Measurement Reliability of Swelling in the Acute Ankle Sprain

by Cameron P. Watson, MAppSc\textsuperscript{14}, Robert A. Boland, PhD\textsuperscript{13,1}, Kathryn M. Refshauge, PhD\textsuperscript{2}

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: ±0.2cm for figure of eight, and ±0.4cm for dorsiflexion lunge methods within a session, and ±0.3cm and ±0.4cm 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 (>0.5cm) 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)

Ankle sprain is one of the most common injuries amongst sporting populations\textsuperscript{9}, 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

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.\textsuperscript{14} 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\textsuperscript{18,19}, but can also accumulate around the subtalar, talocrural and inferior tibiofibular joints.
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\textsuperscript{13,16}, 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\textsuperscript{13} 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.\textsuperscript{13, p.131} (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).\textsuperscript{13,16,21} Furthermore, the method correlates highly ($r>0.88$) with the water displacement method for both injured\textsuperscript{13,16} and uninjured\textsuperscript{11} 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\textsuperscript{10,17}, with consequent high impact on functional activities such as walking\textsuperscript{2,3,5,8}, and ascending and descending stairs.\textsuperscript{1,12}

Restoration of ankle dorsiflexion ROM is therefore a priority of early rehabilitation.\textsuperscript{2,3,8}

Measurement of dorsiflexion in standing\textsuperscript{4} simulates the ROM achieved during these functional tasks.\textsuperscript{1,6} 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.\textsuperscript{4,5}

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\textsuperscript{10} 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**

1. **Figure of eight method for measuring swelling**

The protocol used in the current study was based on that described by Mawdsley. 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.

2. **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.,. Each participant stood on an apparatus consisting of a horizontal footplate attached to a vertical board. (Fig. 2)
Table 1 | Intraclass correlation coefficient (ICC(1,1)) (ICC(2,1)) 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).

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

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 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(1,1) and ICC(2,1)) and 95% confidence intervals (CI) 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: 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.
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\(^ {15} \), 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\(_{(2,1)}\) values were > 0.90 (Table 1), consistent with very high correlation.\(^ {15} \)
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 >0.4cm (1.96 x SEM) would be due to the treatment. Alternatively, an increase of ≥0.4cm would indicate that swelling had increased.

For the dorsiflexion lunge method of measuring range of motion, 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 >0.8cm 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, ICC\(_{1,1}\) 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 > 0.7cm between treatments would not be due to error.

For the dorsiflexion lunge method of measuring range of motion, ICC\(_{1,1}\) values 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 >0.8cm after treatment 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, 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 > 0.6cm would not be due to error.

Similarly, for the dorsiflexion lunge method of measuring range of motion, 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 >0.8cm 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 inter-rater 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, 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 as well as asymptomatic ankles. However, previous research has only documented the reliability of the figure of eight method for repeated measurements taken within a single session. 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, 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. While water displacement is the gold standard method for measuring lower limb volume, 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. 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 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. 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