Intra-rater and inter-rater reliability of a weight-bearing lunge measure of ankle dorsiflexion

Kim Bennell
Richard Talbot
Henry Wajswelner
Wassana Techovanich
David Kelly

This study aimed to evaluate the inter-rater and intra-rater reliability of a weight-bearing dorsiflexion (DF) lunge in 13 healthy subjects. Four raters with varying clinical experience tested all subjects in random order. Two of the raters repeated the measurements one week later. Two methods were used to assess the DF lunge: (i) the distance from the great toe to the wall and (ii) the angle between the tibial shaft and the vertical using an inclinometer. The average of three trials was used in data analysis. Intra-rater intraclass correlation coefficients (ICCs) ranged from 0.97 to 0.98. Inter-rater ICC values were 0.97 (angle) and 0.99 (distance). Results indicate excellent reliability for both methods of assessing a DF lunge.

[Bennell KL, Talbot RC, Wajswelner H, Techovanich W, Kelly DH and Hall AJ: Intra-rater and inter-rater reliability of a weight-bearing lunge measure of ankle dorsiflexion. *Australian Journal of Physiotherapy* 44: 175-180]

Key words: Ankle Joint; Range of Motion, articular; Reproducibility of Results

In order to justify intervention and judge its effectiveness, physiotherapists need to incorporate objective measures in their clinical practice. Many measures may be taken by the same and by different therapists in the management of one patient. Therefore it is necessary to determine if the measurements used are reliable both within and between therapists.

Adequate range of ankle dorsiflexion (DF) is necessary for the normal performance of functional activities such as walking, running, stair-climbing, rising from a chair and squatting (Bohannon et al 1989). Clinically, restriction of this movement is often seen post-injury or following immobilisation. Restricted DF has also been implicated as a contributing factor in overuse injuries of the lower limb and foot (Hughes et al 1985, Warren and Jones 1987). The assessment of DF is therefore important for physiotherapists.

Several methods have been reported in the literature for measuring ankle DF. These include electric goniometers (Clapper and Wolf 1988), rulers (Montgomery et al 1989), photography of skin markers and subsequent measurement of an image projection (Bohannon et al 1991, Moseley and Adams 1991), a specifically designed six degrees of freedom fixture (Allinger and Engsberg 1993, Grimston et al 1993), roentgen stereophotogrammetry (Lundberg et al 1989), visual estimation (Youdas et al 1993), as well as the inclinometer and universal goniometer. Use of most of these devices is impractical in the clinical setting.

Physiotherapists commonly use a weight-bearing lunge test to assess DF at the ankle. For this test, the patient is required to place their foot perpendicular to a wall and to lunge their knee toward the wall. The foot is

KL Bennell BAppSc(Phty), PhD is a lecturer in the School of Physiotherapy at The University of Melbourne.

RC Talbot BPhty is a physiotherapist at the Goulburn Valley Base Hospital, Shepparton, Victoria.

H Wajswelner BAppSc(Phty), GradDip ManipPhty is a lecturer in the School of Physiotherapy at The University of Melbourne.

W Techovanich BSc(Phty), MSc(Phty) is a lecturer at Srinakharinwirot University, Bangkok, Thailand.

DH Kelly BSc, BAppSc(Phty), GradDipManipPhty is an associate lecturer in the School of Physiotherapy at The University of Melbourne

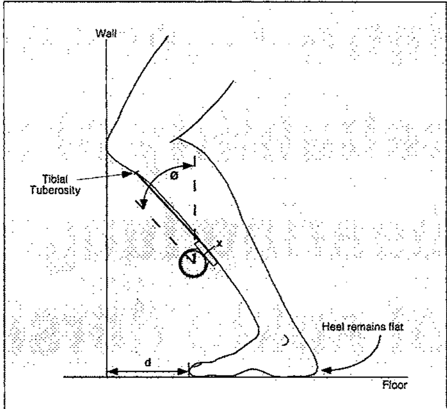
AJ Hall BAppSc(PhysEd) was a student in the School of Physiotherapy at The University of Melbourne at the time of writing this article.

Correspondence: Dr Kim Bennell, School of Physiotherapy, The University of Melbourne, 200 Berkeley Street, Carlton, Victoria 3053. E-mail: k.bennell@physio.unimelb.edu.au.

progressively moved away from the wall until the maximum range of ankle dorsiflexion is reached without the heel lifting. The most frequent measurements taken at this point are the distance from the foot to the wall or the angle of the tibial shaft from the vertical using a gravity goniometer. The benefits of the DF lunge test are that it is cost and time efficient, requires minimal equipment and is performed in weight-bearing. The latter is a particular advantage as the torque applied to the ankle is many times greater than that applied by non-weight bearing methods and hence the resulting measurement may be more indicative of the range available for functional tasks. On the negative side, the test cannot be performed on patients for whom weight-bearing is contraindicated.

Surprisingly few studies have investigated the reliability of measurement of a DF lunge test (Ekstrand et al 1982). Ekstrand et al (1982) used a modified lunge position and measured DF using a gravity-affected flexometer. The results were an intra-tester coefficient of variation of 9.6 per cent which was reduced to 2.6 per cent after more stringent protocols were introduced. Inter-tester reliability was not assessed in this study. Other studies have examined the reliability of measuring DF in a non-weight bearing position (Elveru et al 1988, Grimston et al 1993, Jonson and Gross 1997, Moseley and Adams 1991, Rome and Cowieson 1996, Youdas et al 1993). These studies report ICCs ranging between 0.28 and 0.98 for inter-rater reliability and 0.74 and 0.98 for intra-rater reliability.

The paucity of research into the reliability of measuring DF in weightbearing, particularly the use of a distance to wall measurement, means that such a study is both worthwhile and necessary. The purpose of this study is to examine the inter-rater and intra-rater reliability of measuring DF lunge using two methods: (i) distance from the tip of the great toe to the wall in centimetres, and (ii) angle of the tibial shaft from the vertical in degrees. The significance of this research is that



Where:
x 15 cm along anterior tibial shaft
ø angle between anterior tibial border and the vertical (degrees)
d distance from the wall to big toe (cm)

Figure 1. Diagram depicting the two methods of measuring dorsiflexion lunge.

findings will provide an indication of the amount of error associated with DF lunge measurements.

Method

Subjects

Thirteen subjects (eight males and five females) volunteered for this study in response to an advertisement within the School of Physiotherapy at The University of Melbourne. All were full-time physiotherapy students in the first year of their course. The mean (SD) age of subjects was 18.8 (2.0) years, the mean height was 170.8 (7.5) cm and the mean weight was 62.2 (10.3) kg. None of the subjects reported any current neurological or lower limb musculoskeletal injury. The study was approved by the Human Research Ethics Committee of The University of Melbourne and all subjects provided written informed consent.

Raters

Four raters (A-D) were used in this study. Rater A was a second year physiotherapy student with knowledge of anatomy and basic goniometry use but with minimal clinical experience. Raters B, C and D were all qualified physiotherapists. Rater B had nine years of clinical experience in Thailand. Raters C and D were manipulative physiotherapists with nine and 17 years clinical experience respectively. Raters B, C and D all used the DF lunge test in routine clinical practice.

Procedure

On the first test session, all raters tested all subjects. The order in which raters tested the subjects was randomly assigned from a series of orders derived from a Latin square design. On the second test session, one week later, only Raters A and B were available to test the subjects. The same subject and rater testing order was used for the second test occasion.

As there is no significant difference in mean ankle DF between left and right sides in normal individuals, only the left leg was measured (Stefanyshyn and Engsberg 1994). Two measurements were used to assess the DF lunge: (i) the distance (to the nearest 0.1cm) from the end of the big toe to the wall using a tape measure on the floor and (ii) the angle (to the nearest degree) between the anterior border of the tibia and the vertical using an inclinometer (Isomed, Portland USA) (Figure 1). Only the left leg was tested.

A 10min training session was conducted to familiarise raters with the test procedure. The procedure involved the rater marking a point on the anterior border of the left tibia 15cm below the middle of the tibial tuberosity using a non-permanent felt-tipped pen. This was the point of application of the middle of the inclinometer. A line was drawn on the left heel bisecting the calcaneus to enable consistent foot position on the tape on the floor. The test procedure was demonstrated to subjects prior to commencement and standardised instructions were given.

Subjects positioned their left foot so that the heel line and big toe were aligned on the tape measure on the floor. They lunged forward so that their knee touched a vertical line drawn on the wall. Subjects were allowed to hold onto the wall for balance during the test and were free

Table 1. Mean and standard deviation for DF lunge distance and angle measurements of each rater

	Distance (cm)		Angle (°)	
Rater	Mean	SD	Mean	SD
A	13.9	3.8	50.4	8.1
B	13.8	4.0	50.3	7.9
C	13.8	3.7	50.8	7.7
D	13.6	3.8	46.3 *	7.7

* Rater D significantly different from all other raters, $p < 0.05$ Fisher tests.

Table 2. Intraclass correlation coefficient and standard error of measurement with 95 per cent confidence intervals for inter-rater and intra-rater reliability of dorsiflexion lunge distance and angle measurements.

Measurements	Inter-rater reliability	Intra-rater reliability Rater A	Intra-rater reliability Rater B
Distance			
ICC	0.99	0.98	0.97
95% CL	0.97-0.99	0.93-0.99	0.90-0.99
SE_{meas} (cm)	0.4	0.5	0.6
95% CI (cm)	0.8	1.0	1.2
Angle			
ICC	0.97	0.98	0.98
95% CL	0.90-0.99	0.93-0.99	0.93-0.99
SE_{meas} (degrees)	1.4	1.1	1.1
95% CI (degrees)	2.7	2.2	2.2

$ICC_{(2,3)}$ for inter-rater reliability, $ICC_{(3,3)}$ for intra-rater reliability
95% CL according to method by Rosner (1995).
$SE_{meas} = SD \times \sqrt{1\text{-}ICC}$; 95 per cent CI = 1.96 x SE_{meas}

to rest the untested leg in a comfortable position on the floor. Up to five lunging attempts were allowed to find the maximum distance from the wall where the subject could touch the wall with the knee while maintaining heel contact. During the lunge the subject's heel was held by the rater to prevent lifting from the floor. No attempt was made to limit pronation or supination in the foot. At the maximum lunge point, the rater placed the inclinometer on the tibial mark and recorded the achieved angle. The distance to the wall from the tip of the big toe was then measured from the tape measure. The measurements were made twice more after the subject had stood and resumed a comfortable position. The subject then changed to another rater and the procedure was repeated. Each rater was unaware of the scores of other raters and all marks drawn on the subject were removed following each rater. Subjects were not informed of their scores.

Statistical analysis

All data were analysed using Statview SE+ Graphics software (Abacus Concepts Inc., Berkeley USA). For each subject, the mean of three scores was used in statistical analysis. This was justified as there were no significant differences between trials for any rater. Data were examined for normality and homogeneity of variance to ensure they fulfilled the criteria for parametric tests. Differences between raters for measurements of DF lunge distances and angles were compared using a repeated measures one way analysis of variance (ANOVA) followed by post-hoc Fisher tests. Differences between first and second test occasions for Raters A and B were compared using paired *t*-tests. To establish reliability, ICCs were used to indicate the extent of combined systematic and random error. Based on the method of Shrout and Fleiss (1979), an ICC (model 2,3) was calculated for inter-rater reliability. This was performed for each rater pair combination as well as for the raters as a group. For intra-rater reliability, an ICC (model 3,3) was chosen. Ninety-five per cent confidence limits were calculated for the ICCs using the method of Rosner (1995). To allow the error to be expressed in original units of measurement, the standard error of measurement (SE_{meas}) was calculated using the ICC ($SE_{meas} = SD \times \sqrt{1\text{-}ICC}$). Ninety-five per cent confidence intervals were calculated for each SE_{meas} using the z distribution (95 per cent CI = 1.96 x SE_{meas}). Two-tailed significance levels were set at $p < 0.05$.

Results

Inter-rater reliability

The mean (SD) distance and angle measurements for each rater at the first test occasion are shown in Table 1. There was a significant difference between raters for angle measurements ($F_{(3,36)} = 25.51$, $p < 0.001$). The mean angle for Rater D was approximately 4 degrees lower than each of the other raters. There was no significant difference between raters for the distance measurements ($F_{(3,36)} = 0.69$, $p = 0.56$).

Inter-rater reliability was excellent for both distance and angle measurements (Table 2) with high group ICC (2,3) values and small SE_{meas} indicating minimal random or systematic error. The size of the SE_{meas} as a percentage of the Test 1 means was approximately 3 per cent for both angle and distance measurements.

Table 3 shows the correlation (given as ICCs together with their 95 per cent confidence limits) between rater pairs for both angle and distance measurements. The correlations between raters ranged from 0.89 to 0.99.

Intra-rater reliability

There was no significant difference between the Test 1 and Test 2 angle measurements for either Rater A ($t_{(12)}$ = 0.60, p = 0.56) or Rater B ($t_{(12)}$ = -1.2, p = 0.25). There was also no significant difference between the two distance measurement tests for Rater A ($t_{(12)}$ = 2.0, p = 0.07) or Rater B ($t_{(12)}$ = 0.03, p = 0.99).

Both Rater A and Rater B demonstrated excellent intra-rater reliability for angle and distance measurements performed one week apart (Table 2). ICC values were very high and SE_{meas} values were small, indicating minimal random and systematic error.

Relationship between angle and distance measurements

For each of the raters, the angle and distance measures from the first test occasion were highly correlated (Rater A: r = 0.95; Rater B: r = 0.95; Rater C: r = 0.96; Rater D: r = 0.93; all $p < 0.001$). The relationship between the two methods of measurement is shown in Figure 2.

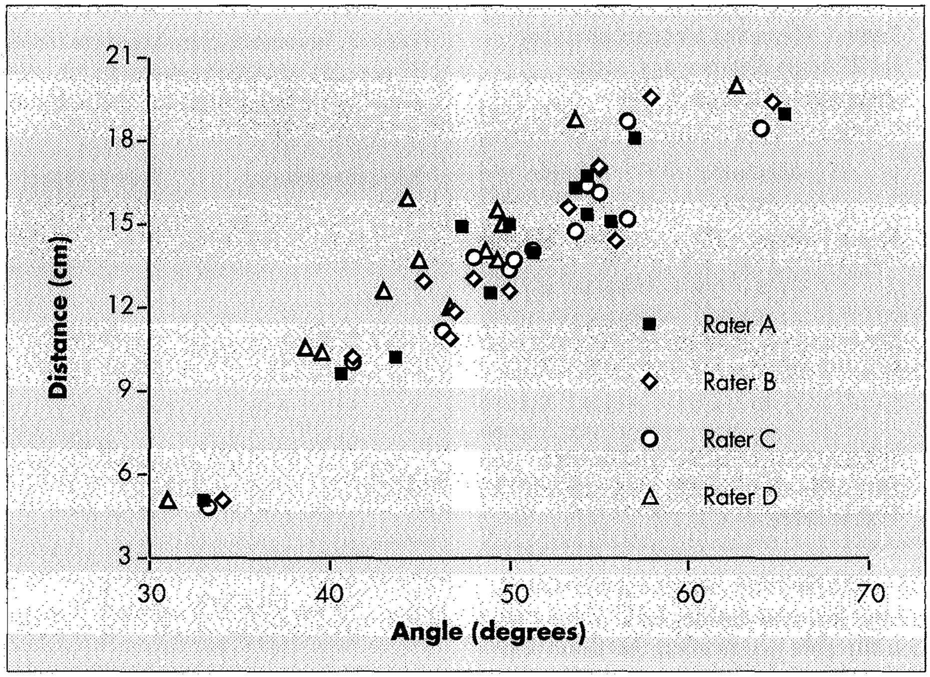


Figure 2. Scatterplot of relationship between angle and distance measurements for each of the raters on the first test occasion.

Discussion

Our results in normal subjects indicate that both angle and distance measurements of a DF lunge test can be reliably performed by the same therapist as well as by different therapists with varying clinical experience. The amount of measurement error calculated from this study enables interpretation of the size of the change in angle or distance measurements required to detect genuine change resulting from an intervention. For example, if a therapist was using a mobilising technique to improve range of ankle DF, a change of greater than approximately 1cm or 3 degrees would be required to confidently attribute the observed change to the technique rather than to measurement error. Inclusion of raters with a range of clinical expertise strengthens our ability to generalise the results to other physiotherapists. This study provides evidence to support the use of a DF lunge as an objective measurement tool in physiotherapy practice.

Differences in skill level and experience of the raters did not appear to influence the repeatability of DF lunge measurements. Although Rater D had significantly lower mean angle measurements than the other raters, the reliability coefficients comparing this rater with the other raters were high. Similarly, when comparing the least experienced rater, the physiotherapy student, with the other raters who were all experienced physiotherapists, the reliability coefficients were also high. These results probably reflect the inherent simplicity of the DF lunge test.

The high reliability found in our study may be due to several factors. First, the wide spread of absolute distance and angle values, with measurements ranging from approximately 5-20cm and from 30-68 degrees, prevents range effects from producing spuriously low reliability coefficients. Second, we tested healthy subjects. In such individuals, dorsiflexion is likely to be limited by a combination of bony (Smith and Reischl 1988), ligamentous (Nigg et al 1990) and muscular (Moseley and Adams 1991) factors. In patients with ankle pathology, pain or swelling may be the limiting factors and these may render the DF lunge test less reliable. Further research is needed to evaluate the reliability of the DF lunge test in a patient population to allow comparison with our results in normal individuals. Third, the DF lunge test is simple to perform. The use of a gravity goniometer requires one point of application, unlike the universal goniometer where error can be introduced from locating the axis of movement and positioning both arms. Last, the mean of three tests was used for data analysis. This may provide a better representation of the subject's performance than a single measurement.

Table 3. Correlation matrix showing intraclass correlation coefficients together with their 95 per cent confidence limits comparing each rater pair for angle and distance measures

	Rater	B	C	D
Angle	A	0.99 (0.97-0.99)	0.99 (0.97-0.99)	0.92 (0.75-0.98)
	B	-	0.99 (0.97-0.99)	0.90 (0.69-0.97)
	C	-	-	0.89 (0.67-0.97)
Distance	A	0.98 (0.93-0.99)	0.99 (0.97-0.99)	0.98 (0.93-0.99)
	B	-	0.99 (0.97-0.99)	0.99 (0.97-0.99)
	C	-	-	0.99 (0.97-0.99)

It is difficult to directly compare our results with those from other studies, due to differences in methodology, particularly the measurement of DF. Most other investigators have assessed DF in a non-weight bearing position and have used a variety of measurement techniques. However, our intra-rater results compare favourably with previous reports where ICC values have ranged from 0.74 to 0.98 (Elveru et al 1988, Grimston et al 1993, Jonson and Gross 1997, Mayerson and Milano 1984, Moseley and Adams 1991, Youdas 1993).

Our inter-rater results are also similar to or better than other studies measuring DF. An ICC of 0.50 was found by Elveru et al (1988) using 14 raters and 43 subjects. Their reliability may have been adversely influenced by a non-standardised measurement position. Jonson and Gross (1997) reported an ICC of 0.65 but attributed the moderate value to a narrow range of scores, as the mean absolute difference in measurement was only 3 degrees. Grimston et al (1993), Mayerson and Milano (1984) and Moseley and Adams (1991) standardised their test positions and found ICC or Pearson r values to be 0.98, 0.97 and 0.97 respectively. This approximates the results of our study.

Reliability is only one aspect of establishing the usefulness of a measurement tool in research and in clinical decision-making. Validity, that is, the degree to which an instrument measures what it is supposed to measure (Portney and Watkins 1993), is another important aspect. The lunge test provides an indication of the amount of DF available at the ankle/foot complex, particularly at the talocrural joint. However, it does not measure actual motion at any one specific joint. Dorsiflexion as measured by this test is a combination of movement at a number of joints including the talocrural, subtalar and midtarsal joints. No effort was made to control foot supination or pronation during data collection, except indirectly via knee placement to the vertical line on the wall. Hence alterations in foot posture may have contributed to the range of DF lunge measured in our test. Future studies are needed to quantify the contribution of various structures, such as gastrocnemius/soleus and talocrural, subtalar and midtarsal joints, on DF lunge measurements.

Another validity issue with the use of the distance measurement is whether foot length and leg length scale proportionally between subjects. If distance measurements are to be compared between individuals, then it is important that there is a consistent foot length and leg length ratio. For example, assuming that two subjects had the same actual range of ankle DF, the person with a shorter foot relative to leg length would score a larger distance value than the person with a longer foot relative to leg length. Conversely, angle measurements would be similar in both these individuals. We did not directly address this issue in our study. However, since angle and distance measurements of DF lunge were highly correlated this suggests that, within the range included in this study, the two methods were providing similar information about ankle/foot complex DF.

Conclusion

This study shows that in healthy subjects, distance and angle measurements of a DF lunge test can be reliably performed by the same therapist as well as by different therapists with varying clinical experience. This study provides evidence to support the use of a DF lunge as an objective measurement tool in physiotherapy practice. Further research is needed to confirm the results in different patient populations and to investigate validity issues associated with this test.

References

Allinger T and Engsberg J (1993): A method to determine the range of motion of the ankle joint complex in vivo. *Journal of Biomechanics* 26: 69-76.

Bohannon R, Tiberio D and Zito M (1989): Selected measures of ankle dorsiflexion range of motion: differences and intercorrelations. *Foot and Ankle* 10: 99-103.

Bohannon R, Tiberio D and Waters G (1991): Motion measured from forefoot and hindfoot landmarks during passive ankle dorsiflexion range of motion. *Journal of Orthopaedic and Sports Physical Therapy* 13: 20-22.

Clapper M and Wolf S (1988): Comparison of the reliability of the Orthoranger and the standard goniometer for assessing active lower extremity range of motion. *Physical Therapy* 68: 214-218.

Ekstrand J, Wiktorsson M, Öberg B and Gillquist J (1982): Lower extremity goniometric measurements: A study to determine their reliability. *Archives of Physical Medicine and Rehabilitation* 63: 171-175.

Elveru R, Rothstein J and Lamb R (1988): Goniometric reliability in a clinical setting. *Physical Therapy* 68: 672-677.

Grimston S, Nigg B, Hanley D and Engsberg J (1993): Differences in ankle joint complex range of motion as a function of age. *Foot and Ankle* 14: 215-222.

Hughes L (1985): Biomechanical analysis of the foot and ankle for predisposition to developing stress fractures. *Journal of Orthopaedic and Sports Physical Therapy* 7: 96-101.

Jonson S and Gross M (1997): Intraexaminer reliability, interexaminer reliability, and mean values for nine lower extremity skeletal measures in healthy naval midshipmen. *Journal of Orthopaedic and Sports Physical Therapy* 25: 253-263.

Lundberg A, Goldie I, Kalin B and Selvik G (1989): Kinematics of the ankle/foot complex: Plantarflexion and dorsiflexion. *Foot and Ankle* 9: 194-200.

Mayerson N and Milano R (1984): Goniometric measurement reliability in physical medicine. *Archives of Physical Medicine and Rehabilitation* 65: 92-94.

Montgomery L, Nelson F, Norton J and Deuster P (1989): Orthopedic history and examination in the etiology of overuse injuries. *Medicine and Science in Sports and Exercise* 21: 237-243.

Moseley A and Adams R (1991): Measurement of passive ankle dorsiflexion: Procedure and reliability. *Australian Journal of Physiotherapy* 37: 175-181.

Nigg BM, Skarvan GS, Frank CB and Yeadon MR (1990): Elongation and forces of ankle ligaments in a physiological range of motion. *Foot and Ankle* 11: 30-40.

Portney L and Watkins M (1993): Foundations of Clinical Research: Applications to Practice. Norwalk: Appleton and Lange, p. 69.

Rome K and Cowieson F (1996): A reliability study of the universal goniometer, fluid goniometer, and electrogoniometer for the measurement of ankle dorsiflexion. *Foot and Ankle International* 17: 28-32.

Rosner B (1995): Fundamentals of Biostatistics. (4th ed.) New York: Duxbury Press Publishers, p. 512.

Shrout P and Fleiss J (1979): Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin* 86: 420-428.

Smith RW and Reischl BS (1988): The influence of dorsiflexion in the treatment of severe ankle sprains: An anatomical study. *Foot and Ankle* 9: 28-33.

Stefanyshyn D and Engsberg J (1994): Right to left differences in the ankle joint complex range of motion. *Medicine and Science in Sports and Exercise* 26: 551-555.

Warren B and Jones CJ (1987): Predicting plantar fasciitis in runners. *Medicine and Science in Sports and Exercise* 19: 71-73.

Youdas J, Bogard C and Suman V (1993): Reliability of goniometric measurements and visual estimates of ankle joint active range of motion obtained in a clinical setting. *Archives of Physical Medicine and Rehabilitation* 74: 1113-1118.