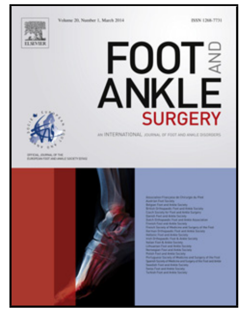
Accepted Manuscript



Title: Inter- and intraclass correlations for three standard foot radiographic measurements for plantar surface angles. Which measure is most reliable?

Authors: Erik Hohmann, Jolandie Myburgh, Natalie Keough, Kevin Tetsworth, Vaida Glatt

PII: S1268-7731(18)30138-3
DOI: https://doi.org/10.1016/j.fas.2018.07.007
Reference: FAS 1204

To appear in: *Foot and Ankle Surgery*

Received date: 11-4-2018
Revised date: 1-7-2018
Accepted date: 16-7-2018

Please cite this article as: Hohmann Erik, Myburgh Jolandie, Keough Natalie, Tetsworth Kevin, Glatt Vaida.Inter- and intraclass correlations for three standard foot radiographic measurements for plantar surface angles.Which measure is most reliable?.*Foot and Ankle Surgery* (2018), https://doi.org/10.1016/j.fas.2018.07.007

Inter- and intraclass correlations for three standard foot radiographic measurements for plantar surface angles. Which measure is most reliable?

Erik Hohmann* , Jolandie Myburgh , Natalie Keough , Kevin Tetsworth , Vaida Glatt

Erik Hohmann*

FRCS, FRCS (Tr&Orth), MD, PhD

Medical School, University of Pretoria, South Africa

Valiant Clinic/Houston Methodist Group

Jolandie Myburgh

PhD

Department of Anatomy, School of Medicine, University of Pretoria, South Africa

Natalie Keough

PhD

Department of Anatomy, School of Medicine, University of Pretoria, South Africa

Kevin Tetsworth

MD, FRACS

Department of Orthopaedic Surgery, Royal Brisbane Hospital, Herston, Australia

Department of Surgery, School of Medicine, University of Queensland, Australia

Orthopaedic Research Centre of Australia, Brisbane, Queensland, Australia

Vaida Glatt

PhD

Department of Orthopaedic Surgery, University of Texas Health Science Center, San Antonio, TX, USA

***Corresponding Author:**

Erik Hohmann

ehohmann@houstonmethodist.org

Valiant Clinic/Houston Methodist Group

PO Box 414296, Dubai, United Arab Emirates

Declarations of interest:
none

Highlights

- Intra- and inter-class correlations for three commonly used for plantar surface angles were high.
- For intra-and inter-class correlations the length-height index is the most reliable variable
- LHI represents the best, most reliable, and reproducible measure of arch height

- Calcaneal and calcaneal pitch angle were only of moderate value because of observer error

Abstract

The purpose of this study was to evaluate the reliability and reproducibility of three commonly used radiographic measures for plantar surface angles on 10 healthy male volunteers. The calcaneal angle (CA), calcaneal pitch angle (CPA), and length-height index (LHI) was measured by three independent examiners on two occasions on lateral foot radiographs. Intra- and inter-rater correlations were calculated using a general linear estimate model and post-hoc tests for repeated measures. Bland-Altman's Plots with limits of agreement were used for observer differences in scores. The intra-class correlations for the CA ranged from 0.91-0.94, for the CPA from 0.93-0.98, and for the LHI from 0.96-0.97. The inter-class correlations were 0.80 for CA, 0.83 for CPA and 0.93 for LHI. The results of this study strongly suggest that the length-height index was the most consistent and reliable measure for arch height.

Level of Evidence:

Diagnostic Level II, validity

Key Words:

Arch height; radiological arch measures; **calcaneal angle; calcaneal pitch angle**

Introduction

Structural abnormalities, both congenital and acquired, are considered predisposing factors to injury of the lower limb, particularly the foot. The biomechanical function of the foot is highly dependent upon and influenced by the structure of the foot (1). The medial arch is an important concept, and arch height has considerable functional relevance (2). Plantar loading shows significant differences between flat and normal feet (3), and this could result in injury (4,5,6). Static assessment of the foot is common clinical practice, but the lack of repeatability and reliability are a concern (7,8,9). Numerous anthropometric measures have been suggested over the last three decades ranging from rearfoot angle, medial longitudinal arch angle, navicular drop, and footprint, but the 'gold standard' remains the lateral weight-bearing radiograph (2,7,10). Digital photographic measurements have recently been shown to be highly reliable and valid (11). However, this study failed to compare its findings against the gold standard, and therefore lacks construct validity.

Corrective procedures are commonly performed for a variety of disorders affecting the foot, including acquired and congenital flatfeet in both children and adults, as well as post-traumatic, neurological, and other conditions.. For surgical planning prior to arthrodesis and deformity correction it is essential to have reliable and valid radiological measures available (12,13,14,15). The accuracy and reproducibility of these measurements not only aid in monitoring the progression of a particular deformity, but also allow the surgeon to decide on the indications for a procedure, select the type of treatment and then evaluate the success of the intervention [9,27]. Test-retest reliability for various radiographic and anthropometric variables has been published by several authors (2,10,13,15,16), and showed only moderate correspondence with large errors for several variables. In general, intra-observer reliability is consistently stronger than inter-observer reliability (13,16,17,18,19).

When performing a lateral foot radiograph the position of the foot, beam angle, distance, and plate position are all critical factors, and the tibiotalar articular surfaces should be parallel with no extrusion of the talus (19,20). The purpose of this study was to therefore evaluate the reliability and reproducibility of three commonly used radiographic measures for plantar surface angles. The hypothesis was that all measured measures would have very similar inter-observer and intra-observer reliability.

Methods

The study protocol was approved by the hospital IRB, and conforms to the international standards of the Helsinki declaration. Ten healthy male volunteers gave their informed consent to participate. Their mean age was 23.6 ± 4.3 years, their mean height 1.81 ± 0.06 m, and their mean weight 80 ± 11.5 kg. Females were specifically excluded to avoid bias introduced by the hormonal effect on soft-tissue during the different phases of the menstrual cycle (21). Individuals with a history of foot and ankle injuries, previous surgery to the lower extremity, or known congenital or acquired deformity such as club foot or hallux valgus were excluded.

The method involved positioning the right leg in a custom-made lower limb positioning device (LLPD) (Figure 1). This device was constructed of wood and nylon, and was developed specifically to position the foot and lower limb. **The right leg was placed in the device and adjustable lateral and medial knee supports were used to position the foot (Figure 1).** When situated, a lateral radiograph of the foot was taken by the same radiographer for all ten subjects. This device was used to reduce measurement bias and allow reproducible positioning for all radiographs. Distance and angle of exposure were standardised to minimise parallax errors, and the location of the positioning device remained unchanged throughout acquisition of all images.

The images of each volunteer were digitally duplicated ten times, and then randomly sorted into an image file by an independent research associate. Three independent examiners were then presented a total of 100 randomly sorted radiographs, and instructed to measure the calcaneal angle (CA), the calcaneal pitch angle (CPA), and the length height index (LHI) on each radiograph using AutoCAD™ 2000 and Digimizer. Each examiner was asked to perform these measures independently. Each examiner measured 100 randomly sorted radiographs on three separate occasions. Written step-by-step instructions were given to each examiner to reduce error and measurement bias, and the three assessments were separated by seven days between occasions.

Three specific measures were selected from the standard German radiology reference (22):

Calcaneal Angle (CA): A line was drawn from the most inferior point of the calcaneus to the most inferior point of the first metatarsal bone (weigh-bearing line - WBL). The second line was drawn line midway through the long axis of the body of the calcaneus. The angle (alpha) between the intersection of these two lines was determined by the software and recorded (Figure 2). **To construct the midway line, an initial line was drawn from the most cephalic part of the posterior process of the calcaneus to the lower border of the calcaneal tubercle. A second line was drawn between the highest and lowest point of the calcaneal articular surface of the calcaneo-cuboid joint. Software then established the midline axis between these two preliminary lines defining the midway line.**

Calcaneal Pitch Angle (CPA): A line was drawn along the inferior border of the calcaneus. The angle (beta) between intersection of this line and the mid-calcaneal line drawn previously was determined by the software and recorded (Figure 2)

Length Height Index (LHI): Two lines were drawn perpendicular to the WBL, through the most posterior aspect of the calcaneus and the most anterior aspect of the first **metatarsal**. The distance between the intersections of these two lines on the WBL was measured and defined as the length. A third vertical perpendicular line to the WBL was drawn from the most superior aspect of the navicular bone, in contradistinction to Hellinger who recommends using the midpoint of the navicular (12). The most superior point of the navicular is easier to identify and thus reduces bias, while not substantially influencing the relationship between length and height. The distance between the intersection with the WBL and the most superior aspect of the navicular was therefore defined as the height (Figure 2). Length-height ratio was calculated by dividing foot length by navicular height (a/b).

Descriptive statistics (means and standard deviations) for the three independent variables were calculated for all test events. Normality was analysed using the Shapiro-Wilks test. Intra- and inter-rater correlations were assessed by using a general linear estimate model and post-hoc tests for repeated measures, and the 95% confidence intervals for intra- and inter-rater reliability (ICC) were calculated (23,24,25). Bland-Altman's Plots with limits of agreement (LOA) were then utilized to evaluate the agreement between observers' scores (26). The algorithm of Landis and Koch was used to assess the rate of agreement. Values above 0.80 represented excellent agreement, values between 0.62-0.79 were considered good agreement, values between 0.41-0.61 indicated moderate agreement, and values below 0.4 suggested fair to poor agreement (25). All analyses were conducted using STATA SE (Version 12.0; StataCorp, College Station, Texas, USA) for Windows, and the comprehensive meta-analysis software package (CMA), version 3 (Biostat Inc, Englewood, NJ, USA).

Results

Intra-class correlation and intra-observer error analysis

Calcaneal Angle (CA): Intraclass correlation coefficients for intraobserver measurements for CA were high and ranged from 0.91to 0.94, and the 95% confidence intervals were less than 30% indicating a good model fit (Table 1). For observer 2 a lower confidence interval (36%) was noted. After removing one outlier, the lower level confidence interval increased to 0.69 and was now within the 30% range of a good model fit (26%). With regard to the Bland-Altman plot, it demonstrates the limits of agreement are wide and there were three outliers, indicating moderate to ambiguous results. However, the even distribution of variables around the bias line displays consistent variablity, signifying a moderately useful measurement (Figure 3).

Calcaneal Pitch Angle (CPA): Intraclass correlation coefficients for intraobserver measurements were high and ranged between 0.93-0.98, and the 95% confidence intervals were less than 30% indicating a good model fit (Table 2). Similar to the calcaneal angle measures, one observer had a lower confidence interval (36%). There was no obvious outlier and the Bland-Altman plot illustrates narrow limits of agreement with no outliers and an even distribution of variables around the bias line, indicative of a repeatable measurement (Figure 4).

Length Height Index (LHI): Intraclass correlation coefficients for intraobserver measurements were high and ranged between 0.96-0.97, and the 95% confidence intervals were 21% and less indicating a good model fit (Table 3). The Bland-Altman plot indicates a very narrow limit of agreement, no outliers, and an even distribution around the bias line (Figure 5). In contrast to both CA and CPA, this measurement was the most reliable of the three tested.

Inter-class and inter-observer error analysis

The inter-class correlation coefficients for inter-observer measurements had values above 0.80 representing excellent agreement. They ranged from 0.80 for CA, to 0.83 for CPA, and 0.93 for LHI (Table 4). However, the 95% confidence intervals for CA were wide suggesting poor agreement. The Bland-Altman graphs also indicate wide limits of agreement for the CA, an uneven distribution of the variables around the bias line, and outliers, indicating a high likelihood of inter-observer error (Figure 6). The 95% confidence intervals for CPA were narrow (<15%), indicating a good model fit and excellent agreement. However, the Bland-Altman graphs demonstrate wide limits of agreement and uneven distribution of the variables around the bias line, indicating inter-observer error (Figure 7). The highest inter-class correlation coefficient was observed for LHI.(0.93), with narrow 95% confidence intervals of less than 10% (Table 4). The Bland-Altman graphs demonstrate a narrow limit of agreement and even distribution of the variables around the bias line, indicating low inter-observer error (Figure 8). The LHI proves to be the best measure, as it demonstrates a narrow agreement limit and an even distribution of variables around the bias line.

Discussion

The most important finding of this study was that the intra-class correlation for all three measures were high (>0.9), with narrow confidence intervals indicating a good model fit. The intra-observer error was lowest for the length-height-index (LHI), suggesting LHI is the most reliable variable by intra-class criteria. Similarly, for the inter-observer correlations all three measures demonstrated ICC values above 0.8, and was again highest for LHI with narrow agreement limits. Therefore, LHI represents the best, most reliable, and reproducible of the three variables to measure arch height in the foot on weight-bearing radiographs.

Saltzman et al. investigated the reliability of several radiographic measures, and demonstrated very high intra- and inter-rater correlation coefficients (10). They reported an ICC of 0.99 for the calcaneal pitch angle with very narrow 95% confidence intervals (10). These results are similar to our findings, and confirm that the calcaneal pitch angle the most reliable and reproducible measurement tested. Menz et al. also demonstrated very high ICC's for CPA of 0.99 with very narrow confidence intervals using 20 randomly selected radiographs in older individuals with a 1 month re-test interval (2). However, both Menz et al. and Saltzman et al. (2,10) failed to account for observer error, and Saltzman et al. (10) measured all radiographs only twice with a thirty minute interval between sessions. In contrast, our study randomly measured all radiographic variables ten times and increased the between-session interval to seven days. This approach is most likely more sensitive, and reduces both measurement and examiner bias. The intra- and inter- observer error analysis demonstrated that CPA is a good measure if one individual examiner repeatedly uses this variable. However, observer error for inter-rater correlations was present, and CPA is perhaps less useful than previously believed. Saltzman et al. have also measured arch height and foot length but have not calculated LHI (10). Similar to CPA, very high coefficients at 0.99 were observed. The LHI was the most reliable measure in our study, and the observer error analysis indicated low intra- and inter-observer error. The LHI may therefore be considered the most useful, reliable, and reproducible variable.

Despite applying rigid criteria and reducing positional error by placing the foot into the LLPD for the acquisition of the radiographs, a significant difference in ICC was not observed when comparing our findings to Saltzman et al. and Menz et al. (2,10). Tochigi et al. have demonstrated a measurement error of 2.2% with every 10 degrees of ankle malposition for lateral ankle radiographs (27). Miller et al. demonstrated there is substantial variability with weight-bearing applied during radiographs, and questioned the reliability and interpretation when interpreting these radiographs (28). A positioning device could reduce variations during acquisition of weight-bearing radiographs, but it appears other factors also influence the reliability and reproducibility of radiographic arch height measures. Limb position may therefore be less important than previously believed, provided it is within acceptable limits. It has also been suggested that experienced surgeons may reach a higher level of agreement, and Guo et al. could not demonstrate that precise measurements of weight-bearing lateral ankle views were not influenced by the surgeon’s experience (19).

Several clinical variables to characterize the foot arch have been described previously, ranging from static ink measures to the foot posture index as a simple quantification of static foot alignment (7,29). However, static footprint measures are controversial and lack validity (30,31), and the clinical measures for static foot posture do not correlate well between classifications (7). In fact Cornwall et al. suggested that one commonly used index, the modified foot posture index, should be used with extreme caution and may have limited value (32).

Static foot alignment is one of the factors that could increase the risk of injury and is frequently assessed in clinical practice (7). The correlations between the clinical and radiographic measures are at best moderate, and range from 0.24 to 0.74 (7,31). Radiographic assessment requires specialized equipment and exposure to radiation, but remains the current 'gold standard'. The reliability of this standard, the weight-bearing lateral radiograph, is crucial for the interpretation of the findings for both clinical practice and research (29). Previous authors have demonstrated very high ICC values, and these were confirmed by the current study. However, as the intra- and inter-observer agreement for these measures only demonstrates low errors for the LHI, the reliability and validity of the other two angular measurements (CA and CPA) should be viewed with caution.

At this stage, the use of positioning devices for reproducible and reliable lateral weight-bearing radiographs is a research tool only, and is not yet being routinely used in clinical medicine. Lateral weight-bearing foot radiographs are typically obtained with the patient standing erect, supporting the cassette. However both foot and cassette position are not standardised within and between radiology units, and one can safely assume that the reliability and reproducibility of these radiographs is rather low. Our results have demonstrated very high intraclass correlations with narrow 95% confidence intervals, and lower interclass correlations with wider 95% confidence intervals. If radiographs are not obtained in a standardised, reproducible, and reliable fashion the ICC measures will most likely be very low, not allowing between and within

patients comparisons as the errors introduced by variable cassette and foot position will not allow reliable data collection. We would, therefore, encourage the introduction of a reliable foot positioning device for all clinical applications.

This study has several limitations. The study was limited to three variables and did not include other commonly described measures such as arch index, calcaneal metatarsal angle, navicular height, and talonavicular coverage angle. However, the ICC and limits of agreement for most of these measures were reported to be moderate 931). The included participants were recruited from a young and healthy population, and it is possible these findings may not apply to individuals with intrinsic pathology and older patients with other diseases, limiting the external validity. Radiographic techniques were standardized and an experienced radiographer performed all examinations, and poor technique may contribute to diagnostic error. The experience of the individual researcher also could have influenced inter-rater correlations. However, McLaughlin et al. demonstrated there were no differences between novice and experienced examiners when measuring the foot posture index (33). Finally, use of the LLPD has yet to be validated, and the improved results here could be interpreted as the ability of the LLPD to assist in producing reliable and reproducible radiographs. However, it is already established that good quality radiographs, with the tibiotalar articular surfaces overlapping with no talar extrusion, are critical to perform high quality studies (11).

Conclusions

The results of this study strongly suggest that the length-height index was the most consistent and reliable measure for arch height. The correlations were above 0.9 for both intra- and interclass reliability, with narrow confidence intervals and low observer error. For the other measures the correlations were also excellent, but the calculated observer error suggested only a moderate value for intra- and interclass reliability for both the calcaneal angle and the calcaneal pitch angle.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

[1] McCrory JL, Young MJ, Boulton AJM, Cavanagh PR. Arch index as a predictor of arch height. The Foot, 7: 79-81, 1997

[2] Menz HB, Munteanu SE. Validity of 3 clinical techniques for the measurement of static foot posture in older people. J Orthop Sports Phys Ther; 35:479-486, 2005

[3] Queen RM, Mall NA, Nunley JA, Chuckpaiwong B. Differences in plantar loading between flat and normal feet during different athletic tasks. Gait Posture; 29 (4):582-586, 2009

[4] Kaufman KR, Brodine SK, Shaffer RA, Johnson CW, Cullison TR. The effect of foot structure and range of motion on musculoskeletal overuse injuries. Am J Sports Med; 27 (5):585-593, 1999

[5] Michelson JD, Durant DM, McFarland E. The injury risk associated with pes planus in athletes. Foot Ankle Int; 23 (7):629-633, 2002

[6] Tong JW, Kong PW. Association between foot type and lower extremity injuries: systematic literature review with meta-analysis. J Orthop Sports Phys Ther; 43 (10):700-714, 2013

[7] Langley B, Cramp M, Morrison SC. Clinical measures of static foot posture do not agree. J Foot Ankle Res; 9:45, 2016

[8] Razeghi M, Batt ME. Foot type classification: a critical review of current methods. Gait Posture; 15 (3):282-291, 2002

[9] Williams DS, McClay IS. Measurements used to characterize the foot and the medial longitudinal arch: reliability and validity. Phys Ther; 80 (9):864-871, 2000

[10] Saltzman CL, Nawoczenski DA, Talbot KD. Measurement of the medial longitudinal arch. Arch Phys Med Rehabil;76:45-49, 1995

[11] Cobb SC, James CR, Hjertsedt M, Kruk J. A digital photographic measurement method for quantifying foot posture: validity, reliability and descriptive data. J Athl Train; 46 (1):20-30, 2011

[12] Bourdet C, Seringe R, Adamsbaum C, Glorion C, Wicart P. Flatfoot in children and adolescents: analysis of imaging findings and therapeutic implications. Orthop Traumatol Surg Res; 99 (1):80-87, 2013

[13] Evans AM, Copper AW, Scharfbiling RW, Scutter SD, Williams MT. Reliability of the foot posture index and traditional measures of foot position. J Am Podiatric Med Assoc, 93(3): 203-213, 2003

[14] Lin YC, Mhuircheartaigh JN, Lamb L, Kung J. Imaging of adult flatfoot: correlation of radiographic measurements with MRI. Am J Radiol; 204:354-359, 2015

[15] Saltzman Cl, Brandser EA, Berbaum KS, DeGnore L, Holmes JR, Katcherian DA, Teasdall RD, Alexander IJ. (1994) Reliability of standard foot radiographic measurements. Foot Ankle Int; 15(12): 661-665, 1994

[16] Sell KE, Verity TM, Worrell TW, Pease BJ, Wigglesworth J. Two measurement techniques for assessing subtalar joint position: a reliability study. J Orthop, Sports Phys Ther; 19:162, 1994

[17] Astrom M, Arvidson T. Alignment and joint motion in the normal foot. J Orthop Sport Phys Ther, 22:216, 1995

[18] Diamond JE, Mueller MJ, Delitto A, Sinacore DR. Reliability of a diabetic foot evaluation. Physical Therapy, 69:797-802, 1989

[19] Gou C, Zhu Y, Deng L, Xu X. Reliability of measurements on lateral ankle radiographs. BMC Musculoskelet Disord; 17:297, 2016

[20] Knupp M, Bollinger L, Hintermann B. Treatment of posttraumatic varus ankle deformity with supramalleolar osteotomy. Foot Ankle Clin. 17 (1):95-102, 2012

[21] Eiling E, Bryant AL, Petersen W, Murphy A, Hohmann E. Effects of menstrual-cycle hormone fluctuations on musculotendinous stiffness and knee joint laxity. Knee Surg Sports Traumatol Arthrosc; 15:126-132, 2007

[22] Hellinger J. Messmethoden in der Skelettradiologie: Linien, Distanzen, Winkel und ihre klinische Bedeutung. Stuttgart; New York: Thieme 1995. ISBN 3-13-119701-3

[23] De Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol 59:1033–1039, 2006

[24] Fleiss JL: Statistical Methods for Rates and Proportions. 3rd edition. Hoboken, New York, John Wiley & Sons; 2003.

[25] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics, 33(1):159-174, 1977

[26] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet;1:307–310, 1986

[27] Tochigi Y, Suh JS, Amendola A, Pedersen DR, Saltzman CL. Ankle alignment on lateral radiographs. Part 1: sensitivity of measures to perturbations of ankle positioning. Foot Ankle Int; 27 (2):82-87, 2006

[28] Miller CP, Ghorbanhoseini M, Ehrlichmann LK, Walley KC, Ghaheri A, Kwon JY. High variability of observed weight bearing during standing foot and ankle radiographs. Foot Ankle Int; 38 (6):690-693, 2017

[29] Terada M, Wittwer AM, Gribble PA. Intra-rater and inter-rater reliability of the five image-based criteria of the foot posture index. Int J Sports Phys Ther; 9 (2):187-194, 2014

[30] Mathieson I, Upton D, Prior TD. Examining the validity of selected measures of foot type: a preliminary study. J Am Podiatr Med Assoc; 94:275-281, 2004

[31] Murley GS, Menz HB, Landorf KB. A protocol for classifying normal- and flat-arched foot posture for research studies using clinical and radiographic measurements. J Foot Ankle Res; 2:22, 2009

[32] Cornwall MW, McPoil TG, Lebec M, Vicenzino B, Wilson J. Reliability of the modified foot posture index. J Am Podiatr Med Assoc; 98 (1):7-13, 2008

[33] McLaughlin P, Vaughan B, Shanahan J, Martin J, Linger G. Inexperienced examiners and the foot posture index: a reliability study. Man Ther; 26:238-240, 2016

Figure Legends

Figure 1

Lower limb positioning device (LLPD): A) lateral view of the adjustable medial and lateral supports to position the tibia with the knee at 30º of flexion; B) posterior oblique view of the adjustable supports at the foot (on posterior and medial aspects on right foot); C) lateral view of right foot indicating the medial and posterior supports for the foot. **The figure here is used for demonstration purposes only; females subjects were not included in this study.**

Figure 2

Lateral radiograph taken of the right foot positioned in the LLPD (lower limb positioning device) indicating the three measurements taken: α - calcaneal angle (CA); β - calcaneal pitch angle CPA); a/b – length-height index (LHI)

Figure 3

Bland-Altman graph indicating the limits of agreement between each observer for the Calcaneal Angle (CA) measurement. Three outliers indicate moderate to ambiguous results and the even distribution the bias line displays consistent variablity, signifying a moderately useful measurement.

Figure 4

Bland-Altman graph indicating narrow limits of agreement between each observer for the Calcaneal Pitch Angle (CPA) measurement. The even distribution around the bias line indicate repeatable measurement.

Figure 5

Bland-Altman graph showing very narrow limits of agreement between each observer for the Length-Height Index (LHI) measurement. This measurement was the most reliable of the three tested.

Figure 6

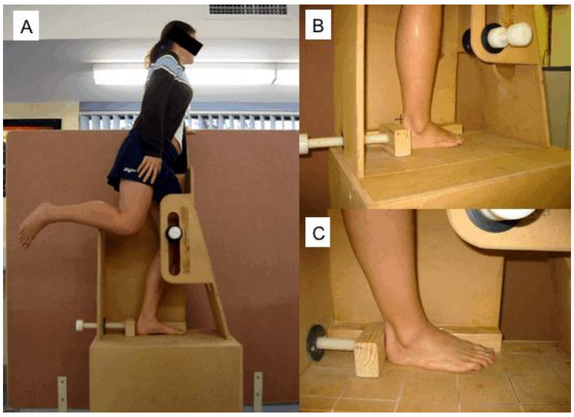
Bland-Altman graph demonstrate wide limits of agreements between the observers for the Calcaneal Angle (CA) measurement. The uneven distribution around the bias line indicate a high likelihood of inter-observer error.

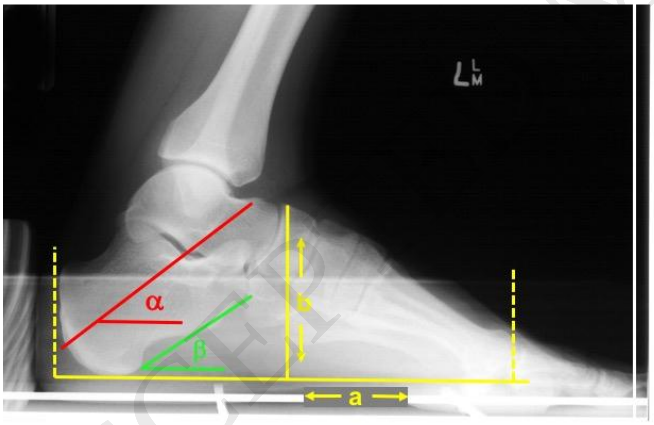
Figure 7

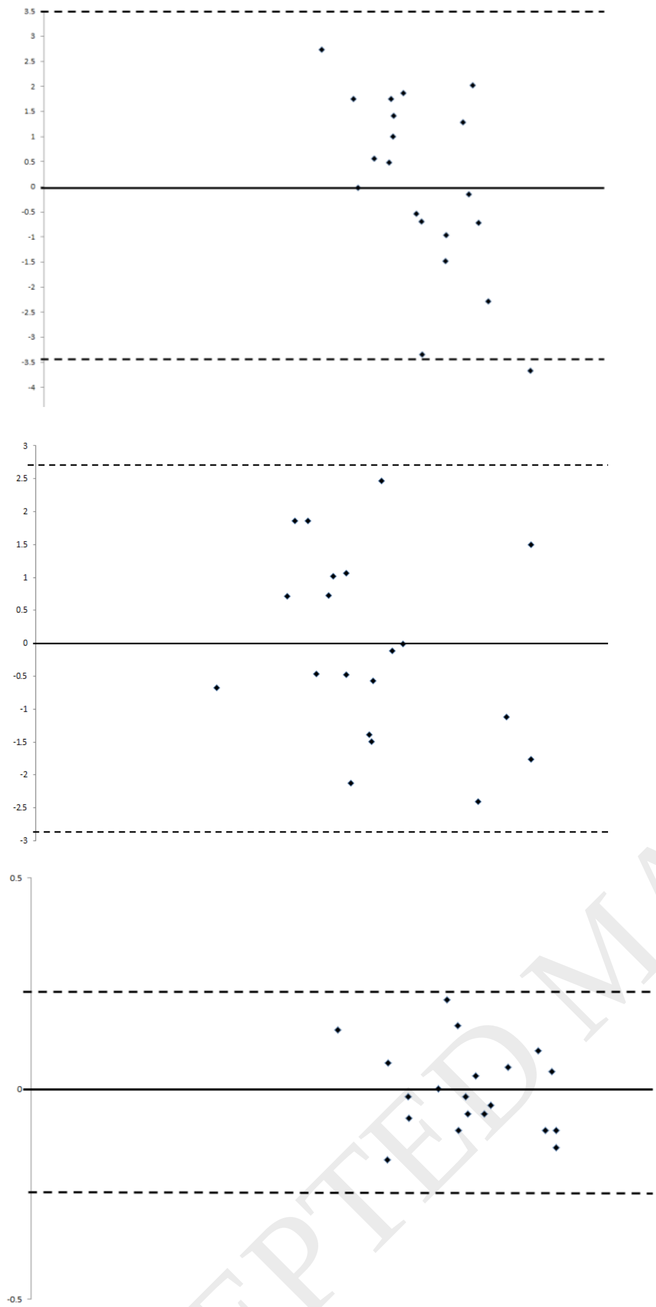
Bland-Altman graph showing wide limits of agreement between the observers for the Calcaneal Angle (CA) measurement. The uneven distribution around the bias line indicate inter-observer error.

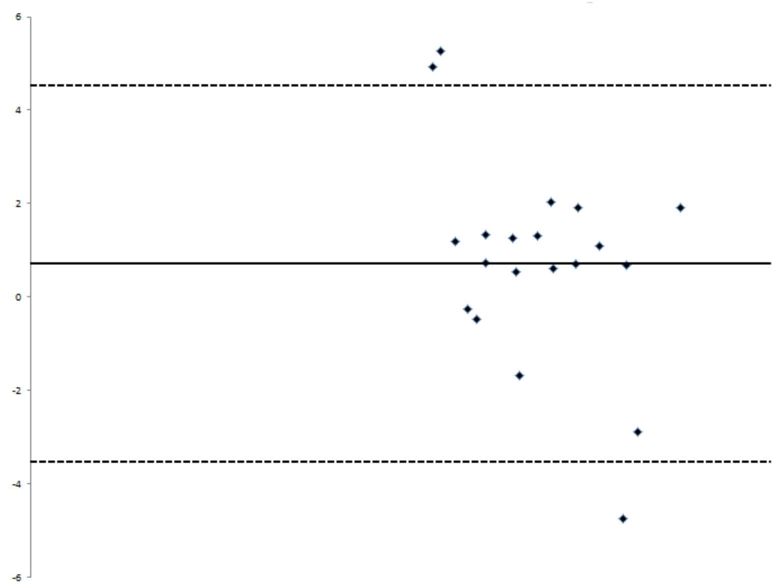
Figure 8

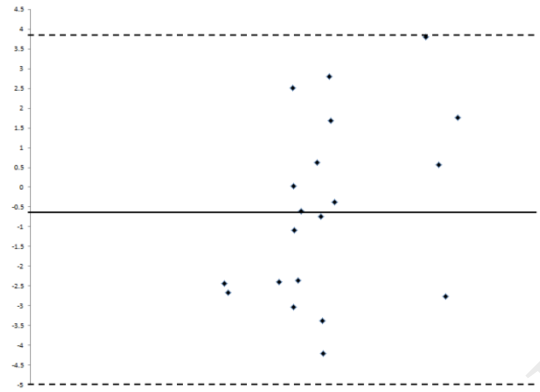
Bland-Altman graph demonstrate a narrow limit of agreement between the observers for the Calcaneal Angle (CA) measurement. The uneven distribution around the bias line indicate inter-observer error.

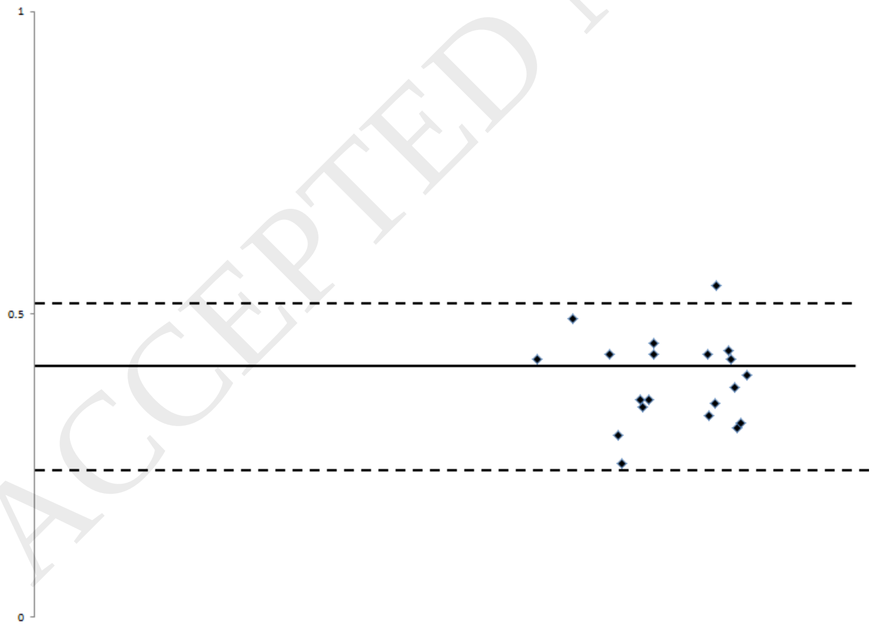





3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
-4
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
0.5
0
-0.5


6
4
2
0
-2
-4
-6


4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
-4
-4.5
-5


1
0.5
0

Tables

Table 1: ICC Results for the calcaneal angle measurements

	Intraclass	95% CI	F	p
Observer 1	0.94	0.77-0.99	133.73	0.01
Observer 2	0.93	0.59-0.97	118.8	0.0001
Observer 3	0.91	0.84-0.99	291.81	0.0001

Table 2: ICC results for the calcaneal pitch angle measurements

	Intraclass	95% CI	F	p
Observer 1	0.95	091-0.99	185.37	0.0001
Observer 2	0.98	0.88-0.99	446.28	0.0001
Observer 3	0.93	0.61-0.97	114.92	0.0001

Table 3: ICC results for the length-height-ratio

	Intraclass	95% CI	F	p
Observer 1	0.96	0.77-0.99	233.95	0.0001
Observer 2	0.97	0.76-0.99	314.43	0.0001
Observer 3	0.97	0.75-0.98	294.1	0.0001

Table 4: ICC results for inter-observer reliability for the three measurements between the 3 independent observers

	Interclass	95% CI	F	p
Calcaneal Navicular	0.80	0.39-0.87	78.19	0.0001
Calcaneal Pitch	0.83	0.75-0.95	245.3	0.0001
Length Height Ratio	0.93	0.85-0.98	245.3	0.0001