

The intra and inter-rater reliability of a modified weight-bearing lunge measure of ankle dorsiflexion

GRAFTON, Kate and O'SHEA, Simon

Available from Sheffield Hallam University Research Archive (SHURA) at:

<http://shura.shu.ac.uk/10374/>

This document is the author deposited version. You are advised to consult the publisher's version if you wish to cite from it.

Published version

GRAFTON, Kate and O'SHEA, Simon (2013). The intra and inter-rater reliability of a modified weight-bearing lunge measure of ankle dorsiflexion. *Manual Therapy*, 18 (3), 264-268.

Copyright and re-use policy

See <http://shura.shu.ac.uk/information.html>

Title

The intra and inter-rater reliability of a modified weight-bearing lunge measure
of ankle dorsiflexion.

Simon O'Shea MSc.

Kate Grafton MSc.

Email: simoncoshea@hotmail.com.

Correspondence: Physiotherapy Department, The Rotherham NHS Foundation
Trust, Moorgate Road, Rotherham, South Yorkshire, S60 2UD. Tel 01709 424400

Keywords: Reliability, lunge, ankle, dorsiflexion.

Abstract

This study assessed the intra and inter-rater reliability of a modified weight-bearing lunge measure of ankle dorsiflexion range of movement. Thirteen healthy subjects were recruited. Each subject performed 3 repetitions of the lunging method with one rater and 3 more repetitions with a second rater within 30 minutes. The process was repeated within 3 hours. Intra-rater reliability results indicated excellent correlation of measurements (Intraclass Correlation Coefficients (ICCs) of 0.98 to 0.99). Standard Error of Measurement (SEM), 95% Limits of Agreement (LOA) and Coefficient of Repeatability (CR) calculations indicated suitably low ranges of measurement variance (SEM = 0.4cm, LOA = ± 1.28 to ± 1.47 cm and CR = 1.21 to 1.35cm). Inter-rater reliability was also deemed excellent (ICC = 0.99, SEM = 0.3cm, LOA = ± 0.83 to ± 1.47 cm, CR = 1.44cm). The modified lunge technique therefore demonstrates excellent intra and inter-rater reliability.

Text

1

Introduction

2

3 Suitable ankle dorsiflexion range of movement (DFR) is needed for efficient
4 walking (Magee, 2008). Hypomobility of DFR is associated with pathologies including
5 tendonopathies (Kaufman et al., 1999) and fractures (Agosta and Morarty, 1999); and
6 restoring DFR is a common aim of rehabilitation following ankle fractures (Lin et al.,
7 2009) and sprains (Collins et al., 2004). Consistent measurement before and after
8 treatment is important so progress can be monitored.

9 Weight-bearing DFR measurements have demonstrated greater reliability and
10 are more functionally orientated than non-weight bearing alternatives (Bennell et al.,
11 1998; Aitkenhead 2002; Jones et al., 2005; Munteanu et al., 2009). Greater sensitivity
12 (Bagget and Young, 1993) and superior cost and time effectiveness of functional
13 weight-bearing methods have been claimed (Bennell et al., 1998; Jones et al., 2005).

14 A weight-bearing DFR measurement method that has demonstrated excellent
15 reliability (Intraclass Correlation Coefficients (ICCs) of 0.97 to 0.99 for intra and inter-
16 rater reliability respectively) involves lunging towards a wall (Bennell et al., 1998). The
17 lunge is repeated up to 5 times to enable the foot to be moved away or towards the
18 wall until the 'end range' is found.

19 An adapted version of the technique (Jones et al., 2005) involving pushing a
20 moveable datum with the lunging knee has also shown good reliability (ICCs 0.82 to
21 0.99). However, use of customised equipment makes this technique less practical and
22 more expensive.

23 A modified DFR measurement technique has been developed that can be
24 viewed as a clinically simplified version of that proposed by Jones et al. (2005).

25 Instead of pushing a custom-made datum with the knee, the new technique uses the
26 upright leg of a clinic table (see figure 1, table length 61cm, width 30cm and height
27 71cm). Three repetitions of the test are performed and the mean figure used. The
28 benefits of this method above others are the speed of the test and simplicity of
29 explanations to patients. Also, varied foot positioning may change the amount of
30 pronation and subsequently affect DFR (Pope et al., 1998). With the modified
31 technique the foot position can remain unaltered which improves standardisation of the
32 technique. The modified technique may therefore be less prone to variation.
33 Establishing the intra and inter-rater reliability is needed before this modified lunge
34 DFR measurement technique can be recommended. Direct comparisons with Bennell
35 et al. (1998) and Jones et al. (2005) would need specific equipment and more
36 repetitions that may lead to mobilisation effects or prolong the study duration and
37 introduce potential variance of DFR if measured on different days. The proposed lunge
38 measure will therefore be compared to previous results of the aforementioned studies
39 instead.

40

41 **Method**

42 Pilot Study

43 A pilot study ($n = 5$) was undertaken to refine instructions and inform a power
44 calculation (Walter et al., 1998). The pilot study generated ICC scores of > 0.9 . Type I
45 and II error probability selected was 0.05 and 0.2 respectively. The ICC parameter was
46 therefore set at 0.9 (Walter et al., 1998, table 2) giving a calculated sample size of
47 thirteen.

48

49 Subjects

50 Thirteen volunteers (6 males, 7 females), mean age of 39 (standard deviation
51 (SD) 14.5) and height of 168cm (SD 10.1) were recruited from staff at the Chesterfield
52 Royal Hospital. Exclusion criteria (expanded from Munteanu et al., 2009) included
53 acute or chronic lower limb pathology in the past year, previous lower limb surgery,
54 neurological or balance deficits or an inability to perform or sustain a lunge for any
55 reason.

56 Recruitment included verbal and emailed presentations to staff members.
57 Written consent was gained and data was anonymised then securely stored. Sheffield
58 Hallam University Research Ethics Committee gave ethical approval.

59

60 Raters

61 Two raters were used for all measurements. Rater 1 had 5 years clinical
62 Physiotherapy experience and devised the modified technique. Rater 2 had 15 years
63 of experience and was provided with a 15 minute training session to ensure
64 standardisation between the raters.

65

66 Procedure

67 The full procedure and rationale is detailed in figures 1 and 2. Subjects looked
68 forwards at all times and the tape measure was covered to blind the subjects from their
69 performance. Raters measured many subjects in succession and had no access to
70 previous measurements to minimise recall of data.

71

72 Data Analysis

73 Raw data was screened for anomalies. Bland and Altman plots (Bland and
74 Altman, 1999), box plots and histograms assessed whether data was homoscedastic,
75 normally distributed and not dependent upon the mean, which would affect statistical
76 power (Atkinson and Neville, 1998; Bland, 2000).

77 Correlations were used to assess if age, height, or gender corresponded with
78 measurements. Differences between the first and second measurement sessions, and
79 between the two raters were evaluated using repeated ANOVA calculations. Post hoc
80 statistical tests (Bonferroni) were performed where differences were identified.

81 Methods for assessing reliability have varied rationales and limitations;
82 combinations of statistical methods are therefore suggested (Atkinson and Nevill, 1998;
83 Rankin and Stokes, 1998).

84 ICC (3,k) was utilised (Shrout and Fleiss, 1979). Error range and repeatability
85 was calculated with standard error of measurement (SEM), 95% confidence intervals
86 (CI), 95% limits of agreement (LOA) and the coefficient of repeatability (CR) (British
87 Standards Institute, 1979; Denegar and Bull, 1993; Atkinson and Nevill, 1998; Rankin
88 and Stokes, 1998; Bland and Altman, 1999; Bland, 2000). 95% LOA demonstrate the
89 range of measurement error within the sample and CR extrapolate a predictive figure
90 for future measurement variance to 95% probability (British Standards Institution,
91 1979). The significance level was set at $p < 0.05$. SPSS version 16 software was
92 used.

93

94 Results

95 Thirteen volunteers completed the study. No gender bias was evident.
96 Histograms plus Bland and Altman plots confirmed that the data was homoscedastic
97 (see figure 3). No dependence upon the mean and minimal measures beyond 95%
98 LOA were evident (see figure 3) confirming a lack of anomalies or systematic bias.

99

100 *Intra-rater Reliability*

101 Excellent intra-rater correlation was found for rater 1 (ICC = 0.98) and rater 2
102 (0.99). See tables 1 and 2 for statistical analysis results. The level of error was also
103 good (SEM = 0.4cm) for both raters. The spread of this error was small (95% CI =
104 0.8cm for both raters) and the maximum 95% LOA was ± 1.47 cm for rater 1 and
105 ± 1.28 cm for rater 2, indicating a narrow band of difference between a raters first and
106 second measurement session (see table 2 for all LOA data). CR indicated suitably
107 small differences between repeated measurements (CR = 1.35cm for rater 1 and
108 1.21cm for rater 2).

109

110 *Inter-rater reliability*

111 Box plots demonstrated no significant anomalies (see figure 4) but rater 1
112 appeared to provide shorter measurements. Repeated ANOVA outcomes confirmed
113 this and Bonferroni results show the second measurement session by rater 1
114 measured significantly lower ($P = 0.01$) than rater 2's second session with mean figures
115 of 9.1cm (range 4.2 to 13.3) versus 9.5cm (range 4.8 to 14.1) respectively. No
116 difference was demonstrated between rater 1 and rater 2 at the first session.

117 Despite the difference between the raters second session measurements,
118 excellent inter-rater correlation was found (ICC = 0.99). The level of error was also

119 good (SEM = 0.3cm). The spread of this error was small (95% CI = 0.6cm). LOA
120 ranged from ± 0.83 to ± 1.47 cm. A CR of 1.44cm also indicated a small difference
121 between raters measurements.

122

123 Discussion

124 The results indicate the modified lunge DFR measurement technique is reliable
125 with a healthy sample. ICC figures of 0.98 to 0.99 demonstrates the technique
126 generates correlated repeated measurements (Bruton et al., 2000).

127 SEM and LOA figures evaluate the range of measurement variation and are
128 recommended alongside ICCs (Denegar and Bull, 1993; Atkinson and Nevill, 1998;
129 Rankin and Stokes, 1998; Bland and Altman, 1999). A maximum SEM of 0.4 (95% CI
130 = 0.8) further supports the modified technique. The LOA indicate that a difference
131 beyond ± 1.47 cm is needed to ensure that changes in measured distances are not the
132 result of measurement variation with 95% confidence.

133 The CR provides the minimum detectable distance to 95% probability (British
134 Standards Institute, 1979; Bland, 2000). CR figures of 1.21 to 1.35cm (intra-rater) and
135 1.44cm (inter-rater) are in accordance with LOA data. A difference of 0.03cm exists
136 between the upper LOA and CR findings. If the conservative, larger figure is used, a
137 measurement difference less than 1.47cm may be a result of measurement error. A
138 difference greater than 1.47cm is deemed clinically significant and not attributable to
139 measurement variability. Clinical responses to injury and treatments lead to changes
140 that far exceed these ranges of 'error' and enable the technique to detect relevant
141 changes. For example, a difference of 6.2cm has been noted between sprain injury
142 patients and asymptomatic subjects (Collins et al., 2004).

143 Other studies used SEM (Bennell et al., 1998), CI or 75 percentiles (Hoch and
144 McKeon, 2011) to provide ranges beyond which variance is thought to be absent.
145 These methods do not provide 95% confidence or probability that the difference
146 between two measurements is not attributable to error, unlike LOA (Rankin and Stokes,
147 1998; Bland and Altman, 1999) and CR (British Standards Institution, 1979). Previous
148 lunge measurement reliability studies (Bennell et al., 1995; Jones et al., 2005) did not
149 use a predictive statistic such as the CR but this enhances statistical analysis.

150 Claims of high reliability have been made by authors of other lunge DFR
151 measurement techniques. Bennell et al. (1998) achieved similar ICCs (0.97 to 0.99) to
152 the present study and a SEM (0.4 to 0.6) that was marginally larger, however no clear
153 exclusion criteria was applied which may explain the greater variation of measures (SD
154 = 3.7 to 4 compared with 2.8 in this study). Increased variation has been associated
155 with inflated ICC figures (Denegar and Ball, 1993; Bland and Altman, 1999) because
156 the calculation generates a relative index of variance. Increased variation in the range
157 of measures enhances the power of such formulae to detect patterns in the
158 calculations and vice versa (Mitchell, 1979; Haas, 1991; Atkinson and Nevill, 1998).
159 Altering foot position every time may explain the greater variation.

160 The ICCs suggest the methods of Jones et al. (2005) are inferior to the modified
161 technique (lowest figure of 0.66 compared to 0.98). Wider LOA of $\pm 3.85\text{cm}$ compared
162 to $\pm 1.47\text{cm}$ with the modified technique strengthens this argument. Using specialist
163 equipment also makes the datum method (Jones et al., 2005) more time intensive and
164 expensive.

165 Better ICCs, SEM and LOA have been shown with the proposed modified DFR
166 technique compared to alternatives (Bennell et al., 1995; Jones et al., 2005). The CR
167 data gives further weight to these findings and a predictive confidence of 95%.

168

169 Limitations

170 Blinding of subjects to all results and the raters to previous measurements was
171 undertaken but the potential for bias was high as the technique was devised by rater 1.
172 The second rater, with no involvement with the technique or study, generated better
173 ICC and LOA findings. This suggests the potential researcher bias was not present.

174 A significant inter-rater difference was found between the second session of
175 measurements with rater 1 measuring shorter distances. This could be due to
176 interpretations of heel lifting. Excessive grasping of the heel or over vigilance
177 preventing pronation in an attempt to ensure strict standardisation may have altered
178 the movement and explain a reduced score. One explanation for this may have been
179 over-eagerness to limit any mobilisation effect as the second session was performed
180 up to 3 hours later using subjects who were mobilising during this time. Kinematic and
181 pressure sensor technology would enable assessment of this but would incur greater
182 cost so was not available. Some studies have utilised an electromechanical lever
183 (Aitkenhead, 2002) or a restraining strap placed over the mid foot region (Jones et al.,
184 2005) but this was thought contrary to the clinically orientated aims of the present
185 technique and difficult to standardise. Despite this discrepancy excellent inter-rater
186 ICC results were evident.

187 Anecdotally, clinical use of the modified DFR measurement technique often
188 results in patients being unable to touch the table leg with their knee due to
189 hypomobility of the ankle. In these cases the shortest distance between the patella
190 and the table leg is used as the measured distance. Similar methods have proven
191 reliable with ankle fracture patients (Simondson et al., 2012). The use of the modified

192 technique with patients requires further reliability assessment before it can be
193 advocated widely.

194

195 Conclusion

196 This study demonstrated the proposed modified weight-bearing DFR lunge
197 technique is reliable when used with a healthy sample. A difference greater than
198 1.47cm represents a meaningful difference beyond the variation of the technique.

199

References

200

201 Agosta J, Morarty R. Biomechanical analysis of athletes with stress fractures of the
202 tarsal navicular bone: a pilot study. Australian Journal of Podiatric Medicine 1999;
203 33(1): 13-18.

204

205 Aitkenhead I. Ankle joint dorsiflexion assessment: the development of a new weight-
206 bearing method. British Journal of Podiatry 2002; 5(2): 32-35.

207

208 Atkinson G, Nevill A. Statistical methods for assessing measurement error (reliability)
209 in variables relevant to sports medicine. Sports Medicine 1998; 26(4): 217-238.

210

211 Bagget B, Young G. Ankle joint dorsiflexion establishment of normal range. Journal of
212 the American Podiatric Medical Association 1993; 85(5): 251-254.

213

214 Bennell K, Talbot R, Wajswelner H, Techovanich W, Kelly D. Intra-rater and inter-rater
215 reliability of a weight-bearing lunge measure of ankle dorsiflexion. Australian Journal of
216 Physiotherapy 1998; 44(3): 175-180.

217

218 Bland M, Altman D. Measuring agreement in method comparison studies. Statistical
219 Methods in Medical Research 1999; 8(2): 135-160.

220

221 Bland M. An introduction to medical statistics, 3rd ed. Oxford University Press, 2000.
222 p. 123-343.

223

224 British Standards Institute. Precision of test methods. Part 1: Guide for determination
225 and reproducibility for a standard test method. British Standards Institute, London,
226 1979. p. 1-22.

227

228 Bruton A, Conway J, Holgate S. Reliability: What is it, and how is it measured?
229 Physiotherapy 2000; 86(2): 94-99.

230

231 Collins N, Teys P, Vicenzino B. The initial effects of a Mulligan's mobilization with
232 movement technique on dorsiflexion and pain in subacute ankle sprains. Manual
233 Therapy 2004; 9(2): 77-82.

234

235 Denegar C, Ball D. Assessing reliability and precision of measurement: An introduction
236 to intraclass correlation and standard error of measurement. Journal of Sport
237 Rehabilitation 1993; 2(1): 35-42.

238

239 Haas M. Statistical methodology for reliability studies. Journal of Manipulative and
240 Physiological Therapeutics 1991; 14(2): 119-132.

241

242 Hoch M, McKeon P. Normative range of weight-bearing lunge test performance
243 asymmetry in healthy adults. *Manual Therapy* 2011; 16(5): 516-519.

244

245 Jones R, Carter J, Moore P, Wills A. A study to determine the reliability of an ankle
246 dorsiflexion weight-bearing device. *Physiotherapy* 2005; 91(4): 242-249.

247

248 Kaufman K, Brodine S, Shaffer R, Johnson C, Cullison T. The effect of foot structure
249 and range of motion on musculoskeletal overuse injuries. *American Journal of Sports*
250 *Medicine* 1999; 27(5): 585-593.

251

252 Lin C, Moseley A, Herbert R, Refshauge K. Pain and dorsiflexion range of motion
253 predict short and medium-term activity limitation in people receiving physiotherapy
254 intervention after ankle fracture: An observational study. *Australian Journal of*
255 *Physiotherapy* 2009; 55(1): 31-37.

256

257 Magee D. *Orthopedic physical assessment*, 5th ed. Saunders, Elsevier, 2008. p. 881.

258

259 Mitchell S. Interobserver agreement, reliability and generalizability in data collected in
260 observational studies. *Psychological Bulletin* 1979; 86(2): 376-390.

261

262 Munteanu S, Strawhorn A, Landorf K, Bird A, Murley G. A weightbearing technique for
263 the measurement of ankle dorsiflexion with the knee extended is reliable. *Journal of*
264 *Science and Medicine in Sport* 2009; 12: 54-59.

265

266 Pope R, Herbert R, Kirwan J. Effects of ankle dorsiflexion range and pre-exercise calf
267 muscle stretching on injury risk in army recruits. *Australian Journal of Physiotherapy*
268 1998; 44(3): 165-172.

269

270 Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of
271 appropriate statistical analyses. *Clinical Rehabilitation* 1998; 12: 187-199.

272

273 Shrout P, Fleiss J. Intraclass correlations: Uses in assessing rater reliability.
274 *Psychological Bulletin* 1979; 86(2): 420-428.

275

276 Simondson D, Brock K, Cotton S. Reliability and smallest real difference of the ankle
277 lunge test post ankle fracture. *Manual Therapy* 2012; 17(1): 34-38.

278

279 Walter S, Eliasziw M, Donner A. Sample size and optimal designs for reliability
280 studies. *Statistics in Medicine* 1998; 17(1): 101-110.

Table1: Means, SD and statistical results of each rater and both raters combined.

Rater	1st session Mean Distance & SD (cm)	2nd session Mean Distance & SD (cm)	ICC	SEM (cm)	95% CI (cm)	CR (cm)
1	9.1 (2.7)	9.1 (2.8)	0.98	0.4	-0.4 – 1.2	1.35
2	9.5 (2.9)	9.5 (2.9)	0.99	0.4	-0.4 – 1.2	1.21
1 + 2	9.3 (2.8)	9.3 (2.8)	0.99	0.3	-0.3 – 0.9	1.44

Table2: Differences between measurement sessions and between raters. R1S1 = rater 1, 1st measurement session ; R1S2 = rater 1, 2nd session; R2S1 = rater 2, 1st session; R2S2 = rater 2, 2nd session.

Rater & Session	Mean Difference	95% LOA	±
R1S1 Vs R1S2	-0.07	-1.57 to 1.43	1.47
R2S1 Vs R2S2	-0.04	-1.34 to 1.26	1.28
R1S1 Vs R2S1	0.48	-0.96 to 1.92	1.44
R1S1 Vs R2S2	0.45	-1.02 to 1.92	1.47
R1S2 Vs R2S1	0.49	-0.74 to 1.72	1.23
R1S2 Vs R2S2	0.45	-0.38 to 1.28	0.83

Figure 1a Starting position with foot placed on tape and toe against upright of table. (b) Final lunge position.



Figure 1b

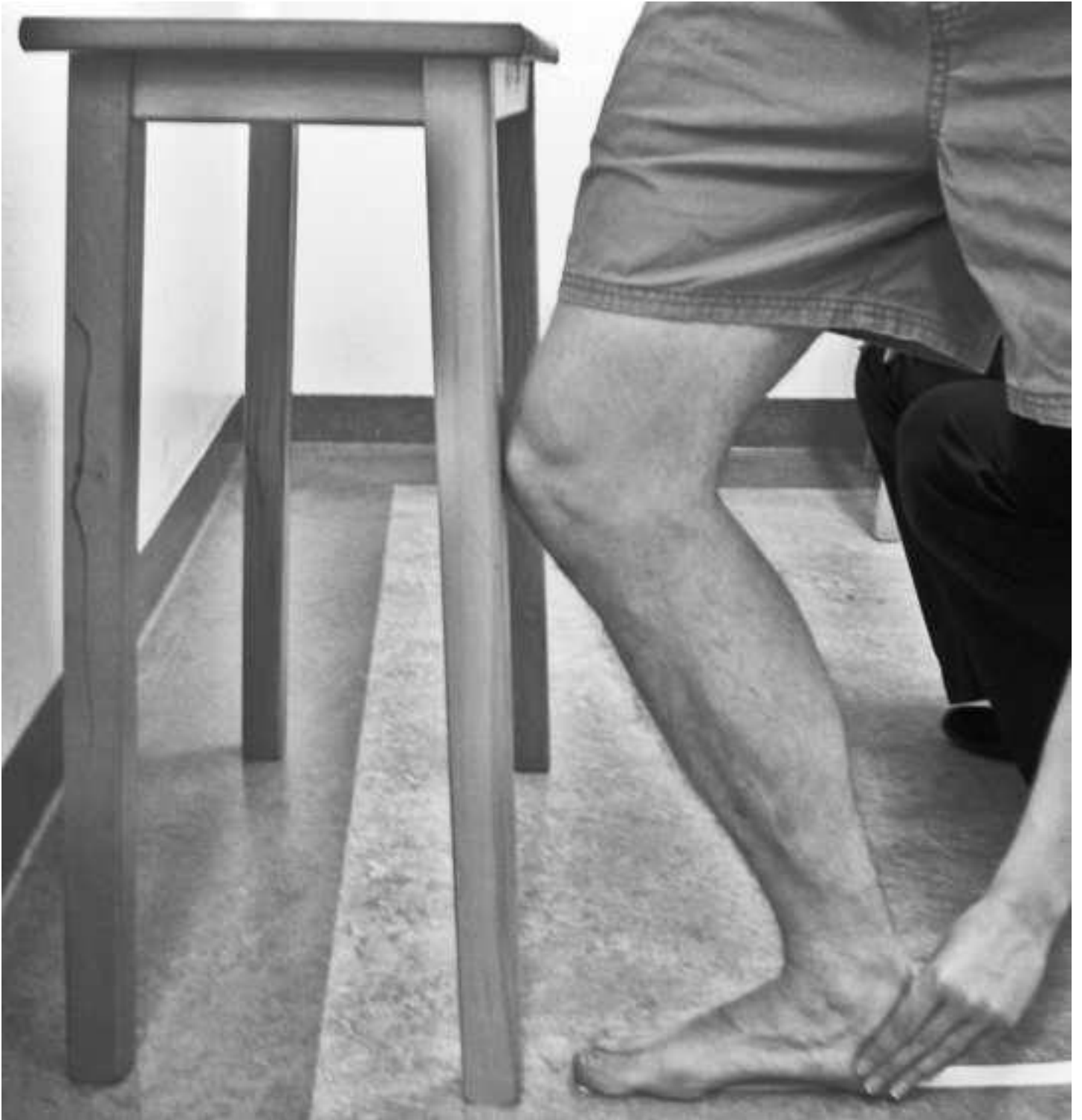


Figure 2: Sequence of measurement sessions and rationale of standardised procedure.

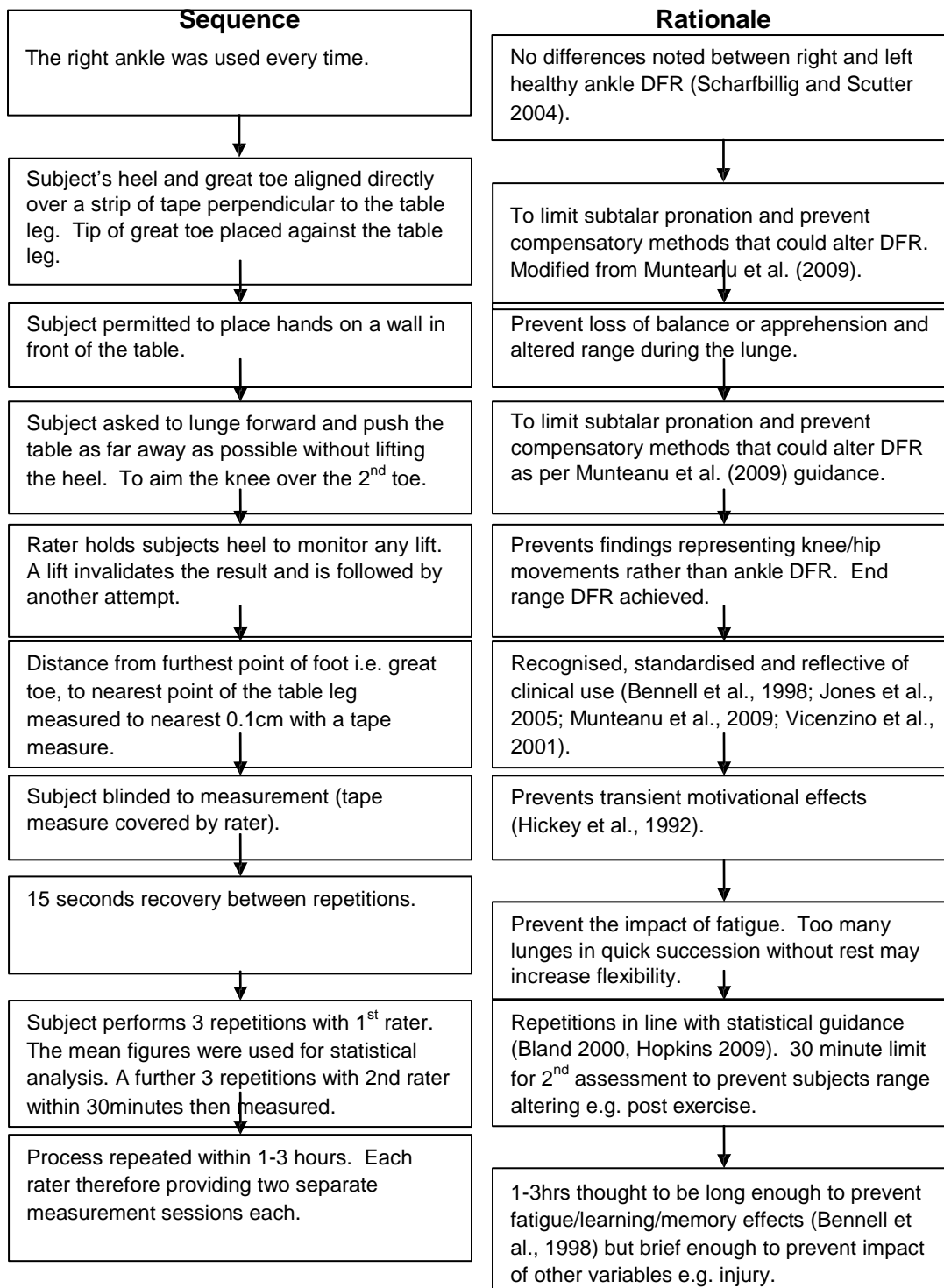


Figure3: Bland and Altman plot demonstrating the difference between measures taken by rater 1 and rater 2 against actual measurements. Includes mean difference and 95% LOA.

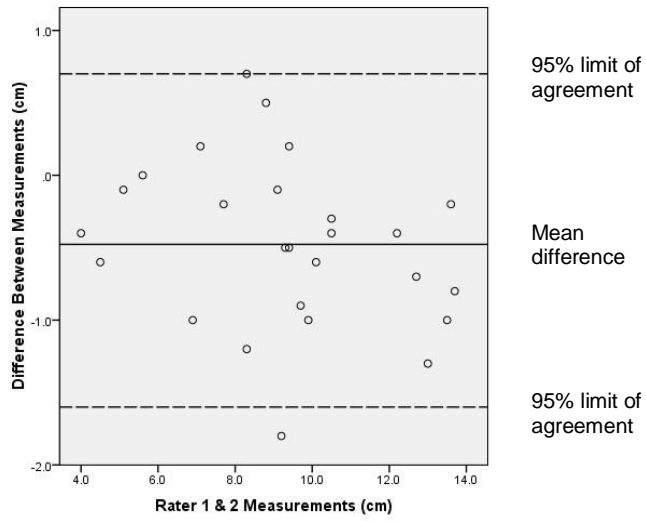


Figure 4: Box plot showing all measurements taken by rater 1 and rater 2. Means, upper/lower limits and interquartile ranges shown.

