Intrarater Reliability and Agreement of Linear Encoder Derived Heel-Rise Endurance Test Outcome Measures in Healthy Adults

Christopher Byrne[1,2], David J Keene[1], Sarah E Lamb[1], Keith Willett[1]

[1]Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, UK. [2]School of Health Professions, Faculty of Health and Human Sciences, Plymouth University, UK.

Acceptance date: 10th July 2017

Key words: Measurement error; outcome measure; plantar flexion; Achilles tendon.

Conflict of Interest and Financial Disclosure Statement: This report was developed in association with the National Institute for Health Research (NIHR) Oxford Musculoskeletal Biomedical Research Unit funding scheme and the PATH-2 trial, funded by the Efficacy and Mechanism Evaluation Programme, an MRC and NIHR partnership. The PATH-2 trial is registered at ClinicalTrials.gov with registration number NCT02302664. The views expressed in this publication are those of the authors and not necessarily those of the MRC, NHS, NIHR or the Department of Health. The bespoke software was developed in association with Ergotest Innovation A.S., Porsgrunn, Norway, who had no role in the study design; data collection, access, analysis, or interpretation; writing of the report; or the decision to publish. The authors certify they have no affiliations with or financial interest in the subject matter or materials discussed in the report and as such have no conflicts of interest.

Corresponding Author: Dr Chris Byrne, Peninsula Allied Health Centre, Derriford Road, Plymouth, PL6 8BH, UK. Email: chris.byrne@plymouth.ac.uk

Abstract

A linear encoder measuring vertical displacement during the heel-rise endurance test (HRET) enables the assessment of work and maximum height in addition to the traditional repetitions measure. We aimed to compare the test-retest reliability and agreement of these three outcome measures. Thirty-eight healthy participants (20 females, 18 males) performed the HRET on two occasions separated by a minimum of seven days. Reliability was assessed by the intraclass correlation coefficient (ICC) and agreement by a range of measures including the standard error of measurement (SEM), coefficient of variation (CV), and 95% limits of agreement (LoA). Reliability for repetitions (ICC = 0.77 (0.66, 0.85)) was equivalent to work (ICC = 0.84 (95% CI 0.76, 0.89)) and maximum height (ICC = 0.85 (0.77, 0.90)). Agreement for repetitions (SEM = 6.7 (5.8, 7.9); CV = 13.9% (11.9, 16.8%); LoA = -1.9 ± 37.2%) was equivalent to work (SEM = 419 J (361, 499 J); CV = 13.1% (11.2, 15.8%); LoA = 0.1 ± 34.8%) with maximum height superior (SEM = 0.8 cm (0.6, 1.0 cm); CV = 6.6% (5.7, 7.9%); LoA = 1.3 ± 17.1%). Work and maximum height demonstrated acceptable reliability and agreement that was at least equivalent to the traditional repetitions measure.

Introduction

The heel rise endurance test (HRET) is a popular method of assessing ankle function in research and clinical practice (Hebert-Losier et al., 2009a, Hebert-Losier et al., 2009b). The HRET involves repetitive concentric-eccentric muscle action of the plantar flexors in unipedal stance until volitional task failure with a side-to-side comparison of the maximum number of repetitions defining the outcome measure (Hebert-Losier et al., 2009a). Maximum repetitions demonstrates acceptable test-retest reliability and agreement (Moller et al., 2005) and is consistently employed as an outcome measure in rehabilitation studies of Achilles tendon rupture (ATR) (Bostick et al., 2010, Buchgraber and Passler, 1997, Moller et al., 2002, Weber et al., 2003).

Silbernagel et al. (Silbernagel et al., 2006, Silbernagel et al., 2010) recently introduced a new HRET measuring device in the form of a linear displacement sensor attached to the heel enabling the height of each repetition to be measured and three outcome measures quantified i.e. number of repetitions, total work in joules, and maximum heel rise height in cm. The authors reported that the two novel outcome measures of work and maximum height were more sensitive than repetitions in detecting functional impairment at 6, 12, and 24 months following ATR and recommended their use as outcome measures in future research (Silbernagel et al., 2010, Olsson et al., 2011). Whilst work and maximum height have demonstrated good criterion validity and responsiveness (Silbernagel et al., 2010, Nilsson-Helander et al., 2010, Olsson et al., 2011), the measurement properties of reliability and agreement for these two novel indices have yet to be determined in either healthy or clinical populations. Our purpose was to evaluate these measurement properties, firstly in healthy participants, with a view to employing the HRET as the primary outcome measure in a large multi-centre randomised controlled trial comparing treatment with platelet-rich plasma injection versus placebo in acute Achilles tendon rupture (PATH-2 Trial, ClinicalTrials.gov registration number NCT02302664).

Therefore, the aim of the current study was to measure and compare the intrarater test-retest reliability and measurement agreement of the three HRET outcome measures in healthy adult participants during a standardised and computerised HRET employing a linear displacement sensor. The findings are reported in accordance with recent guidelines for reporting reliability and agreement studies (Kottner et al., 2011).

Methods

Participants

Participants were recruited through advertisement posters on University research community notice boards. Inclusion criteria were age 18 years and above, able to give informed consent and follow instructions. Exclusion criteria were history (in either leg) of Achilles tendon pain, previous Achilles tendon rupture, previous major ankle injury or deformity, and recent lower limb injury. Forty healthy participants (20 males & 20 females) volunteered with written informed consent to participate in this study, which was approved by the Institutional ethics committee. Data are presented on 38 participants (18 males & 20 females; mean ± SD age 36 ± 9 years, body mass 71.5 ± 15.3 kg) because two participants withdrew from the study following the first test.

Heel-Rise Endurance Test

To evaluate test-retest reliability and agreement, the HRET was performed on two separate occasions separated by an interval of at least seven days. Since this form of exercise has a large eccentric component and is unaccustomed for most people it often produces the symptoms of exercise-induced muscle damage (e.g. muscle weakness and delayed-onset muscle soreness) and therefore a minimum seven-day recovery period was chosen to allow full recovery between the test and retest (Byrne et al., 2004). To determine the intrarater reliability and agreement of the procedures in our own hands (Weir, 2005), a single trained outcome assessor performed all HRET measurements and was therefore an unblinded assessor. Participants were instructed to maintain their normal levels of physical activity between tests and to avoid any intense physical activity in the hours before testing. Before testing, participants completed the Lower Extremity Functional Scale (scored from 0-80 with higher scores indicating better function) (Binkley et al., 1999), had their body mass measured on calibrated class III scales, watched a video demonstration of the HRET, read standardised written instructions detailing their expected conduct during the test, and completed a standardised warm-up. The warm-up consisted of five minutes continuous walking at usual pace followed by 10 double leg heel rises on a 10° incline box guided by a digital metronome at a rate of 30 heel rises$\cdot$min^{-1}.

During the test, participants were instructed to adopt a single leg stance with full knee extension on a 10° incline box facing a wall with only fingertip support; to raise the heel as high as possible on each repetition at a rate of 30 rises$\cdot$min^{-1} guided by a digital metronome; and to perform as many heel raises as possible (Hebert-Losier et al., 2009a, Silbernagel et al., 2010). The dominant limb was tested first and then the non-dominant limb after three minutes of recovery. The height of each heel-rise was measured by a spring-loaded cord attached to the bare heel of the participant and connected to a linear displacement sensor with a measurement resolution and sample rate of 0.019 mm and 200 Hz, respectively (Encoder, MUSCLELAB™, Ergotest Innovation A.S., Porsgrunn, Norway). Each test was video recorded and a bespoke software integrated encoder and video data (PATH-2, MUSCLELAB™, Ergotest Innovation A.S., Porsgrunn, Norway). Figure 1 illustrates the experimental set-up. The software was programmed with 1.0 cm concentric (upward) and eccentric (downward) thresholds to provide tolerance for minor movements and signal directional changes (eccentric ≥1.0 cm) and new repetitions (concentric ≥1.0 cm). Participants either stopped (i.e. volitional task failure) or were audibly instructed to stop with both feet flat on the box whenever any of the following test termination criteria were observed: inability to keep pace with the metronome; inability to maintain full knee extension of the standing leg; or using more than fingertip support. The desired endpoint was volitional task failure, however the outcome assessor used verbal prompts whenever the termination criteria were observed and stopped the test if the participant did not respond to two consecutive prompts.

**** Insert Figure 1 Here ****

Data Processing

Displacement and video data were reviewed after each HRET to identify and eliminate any movement artefacts from the data that occurred after test termination but before the linear encoder had stopped recording data. Large amplitude movement artefacts (mean ± SD height = 29.6 ± 9.0 cm) were observed and removed in 4.6% of tests due to participants moving their leg to alleviate discomfort immediately after test termination. For comparison to previous research, only repetitions with height ≥5.0 cm were included in the analysis (Moller et al., 2005, Svantesson et al., 1998). Repetitions <5.0 cm were observed in 31.6% of tests resulting in a (mean ± SD) minor loss of 1.8 ± 0.9 (range 1-5) repetitions per affected test. These repetitions were removed from individual datasets during the post-HRET data and video review. Additionally, two consecutive repetitions <5.0 cm was considered a final test termination criteria in the post-HRET data and video review and was observed in 11.2% of tests (Buchgraber and Passler, 1997). Three HRET outcome measures were defined as: number of repetitions (n); work (J) as the product of body mass (kg), total vertical displacement (m), and the constant 9.807 converting kilopond-metres to joules; and maximum height (cm) as the greatest height of a single repetition (Silbernagel et al., 2010). Additionally, a limb symmetry index was computed for each measure as the performance of the left leg as a percentage of the right leg (Silbernagel et al., 2010).

Statistical Analysis

A number of statistical methods for assessing measurement error were employed (Kottner et al., 2011). Descriptive data are presented as mean and standard deviation. Evidence of test-retest systematic bias was analysed with paired sample t-tests to investigate changes in the mean. The mean differences are presented with the 95% confidence interval (95% CI), and the standardized mean difference effect size (*d*) was calculated and interpreted as: trivial (<0.2); small (≥0.2); medium (≥0.5); or large (≥0.8). The intraclass correlation coefficient (ICC) type 2,1 (two-way random, consistency definition) with 95% CI examined intrarater reliability as the relative consistency of individuals across days (Weir, 2005). Measurement agreement (i.e. the degree to which scores are identical) was examined with four methods. Firstly, the standard error of measurement (SEM) with 95% CI described the variation in the same units as the original measurement (Weir, 2005, Hopkins, 2000). Secondly, following logarithmic (natural) transformation of the data, the coefficient of variation (CV) with 95% CI described the variation in percent (Hopkins, 2000). Thirdly, the minimal detectable change at the 90% confidence level (MDC_{90}) was computed to estimate values representing a meaningful and true change (Weir, 2005). Finally, 95% limits of agreement (LoA) were expressed in original measurement units and as percent differences (Bland and Altman, 1986). During LoA analysis, Pearson's correlation coefficient (*r*) was employed to determine if absolute differences were proportional to the magnitude of measurement (i.e. heteroscedasticity). If heteroscedasticity was still present (i.e. $P < 0.05$) after expressing absolute differences as a percentage of the individual mean values, a logarithmic (natural) transformation of the original data was performed followed by antilogs to express the LoA as ratio percentages (Bland and Altman, 1986). Statistical analysis was performed with IBM SPSS Statistics 23.

Results

The 38 participants completed the study with a test-retest interval of 9 ± 2 (range 7-14) days. No changes in body mass (test = 71.5 ± 15.5 kg; retest = 71.5 ± 15.4 kg, $P = 0.745$) or LEFS (test = 78.5 ± 4.4; retest = 78.7 ± 4.3, $P = 0.071$) were observed. Table 1 illustrates that no significant test-retest changes in the mean were observed for each outcome measure with all effect sizes being trivial. Table 2 illustrates reliability and agreement data for each outcome measure. Heteroscedasticity was observed in repetitions and work LoA data. Expressing LoA as percentages resolved the issue for work but not repetitions. Logarithmic transformation and expression as ratio LoA also failed to resolve the heteroscedasticity issue for repetitions. Figure 2 illustrates LoA absolute and percent differences plots for the three outcome measures.

**** Insert Table 1 Here ****

**** Insert Table 2 Here ****

**** Insert Figure 2 Here ****

Fatigue was evident during the tests, as illustrated by the 5.8 ± 2.2 (1.5-12.0) cm or 42.3 ± 13.9 (12.0-69.0) % reduction from maximum to minimum repetition height. Maximum height (13.6 cm) occurred on average on the 4th repetition (12% of total) with minimum height (7.8 cm) occurring on the 32nd repetition (93% of total).

Discussion

With healthy adult participants, a standardised protocol, and a single outcome assessor, the novel HRET outcome measures of work and maximum height demonstrated acceptable reliability and measurement agreement that was at least equivalent to the traditional repetitions measure. No evidence of systematic bias was observed for the three measures on each limb suggesting the absence of any learning effect on performance. ICC, representing the test-retest relative consistency of individuals was good (i.e. >0.7) for all measures (de Vet et al., 2006). Although ICC estimates for work (ICC = 0.84 (0.76 to 0.89)) and maximum height (ICC = 0.85 (0.77 to 0.90)) appeared better than repetitions (ICC = 0.77 (0.66 to 0.85)), we interpreted reliability as equivalent between the three measures due to the overlap in 95% CI. ICC for limb symmetry indices was understandably poor due to the low range of values in our healthy participants (de Vet et al., 2006).

The SEM and CV agreement data in Table 2 indicate the typical error expected from test to test for any one participant (Hopkins, 2000). It is expected that 68% of differences between tests will lie within the SEM or CV. The CV for repetitions (13.9% (11.9, 16.8%)) was equivalent to work (13.1% (11.2, 15.8) with maximum height (6.6% (5.7, 7.9%)) demonstrating superior agreement than both repetitions and work. The MDC_{90} data in Table 2 indicate that 90% of individuals will demonstrate variation less than this magnitude when retested and that a true change in performance is one that exceeds the MDC_{90} value (Weir, 2005). The 95% LoA results illustrated in Figure 2 and Table 2 confirm the absence of systematic bias for all measures and indicate that for a new individual from the studied population, it would be expected with 95% probability, that test-retest differences would fall within the LoA (Bland and Altman, 1986). Measurement error was proportional to the mean for repetitions and work, and although percent LoA resolved this issue for work, heteroscedasticity remained for repetitions even after logarithmic transformation of the data (Bland and Altman, 1986). As illustrated in Figure 2 (B, D, F), percent LoA revealed that random error for maximum height (± 17.1%) was less than half that of work (± 34.8%) and repetitions (± 37.2%). In summary, measurement agreement for repetitions and work is generally equivalent although repetitions data exhibit a complex relationship between the magnitude of measurement and test-retest difference. Additionally, maximum height exhibits superior measurement agreement than both repetitions and work.

The number, reliability, and agreement of our repetitions data are consistent with previously reported data for healthy participants (Svantesson et al., 1998, Moller et al., 2005, Hebert-Losier et al., 2009a) . For example, Moller et al.(Moller et al., 2005) reported test-retest means of 29.2-30.4, a non-significant mean difference of 1.2 ($P = 0.71$), ICC 0.84, CV 19.1%, and LoA 1.2 ± 16.5 repetitions for 10 healthy males. Our data are also very consistent with the performance of the healthy uninvolved limb of Achilles tendon rupture patients (Silbernagel et al., 2010, Olsson et al., 2011). For example, Olsson et al. (Olsson et al., 2011) employed the same measurement equipment and protocol as the current study and observed 35 ± 12 repetitions, 3043 ± 1087 J of work, and 13.6 ± 2.5 cm maximum height for the uninvolved limb two years after conservatively managed Achilles tendon rupture.

We have reported the six key testing parameters recommended by Hebert-Losier et al. (2009a) to aid standardisation of the HRET (i.e. ankle starting position (10° dorsiflexion); knee starting position (full extension); height (as high as possible); pace (30 raises·min^{-1}); balance support (finger tips); outcome measures (repetitions, work, maximum height); and test termination criteria (see Methods)). The linear encoder employed a minimum 1.0 cm threshold height that required attainment for the repetition to be counted. In addition, our data processing employed a further minimum height threshold of 5.0 cm that required attainment for a repetition to be counted and two consecutive repetitions <5.0 cm served as a test termination criteria. The 5.0 cm threshold was employed for consistency with previous research (Buchgraber and Passler, 1997, Moller et al., 2005) and to enable the comparison of our reproducibility data with previous studies (Moller et al., 2005). Despite almost a third of tests including repetitions <5.0 cm, the average number per affected test was only 1.8 with a maximum of five repetitions, and only 11.2% of tests were terminated on the basis of two consecutive <5.0 cm repetitions. Whilst we recommend use of the 1.0 cm threshold, the 5.0 cm threshold does not warrant recommendation. In addition, we recommend video recording the HRET to identify movement artefacts at the end of testing, which occurred in 4.6% of our tests, and could lead to erroneous values, particularly for maximum height and work, if not identified.

Our instruction to raise the heel as high as possible with each repetition is the most frequent height criterion reported in HRET research (Hebert-Losier et al., 2009a, Hebert-Losier et al., 2009b) and produced fatigue (42.3 % reduction in height) and task failure on average after 33.8 repetitions. Whilst acceptable reliability and agreement have been reported with alternative sub-maximal approaches, such as attaining a 5 cm height threshold on each repetition until task failure (Haber et al., 2004), we favour the current approach employing full range of movement at the ankle.   A further approach has been described in which the test is terminated when near maximum height can no longer be attained (Sman et al., 2014). This resulted in fewer repetitions (i.e. 23 ± 13.3) than the grand mean of 33.8 in our study and we argue that our protocol represents a better test of the endurance property of the muscle-tendon unit.

In addition to acceptable measurement error, recent evidence indicates that work and maximum height are more sensitive impairment measures than repetitions following ATR (Silbernagel et al., 2010, Olsson et al., 2011). For example, repetitions classified the percentage of patients having normal function at six and twelve months after ATR as 38% and 63%, respectively (Silbernagel et al., 2010). This compared to 9% and 23% for work and 6% and 22% for maximum height at six and twelve months, respectively (Silbernagel et al., 2010). Work and maximum height (but not repetitions) also demonstrated significant positive associations at six months post ATR with the Achilles tendon Total Rupture Score (Silbernagel et al., 2010), a validated patient reported outcome measure (Nilsson-Helander et al., 2007). Work quantifies the positive vertical displacement of each repetition and therefore brings greater scientific rigour and accuracy to the quantification of muscle-tendon unit endurance performance. Maximum height appears reflective of reduced end-range strength (Mullaney et al., 2006) and increased tendon compliance and length (Silbernagel et al., 2012, Schepull et al., 2007) following ATR. We therefore support the adoption of work and maximum height as objective outcome measures of muscle-tendon function in ATR research and clinical practice.

Study Limitations

The HRET described in this study is designed for use as an objective outcome measure in ATR rehabilitation. Our use of a healthy population limits the generalisability of our findings to a clinical population. Ideally, this study should be replicated in an ATR population. This would provide important information on reliability and agreement (particularly for the limb symmetry index) and enable the computation of the smallest detectable change for treatment evaluation. The use of a single trained outcome assessor (intrarater reliability) also limits the generalisability of our findings to interrater reliability. Nevertheless, we would anticipate similar interrater reliability and agreement if the standardisation procedures described in this study are adhered to.

Conclusion

The reliability and measurement agreement of novel HRET work and maximum height outcome measures were at least equivalent to the traditional repetitions measure in healthy adults. Maximum height, in particular, demonstrated the best measurement agreement. These novel indices are supported by a stronger scientific rationale than repetitions and their use should add value to the ankle function assessment of ATR patients in research and clinical practice.

Acknowledgements

This report was developed in association with the National Institute for Health Research (NIHR) Oxford Musculoskeletal Biomedical Research Unit funding scheme and the PATH-2 trial, funded by the Efficacy and Mechanism Evaluation Programme, an MRC and NIHR partnership. The PATH-2 trial is registered at ClinicalTrials.gov with registration number NCT02302664. The views expressed in this publication are those of the authors and not necessarily those of the MRC, NHS, NIHR or the Department of Health. The bespoke software was developed in association with Ergotest Innovation A.S., Porsgrunn, Norway, who had no role in the study design; data collection, access, analysis, or interpretation; writing of the report; or the decision to publish. The authors certify they have no affiliations with or financial interest in the subject matter or materials discussed in the report.

Table 1: Assessment of test-retest systematic bias for three heel-rise endurance test outcome measures in 38 healthy adults.

	Test (mean ± SD)	**Retest (mean ± SD)**	**Mean Difference (95% CI)**	***P* Value**	**Effect Size (*d*)**
Repetitions					
Left (n)	32.6 ± 11.4	32.6 ± 13.2	0.0 (-2.8, 2.9)	.985	0.0
Right (n)	34.7 ± 12.1	35.1 ± 18.0	-0.4 (-3.7, 3.0)	.826	-0.02
LSI (%)	94.6 ± 13.0	95.9 ± 15.2	-1.3 (-6.7, 4.1)	.624	-0.09
Work					
Left (J)	2436 ± 908	2473 ± 1125	-37 (-211, 137)	.666	-0.04
Right (J)	2583 ± 863	2681 ± 1224	-98 (-314, 116)	.358	-0.09
LSI (%)	94.5 ± 13.8	93.6 ± 14.2	0.9 (-4.5, 6.2)	.745	-0.06
Maximum Height					
Left (cm)	13.5 ± 2.1	13.7 ± 2.0	-0.1 (-0.5, 0.2)	.448	-0.10
Right (cm)	13.6 ± 1.8	13.8 ± 1.9	-0.2 (-0.6, 0.2)	.276	-0.11
LSI (%)	99.8 ± 9.7	99.3 ± 7.4	0.4 (-3.2, 4.1)	.812	0.06

Abbreviations: LSI, limb symmetry index; 95% CI, 95% confidence interval.

Table 2: Intrarater test-retest reliability and agreement for three heel-rise endurance test outcome measures in 38 healthy adults.

	ICC	**SEM**	**CV**	**MDC$_{90}$**	**LoA**	**Percent LoA**
	(95% CI)	**(95% CI)**	**(95% CI)**		**(bias ± 95%)**	**(bias ± 95%)**
Repetitions						
Left	0.75 (0.58, 0.86)	6.1 (5.0, 7.9)	15.0 (12.1, 19.9)	14.2	0.0 ± 16.9*	-1.3 ± 37.9*
Right	0.78 (0.61, 0.88)	7.4 (6.0, 9.6)	12.6 (10.1, 16.7)	16.9	0.4 ± 20.1*	-2.4 ± 37.0*
Pooled	0.77 (0.66, 0.85)	6.7 (5.8, 7.9)	13.9 (11.9, 16.8)	15.5	0.2 ± 18.5*	-1.9 ± 37.2*
LSI	0.33 (0.03, 0.74)	11.6 (9.5, 15.0)	11.8 (9.5, 15.5)	27.0	1.3 ± 32.1*	1.0 ± 32.8
Work						
Left	0.87 (0.76, 0.93)	375 (305, 485)	13.5 (10.9, 17.9)	874	37 ± 1038*	-0.5 ± 33.5
Right	0.81 (0.66, 0.90)	463 (377, 599)	13.4 (10.8, 17.7)	1080	99 ± 1283*	0.6 ± 36.4
Pooled	0.84 (0.76, 0.89)	419 (361, 499)	13.1 (11.2, 15.8)	977	68 ± 1161*	0.1 ± 34.8
LSI	0.34 (0.02, 0.59)	11.4 (9.3, 14.8)	13.2 (10.6, 17.3)	26.7	-0.9 ± 31.7	-1.1 ± 34.3
Maximum Height						
Left	0.87 (0.76, 0.93)	0.8 (0.6, 1.0)	6.3 (5.1, 8.2)	1.8	0.1 ± 2.1	1.2 ± 16.4
Right	0.83 (0.69, 0.91)	0.8 (0.6, 1.0)	7.2 (5.9, 9.5)	1.8	0.2 ± 2.1	1.4 ± 17.9
Pooled	0.85 (0.77, 0.90)	0.8 (0.7, 0.9)	6.6 (5.7, 7.9)	1.8	0.2 ± 2.1	1.3 ± 17.1
LSI	0.17 (-0.15, 0.46)	7.9 (6.4, 10.2)	9.0 (7.3, 11.8)	18.3	-0.4 ± 21.8	-0.2 ± 22.1

Abbreviations: CV, percent coefficient of variation; ICC, intraclass correlation coefficient; LoA, limits of agreement; LSI, limb symmetry index; MDC$_{90}$, minimal detectable change at the 90% confidence level; SEM, standard error of measurement; 95% CI, 95% confidence interval. *Indicates presence of heteroscedasticity in data (Pearson's r, $P < 0.05$).

Figure Captions

Figure 1: Participant performing the heel-rise endurance test on a 10° incline box, raising the heel as high as possible on each repetition, and with the linear encoder measurement device attached to the heel.

Figure 2: Limits of agreement plots for (**A**) repetitions, (**B**) repetitions percent differences, (**C**) work, (**D**) work percent differences, (**E**) maximum height, and (**F**) maximum height percent differences. Plots illustrate individual difference in test-retest performance plotted against the mean of the individual's two performances (**A, C, E**) or the difference expressed as a percentage of the mean of the individual's two performances (**B, D, F**). Data represent the pooled data (i.e. left and right legs, n = 76) of the 38 participants with the solid line representing the mean difference and the dashed lines representing the upper and lower 95% limits of agreement.

Figure 1

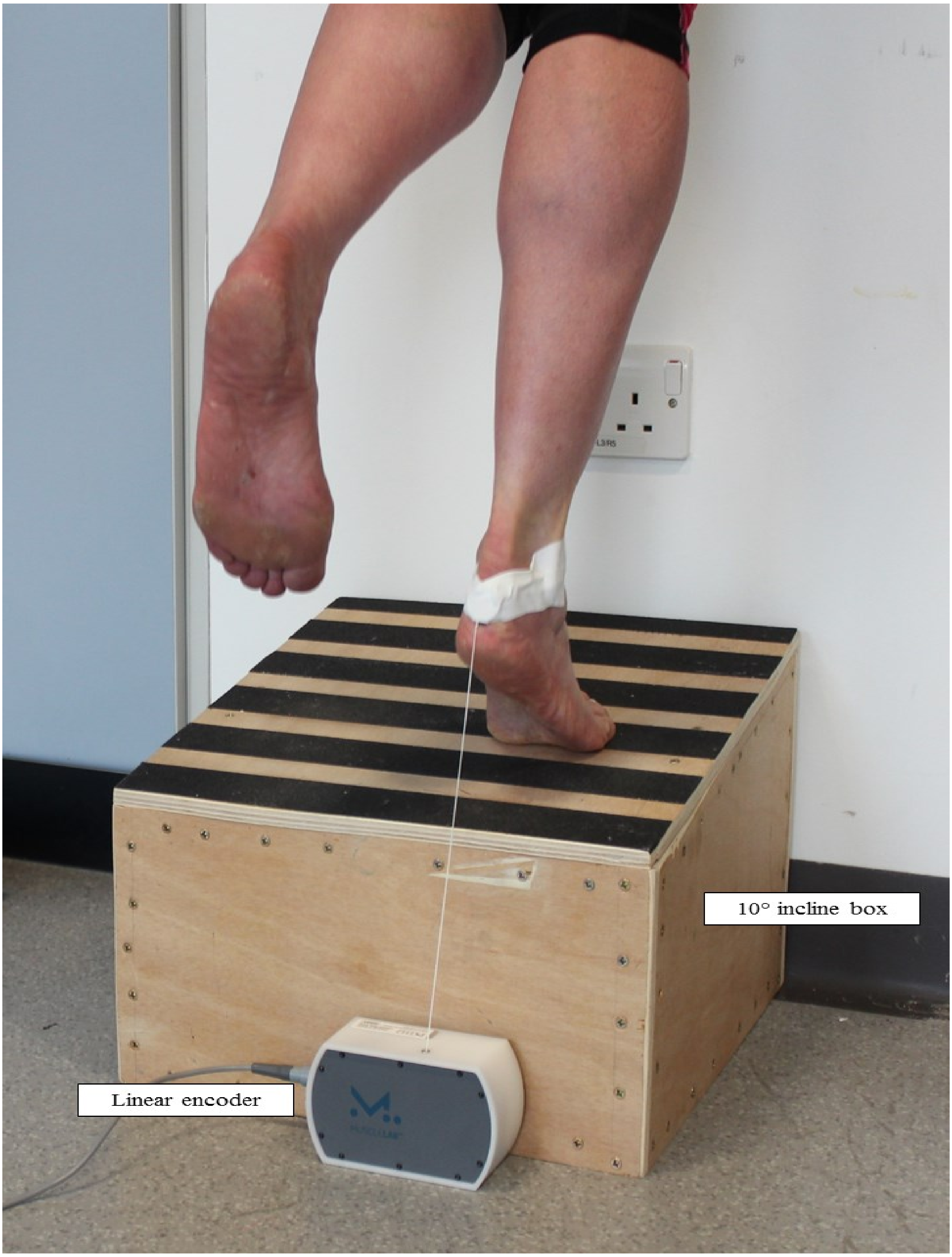
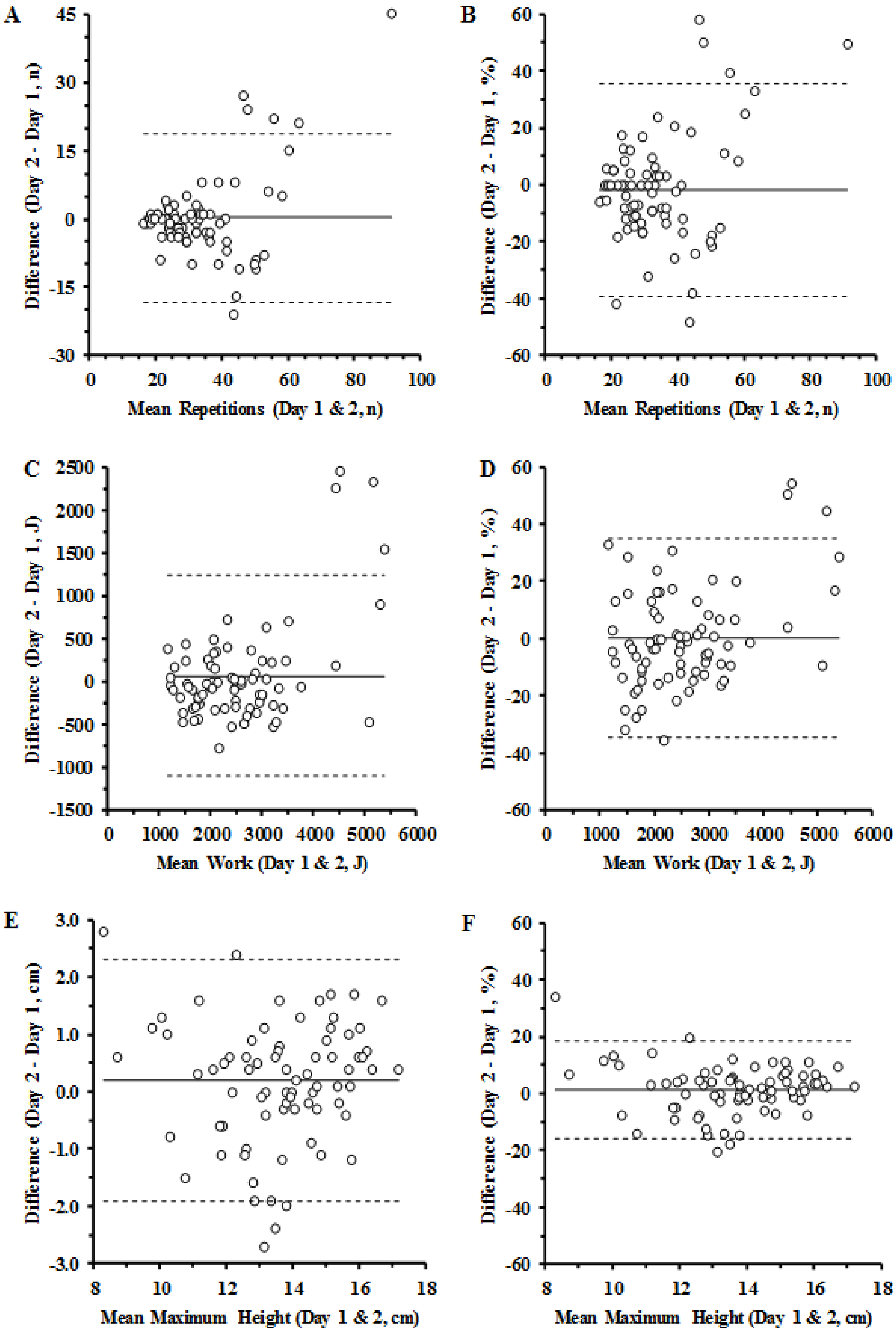


Figure 2


A
45
30
15
0
-15
-30
0
20
40
60
80
100
Difference (Day 2 - Day 1, n)
Mean Repetitions (Day 1 & 2, n)
B
60
40
20
0
-20
-40
-60
0
20
40
60
80
100
Difference (Day 2 - Day 1, %)
Mean Repetitions (Day 1 & 2, n)
C
2500
2000
1500
1000
500
0
-500
-1000
-1500
0
1000
2000
3000
4000
5000
6000
Difference (Day 2 - Day 1, J)
Mean Work (Day 1 & 2, J)
D
60
40
20
0
-20
-40
-60
0
1000
2000
3000
4000
5000
6000
Difference (Day 2 - Day 1, %)
Mean Work (Day 1 & 2, J)
E
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
8
10
12
14
16
18
Difference (Day 2 - Day 1, cm)
Mean Maximum Height (Day 1 & 2, cm)
F
60
40
20
0
-20
-40
-60
8
10
12
14
16
18
Difference (Day 2 - Day 1, %)
Mean Maximum Height (Day 1 & 2, cm)

References

BINKLEY, J. M., STRATFORD, P. W., LOTT, S. A. & RIDDLE, D. L. 1999. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. . *Phys Ther,* 79**,** 371-83.

BLAND, J. M. & ALTMAN, D. G. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet,* 1**,** 307-10.

BOSTICK, G. P., JOMHA, N. M., SUCHAK, A. A. & BEAUPRE, L. A. 2010. Factors associated with calf muscle endurance recovery 1 year after achilles tendon rupture repair. *J Orthop Sports Phys Ther,* 40**,** 345-51.

BUCHGRABER, A. & PASSLER, H. H. 1997. Percutaneous repair of Achilles tendon rupture. Immobilization versus functional postoperative treatment. *Clin Orthop Relat Res***,** 113-22.

BYRNE, C., TWIST, C. & ESTON, R. 2004. Neuromuscular function after exercise-induced muscle damage: theoretical and applied implications. *Sports Med,* 34**,** 49-69.

DE VET, H. C., TERWEE, C. B., KNOL, D. L. & BOUTER, L. M. 2006. When to use agreement versus reliability measures. *J Clin Epidemiol,* 59**,** 1033-9.

HABER, M., GOLAN, E., AZOULAY, L., KAHN, S. R. & SHRIER, I. 2004. Reliability of a device measuring triceps surae muscle fatigability. *Br J Sports Med,* 38**,** 163-7.

HEBERT-LOSIER, K., NEWSHAM-WEST, R. J., SCHNEIDERS, A. G. & SULLIVAN, S. J. 2009a. Raising the standards of the calf-raise test: a systematic review. *J Sci Med Sport,* 12**,** 594-602.

HEBERT-LOSIER, K., SCHNEIDERS, A. G., NEWSHAM-WEST, R. J. & SULLIVAN, S. J. 2009b. Scientific bases and clinical utilisation of the calf-raise test. *Phys Ther Sport,* 10**,** 142-9.

HOPKINS, W. G. 2000. Measures of reliability in sports medicine and science. *Sports Med,* 30**,** 1-15.

KOTTNER, J., AUDIGE, L., BRORSON, S., DONNER, A., GAJEWSKI, B. J., HROBJARTSSON, A., ROBERTS, C., SHOUKRI, M. & STREINER, D. L. 2011. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol,* 64**,** 96-106.

MOLLER, M., LIND, K., MOVIN, T. & KARLSSON, J. 2002. Calf muscle function after Achilles tendon rupture. A prospective, randomised study comparing surgical and non-surgical treatment. *Scand J Med Sci Sports,* 12**,** 9-16.

MOLLER, M., LIND, K., STYF, J. & KARLSSON, J. 2005. The reliability of isokinetic testing of the ankle joint and a heel-raise test for endurance. *Knee Surg Sports Traumatol Arthrosc,* 13**,** 60-71.

MULLANEY, M. J., MCHUGH, M. P., TYLER, T. F., NICHOLAS, S. J. & LEE, S. J. 2006. Weakness in end-range plantar flexion after Achilles tendon repair. *Am J Sports Med,* 34**,** 1120-5.

NILSSON-HELANDER, K., SILBERNAGEL, K. G., THOMEE, R., FAXEN, E., OLSSON, N., ERIKSSON, B. I. & KARLSSON, J. 2010. Acute achilles tendon rupture: a randomized, controlled study comparing surgical and nonsurgical treatments using validated outcome measures. *Am J Sports Med,* 38**,** 2186-93.

NILSSON-HELANDER, K., THOMEE, R., SILBERNAGEL, K. G., THOMEE, P., FAXEN, E., ERIKSSON, B. I. & KARLSSON, J. 2007. The Achilles tendon Total Rupture Score (ATRS): development and validation. *Am J Sports Med,* 35**,** 421-6.

OLSSON, N., NILSSON-HELANDER, K., KARLSSON, J., ERIKSSON, B. I., THOMEE, R., FAXEN, E. & SILBERNAGEL, K. G. 2011. Major functional deficits persist 2 years after acute Achilles tendon rupture. *Knee Surg Sports Traumatol Arthrosc,* 19**,** 1385-93.

SCHEPULL, T., KVIST, J., ANDERSSON, C. & ASPENBERG, P. 2007. Mechanical properties during healing of Achilles tendon ruptures to predict final outcome: a pilot Roentgen stereophotogrammetric analysis in 10 patients. *BMC Musculoskelet Disord,* 8**,** 116.

SILBERNAGEL, K. G., GUSTAVSSON, A., THOMEE, R. & KARLSSON, J. 2006. Evaluation of lower leg function in patients with Achilles tendinopathy. *Knee Surg Sports Traumatol Arthrosc,* 14**,** 1207-17.

SILBERNAGEL, K. G., NILSSON-HELANDER, K., THOMEE, R., ERIKSSON, B. I. & KARLSSON, J. 2010. A new measurement of heel-rise endurance with the ability to detect functional deficits in patients with Achilles tendon rupture. *Knee Surg Sports Traumatol Arthrosc,* 18**,** 258-64.

SILBERNAGEL, K. G., STEELE, R. & MANAL, K. 2012. Deficits in heel-rise height and achilles tendon elongation occur in patients recovering from an Achilles tendon rupture. *Am J Sports Med,* 40**,** 1564-71.

SMAN, A. D., HILLER, C. E., IMER, A., OCSING, A., BURNS, J. & REFSHAUGE, K. M. 2014. Design and reliability of a novel heel rise test measuring device for plantarflexion endurance. *Biomed Res Int,* 2014**,** 391646.

SVANTESSON, U., CARLSSON, U., TAKAHASHI, H., THOMEE, R. & GRIMBY, G. 1998. Comparison of muscle and tendon stiffness, jumping ability, muscle strength and fatigue in the plantar flexors. *Scand J Med Sci Sports,* 8**,** 252-6.

WEBER, M., NIEMANN, M., LANZ, R. & MULLER, T. 2003. Nonoperative treatment of acute rupture of the achilles tendon: results of a new protocol and comparison with operative treatment. *Am J Sports Med,* 31**,** 685-91.

WEIR, J. P. 2005. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res,* 19**,** 231-40.