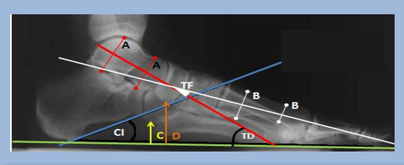
A reliability study to investigate inter- and intra-rater repeatability of plain film radiographic measurements pertinent to foot posture assessment

Lisa H *Westcott* Chandler BSc(Hons) Cert. Ed. MSc; The University of Northampton
Dr Simon Otter BSc(Hons) PGCert MSc PhD; The University of Brighton

Statement of Purpose



- This study aimed to investigate the reliability of a range of radiographic foot measurements reported in the literature.
- The reliability of a series of lateral view and anteroposterior (AP) radiographic measurements in relation to foot posture assessment was investigated.
- The principle objective for the study was to validate a repeatable method of radiograph charting for future research in determining post surgical outcomes following arthroereisis surgery for adult acquired flatfoot.

Literature Review

Radiographic studies are widely reported in the literature in relation to surgical outcomes for deformities including adult-acquired flatfoot. The multiple radiographic measurements used to quantify flatfoot deformity vary amongst authors and 'normal' mean and range values in adult populations have been shown to differ [1-4]. There is no consistently used measurement approach for flatfoot and a lack of a definitive classification system has previously been blamed [5].

Radiographic measurement repeatability is affected by factors including: radiographer technique, consistency of measurement, style of measurement, number of measurers/observers involved, time delay between measurements, complexity of angle measurement, X-ray resolution, and clarity of anatomical reference points. Metcalfe *et al.* [6] highlight that subjective human error always exists in measuring radiographic angles regardless of the use of plain film or digital images.

Methodology

Five AP and five lateral view electronic radiographs were printed and duplicated, yielding 3 copies of each radiograph for two measurers. Each was charted independently with a minimum of a seven day period between charting and measurers were blinded to each other's results.

Radiographic measurements were recorded in separate tables at each occasion to blind measurers from their own previous measurements and those of the other measurer.

- **Null hypothesis 1:**
  H0: the gauge cannot produce the same results when measurements are taken on separate occasions.
- **Alternative hypothesis 2:**
  H1: the gauge can produce the same measurement when taken at different times by the same measurer.

Quantitative analysis aimed to investigate intra-rater (between the same measurer) and inter-rater reliability (between different measurers) of each of the measurements listed in Figure 1. Each measurer was provided with a description and schematic of each measurement (Figure 2) to ensure a consistent approach to charting.

X Ray Projection	Measurements Tested	Abbreviation	Gauge
Lateral	Calcaneal inclination angle [°]	CI	Tractograph
	Talar declination angle [°]	TD	Tractograph
	Talo-calcaneal angle [°]	TC	Tractograph
	Talar-first metatarsal angle [°]	T1st	Tractograph
	Calcaneal-first metatarsal angle [°]	C1st	Tractograph
	Navicular height [mm]	Nav	Ruler
	Cuneiform height [mm]	Cn	Ruler
	Cuboid height [mm]	Cb	Ruler
	Foot length [mm]	Foot len	Ruler
AP	Talonavicular angle [°]	TN	Tractograph
	Talar-first metatarsal angle [°]	T1 (AP)	Tractograph
	Calcaneal-fifth metatarsal angle [°]	C5th	Tractograph

Figure 1: Lateral and anteroposterior view measurements tested

Results were imported into an Excel spreadsheet. Microsoft Excel 2007 and IBM SPSS Version 20 were used for quantitative data analysis.

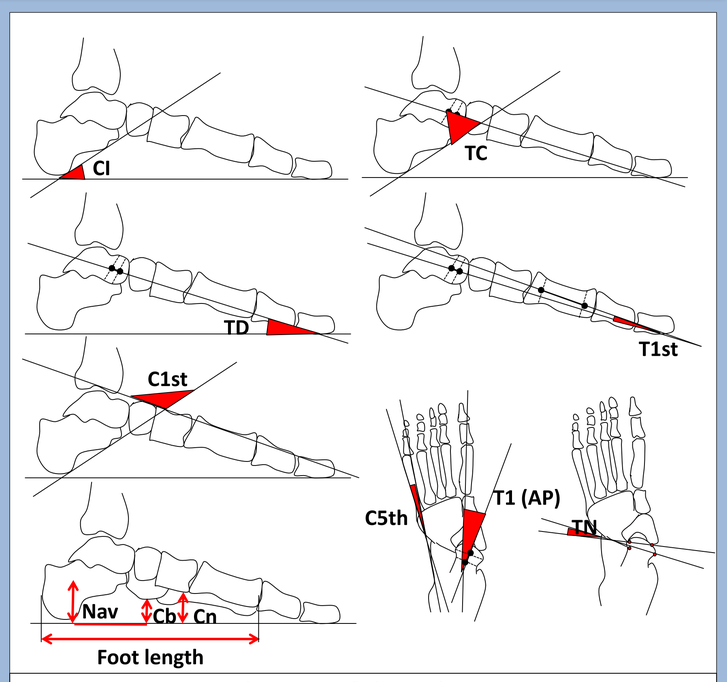


Figure 2: Schematic of charted measurements

Results

Coefficient of determination (R^2) values were calculated for each measurement to quantify the degree of relationship between Measurer X compared to Measurer Y. The ideal relationship would yield $R^2 = 1$.

In a repeatability and reproducibility (gauge R&R) context, analysis of variance (ANOVA) was used to identify the relative variation caused by the gauge and the measurer. Gauge R&R studies quantify the precision errors of a measurement system to determine its acceptability. This ensured any variation in results was representative of either the *repeatability* of the gauge (measuring equipment), the *reproducibility* of the measurements by different people, or natural (random) variation.

Figure 3: R^2 and Gauge R&R results

Measurement	Inter-rater coefficient of determination R^2	Gauge R&R
CI	0.817	28%
TD	0.574	51%
TC	0.643	47%
T1st	0.518	49%
C1st	0.958	15%
Nav	0.96	12%
Cn	0.921	39%
Cb	0.903	20%
Ft len	0.993	7%
TN	0.621	56%
T1 (AP)	0.63	54%
C5th	0.022	91%

Acceptability criteria for a gauge R&R study [7]:
- **< 10% error:** gauge system ideal (excellent)
- **10-20% error:** gauge system may be acceptable based upon
- importance of application (good)
- **20-30% error:** gauge system is marginal and will cause problems
- **>30% error:** gauge system is poor and needs improvement

Calcaneal inclination scored R^2 of 0.817 (p=0.01) and ANOVA gave acceptable reproducibility between measurers. Calcaneal-first metatarsal angle (Figure 4) scored favourably on gauge R&R and strong correlation was observed between measurers (R^2 = 0.958).

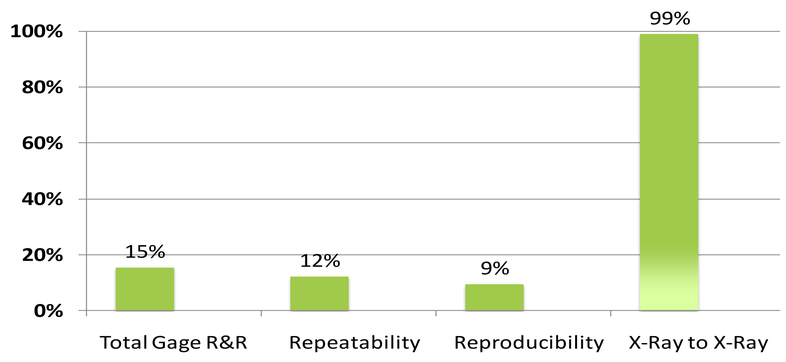
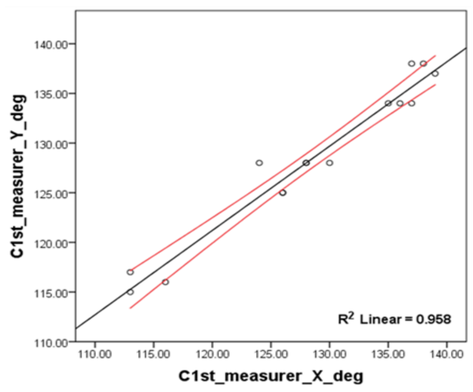


Figure 4: Calcaneal first metatarsal lateral view
LEFT: analysis of variance
RIGHT: inter measurer correlation using R^2



Navicular height between measurers correlated strongly (R^2 = 0.960) as did foot length (R^2 = 0.993), cuneiform height (R^2 = 0.921) and cuboid height (R^2 = 0.903), exhibiting excellent intra- and inter-rater reliability. C5th angle (Figure 5) produced a poor correlation between measurers (R^2 = 0.022) and a high gauge R&R reading (91%) indicating that the measurement system is unreliable. All other lateral view measurements and AP view measurements were found to be unreliable based on these indicators (Figure 3).

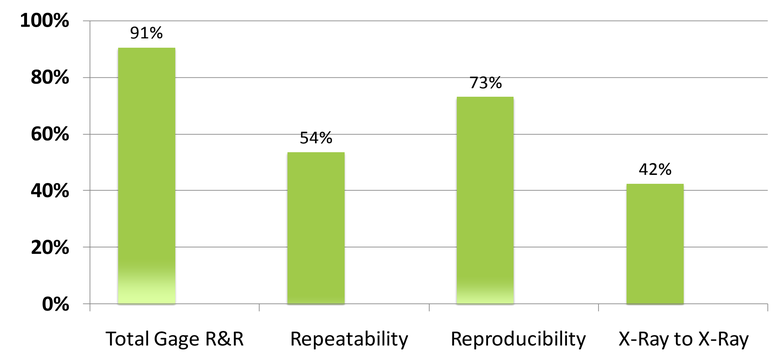
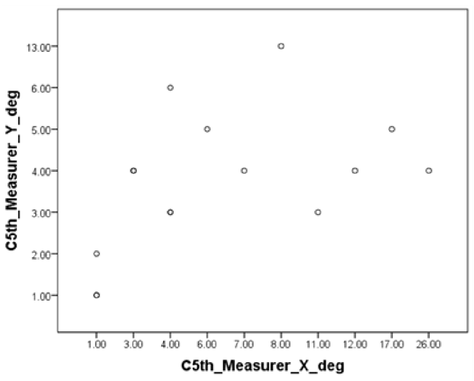


Figure 5: Calcaneal fifth metatarsal lateral view
LEFT: analysis of variance
RIGHT: inter measurer correlation using R^2



Discussion

- The measurement of calcaneal inclination, navicular height, cuboid height and foot length were reliable indicators for intra-rater repeatability and inter-rater (reproducibility). R^2 values for cuneiform height conform to published literature [3]. Foot length and navicular height were similarly reported as having high repeatability in a recent study [6].
- Angles from AP radiographs showed unacceptable variance between the same and different measurers; these angles are not recommended as reliable indicators in foot profile assessment. Talonavicular angle (R^2 = 0.621) although higher than in other research [3], still was not deemed high enough to demonstrate good inter-measurer correlation. These findings are similarly reported in research by Metcalfe *et al.* [6].
- Any angle requiring talus bisection scored unfavourably on inter- and intra-rater repeatability and gave poor R^2 scores amongst measurers.
- Angles with multiple stages did not produce favourable repeatability between the same and different measurers, conforming to the findings of other research [6].
- A new approach to talus bisection warrants further investigation and caution should be applied in reliance of such measurements, particularly to post-surgical outcome.

References

[1] DiGiovanni JE. Smith SD. Normal biomechanics of the adult rearfoot: a radiographic analysis. JAPMA 1976; 66(11): 812-24.
[2] 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 2009;2:22.
[3] Younger AS, Sawatzky B, Dryden P. Radiographic assessment of adult flatfoot. Foot Ankle Int 2005;26(10):820–5.
[4] Thomas JL, Kunkel MW, Lopez R, Sparks D. Radiographic values of the adult foot in a standardized population. J Foot Ankle Surg 2006;45(1):3–12.
[5] Haddad SL, Myerson MS, Younger A, Anderson, RB, Davis WH, Manoli A 2nd. Symposium: Adult Acquired Flatfoot Deformity. Foot Ankle Int 2011 32(1): 95-111.
[6] Metcalfe SA, Bowling FL, Baltzopolous V, Maganaris C, Reeves ND. The reliability of measurements taken from radiographs in the assessment of paediatric flat foot deformity. The Foot 2012 22(3):156-62.
[7] Brook Q. Six Sigma and MINITAB – A Tool Box Guide for Managers, Black Belts and Green Belts. QSB Consulting Ltd. 2004.

This study was granted ethical approval by the Faculty of Health and Social Science Research Ethics & Governance Committee, The University of Brighton.

FIP World Congress of Podiatry, Rome 17-19 October 2013

Contact: lisa.westcott@northampton.ac.uk
University of Northampton: 01604 892969