.54 (1.28, 3.37)	9.66 (7.18, 11.63)	9.01 (5.86, 11.32)
2	0.78 (0.61, 0.91)	0.80 (0.63, 0.94)	4.21 (3.35, 4.92)	5.11 (-1.04, 7.30)	7.31 (5.31, 8.87)	7.89 (5.15, 9.90)
3	1.04 (0.58, 1.36)	1.16 (0.92, 1.36)	5.65 (4.79, 6.41)	3.19 (2.08, 4.00)	12.95 (5.39, 17.51)	10.25 (5.74, 13.31)

*within subject root-mean-square coefficient of variation in percentage (95% confidence intervals)

Table 4 Interrater reliability for the three raters for each TAR position variable (within subject root-mean-square coefficient of variation and 95% confidence intervals)

Rater	Alpha	Beta	Gamma	Length *a*	Talar Tilt Length *b*	Talar Tilt Length *c*
1 v 2	1.39 (1.18, 1.57)*	1.52 (1.11, 1.84)	6.53 (5.29, 7.57)	6.66 (4.37, 8.35)	16.87 (12.82, 20.12)	17.65 (9.16, 23.22)
2 v 3	1.65 (1.20, 2.00)	2.19 (1.87, 2.47)	6.96 (4.91, 8.53)	7.92 (5.75, 9.60)	13.71 (10.63, 16.22)	16.91 (11.26, 21.10)
1 v 3	1.48 (1.11, 1.77)	1.73 (1.40, 2.00)	7.96 (6.19, 9.40)	3.59 (2.53, 4.40)	15.47 (10.15, 19.39)	14.97 (6.60, 20.11)

* within subject root-mean-square coefficient of variation in percentage (95% confidence intervals)

FIGURES

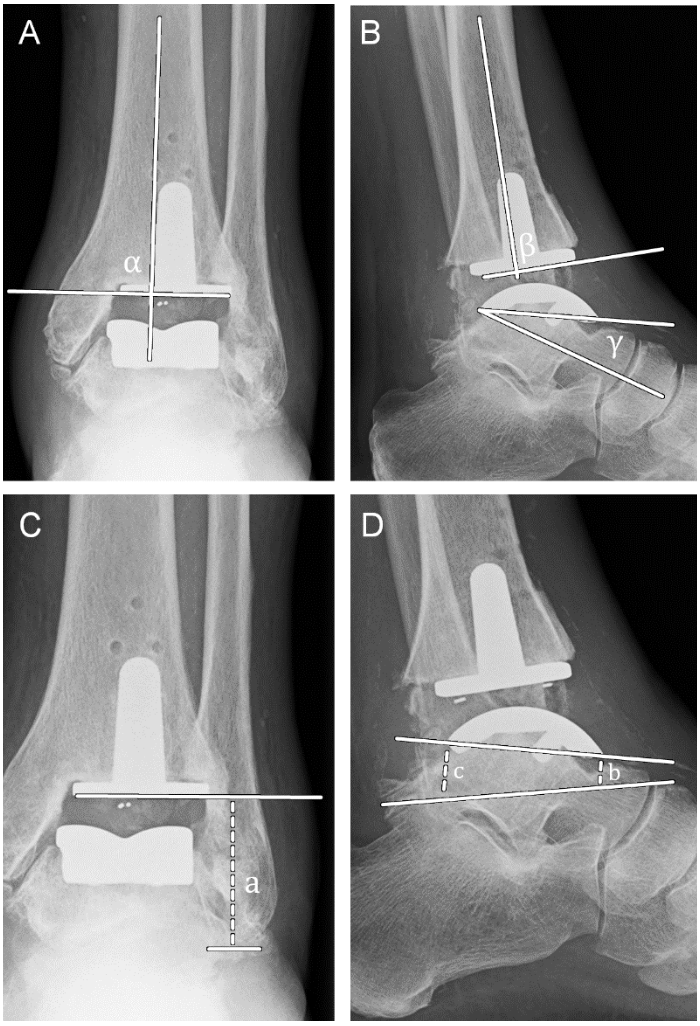

Figure 1

Anteroposterior and lateral ankle radiographs demonstrating the lines drawn to obtain the variables used to describe the position of the TAR. The alpha (α) angle and beta (β) angles are subtended between a line drawn along the long axis of the tibia and the articular surface of the tibial component on the AP and lateral projections respectively (A, B). The gamma (γ) angle is subtended by a line between the most anterior and posterior points of the talar component, on the lateral projection, and a line drawn from the most posterior point of the prosthesis along the middle of the neck of the talus (B). Length *a* is a perpendicular measure from a line drawn laterally from the articular surface of the tibial component to the tip of the lateral malleolus (C). Talar tilt measures *b* and *c* are drawn perpendicular to a line connecting the anterior and posterior margins of the talar component, on the lateral projection, to a line connecting the anterosuperior corner of the head of the talus to the posterosuperior margin of the posterior subtalar facet of the calcaneus (D).

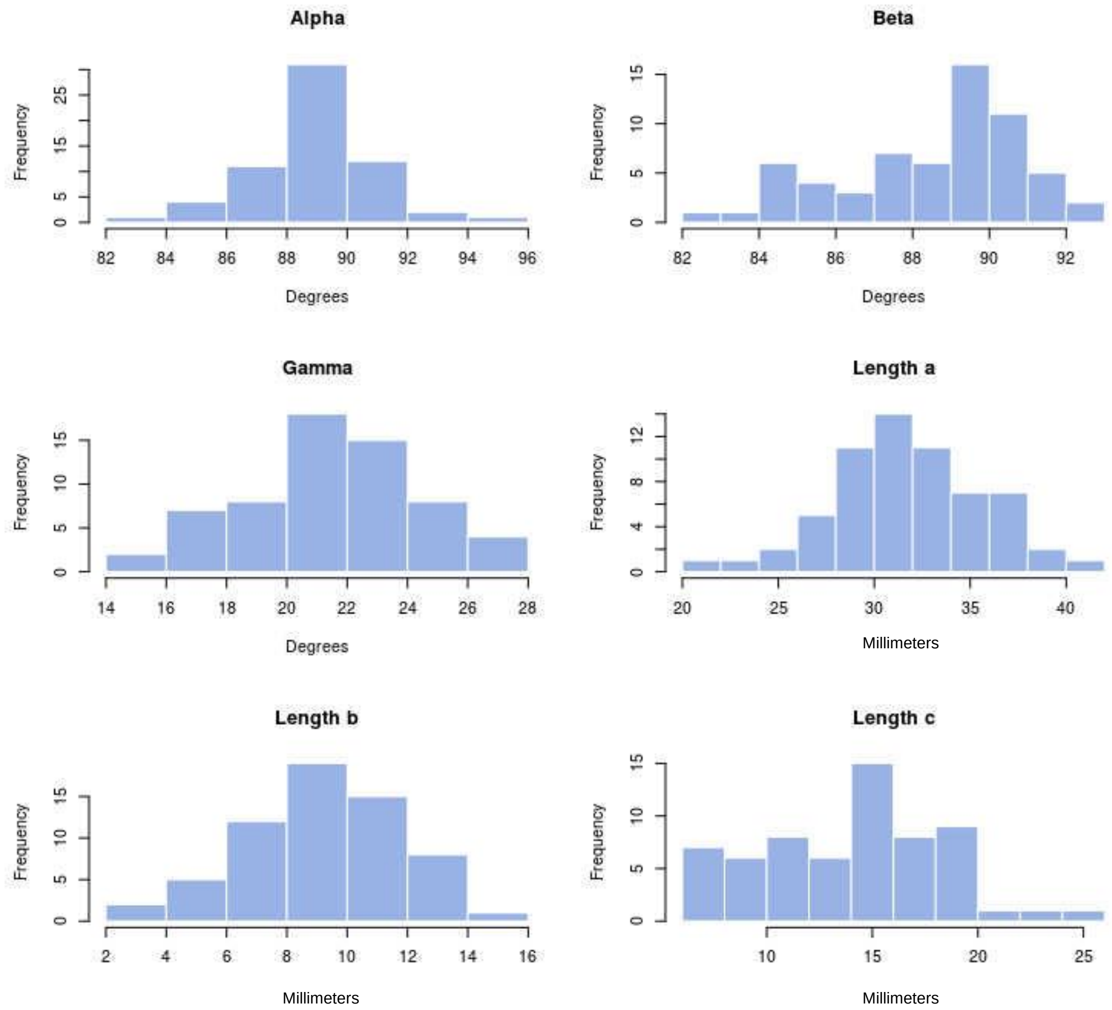

Figure 2

Frequency histograms of the average of 6 measurements (two from each rater) for each variable describing the position of the TAR. All but the beta angle conforms to a normal distribution according to the Shapiro-Wilk of normality.

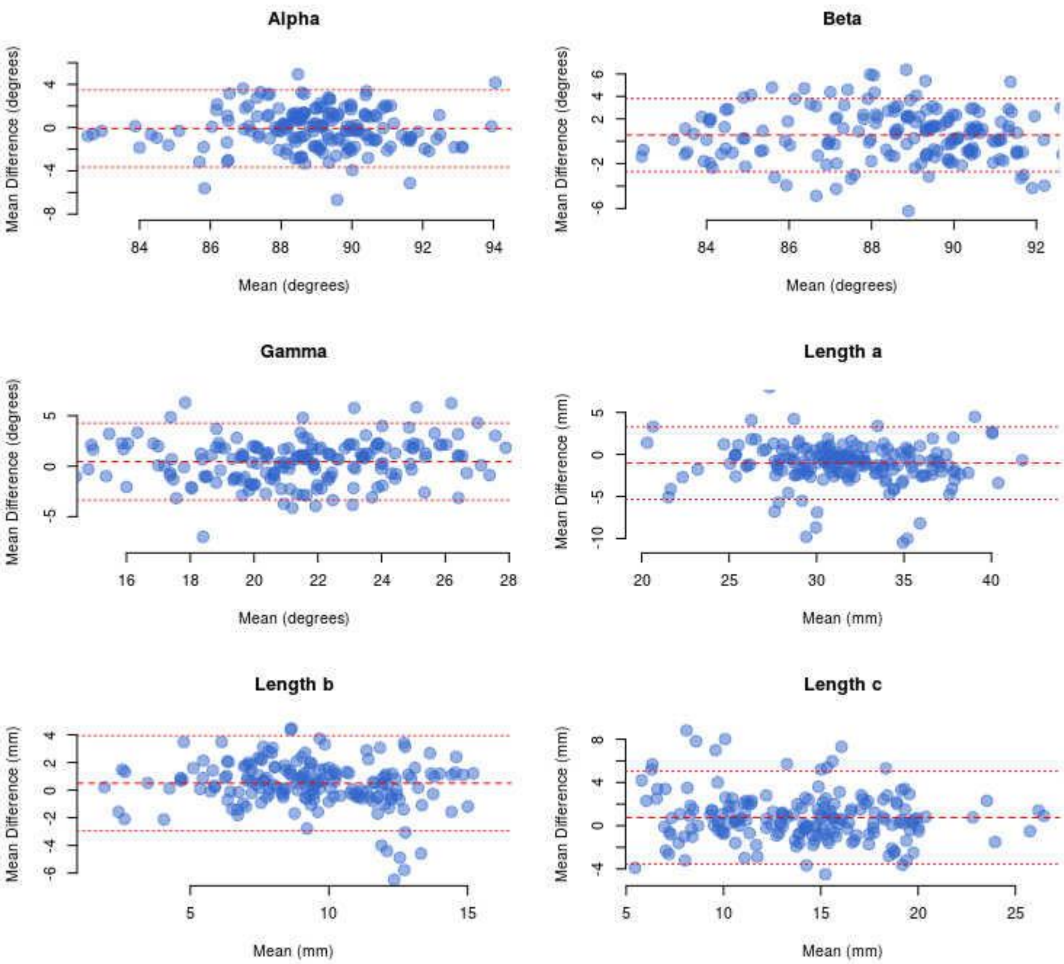

Figure 3

Bland-Altman (Tukey mean-difference) plots illustrating the interrater reliability between the raters' first observation of all six variables. There is no funneling of data which suggests that interrater variance is not affected by the magnitude of the measurement. The electronic goniometer measures the angular measurements to the nearest whole degree which is why the data points for α, β and γ are all evenly spaced. The linear measurements were recorded to 2 decimal points producing overlapping data points.