

# Measurement Reliability of Swelling in the Acute Ankle Sprain

by Cameron P. Watson, MAppSc<sup>1,4</sup>, Robert A. Boland, PhD<sup>1,3</sup> , Kathryn M. Refshauge, PhD<sup>2</sup>

The Foot & Ankle Journal 1 (12): 4

**Background:** Swelling and painful restriction of dorsiflexion characterize acute ankle sprain, and require accurate measurement to monitor effectiveness of intervention. Reliability of the figure of eight tape method for swelling and the weight-bearing lunge for dorsiflexion are highly reliable in the laboratory, but untested in the less predictable clinical setting.

**Materials and Methods:** We determined intra and interrater reliability and standard error of measurement (SEM) of both methods in the clinical environment, using 4 physiotherapists as raters. Measurements were taken twice within a session and at a follow-up session from the uninjured ankle in 22 participants with unilateral ankle sprain, and from a randomly selected ankle in 11 uninjured participants.

**Results:** Within session intrarater reliability was very high for both figure of eight (Intraclass correlations coefficients [ICC] = 0.99) and weight-bearing lunge (ICC = 0.97) methods. Between-session inter-rater reliability was also very high (ICC > 0.99). The SEM was small for all measurements:  $\pm 0.2\text{cm}$  for figure of eight, and  $\pm 0.4\text{cm}$  for dorsiflexion lunge methods within a session, and  $\pm 0.3\text{cm}$  and  $\pm 0.4\text{cm}$  respectively for between-session measurements.

**Conclusions:** Using simple techniques, swelling and dorsiflexion can be measured with high reliability in the clinic by different clinicians and can detect small changes in status between and within treatments.

**Clinical Relevance:** Clinically meaningful changes ( $>0.5\text{cm}$ ) can be detected by clinicians with varying levels of expertise and can confidently be attributed to the intervention rather than measurement error.

**Key words:** Swelling, ankle, inversion, sprain, standard error of measurement

**Accepted:** November, 2008

**Published:** December, 2008

This is an Open Access article distributed under the terms of the Creative Commons Attribution License. It permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©The Foot & Ankle Journal ([www.faoj.org](http://www.faoj.org))

Ankle sprain is one of the most common injuries amongst sporting populations<sup>9</sup>, and possibly in the general community.

**Address correspondence to:** Robert A. Boland, PhD  
University of Sydney, Faculty of Health Sciences, Discipline of Physiotherapy, PO Box 170, Lidcombe, NSW 1825, Australia.  
Email: [R.Boland@usyd.edu.au](mailto:R.Boland@usyd.edu.au)

<sup>1</sup> University of Sydney, Faculty of Health Sciences, Discipline of Physiotherapy, East St., Lidcombe, NSW, Australia.

<sup>2</sup> University of Sydney, Faculty of Health Sciences, Lidcombe, NSW, Australia.

<sup>3</sup> Prince of Wales Medical Research Institute, Randwick, NSW, Australia.

<sup>4</sup> ELITE Physiotherapy Exercise & Rehabilitation, Menai, NSW, Australia.

In addition to pain, the earliest symptoms are swelling and restricted dorsiflexion range of motion, and these symptoms can persist for years in up to 70% of people after a sprain.<sup>14</sup> To determine efficacy of treatment and monitor progress, it is essential that these impairments be measured reliably and accurately in the clinic.

Swelling secondary to sprain of the lateral ankle ligaments is commonly localized, usually around the lateral malleolus<sup>18,19</sup>, but can also accumulate around the subtalar, talocrural and inferior tibiofibular joints.



**Figure 1** The figure of eight tape method for measuring ankle swelling.

Measurement of swelling therefore needs to specifically include measurement of volume in these areas. The gold standard for such measurement is the water displacement method<sup>13,16</sup>, but this method may be too time-consuming to use efficiently in the clinic. Although an indirect method of measuring ankle swelling, the figure of eight method<sup>13</sup> is time-efficient, cost-effective, and easy to apply in clinical settings. The therapist wraps a tape measure over standardized anatomical landmarks around the ankle and the distance provides a circumferential estimate of volume.<sup>13, p. 131</sup> (Fig. 1)

The figure of eight method is highly reliable for measuring ankle swelling in the laboratory: within-session intra and interrater reliability ranged from 0.98 to 0.99 (intraclass correlation coefficients [ICC] for asymptomatic and swollen ankles.<sup>13,16,21</sup> Furthermore, the method correlates highly ( $r > 0.88$ ) with the water displacement method for both injured<sup>13,16</sup> and uninjured<sup>11</sup> ankles, thereby conferring some validity for the figure of eight method.

Ankle dorsiflexion range of motion (ROM) during weight bearing is also commonly limited following ankle sprain<sup>10,17</sup>, with consequent high impact on functional activities such as walking<sup>2,3,5,8</sup>, and ascending and descending stairs.<sup>1,12</sup>

Restoration of ankle dorsiflexion ROM is therefore a priority of early rehabilitation.<sup>2,3,8</sup>

Measurement of dorsiflexion in standing<sup>4</sup> simulates the ROM achieved during these functional tasks.<sup>1,6</sup> This is particularly relevant because the torques applied to the ankle in weight-bearing are clearly greater than in non-weight-bearing, and the resultant measurement may be more indicative of the range available for functional activities.<sup>4,5</sup>

Measurement of ankle dorsiflexion ROM using the weight bearing lunge method has been shown<sup>4</sup> to be highly reliable in the laboratory (between-session intra-rater reliability  $ICC_{(3,3)}$  0.97 to 0.98, within-session inter-rater reliability  $ICC_{(2,3)}$  0.99). However, it is unclear whether these methods are efficient and reliable in uncontrolled clinical environments, where high reliability is essential for monitoring of progress and treatment effects.

The aim of the current study therefore, was to assess in the clinical environment the reliability of: i) the figure of eight method; and ii) the weight bearing lunge method for measurements taken within- and between-sessions. The unaffected ankle of participants was investigated because it is not possible to determine between-session reliability on the injured ankle, given the expected rapid changes in swelling and dorsiflexion range<sup>10</sup> and associated confounding effects of intervention on repeatability. Nevertheless, this information is important clinically, because, similar to Phase I and II trials, laboratory results cannot necessarily be generalized to the clinical environment.

## Methods

### Design

A repeated measures design was used to test reliability. When a participant was attending for treatment of an injured ankle, the treating therapist and a second rater measured outcomes on the uninjured ankle before treatment commenced for the affected ankle.

On the first test occasion, raters took two measurements of both ROM and swelling within a 1-hour period. Repeat measurements for between-session reliability were taken approximately one week later. To minimize unblinding, raters used a new data sheet to record each measurement before sealing each data sheet in an opaque envelope for later analysis.

### ***Participants***

The raters were four physiotherapists of varying post-graduate experience (range 4 -15 years, mean 8.2 years). The participants (patient group) consisted of 15 males and 18 females aged 10 to 76 years (mean, standard deviation [SD]: 28, 14.3 years), recruited from staff and patients attending a physiotherapy and sports injury clinic in Sydney, Australia. Eleven participants (3 male, 8 female) were injury-free and asymptomatic. Twenty-two (12 male, 10 female) had sustained a recent unilateral ankle sprain, fifteen while participating in competitive sport, and seven while walking on an uneven surface. The asymptomatic contralateral ankle was measured in injured participants and a randomly selected ankle in healthy participants. While raters routinely used the figure of eight and dorsiflexion lunge methods in their clinical practice, they were familiarized with the standardized protocols for measurement before undertaking data collection.

### ***Protocols***

#### *i) Figure of eight method for measuring swelling*

The protocol used in the current study was based on that described by Mawdsley.<sup>13</sup> Participants were positioned in long-sitting on a bed with the experimental foot resting over the end. (Fig. 1) The following standardized landmarks were marked with a pen prior to measurement: a) the point midway over the anterior ankle between the tibialis anterior tendon and lateral malleolus, b) the navicular tuberosity, c) the base of the fifth metatarsal, and d) the inferior tip of the medial malleolus.



**Figure 2** The weight-bearing lunge method for measuring ankle dorsiflexion.

To blind therapists during measurements, one surface of a double-sided retractable plastic tape measure was blackened leaving the zero point visible. The rater placed the zero point over the mark on the anterior aspect of the ankle and pulled the tape medially over the navicular tuberosity, and then infero-laterally across the medial arch to the proximal aspect of the base of the fifth metatarsal. The tape was then pulled superiorly and medially over the tarsal bones across the inferior aspect of the medial malleolus, and postero-laterally around the Achilles tendon over the distal lateral malleolus to finish at the zero point. The rater tightened the tape measure and then released tension slightly to ensure there was no indentation of soft tissue. To obtain the measurement, a clip was placed at the point of intersection between the zero and finish points of the tape. (Fig. 1) The examiner removed and turned over the tape and recorded the result to the nearest millimeter.

#### *ii) Weight-Bearing lunge for range of ankle dorsiflexion*

The weight-bearing lunge used to measure ankle dorsiflexion range was based on that described by Bennell, et al.,<sup>4</sup>. Each participant stood on an apparatus consisting of a horizontal footplate attached to a vertical board. (Fig. 2)

	Figure of 8 (swelling)	Lunge (Dorsiflexion)
<b>Intra-rater</b> <b>(within-session)</b>	ICC <sub>(2,1)</sub> = 0.99 95% CI: 0.99 to 1	ICC <sub>(1,1)</sub> = 0.97 95% CI: 0.94 to 0.99
<b>Number of Participants</b>	<i>n</i> = 20	<i>n</i> = 20
<b>Intra- rater</b> <b>(between-session)</b>	ICC <sub>(2,1)</sub> = 0.99 95% CI: 0.99 to 1	ICC <sub>(1,1)</sub> = 0.95 95% CI: 0.90 to 0.98
<b>Number of Participants</b>	<i>n</i> = 27	<i>n</i> = 26
<b>Inter- rater</b> <b>(between-session)</b>	ICC <sub>(1,1)</sub> = 0.99 95% CI: 0.98 to 1	ICC <sub>(1,1)</sub> = 0.95 95% CI: 0.88 to 0.98
<b>Number of Participants</b>	<i>n</i> = 25	<i>n</i> = 24

**Table 1** Intraclass correlation coefficient (ICC<sub>(1,1)</sub>) (ICC<sub>(2,1)</sub>) with 95% Confidence Interval (CI) for intra- and inter-rater reliability of figure of eight and weight bearing lunge measurements respectively. Data are shown for within and between sessions (*n* = 4 raters).

Participants aligned the great toe and heel of the test leg over a line marked along the center of the footplate. Participants were instructed not to lift the test heel, checked by the examiner who gently palpated for lifting<sup>4</sup> while the participant moved the knee forward into a lunge position, until the patella touched the midline of the vertical board. To prevent forward movement of the great toe as the knee moved forward over the foot, a block was placed in front of the great toe. The measurement recorded was the distance (cm) from the vertical board to the great toe. Participants were given up to five attempts and the best performance was used for further analysis.

### Data Analysis

SPSS for Windows™ was used to calculate ICCs (ICC<sub>(1,1)</sub> and ICC<sub>(2,1)</sub>) and 95% confidence intervals (CI)<sup>20</sup> for each method, within- and between-raters and within- and between-sessions. The ICC values were interpreted according to the definition of Munro and Page<sup>15</sup>: ICC values 0.00 to 0.25 indicated little, if any correlation; 0.26 to 0.49 low correlation; 0.50 to 0.69 moderate correlation; 0.7 and 0.89 high correlation, and 0.9 to 1.0 indicated very high correlation.

	Measurement occasion	Method	
		Figure of 8 swelling (cm)	Lunge dorsiflexion (cm)
<b>Intra- rater (within-session)</b>	1	51.7 (3.2)	12.2 (2.2)
	2	51.6 (3.2)	12.2 (2.1)
	SEM	(± 0.2)	(± 0.4)
<b>n = 20</b>			
<b>Intra- rater (between-session)</b>	1	52.1 (4.0)	115 (1.9)
	2	52.0 (4.1)	117 (2.0)
	SEM	(± 0.3)	(± 0.4)
<b>n = 27</b>			
<b>Inter- rater (between-session)</b>	1	51.8 (3.4)	120 (2.0)
	2	51.7 (3.6)	122 (2.0)
	SEM	(± 0.3)	(± 0.4)
<b>n = 25</b>			

**Table 2** Means (standard deviation) for the 2 measurement occasions and standard error of the measurement (SEM) for figure of eight swelling and weight bearing lunge dorsiflexion lunge measurements ( $n = 4$  raters).

The SEM<sup>15</sup>, was calculated for repeated measurements on the same participant, within- and between-sessions, and was expressed in the original units of measurement to provide error data in clinically relevant terms. Paired t-tests were used to compare the means for each measurement occasion.

## Results

### *Intra-rater reliability: within session*

Measurements of swelling using the figure of eight method and ankle dorsiflexion using the weight-bearing lunge were taken by the same rater one hour apart. For the figure of eight method, ICC<sub>(2,1)</sub> values were > 0.90 (Table 1), consistent with very high correlation.<sup>15</sup>

There was no difference in mean swelling between the first and second measurements ( $p = 0.32$ ). The SEM was 0.2cm (Table 2) indicating that a therapist taking a repeat measurement of swelling after treatment could be confident on 95% of occasions that any reduction  $>0.4$ cm ( $1.96 \times \text{SEM}$ ) would be due to the treatment. Alternatively, an increase of  $\geq 0.4$ cm would indicate that swelling had increased.

For the dorsiflexion lunge method of measuring range of motion,  $\text{ICC}_{(1,1)}$  values were also  $>0.90$  (Table 1), consistent with very high correlation. There was no difference in mean dorsiflexion range between the first and second measurements ( $p=0.5$ ). The SEM was 0.4cm (Table 2). Thus, a therapist taking a repeat measurement after treatment using the weight-bearing lunge could be confident on 95% of occasions that any change in range of motion of  $>0.8$ cm could be attributed to treatment.

#### *Intra-rater reliability: between session*

Measurements of ankle swelling and ankle dorsiflexion were repeated, on average, 6.8 days (range 2 – 28 days) after the first measurement occasion. For the figure of eight method,  $\text{ICC}_{(2,1)}$  values were  $> 0.90$ , consistent with very high correlation (Table 1). There was no difference in mean swelling between measurement occasions ( $p = 0.29$ ). The SEM was 0.3cm (Table 2), indicating that a therapist taking measurements between treatment sessions could be confident on 95% of occasions that a difference in swelling of  $> 0.7$ cm between treatments would not be due to error.

For the dorsiflexion lunge method of measuring range of motion,  $\text{ICC}_{(1,1)}$  values were  $> 0.90$ , consistent with very high correlation (Table 1). There was no difference in mean dorsiflexion range of motion between sessions ( $p = 0.2$ ). The SEM was 0.4cm (Table 2) indicating that a therapist taking repeat measurements between occasions could be confident on 95% of occasions that a difference in ROM of  $> 0.8$ cm would not be due to error.

#### *Inter-rater reliability: between sessions*

To determine inter-rater reliability, two different raters made the repeat measurements of swelling and range of motion, on average 6.8 days (range 2 – 28 days) apart. For the figure of eight method,  $\text{ICC}_{(1,1)}$  values were  $> 0.90$ , indicating very high reliability (Table 1). There was no difference in mean swelling ( $p = 0.2$ ) between the two measurement occasions. The SEM was 0.3cm (Table 2), indicating that a different therapist repeating the measurement one week later could be confident on 95% of occasions that a change in swelling of  $> 0.6$ cm would not be due to error.

Similarly, for the dorsiflexion lunge method of measuring range of motion,  $\text{ICC}_{(1,1)}$  values for inter-rater reliability were  $> 0.90$ , consistent with very high correlation (Table 1). There was no difference in mean range of motion for dorsiflexion ( $p = 0.09$ ). The SEM was 0.4cm (Table 2), indicating that a therapist taking a repeat measurement one week after the first occasion could be confident on 95% of occasions that any difference in ROM of  $>0.8$ cm after treatment would not be due to error.

## Discussion

The current results indicate that intra- and inter-rater reliability were very high for measurements taken in the clinic for both the figure of eight method (ankle swelling) and the weight-bearing lunge method (ankle dorsiflexion). Whilst previous research has reported acceptable reliability in a well-controlled laboratory environment, the current study demonstrated reliability of these methods in the variable clinical environment. Very high intra and interrater reliability was observed for measurements repeated within a single session, and after a one-week interval. Therefore, clinicians can use both techniques with confidence within and between sessions to determine the effects of interventions to improve ankle swelling and dorsiflexion range following ankle sprain. The results presented here will assist clinicians with decisions regarding the management of ankle sprain and monitoring progress with treatment.

Despite the inherent differences between the demands of the clinical environment and the laboratory environment, such as time constraints during measurement procedures and a more unpredictable environment<sup>7</sup>, the current findings for the figure of eight tape method are comparable to data derived from laboratory studies. Very high intra- and inter-rater reliability (ICC values 0.98 - 0.99) have been reported using the figure of eight method for injured<sup>13,16</sup> as well as asymptomatic<sup>11,21</sup> ankles. However, previous research has only documented the reliability of the figure of eight method for repeated measurements taken within a single session.<sup>11,13,16</sup> The current study observed very high intra and interrater reliability both within and between-sessions, with a comparable SEM of 0.4 to 0.5cm<sup>11,13</sup>, and therefore has demonstrated that different therapists can treat the same patient on different occasions and use the figure of eight method to confidently determine treatment effects. Furthermore, the small SEM observed in the current study informs clinicians that changes in swelling of greater than 0.7cm are more likely due to intervention effects than error. This suggests that the error is considerably less than changes that would be considered clinically worthwhile.

The figure of eight method has been reported to correlate well with water displacement methods for measurements of ankle swelling after lower limb injury.<sup>13,16</sup> While water displacement is the gold standard method for measuring lower limb volume<sup>11,13,16</sup>, the method is time consuming, requiring between 5 and 6 minutes to perform, whereas the figure of eight method requires approximately 30 seconds to perform.<sup>11</sup> It also requires relatively sophisticated equipment, unlike the tape measure method. Therefore, the figure of eight method may not only be a more time-efficient method for measuring ankle swelling, but also can be used without sacrificing reliability; even between raters, and between sessions.

Similarly, very high<sup>15</sup> intrarater reliability results have been reported for the weight bearing lunge method of measuring dorsiflexion in participants with asymptomatic ankles, within and between sessions, and for different raters taking repeat measurements within a single session.<sup>4</sup> The current data recorded in a clinical environment are comparable to previous data collected in a laboratory environment, and again indicate that use of a cost and time-efficient method does not sacrifice reliability. Clinicians can therefore be confident that the weight-bearing lunge method for measuring dorsiflexion range is robust in the clinical environment. Furthermore, the small SEM suggests that clinicians detecting changes in ROM of greater than 1cm are more likely to be observing intervention effects than error.

## Conclusions

Whereas previous studies using the figure of eight method for measuring ankle swelling and weight bearing lunge for measuring dorsiflexion have been conducted in laboratory settings, the current study was conducted in a clinical setting characterized by more variable environmental conditions and constraints that replicated conditions likely to be encountered by clinicians during rehabilitation of ankle sprain. In the current study, intra and interrater reliability for each method was observed to remain very high for both within and between sessions data. Therefore, with adequate familiarization, these simple, reliable, and time efficient methods can be used with confidence by clinicians with varying levels of expertise to assess treatment effects on swelling and ankle dorsiflexion in clinical populations.

---

**References**

1. Andriacchi TP, Andersson GB, Fermier RW, Stern D, Galante JO: A study of lower-limb mechanics during stair-climbing. *J Bone Joint Surg* 62A: (5) 749 - 757, 1980.
2. Bahr R, Engebretsen L: Acute ankle sprains: a functional treatment plan for injured athletes. *Consultant* 36 (4): 675 - 680, 1996.
3. Balduini FC, Vegso JJ, Torg JS, Torg E: Management and rehabilitation of ligamentous injuries to the ankle. *Sport Med* 4 (5): 364 - 380, 1987.
4. Bennell K, Talbot R, Wajswelner H, Techovanich W, Kelly D: Intra-rater and inter-rater reliability of a weight-bearing lunge measure of ankle dorsiflexion. *Aust J Physiother* 44 (3): 175 - 180, 1998.
5. Bohannon RW, Tiberio D, Zito M: Selected measures of ankle dorsiflexion range of motion: differences and intercorrelations. *Foot Ankle Int* 10 (2): 99 - 103, 1989.
6. Bohannon RW, Tiberio D, Zito MA: Improving ankle dorsiflexion. *Phys Ther* 77 (9): 982 - 983, 1997.
7. Carr JH, Shepherd RB. *A Motor Relearning Programme for Stroke*. 2nd ed. London: Butterworth-Heinemann Ltd, 1982.
8. Crosbie J, Green T, Refshauge K: Effects of reduced ankle dorsiflexion following lateral ligament sprain on temporal and spatial gait parameters. *Gait Post* 9 (3): 167-172, 1999.
9. Garrick JG, Requa RK. The epidemiology of foot and ankle injuries in sports. *Clinics in Podiatric Medicine & Surgery* 6 (3): 629 - 637, 1989.
10. Green T, Refshauge K, Crosbie J, Adams R: A Randomised controlled trial of a passive accessory joint mobilization on acute ankle inversion sprains. *Phys Ther* 81 (4): 984 - 994, 2001.
11. Henschke N, Boland RA, Adams RD: Responsiveness of two methods for measuring foot and ankle volume. *Foot Ankle Int* 27 (10): 826-832, 2006.
12. Lindsjö U, Danckwardt-Lillieström G, Sahlstedt B: Measurement of the motion range in the loaded ankle. *Clin Orthop Rel Res* 199: 68 - 71, 1985.
13. Mawdsley RH, Hoy DK, Erwin PM: Criterion-related validity of the figure-of-eight method of measuring ankle edema. *J Orthop Sports Phys Ther* 30 (3): 149 - 153, 2000.
14. McKay GD, Goldie PA, Payne WR, Oakes BW: Ankle injuries in basketball: injury rate and risk factors. *British Journal of Sports Medicine*. 35 (2): 103 - 108, 2001.
15. Munro BH, Page EB. *Statistical Methods for Health Care Research*. 2nd ed. J.B. Lippincott Co, Philadelphia, Pennsylvania, 1993.
16. Petersen EJ, Irish SM, Lyons CL, Mikaski SF, Bryan JM, Henderson NE, Masullo LN. Reliability of water volumetry and the figure of eight method on subjects with ankle joint swelling. *J Orthop Sports Phys Ther* 29 (10):609-615, 1999.
17. Reid DC. *Sports Injury Assessment and Rehabilitation*. Churchill Livingstone, New York, 1992.
18. Schneck CD, Mesgarzadeh M, Bonakdarpour A, Ross GJ: MR imaging of the most commonly injured ankle ligaments. *Radiology* 184 (2): 499 - 512, 1992.
19. Shrier I: Treatment of lateral collateral ligament sprains of the ankle: a critical appraisal of the literature. *Clin J Sport Med* 5 (3): 187 - 195, 1995.
20. Shrout P, Fleiss J: Intraclass correlations: Uses in assessing rater reliability. *Psych Bull* 86 (3): 420 - 428, 1979.
21. Tatro-Adams D, McGann SF, Carbone W: Reliability of the figure-of-eight method of ankle measurement. *J Orthop Sports Phys Ther* 22 (4): 161 - 163, 1995.