Manual Therapy 18 (2013) 264–268

Contents lists available at SciVerse ScienceDirect

Manual Therapy

journal homepage: www.elsevier.com/math

Technical and measurement report

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

Simon O'Shea [a,*], Kate Grafton [b]

[a] Chesterfield Royal Hospital, Calow, Chesterfield S44 5BL, UK
[b] Sheffield Hallam University, Sheffield S10 2BP, UK

ARTICLE INFO

Article history:
Received 27 March 2012
Received in revised form
30 August 2012
Accepted 31 August 2012

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 min. The process was repeated within 3 h. Intra-rater reliability results indicated excellent correlation of measurements (intraclass correlation coefficients (ICCs) of 0.98–0.99). Standard error of measurement (SEM), 95% limits of agreement (LOA) and coefficient of repeatability (CR) calculations indicated suitably low ranges of measurement variance (SEM = 0.4 cm, LOA = ±1.28 to ±1.47 cm and CR = 1.21–1.35 cm). Inter-rater reliability was also deemed excellent (ICC = 0.99, SEM = 0.3 cm, LOA = ±0.83 to ±1.47 cm, CR = 1.44 cm). The modified lunge technique therefore demonstrates excellent intra and inter-rater reliability.

1. Introduction

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

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

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

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

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

* Corresponding author. Tel.: +44 01709 424400.
E-mail address: simoncoshea@hotmail.com (S. O'Shea).

1356-689X/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.math.2012.08.007

2. Method

2.1. Pilot study

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

2.2. Subjects

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

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

2.3. Raters

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

2.4. Procedure

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

2.5. Data analysis

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

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

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

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

3. Results

Thirteen volunteers completed the study. No gender bias was evident. Histograms plus Bland and Altman plots confirmed that the data was homoscedastic (see Fig. 3). No dependence upon the mean and minimal measures beyond 95% LOA were evident (see Fig. 3) confirming a lack of anomalies or systematic bias. All raw data and SPSS outcomes are available as electronic files.

3.1. Intra-rater reliability

Excellent intra-rater correlation was found for rater 1 (ICC = 0.98) and rater 2 (0.99). See Tables 1 and 2 for statistical analysis results. The level of error was also good (SEM = 0.4 cm) for both raters. The

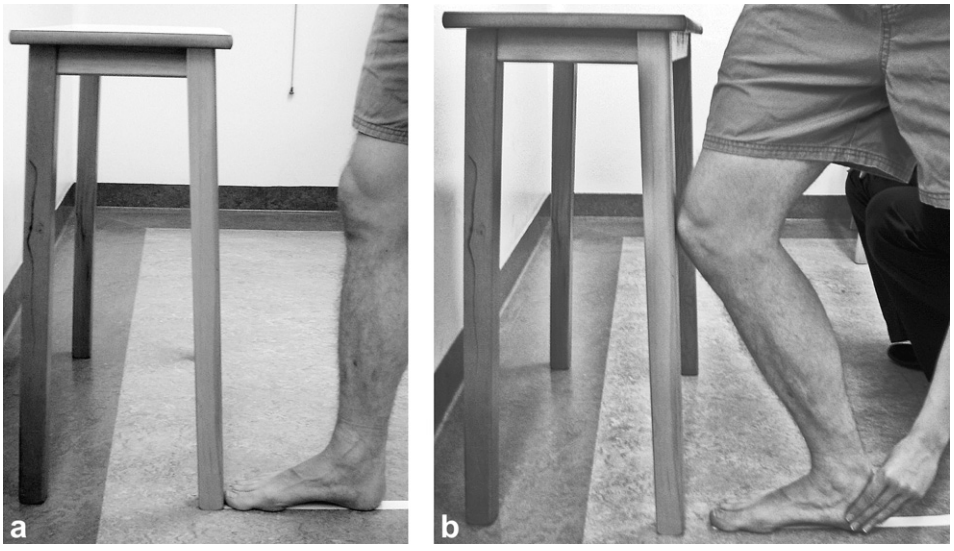

Fig. 1. (a) Starting position with foot placed on tape and toe against upright of table. (b) Final lunge position.