J Rehab Med 2001; 33: 90–96

ANKLE DORSIFLEXOR MUSCLE PERFORMANCE IN HEALTHY YOUNG MEN AND WOMEN: RELIABILITY OF ECCENTRIC PEAK TORQUE AND WORK MEASUREMENTS

Anna Maria Holmbäck,[1] Michelle M. Porter,[2] David Downham[3] and Jan Lexell[4]

From the [1]Department of Physical Therapy, Lund University, Sweden, [2]Faculty of Physical Education and Recreation Studies, University of Manitoba, Canada, [3]Department of Mathematical Sciences, University of Liverpool, UK and [4]Department of Rehabilitation, Lund University Hospital, Sweden

The aims of this study were: (i) to assess the test–retest intrarater reliability of eccentric ankle dorsiflexor muscle performance in young healthy men and women using the Biodex dynamometer; and (ii) to examine different statistical indices for the interpretation of reliability. Thirty men and women (age 22.5 ± 2.5 years, mean ± S.D.) performed three maximal eccentric contractions at 30°/second and 90°/second, with 7–10 days between test sessions. Reliability was evaluated with three intraclass correlation coefficients ($ICC_{1,1}$, $ICC_{2,1}$ and $ICC_{3,1}$), and was excellent for peak torque (ICC 0.90–0.96) and good to excellent for work (ICC 0.69–0.83), with no discernible differences among the three ICCs. Method errors, assessed by the standard error of the measurement (S.E.M.) and S.E.M.%, were low. The Bland & Altman graphs and analyses indicated no significant systematic bias in the data. In conclusion, measurements of eccentric ankle dorsiflexor muscle performance in young healthy individuals using the Biodex are highly reliable.

Key words: ankle joint, biomechanics, movement, muscle contraction, muscles, physical medicine, reference values, reproducibility of results, research design, statistics.

J Rehab Med 2001; 33: 90–96

Correspondence address: Anna Maria Holmbäck, PT, MSc, Department of Physical Therapy, Lund University, P.O. Box 5134, SE-220 05 Lund, Sweden

(Accepted July 17, 2000)

INTRODUCTION

In everyday physical activities, such as walking, running and maintaining balance, both concentric (shortening) and eccentric (lengthening) ankle dorsiflexor muscle contractions are performed. Reduced ankle dorsiflexor strength, following a disorder in the peripheral nervous system, a musculoskeletal injury or increasing age, leads to decreased mobility and can increase the risk of falls (1). To evaluate the effectiveness of rehabilitation interventions to improve ankle dorsiflexor muscle strength, appropriate assessment of the function of the muscles contributing to dorsiflexion performance is essential.

One commonly used method to assess limb muscle function in clinical rehabilitation is isokinetic dynamometry. Isokinetic dynamometers enable the measurement of muscular moment under concentric as well as eccentric actions, and at different angular velocities (2). The usefulness of any device, though, depends on its reliability, the test protocol and the measurements obtained (3). Reliability, often defined by the degree to which test scores are free from errors of measurement (4, 5), can be determined by several complementary statistical methods (6, 7). More than 10 studies have evaluated the test–retest reliability of concentric ankle dorsiflexion strength and shown that it is high or excellent (for ref. see 6). However, only one study (8) has evaluated the test–retest reliability of eccentric ankle dorsiflexor strength, despite this contraction mode being as important as concentric contractions in everyday functional activities. In their study, Porter et al. (8) used a Kin–Com dynamometer and assessed the reliability of eccentric dorsiflexion peak torque measurements at 30°/second in a standing position (instead of the standard seated position), in a group of young and old men and women. Reliability, as assessed by the intraclass correlation coefficient ($ICC_{2,1}$), was fair to good (0.55); when two outliers in the sample of 12 subjects were removed, the ICC increased to 0.86. Because of the increased interest in eccentric ankle dorsiflexor strength measurement, there is a need to establish the reliability of these measurements in the standard seated position.

The most commonly used measurement of muscle strength, both in research and in clinical rehabilitation, is peak torque. In various situations, work (calculated as the area under the curve or as peak torque multiplied by the angular displacement) (9, 10) could be more representative of functional muscle performance than peak torque, as work accommodates the force output throughout the range of movement. Reliability of concentric ankle dorsiflexion work has also been found to be excellent (6), but no data on the reliability of eccentric work measurements have been presented. To establish fully the reliability of ankle dorsiflexor eccentric measurements, peak torque and work need to be measured at different angular velocities.

Test–retest reliability (or intrarater reliability) can be assessed by intraclass correlations coefficients (ICCs) (11), Bland & Altman graphs to indicate any systematic variations in the data (12–14), and indices of measurement errors (15). Several ICCs are available, and the rationale for the selection of a specific equation or formula should be given. Whether different ICCs lead to materially different results is not fully established, and so different ICCs need to be calculated and compared.

 ISSN 1650–1977

All measurements have an inherent variability that must be accommodated before reliable clinical decisions can be made. A central issue in muscle strength measurements in clinical rehabilitation is: if two measurements are taken, before and after a rehabilitation intervention, does the difference between the measurements indicate a "true" difference? Several methods can be used to address this issue (7, 14, 15), but few, if any, have been fully exploited for muscle performance measurements.

The overall aim of this study was to assess the test–retest intrarater reliability of eccentric ankle dorsiflexor muscle performance in young healthy men and women, using a Biodex dynamometer. The reliability of eccentric peak torque and work measurements at two angular velocities was evaluated. In addition, three different ICCs were compared, and the use of different measurement error statistics for making inferences regarding outcome measurements in clinical rehabilitation is described.

SUBJECTS AND METHODS

Subjects

Fifteen men (age 23.8 ± 3.1 years, mean ± S.D., height 180.0 ± 6.1 cm, weight 76.5 ± 7.7 kg) and 15 women (age 22.1 ± 2.0 years, height 170.8 ± 4.4 cm, weight 60.7 ± 5.4 kg) volunteered for the study. They were all students in the Department of Physical Therapy at Lund University, Sweden. None of the subjects showed any signs or symptoms of disease, and none reported any neuromusculoskeletal dysfunction in the tested leg within the past year. All subjects were engaged one to three times per week in recreational sports. Most of the subjects (80%) were familiar with isokinetic testing of the ankle and had been tested before on the Biodex equipment. Before the start of the study, each participant was informed of the testing procedures, and thereafter gave written consent. The project was approved by the Ethics Research Committee of Lund University.

Procedures

Measurements were performed on a Biodex® Multi-Joint System II isokinetic dynamometer (Biodex Medical Systems, Shirley, New York, U.S.A.) with the Biodex Advantage software version 4.0. The standard Biodex ankle unit attachment was used. Before testing each subject, the system was calibrated to be within allowable limits set by the Biodex. To perform a calibration verification, the Biodex controller is set to SETUP mode and sensitivity dial "C" is chosen. The knee attachment, without the knee pad, is secured to the powerhead shaft and the range of motion (ROM) limits are chosen so that the attachment can move slightly above horizontal and slightly past vertical. Using a small level supplied with the system as a guide, the attachment is positioned and secured vertical to the floor, and locked in this position. This position is compiled as reference data. After a free-up of the powerhead shaft, the attachment is positioned, secured and locked horizontal to the floor. The calibration weight is inserted into the attachment and locked to be within a known arm length. The system then compiles this position and the known torque reference data (i.e. a known torque of 36.6 ± 3 Nm).

All tests were conducted by the same person (AMH). Only the dominant leg was tested. Leg dominance was determined by asking which leg was habitually used for one leg hopping and/or kicking a ball. For all 30 subjects, testing was done on the right side. Each subject underwent two identical test sessions scheduled approximately at the same time of the day, with 7–10 days between test sessions. Subjects who had not been tested previously were familiarized with the Biodex equipment and performed sub-maximal and maximal eccentric contractions 1–2 days before the first test session. Both test sessions started with 5 minutes of stationary cycling at a load of 1 W/kg body weight. After the warm-up, each subject was seated in the Biodex chair, with the angles of the hip and the knee joints at 80° flexion (0° neutral position) and 30° flexion (0° straight leg), respectively. Details of the positioning have been presented previously (6). To account for the influence of the gravity effect torque on the test data, each subject's limb was weighed and the data were corrected using the Biodex software.

Start and end-range settings were standardized for all subjects from 10° dorsiflexion (DF) (starting position) to 20° plantar flexion (PF). In the selected eccentric mode the torque limit adjustment dials were used to specify the range of the desired force output. To initiate the shaft motion, subjects were required to exceed a minimum torque threshold corresponding to 10% of the torque limit dial setting. The torque limits were set to 27 or 54 Nm for women and 54 or 81 Nm for men. Warm-up contractions were performed such that all subjects started with three sub-maximal eccentric contractions and one maximal eccentric contraction at 30°/second. After a 1 minute rest, three further sub-maximal eccentric contractions and one maximal eccentric contraction at 90°/second were performed. The torque limits were adjusted after the first sub-maximal contraction at each angular velocity so that the limits were acceptable for the test.

Five minutes after the warm-up contractions, three non-consecutive maximal eccentric contractions were performed, first at 30°/second and then at 90°/second. Each contraction started from a dorsiflexed position. A 30 second rest was allowed between each maximal contraction, and a 2 minute rest between each angular velocity. Each subject was instructed to exert maximal voluntary effort by resisting the movement of the footplate, but they were not verbally encouraged during the contractions. Subjects could not be provided with visual feedback from the monitor, as the eccentric torque curve is not presented until the contraction is completed.

Data

From each of the three contractions at each of the two angular velocities, two measurements were collected: peak torque (Nm) and work (J). Work calculations are numerical integrations of the torque vs angular position curve. Thin rectangles are drawn over the curve and the area of these rectangles is calculated. The current torque value is used for the rectangles' height. One half of the difference between the current and previous torque value is subtracted from the height as an error correction. The Biodex report sheet gives the highest value of the three peak torque (Nm) recordings together with work (J) for this contraction; 47% of the peak torque values were obtained from the first contraction, 28% from the second and 25% from the third. As 82% of the highest work was obtained from the contraction that yielded the highest peak torque, the highest peak torque value and the work from that contraction were used.

Reliability statistics

Reliability of the peak torque and work measurements was assessed by the intraclass correlation coefficients, $ICC_{1,1}$, $ICC_{2,1}$ and $ICC_{3,1}$ (11). If BMS represents the variability in the measurements between subjects, WMS the variability within subjects, JMS the variability between test sessions, EMS the residual mean square and n the number of subjects, then for two test sessions,

(1) $ICC_{1,1} = (BMS - WMS)/(BMS + WMS)$
(2) $ICC_{2,1} = (BMS - EMS)/(BMS + EMS + 2(JMS - EMS)/n)$
(3) $ICC_{3,1} = (BMS - EMS)/(BMS + EMS)$.

For $ICC_{1,1}$ a one-way ANOVA is used, whereas for $ICC_{2,1}$ and $ICC_{3,1}$ a two-way ANOVA is used.

Several statistical measures for analysing method errors were considered: the standard error of the measurement (S.E.M.) and S.E.M.% (16), and the Bland & Altman analyses (12, 13, 14). The S.E.M. is defined by:

(4) $\text{S.E.M.} = \text{S.D.}\ (1 - ICC_{k,1})^{0.5}$

where S.D. is the standard deviation of all the measurements from the two test sessions. The S.E.M.% is defined by:

(5) $\text{S.E.M.\%} = (\text{S.E.M./mean}) \times 100$

where mean is the mean for all the observations from test sessions 1 and 2. The Bland & Altman analyses include the following calculations:

(6) $\bar{d}$ = the mean difference between the two test sessions

Table I. *Peak torque and work at two angular velocities for 15 men and 15 women from the two test sessions*

	Test session	30°/second		90°/second	
		Men	Women	Men	Women
Peak torque (Nm)	1	52.1 ± 9.5	35.5 ± 4.4	53.6 ± 9.2	36.9 ± 4.8
	2	50.7 ± 10.1	35.0 ± 5.0	53.0 ± 9.9	36.1 ± 4.9
Work (J)	1	24.4 ± 6.4	16.1 ± 2.0	24.6 ± 4.4	16.8 ± 2.0
	2	24.5 ± 7.1	16.0 ± 2.5	24.5 ± 5.5	16.4 ± 2.2

Data are means ± S.D.
The highest peak torque and work from each of the three contractions at each angular velocity in test sessions 1 and 2 were obtained from the Biodex report sheet and used to calculate the means and standard deviations for men and women. Work was obtained from the contraction that yielded the highest peak torque.

(7) $S.D._{diff}$ = the standard deviation of the differences between the two test sessions
(8) standard error (S.E.) of $\bar{d} = S.D._{diff}/\sqrt{n}$
(9) 95% confidence intervals (95% CI) $= \bar{d} \pm 2.145 \times S.E.$
(10) 95% limits of agreement (95% LOA) $= \bar{d} \pm 2 \times S.D._{diff}$

In this study $S.D._{diff}$ is used to calculate S.E. and 95% LOA, and S.E. is used to calculate 95% CI for the mean of the differences. The value 2.145 in eqn [9] is obtained from the *t*-table, with 14 ($n - 1$) degrees of freedom (df).

The Bland & Altman analyses also include the formation of graphs, with the difference between test session 1 and test session 2 (1 minus 2) plotted against the mean of the two test sessions for each individual. From these graphs the sizes and ranges of the differences, and their distribution about zero can be discerned.

Throughout the study, significance levels greater than 5% were considered as not significant and 95% CI and 95% LOA were considered. All calculations were performed using the SPSS 9.0 Software for Windows (SPSS, Chicago, IL, U.S.A.).

RESULTS

The means and standard deviations for peak torque and work are presented in Table I. For both angular velocities, peak torque and work were significantly ($p < 0.001$) higher for men than women, whereas no significant difference was found between the two angular velocities for men and women, separately.

The relationships between peak torque and work are shown in Table II. For both angular velocities in both test sessions, there was a highly significant ($p < 0.001$) positive relationship between the measurements.

The values of $ICC_{1,1}$, $ICC_{2,1}$ and $ICC_{3,1}$ for peak torque and work are shown in Table III. For peak torque, there were very small differences between the values of $ICC_{1,1}$, $ICC_{2,1}$ and $ICC_{3,1}$ for both men and women and the two angular velocities, and all ICC values were above 0.9. For work, the values of $ICC_{1,1}$, $ICC_{2,1}$ and $ICC_{3,1}$ were also almost identical, but the actual values of ICC were generally lower than for peak torque. $ICC_{1,1}$, $ICC_{2,1}$ and $ICC_{3,1}$ were consistently higher for men than women at both angular velocities for both peak torque and work.

In Fig. 1A–D, the differences between test session 1 and test session 2 (1 minus 2) for peak torque and work are plotted against the means of the two test sessions for each individual for the two angular velocities. Men had, in general, greater variability for both the mean and the difference values than women. For peak torque at 30°/second for men, the differences between test sessions were asymmetric around the zero line, with only four negative values (Fig. 1A). For work at 30°/second for men, the differences between test sessions were greater for large mean values (Fig. 1C).

The results of the Bland & Altman analyses together with S.E.M. (derived from $ICC_{2,1}$) and S.E.M.% for peak torque and work are presented in Table IV. For peak torque, all $\bar{d}$ values were positive, indicating that the first measurements tended to be larger than the second. For work, the $\bar{d}$ value at 30°/second was negative for men, while all other $\bar{d}$ values were positive. The 95% CI for each $\bar{d}$ value included zero, indicating no significant difference between the two test sessions. For the measurements of peak torque at 30°/second for men the lower limit of the CI was close to zero (−0.18). For both peak torque and work, the width of the 95% CI was wider for men than for women. The 95% LOA was also wider for men than for women, for both peak torque and work at both angular velocities. The S.E.M. values for both peak torque and work were consistently higher for men than for women. The S.E.M.% was consistently higher for work than for peak torque, for both men and women. For peak torque, there were no discernible differences in S.E.M.% between men and women. For work at 30°/second, the S.E.M.% was considerably larger for men than for women, whereas at 90°/second there was a much smaller difference.

To illustrate the use of 95% CI and 95% LOA, the values from Table IV for peak torque recordings at 30°/second are presented for men and women in Fig. 2A and B. The tendency towards a systematic difference between the two test sessions (test 1 being larger than test 2) is shown, with the 95% CI as well as the 95% LOA being asymmetric around the zero line, more so for men than for women.

DISCUSSION

Disorders of the peripheral nervous system, musculoskeletal injuries or increasing age lead to reduced strength of muscles around the ankle and can cause limitations of activities. Isokinetic dynamometry can be used to assess muscle function in the ankle and to evaluate the outcome of a rehabilitation intervention. The usefulness of any assessment tool, though, depends on its reliability. Previously (6), it was shown that concentric ankle dorsiflexor isokinetic muscle performance can

Table II. *Pearson's correlation coefficients between peak torque and work at two angular velocities for 15 men and 15 women from the two test sessions*

	Peak torque vs work at 30°/second		Peak torque vs work at 90°/second	
Test session	Men	Women	Men	Women
1	0.92	0.89	0.94	0.92
2	0.91	0.91	0.92	0.90

The highest peak torque and work from each of the three contractions at each angular velocity in test sessions 1 and 2 were obtained from the Biodex report sheet and were used in the analysis. Work was obtained from the contraction that yielded the highest peak torque.

be reliably examined in young men and women. In the present study, the reliability of eccentric ankle dorsiflexor peak torque and work measurements is shown to be good to excellent. This confirms the reproducibility of isokinetic dynamometry measurements of ankle dorsiflexor muscle performance.

Reliability of eccentric peak torque and work measurements

Peak torque and work were higher for men than women for both angular velocities, but no significant difference was found between the two angular velocities for men and women, separately. Work values were approximately 50% of peak torque values. Similar results for peak torque have been presented previously (8, 17), whereas no comparable data on eccentric ankle dorsiflexor work have been found in the literature.

Table III. *Intraclass correlation coefficients ($ICC_{1,1}$, $ICC_{2,1}$ and $ICC_{3,1}$) for peak torque and work at two angular velocities for 15 men and 15 women*

		30°/second		90°/second	
		Men	Women	Men	Women
Peak torque (Nm)	$ICC_{1,1}$	0.95	0.92	0.96	0.90
	$ICC_{2,1}$	0.95	0.92	0.96	0.90
	$ICC_{3,1}$	0.96	0.92	0.96	0.91
Work (J)	$ICC_{1,1}$	0.79	0.70	0.83	0.79
	$ICC_{2,1}$	0.79	0.70	0.83	0.79
	$ICC_{3,1}$	0.78	0.69	0.82	0.80

The highest peak torque and work from each of the three contractions at each angular velocity in test sessions 1 and 2 were obtained from the Biodex report sheet. Work was obtained from the contraction that yielded the highest peak torque.

Reliability for eccentric ankle dorsiflexion peak torque at both

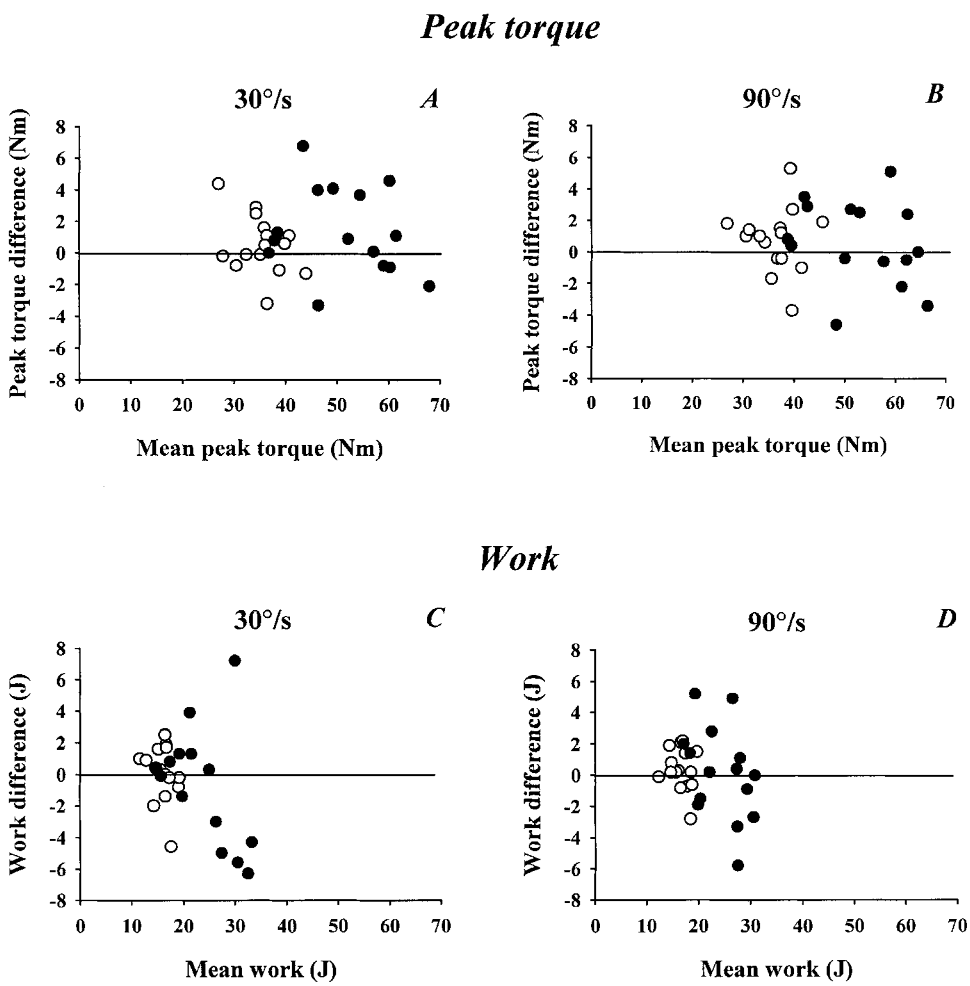

Fig. 1. (A–D) Differences between test session 1 and test session 2 (test 1 minus test 2) for the highest peak torque (Nm) and work (J) plotted against the means for two angular velocities (30°/second and 90°/second) for 15 men (●) and 15 women (○).

Table IV. *Bland & Altman analyses and the standard error of the measurement (S.E.M. and S.E.M.%) for peak torque and work at two angular velocities for 15 men and 15 women from the two test sessions*

		30°/second		90°/second	
		Men	Women	Men	Women
Peak torque (Nm)	$\bar{d}$	1.35	0.53	0.57	0.75
	95% CI	−0.18 to 2.88	−0.52 to 1.58	−0.91 to 2.05	−0.39 to 1.89
	95% LOA	−4.14 to 6.89	−3.23 to 4.29	−4.77 to 5.91	−3.39 to 4.89
	S.E.M.	2.17	1.30	1.90	1.49
	S.E.M.%	4.2	3.7	3.6	4.1
Work (J)	$\bar{d}$	−0.1	0.07	0.13	0.39
	95% CI	−2.57 to 2.37	−0.92 to 1.06	−1.52 to 1.78	−0.34 to 1.12
	95% LOA	−9.04 to 8.84	−3.51 to 3.65	−5.83 to 6.09	−2.27 to 3.05
	S.E.M.	2.93	1.14	1.94	0.91
	S.E.M.%	12.0	7.1	7.9	5.5

The highest peak torque and work from each of the three contractions at each angular velocity in test sessions 1 and 2 were obtained from the Biodex report sheet. Work was obtained from the contraction that yielded the highest peak torque.
$\bar{d}$ = mean difference between the two test sessions; 95% CI = 95% confidence interval for $\bar{d}$; 95% LOA = 95% limits of agreement; S.E.M. = standard error of the measurement based on $ICC_{2,1}$; S.E.M. (%) = 100 × S.E.M. divided by the mean for all observations from the two test sessions.

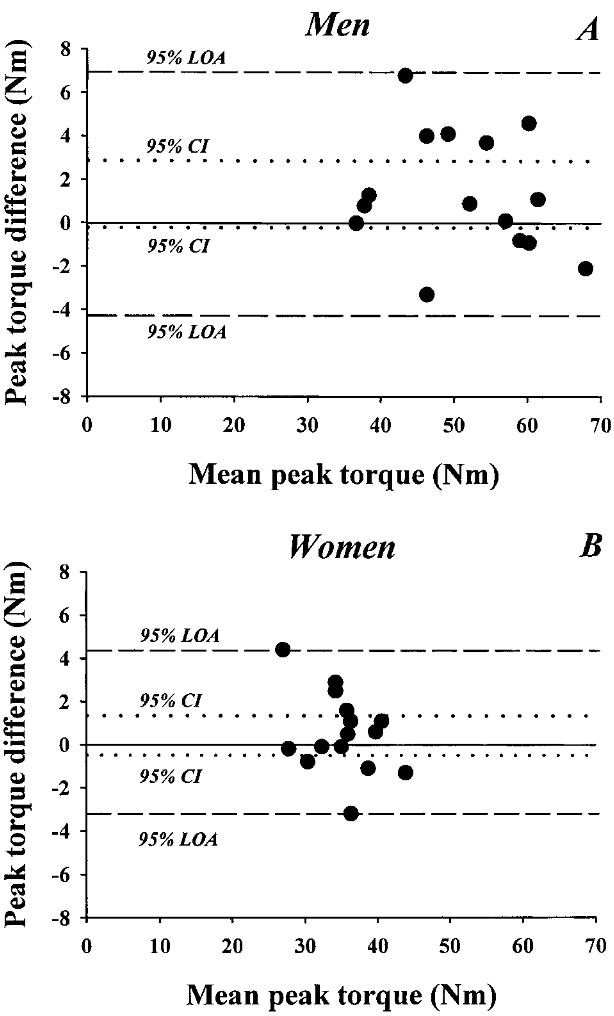

Fig. 2. (A, B) Differences between test session 1 and test session 2 (test 1 minus test 2) for the highest peak torque (Nm) at 30°/second plotted against the means for 15 men (A) and 15 women (B), together with the 95% confidence intervals (CI; ……) and the 95% upper and lower limits of agreements (LOA; ------).

angular velocities was excellent according to the recommendations by Fleiss (18) (ICC 0.90–0.96). In comparison, measurements in the present study were more reliable than those in the only previous study (8) of isokinetic eccentric ankle dorsiflexor muscles, where reliability was fair to good ($ICC_{2,1}$ 0.55). The main difference between the two studies is in the testing position. In the present study, measurements were taken in the standard seated position, while subjects in the other study (8) were tested in a standing position. In addition, the present study used a Biodex dynamometer, while the earlier study used a Kin–Com dynamometer, but it is generally believed that dynamometers do not differ in their mechanical reliability (9, 10).

Reliability for work was good to excellent (ICCs 0.69–0.83). Both the 95% CI and 95% LOA were similar for peak torque and work, but as the values for peak torque were twice as large as the work values, measurements of work were in reality less reliable than those for peak torque, as reflected by the larger S.E.M.% values for work. The lower reliability for work was also illustrated in Fig. 1C, where the size of the mean values for men seemed to affect the differences between test sessions.

In a previous study of concentric ankle dorsiflexor performance (6), reliability for peak torque was also better than for work, although reliability was excellent for both peak torque and work. One reason for the lower reliability of eccentric work measurements could be the difficulty in standardizing the end-range setting of the Biodex equipment. As work is calculated as the area under the curve, it depends on the consistency of the setting of the range of motion. Even though great care was taken in the standardization of the setting, it was difficult to reproduce fully the setting from test session 1 to test session 2, thus influencing work recordings, but not peak torque. In addition, work values were obtained from the contraction that yielded the highest peak torque: 82% of the highest work values were obtained from the contraction that yielded the highest peak torque. Because of lower reliability, and as the relationship between eccentric peak torque and work was strong at both angular velocities (Pearson's r 0.89–0.94), the usefulness of measuring work in young healthy individuals can be questioned.

The ICC values for men were consistently higher than those for women, for both peak torque and work. However, for work the S.E.M.% values were lower for women than for men. For peak torque, the S.E.M.% values were similar for men and women. These results reflect the wider ranges of peak torque and work measurements for men which influence the between-subjects mean square (BMS) values, and therefore give higher ICC values. This emphasizes, yet again, the need for several statistical measures to evaluate reliability.

Visual analysis of the graphs together with the interpretation of the Bland & Altman analyses showed that peak torque measurements from test session 1 tended to be greater than those from test session 2, in particular at 30°/second for men. This tendency was not significant, however, as zero was always included in the 95% CI for $\bar{d}$. Nevertheless, this tendency should be considered in studies of eccentric ankle dorsiflexion muscle performance when repeated measurements are taken.

There were no discernible differences in ICC for the two angular velocities (30°/second and 90°/second). In a previous study of concentric ankle dorsiflexor strength (6), reliability was similar for angular velocities of 30, 60 and 90°/second, but reduced at 120 and 150°/second, for both peak torque and work. Consequently, eccentric strength was measured only at lower velocities. If eccentric strength needs to be measured at higher velocities, reliability has to be established.

Reliability statistics

A common statistical method of assessing reliability is the ICC (11). Different ICC equations have been used widely during the 1990s, but there are no specific recommendations on how to choose the appropriate ICC (7).

The main difference between the three commonly used reliability coefficients ($ICC_{1,1}$, $ICC_{2,1}$ and $ICC_{3,1}$) is how error variance (lack of reliability) is defined. For the same set of data, $ICC_{1,1}$ will usually give lower values than $ICC_{2,1}$, which in most cases will result in lower values than $ICC_{3,1}$ (11), but this was not seen here. Values for the three ICCs were similar for both men and women and the two angular velocities. The marginal differences between calculations of $ICC_{1,1}$ vs $ICC_{2,1}$ and $ICC_{2,1}$ vs $ICC_{3,1}$ have also been found previously (19, 20). In a test–retest reliability study where the differences between mean squares JMS and EMS are reasonably small and the systematic bias is low, the three ICC values will be very similar. Under these circumstances the ICCs are considered interchangeably and the choice between different coefficients can be seen as mostly philosophical (21).

Reliability methods based on correlation coefficients, such as ICC, provide an indication of "relative reliability" (5, 21). However, relative reliability measures are influenced by the range of measured values, give no indication of actual measurement values or any systematic variability in the measures, and cannot be interpreted clinically (5), which was clearly seen in the present study. Reliability studies should therefore always include assessments of measurement errors and analyses of systematic bias, commonly referred to as "absolute reliability" (5, 21).

One common way to assess absolute reliability, and to visualize any systematic variability and possible outliers, is to form the Bland & Altman graphs. From these graphs, random and systematic errors can be detected by examining the direction and magnitude of the scatter around the zero line. By calculating $\bar{d}$ and CI, indices of any systematic bias are obtained. If the value of $\bar{d}$ is positive (or negative), measurements from the first test session tend to be larger (or smaller) than the second measurement. If zero is included within the CI, no significant systematic bias is present in the data. The usefulness of this comprehensive presentation of reliability is clearly seen in the interpretation of the tables and figures in this report, and such presentations should form part of reproducibility assessments in clinical rehabilitation.

In clinical rehabilitation, it is important to detect the magnitude of a "real" change in an individual patient's performance as a result of an intervention. The LOA can be used to assess whether the difference between two measurements from an individual is a true difference. In other words, if the difference before and after an intervention is outside (within) the LOA, it does (does not) represent a real change in performance (15).

Method errors express measurement errors in the same unit as the original measurement. This enables the clinician to interpret reliability in clinically relevant terms: smaller S.E.M. values indicate more reliable measurements (15, 22). A confidence interval for S.E.M. can also be derived and used to estimate sample size (23). S.E.M.% is independent of the units of measurements and could be used as a descriptive tool, for example to compare methods or samples (23).

CONCLUSIONS

This study has determined that eccentric ankle dorsiflexion peak torque in young healthy men and women can be reliably assessed with the Biodex equipment.

Measurements of work are less reliable than those for peak torque and, as for concentric measurements, appear to add no further information than peak torque measurements. A comprehensive assessment of intrarater reliability should be based on specified ICCs, indices of measurement errors and the Bland & Altman graphs.

ACKNOWLEDGMENTS

This study was carried out in the Department of Rehabilitation, Lund University Hospital, Lund, Sweden, and was supported by grants from the Research Council of the Swedish Sports Federation, Gun and Bertil Stohne Foundation, Loo and Hans Osterman Foundation, Magn. Bergvall Foundation, and the Council for Medical Health Care Research in South Sweden. Dr. Michelle Porter was supported by the Medical Research Council of Canada while working in the Department of Rehabilitation, Lund University Hospital.

REFERENCES

1. Wolfson L, Judge J, Whipple R, King M. Strength is a major factor in balance, gait, and the occurrence of falls. J Gerontol 1995; 50A (Special Issue): 64–67.
2. Kellis E, Baltzopoulos V. Isokinetic eccentric exercise. Sports Med 1995; 19: 202–222.
3. Keating JL, Matyas TA. The influence of subject and test design on dynamometric measurements of extremity muscles. Phys Ther 1996; 76: 866–889.
4. Rothstein JM. Measurement and clinical practise: theory and application. In: Rothstein JM, ed. Measurement in physical therapy. New York: Churchill Livingstone; 1985.
5. Domholdt E. Physical therapy research: principles and applications. Philadelphia, PA: WB Saunders; 1993.
6. Holmbäck AM, Porter MM, Downham D, Lexell J. Reliability of isokinetic ankle dorsiflexor strength measurements in healthy young men and women. Scand J Rehab Med 1999; 31: 229–239.
7. Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. Clin Rehab 1998; 12: 187–199.
8. Porter MM, Vandervoort AA, Kramer JF. A method of measuring standing isokinetic plantar and dorsiflexion peak torques. Med Sci Sports Exerc 1996; 28: 516–222.
9. Dvir Z. Isokinetics: muscle testing, interpretation, and clinical applications. New York: Churchill Livingstone; 1995.
10. Perrin DH. Isokinetic exercise and assessment. Champaign, IL: Human Kinetics; 1993.
11. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979; 86: 420–428.
12. Altman DG. Practical statistics for medical research. London: Chapman & Hall; 1991.
13. Bland M. An introduction to medical statistics. Oxford: Oxford University Press; 1995.
14. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1: 307–310.
15. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med 1998; 26: 217–238.
16. Gleeson NP, Mercer TH. The utility of isokinetic dynamometry in the assessment of human muscle function. Sports Med 1996; 21: 18–34.
17. Alexander M. Peak torque values for antagonist muscle groups and concentric and eccentric contraction types for elite sprinters. Arch Phys Med Rehabil 1990; 71: 334–339.
18. Fleiss JL. The design and analysis of clinical experiments. New York: John Wiley & Sons; 1986.
19. Bohannon RW. Commentary. Phys Ther 1989; 69: 190–192.
20. Mosely A, Adams R. Measurement of passive ankle dorsiflexion: procedure and reliability. Aust J Physiother 1991; 3: 175–181.
21. Baumgartner TA. Norm-referenced measurement: reliability. In: Safrit MJ, Wood TM, eds. Measurement concepts in physical education and exercise science. Champaign, IL: Human Kinetics; 1989.
22. Roebroeck ME, Harlaar J, Lankhorst GJ. The application of generalizability theory to reliability assessment: an illustration using isometric force measurements. Phys Ther 1993; 73: 386–400.
23. Stratford PW, Goldsmith CH. Use of the standard error as a reliability index: an applied example using elbow flexor strength data. Phys Ther 1997; 77: 745–750.