Scand J Rehab Med 31: 229–239, 1999

RELIABILITY OF ISOKINETIC ANKLE DORSIFLEXOR STRENGTH MEASUREMENTS IN HEALTHY YOUNG MEN AND WOMEN

Anna Maria Holmbäck, PT, MSc[1], Michelle M. Porter, PhD[2], David Downham, PhD[3] and Jan Lexell, MD, PhD[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, England, [4]Department of Rehabilitation, Lund University Hospital, Sweden

ABSTRACT. The purposes of this study were: (i) to determine the test–retest reliability of isokinetic ankle dorsiflexor strength measurements in young healthy adults using the Biodex dynamometer, and (ii) to examine several statistical measures for the interpretation of reliability. Thirty men and women (mean age 23 ± 3 years) performed three maximal concentric contractions at 30°/s, 60°/s, 90°/s, 120°/s and 150°/s. Reliability of peak torque, work and torque at a specific time were assessed by calculating the intraclass correlation coefficient (ICC $_{2,1}$), Pearson product moment correlation coefficient (r), standard error of the measurement (SEM), method error (ME) and coefficient of variation (CV), and by plotting the differences between observations against their means. Isokinetic tests of ankle dorsiflexor strength in healthy young adults using the Biodex dynamometer were highly reliable (ICC 0.61–0.93). It is recommended that test–retest reliability analyses include the ICC and assessments of measurement errors (SEM, ME or CV), as well as graphs to indicate any systematic variations in the data.

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

INTRODUCTION

Isokinetic dynamometry, using various commercially available equipment, is a frequently applied method for assessing muscle performance, both in research and clinical practice (5). The usefulness of an isokinetic dynamometer depends upon the reproducibility, or reliability, of the equipment, the test protocol and the measurements obtained (12). Reproducibility in this context is defined as the relative absence of measurement errors or consistency of measurement results (15) and is usually evaluated by the intra-rater (one examiner) test–retest reliability.

Many isokinetic reliability studies have focused on knee flexion and extension strength, and have shown the method to be highly reliable (12). In contrast, few studies have assessed the reliability of ankle dorsiflexion strength. The ankle dorsiflexors, in addition to hip and knee muscles, are important for gait and balance tasks (21). Studies of concentric ankle dorsiflexion strength reliability using the KIN-COM 500H (13), Cybex II+ (8), Cybex 6000 (22) and Lido Active Multijoint II (1) dynamometers have all shown high reliability (reliability coefficients above 0.80). Although the Biodex dynamometer is a commonly used device, the only study of concentric ankle dorsiflexion strength (19) has shown low reliability at 30°/s and 120°/s as recalculated by Morris-Chatta et al. (10). Thus, there is a need to establish whether the Biodex dynamometer can be used to reliably measure ankle dorsiflexion concentric strength.

Maximal isokinetic strength is usually determined by peak torque. In the Biodex manual (Biodex Medical Systems, Inc., Shirley, New York, USA) it is suggested that work is "a better indicator of the functional ability of a joint than peak torque, since the muscle must maintain force throughout the range of motion as opposed to force in one instant". The measurement of both peak torque and work in the same individual could therefore address different aspects of neuromuscular performance.

During normal activities in daily life, there is often a need to do episodic high-demand tasks quickly. If such time-critical neuromuscular performance is impaired, for example as a result of a chronic progressive neurological or neuromuscular disorder, the risk for falls and injuries

Fig. 1A, B. Illustrations showing the positioning in the Biodex dynamometer for measuring isokinetic ankle dorsiflexor strength.

increases. A measure of time-critical neuromuscular performance is torque recorded at a specific time. The inclusion of this measure could add to the understanding of functional performance. To the best of our knowledge, no previous study has determined the reliability of isokinetic measurements of time-critical neuromuscular performance in the ankle dorsiflexors.

Reliability is commonly assessed statistically by the Pearson product moment correlation coefficient (r). This coefficient could be misleading, as Pearson's r measures the strength of a relation between two variables, and not the agreement between them (4). The intraclass correlation coefficient (ICC) has been suggested as a more appropriate index, because the systematic variability is also treated as error (16). In clinical settings it has been suggested that measurement errors are calculated to address the inherent variability within the method (14), and that the data are presented graphically to assess any systematic variations (4).

The purposes of this study were: (i) to determine the test–retest intra-rater reliability of isokinetic ankle dorsiflexor peak torque, work and torque at a specific time at different angular velocities, in young healthy men and women using the Biodex dynamometer, and (ii) to examine several statistical measures for the interpretation of reliability. To represent functional angular velocities, for example as in walking, it is suggested that ankle dorsiflexion strength is evaluated at high angular velocities. Wolfson et al. (21) proposed that 120°/s is a functional velocity for ankle dorsiflexion. In the present study, we therefore assessed the reliability at five angular velocities, from 30°/s to 150°/s.

METHODS

Subjects

Fifteen men (age 23 ± 3 years [mean ± SD], height 181 ± 6 cm, weight 74 ± 8 kg) and 15 women (age 23 ± 3 years, height 171 ± 3 cm, weight 65 ± 6 kg) were recruited from the Physical Therapy programme at Lund University. All subjects were in good health with no reported neuromusculoskeletal dysfunction in the tested leg within the past year. None of the participants was training for an athletic event, but participated regularly (1–3 times per week) in recreational sports. Subjects gave written informed consent prior to the study, and the project was approved by the Ethics Research Committee of Lund University.

Equipment

All tests used the Biodex® Multi-Joint System 2 isokinetic dynamometer (Biodex Medical Systems, Inc., Shirley, New York, USA) with the Biodex advantage software version 4.0. The standard Biodex ankle unit attachment with the Biodex provided velcro straps was used. The cushion adjustment knob on the controller was set at "hard" and the highest sensitivity setting ("E") was chosen according to the Biodex manual. The cushion adjustment dial provides a means of varying the point at which deceleration starts. When a "hard" cushion is selected, deceleration begins relatively close to the stopping point. System sensitivity is used to avoid rapid oscillations of the accessories that are linked to the powerhead. The sensitivity setting has no effect on torque data at the preselected testing velocity. Before testing each subject, the system was calibrated according to the procedures in the manual.

Subject positioning

Each subject was seated in the adjustable chair of the dynamometer, with the leg to be tested elevated by a support arm under the knee (Fig. 1A). The subject's ankle was placed on the Biodex foot-plate and the foot was secured with two velcro straps (Fig. 1B). All subjects were tested in stocking feet. Two diagonal standard velcro straps stabilized the trunk, and one strap secured the hip. The arms were crossed over the chest, and the contralateral foot was placed on a support arm attached to the chair.

The transverse axis of the ankle joint was aligned with the rotational axis of the machine. The anatomical reference used to define the transverse axis was a line through the lateral and the medial tibial condyle. A hand-held goniometer was used to set the angles of the hip and the knee joints at 80° flexion (0° neutral position) and 30° flexion (0° straight leg), respectively. The hip and knee joint angles were adjusted by changing the distances between the chair, the foot-plate and the height of the support arm under the knee. This position on the Biodex was recorded for each subject individually and used in the following sessions.

The range of motion of the ankle joint was determined when the subject was positioned in the dynamometer. With the hand-held goniometer, 0° was defined as the tibia being perpendicular to the sole of the foot. End-range setting was standardized for all subjects from 30° plantar flexion (PF) (starting position) to 15° dorsiflexion (DF). To negate the influence of the gravity effect torque on the test data, each subject's limb was weighed and the data were corrected by the Biodex software.

Test protocol

Throughout the study, all tests were done by the same person (AMH). Only the dominant leg was tested. Leg dominance was determined by asking which leg was habitually used for 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 seven days between test sessions. One to three days prior to the first test, a pre-test was done for familiarization with the test procedure and the Biodex equipment. During the pre-test, five submaximal contractions at 30°/s and 120°/s were performed, followed by three maximal contractions at each of the five test velocities (see below).

Both tests started with five minutes of stationary cycling at a load of 1 Watt/kg body weight, followed by five submaximal concentric muscle contractions at 30°/s. After a one minute rest a further five submaximal contractions at 120°/s were performed.

For the test procedure, three non-consecutive maximal concentric contractions were performed for each of five angular velocities in the following order: 30°/s, 60°/s, 90°/s, 120°/s and 150°/s. Each contraction started from a relaxed plantar flexed position without any preload. A 30 second rest was allowed between each maximal contraction, and a two minute rest between each angular velocity. Each subject was instructed to exert maximal voluntary effort by contracting as hard and as fast as possible. Subjects were allowed to view the Biodex computer monitor, which displayed the recorded torque measurement, but nobody was verbally encouraged during the contractions.

Data

From each set of three contractions, three variables were collected for the analysis: (i) peak torque; (ii) work; and (iii) torque–time. The highest peak torque of the three contractions at each angular velocity was obtained from the Biodex report sheet and was used as the criterion score. Work is defined as torque over the distance throughout the entire range of motion and is commonly referred to as 'area under the curve', with its unit being Joule (J) or Newton meter (Nm) (5). For each angular velocity, the work value (Nm) from the contraction that yielded the highest peak torque was obtained from the report sheet. The contraction with the highest torque-time value, i.e. the highest torque recorded at a specific time, was used throughout the analyses. This value was obtained from the analysis test curve provided by the Biodex software, using the cursor on the screen adjusted to 40 ms time blocks. The recording time was calculated from the point the Biodex sensed the velocity input and was chosen to ensure that the maximum torque–time always occurred prior to peak torque. For the five angular velocities tested—30°/s, 60°/s, 90°/s, 120°/s and 150°/s—torque was recorded at 200 ms, 160 ms, 120 ms, 80 ms and 80 ms, respectively. The sampling frequency of the Biodex is set at 100 Hz by the data acquisition hardware.

Statistical analyses

Several measures, or statistiscal tests, were used to assess the reliability of the strength measurements: the intraclass correlation coefficient (ICC), the Pearson product moment correlation coefficient (r), the paired *t*-test and method error statistics. Definitions are given for the intraclass correlation coefficient and method error statistics, as different forms are available. As

Table I. *Two-way analysis of variance (ANOVA) for model [2]*

Source of variation	Degrees of freeedom	Sum of squares	Mean squares	Name
Between subjects	29	2608.25	89.94	BMS
Within subjects	30	119.94	4.00	
Between test sessions	1	0.0135	0.0135	JMS
Residual	29	119.92	4.14	EMS
Total	59	2728.19		

The analysis was based on data for peak torque at 30°/s.

the paired *t*-test did not add to the content of the study, we do not report the results of this test.

As a random effects model in the repeated measures analysis of variance is appropriate, ICC is $ICC_{2,1}$, in the nomenclature of Shrout & Fleiss (16), which is defined by

$$ICC = (BMS-EMS) / (BMS+EMS+2(JMS-EMS)/n) \quad [1]$$

where BMS represents the variability between subjects, JMS the variability in the measurements within subjects, and EMS the variability remaining when the between and within subjects variability have been accommodated. In particular, BMS is the between-subjects mean square, JMS the mean square within subjects, EMS the residual mean square, and n, the number of subjects (15 when the sexes are analysed separately, otherwise 30). To illustrate the terminology, a two-way analysis of variance (ANOVA) for the model

$$\text{Peak torque at } 30°/s = K + \text{subject effect} + \text{test session effect} \quad [2]$$

is presented in Table I.

To investigate why most of the Pearson's r and ICC values were so close in this study, we express r in terms of ICC:

$$r = ICC\ 0.5\ (k+1/k)\ (1 + [2\ (JMS-EMS)/(n(BMS+EMS))]) \quad [3]$$

where k^2 is the ratio of the variances of the measurements in test 1 and test 2. As k is often close to one, the term 0.5 $(k + 1/k)$ is usually close to unity: if the variance of the measurements in test 1 is 50% more than in test 2, then $k = \sqrt{1.5}$ and the value of 0.5 $(k + 1/k)$ is 1.021. For n not small, the value of $(1 + [2(JMS - EMS)/n(BMS + EMS)])$ is close to unity. If JMS exceeds EMS, then r is larger than ICC. If JMS is less than EMS, ICC may be larger than r, but r could be less than ICC for k close to one. In reality, often r and ICC take similar values.

Three forms of method error statistics were considered: the standard error of the measurement (SEM), the method error (ME) and the coefficient of variation (CV). (We have adopted the conventional terminology in publications on this topic, even though we are aware that SEM is sometimes used to represent the standard error of the mean.)

The SEM is defined by

$$SEM = SD_1\ (1 - ICC)^{0.5} \quad [4]$$

where SD_1 is the standard deviation of all the measurements (3).

If SD_2 is the standard deviation of the differences between the two measurements, then the ME is defined by

$$ME = SD_2 / \sqrt{2} \quad [5]$$

It can be shown that if there are n pairs of measurements, then

$$ME^2 = EMS \quad [6]$$

If *n* is sufficiently large and the mean difference small, both highly likely conditions, then ME and SEM take similar values.

The CV of ME is defined by

$$CV = 100\ ME / X_c \quad [7]$$

where X_c is the mean for all the observations from test session 1 and test session 2.

SEM and ME are both expressed in metric units, while CV is here given as a percentage value.

Throughout the study all calculations were performed using the SPSS 6.1 Software (SPSS Inc., Chicago, Ill., USA).

RESULTS

The means and standard deviations for peak torque, work and torque-time are presented in Table II for men and women, and for the two test sessions, separately. The differences between the means of the two sessions were always smaller than 7% and smaller than 3% for 18 of the 30 pairs of means. In Figure 2, the torque–velocity relationships for men and women are illustrated. The absolute peak torque values at the five angular velocities and the torque–velocity relationship are comparable to other studies on healthy young adults (12).

Initially, the reliability statistics for peak torque, work and torque–time—ICC, Pearson's r, SEM, ME and CV—for the five angular velocities were calculated for

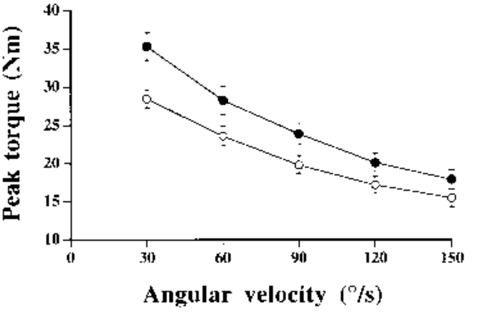

Fig. 2. The torque–velocity relationship (mean ± SE) for men (●) and women (○).

Table II. *Means and standard deviations for peak torque, work and torque-time for five angular velocities*

		Test session	30°/s	60°/s	90°/s	120°/s	150°/s
Men (n = 15)	Peak torque (Nm)	1	35.0 ± 7.5	28.0 ± 7.3	23.9 ± 5.1	20.0 ± 4.5	17.4 ± 4.4
		2	35.5 ± 6.7	28.6 ± 7.2	23.9 ± 6.4	20.2 ± 5.7	18.4 ± 6.0
	Work (Nm)	1	20.2 ± 6.0	16.8 ± 5.1	14.3 ± 4.1	12.4 ± 3.4	10.5 ± 3.0
		2	20.7 ± 5.1	17.8 ± 5.5	15.2 ± 4.6	12.6 ± 4.0	11.2 ± 4.0
	Torque-time* (Nm)	1	30.3 ± 6.0	23.8 ± 5.9	19.8 ± 4.5	15.5 ± 4.2	12.5 ± 4.7
		2	29.2 ± 6.2	25.2 ± 6.8	21.1 ± 4.9	15.4 ± 5.2	12.2 ± 6.3
Women (n = 15)	Peak torque (Nm)	1	28.8 ± 4.8	23.3 ± 4.8	19.5 ± 4.4	7.3 ± 4.0	15.3 ± 4.3
		2	28.3 ± 4.8	23.8 ± 5.1	20.2 ± 4.9	17.1 ± 4.8	15.7 ± 4.9
	Work (Nm)	1	17.1 ± 4.0	14.2 ± 4.0	12.3 ± 3.5	10.7 ± 2.9	9.3 ± 2.8
		2	17.0 ± 4.2	14.9 ± 4.0	12.5 ± 3.6	10.6 ± 3.3	9.5 ± 3.2
	Torque-time* (Nm)	1	23.2 ± 3.8	19.5 ± 3.7	16.1 ± 3.0	12.4 ± 3.8	11.4 ± 3.7
		2	22.2 ± 4.3	19.1 ± 3.4	16.5 ± 4.2	12.9 ± 4.4	12.2 ± 4.0

The highest peak torque, work and torque-time values from each of the three contractions at each angular velocity in test sessions 1 and 2 were used to calculate the means and standard deviations for each group.
*For the five angular velocities tested—30°/s, 60°/s, 90°/s, 120°/s and 150°/s—torque was recorded at 200 ms, 160 ms, 120 ms, 80 ms and 80 ms, respectively.

Table III. *Reliability measures for peak torque at five angular velocities*

	30°/s	60°/s	90°/s	120°/s	150°/s
Intraclass correlation coefficient (ICC)	0.91	0.93	0.90	0.78	0.80
Pearson's correlation coefficient (r)	0.91	0.93	0.88	0.79	0.82
Standard error of the measurement (SEM) (Nm)	2.01	1.69	1.71	2.16	2.15
Method error (ME) (Nm)	2.03	1.71	2.00	2.35	2.25
Coefficient of variation (CV) (%)	6.4	6.6	9.2	12.6	13.5

The highest peak torque value at each angular velocity was used for the reliability analyses.

Table IV. *Reliability measures for work at five angular velocities*

	30°/s	60°/s	90°/s	120°/s	150°/s
Intraclass correlation coefficient (ICC)	0.88	0.83	0.83	0.83	0.76
Pearson's correlation coefficient (r)	0.88	0.84	0.84	0.84	0.78
Standard error of the measurement (SEM) (Nm)	1.70	1.84	1.61	1.36	1.52
Method error (ME) (Nm)	1.76	1.92	1.67	1.44	1.60
Coefficient of variation (CV) (%)	9.4	12.1	12.3	12.5	15.9

For each angular velocity, the highest work value from the contraction that yielded the highest peak torque was used for the reliability analyses.

men and women separately. As no discernible systematic differences between the sexes could be found, the results for men and women are combined throughout the analyses and presentations.

The reliability statistics for peak torque and work for the five angular velocities are presented in Tables III and IV, respectively. For peak torque the values of ICC and Pearson's r are almost identical for the five angular velocities. For work, the values of ICC and Pearson's r are also very similar, but slightly less close to unity than for the peak torque values. For both peak torque and work, the values of SEM and ME are relatively consistent, but the CV increases as the velocity increases. For peak torque and work, 30% of the highest values were obtained from the first contraction, 31% from the second, and 39% from the third.

To illustrate any systematic variability between the two test sessions, the differences between test 1 and test 2 (test 1 minus test 2) are plotted against their mean for each subject for the five angular velocities. As peak torque and work are highly correlated at all five angular velocities (see below), only data for peak torque are presented (Fig. 3A–E). There appears to be no systematic variability within each graph. For 120°/s and 150°/s, the

Table V. *Reliability measures for torque-time* at five angular velocities*

	30°/s	60°/s	90°/s	120°/s	150°/s
Intraclass correlation coefficient (ICC)	0.84	0.89	0.74	0.64	0.61
Pearson's correlation coefficient (r)	0.85	0.90	0.76	0.64	0.61
Standard error of the measurement (SEM) (Nm)	2.40	1.86	2.21	2.48	2.65
Method error (ME) (Nm)	2.42	1.90	2.34	2.76	3.00
Coefficient of variation (CV) (%)	9.2	8.7	12.7	19.7	24.9

*For the five angular velocities tested—30°/s, 60°/s, 90°/s, 120°/s and 150°/s—torque was recorded at 200 ms, 160 ms, 120 ms, 80 ms and 80 ms, respectively. The highest torque-time value at each angular velocity was used for the reliability analyses.

Table VI. *Pearson's correlation coefficients between peak torque and work and between peak torque and torque-time* at five angular velocities*

	Peak torque vs work		Peak torque vs torque-time*	
	Test session 1	Test session 2	Test session 1	Test session 2
30°/s	0.86	0.83	0.94	0.92
60°/s	0.93	0.93	0.93	0.94
90°/s	0.93	0.96	0.88	0.93
120°/s	0.92	0.96	0.75	0.91
150°/s	0.93	0.94	0.74	0.75

*For the five angular velocities tested—30°/s, 60°/s, 90°/s, 120°/s and 150°/s—torque was recorded at 200 ms, 160 ms, 120 ms, 80 ms and 80 ms, respectively. The highest peak torque, work and torque-time values at each angular velocity were used for the analyses.

difference in peak torque has a wider range than for the three lower angular velocities.

The reliability statistics for torque–time are presented in Table V. The values of ICC and Pearson's r are again very similar, but decrease as the velocity increases. The values of SEM, ME and CV for torque–time are slightly larger than those for peak torque and work, with a greater increase at increased angular velocities, in particular for the CV. In Fig. 4A–E, the differences in torque–time measurements are plotted against their mean for each subject for the five angular velocities. The variability between the two test sessions shows a similar pattern as for peak torque (cf. Fig. 3A–E). Eighty-two percent of the highest torque–time values were obtained from the first contraction, 14% from the second contraction and 4% from the third.

In Table VI, the relationships between the highest peak torque and work, and between the highest peak torque and torque–time, for each of the five angular velocities in each of the two test sessions are presented. For all angular velocities in both test sessions there is a highly significant positive relationship ($p < 0.001$) between the measurements. The highest peak torque and torque–time values were obtained from the same contraction in 100 of the 300 test occasions (30 subjects, 5 angular velocities and 2 test sessions), with 75 from the first, 17 from the second and 8 from the third contraction.

DISCUSSION

Factors affecting reproducibility

The usefulness of any assessment method depends upon our knowledge of, and the ability to control, factors that influence the measurements. In recent years, attention to the potential effects of variation in clinical dynamometry has increased considerably (9). For isokinetic dynamometry, four major factors are likely to influence the overall results: the accuracy of the dynamometer, the test protocol, the reproducibility of the measurement parameters, and subject-related factors. The mechanical measurement accuracy of the Biodex and other dynamometers have been assessed and found to be extremely high (18); thus, we did not consider it necessary in this study to address specifically the mechanical measurement accuracy.

Several potential sources of error in the test protocol have to be recognized and their effects reduced to

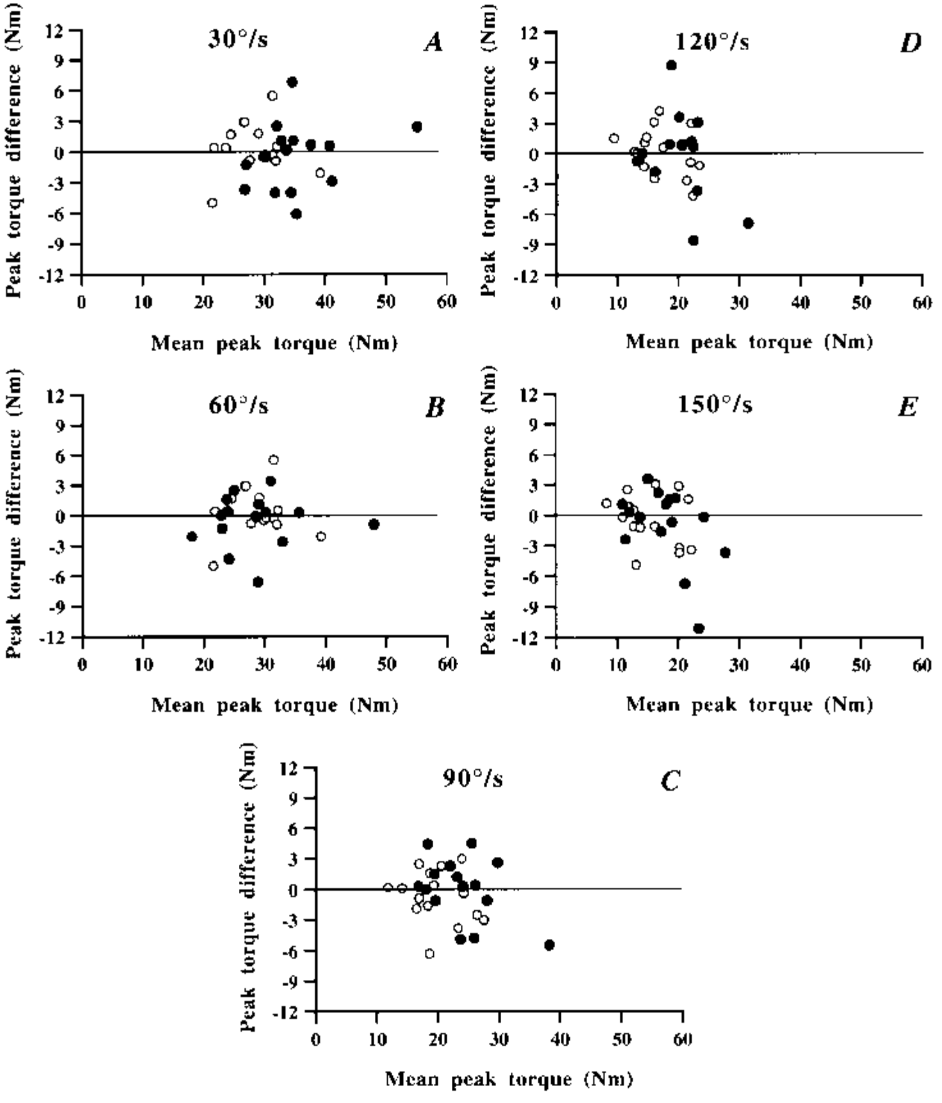

Fig. 3A–E. The differences between test session 1 and test session 2 (test 1 minus test 2) for peak torque plotted against their mean for each subject for the five angular velocities for 15 men (●) and 15 women (○).

optimize the reproducibility. Subject familiarization with the equipment is necessary and a pre-test was done prior to the first test to facilitate coordination of the movements of the ankle joint. Also, the warm-up procedures before each test session followed a strictly standardized protocol with cycling and submaximal contractions at different angular velocities.

Great care has to be taken to correctly position and stabilize the subject and the joint tested. Andersen (1) found that a 1.5 cm displacement of the anatomic axis of the ankle joint caused a 10% change in dorsiflexion and plantar flexion isokinetic peak torque and work. Öberg et al. (23) showed that torques during isokinetic ankle strength testing were significantly higher without upper trunk fixation. In this study, positioning and fixation followed the standard procedures in the Biodex manual, and the position of the subject and the ankle joint were recorded during the first test and carefully reproduced in the second.

The reproducibility during the testing can be influenced by the interval between contractions at a given angular velocity, between trials at different angular velocities, and between test sessions (9). Stratford et al. (17) and Andersen (1) found that a short rest between

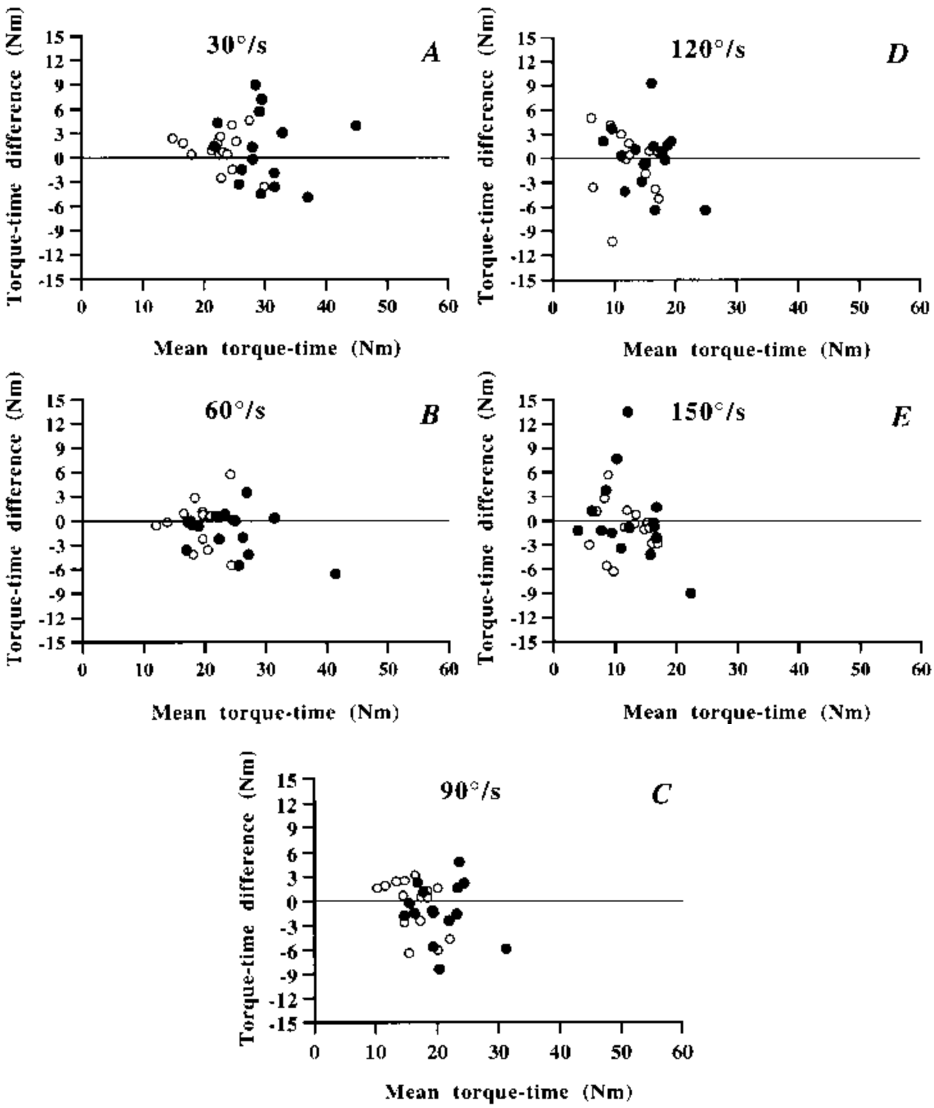

Fig. 4A–E. The differences between test session 1 and test session 2 (test 1 minus test 2) for torque–time plotted against their mean for each subject for the five angular velocities for 15 men (●) and 15 women (○).

contractions resulted in higher isokinetic recordings than with no rest. No study has systematically addressed the effects of a rest period between trials at different angular velocities, but it is likely that a rest will have a similar beneficial effect as a rest between contractions. In our study, the subjects rested 30 seconds between each maximal contraction and 2 minutes between trials at each of the five angular velocities. This was considered to be sufficient to reduce any effects of muscle fatigue on the measurements.

The time between test sessions varied considerably in previous studies of ankle dorsiflexion reliability. The shortest reported time has been 10 minutes in a study that showed low reliability (19). More commonly 24 hours up to 7 days have elapsed between sessions (1, 8, 13, 22), and these studies have generally found higher reliability. To ensure that the effects of both learning and fatigue were eliminated, the retest in this study was performed 7 days after the first test session.

The order of the test velocities can also influence reproducibility (5, 12). Wilhite et al. (20) tested concentric and eccentric knee extension at three angular velocities: 60°/s, 120°/s and 180°/s. The reproducibility of the measurements from subjects who began testing at 180°/s was much lower compared with when this velocity was second or third in the order. Consequently, we tested all subjects in this study at slower velocities first.

Verbal encouragement and visual feedback during the testing can influence the ability to produce maximum effort (9). Peacock et al. (11) described how knee extension performance was significantly enhanced by a combination of visual and auditory feedback, but not by either of them separately. Figoni & Morris (6) found that visual feedback could enhance performance depending on the angular velocity. Because of the difficulties standardizing encouragement, and the seemingly contradictory effects, we decided not to encourage subjects verbally during contractions but to allow them to view the Biodex computer monitor with the torque curve during the contraction.

It should be noted that any reliability study also has to take into account the characteristics of the subject being examined, such as their age and physical activity level, as well as their present and previous medical history (9). The reliability results presented in this study are therefore only applicable to younger healthy men and women, and other groups have to be tested for their reproducibility. However, the method and equipment can be used for other subjects or patients when isokinetic ankle dorsiflexor strength is evaluated.

Reliability statistics

Statistical methods for assessing agreement between repeated quantitative measurements have been considered by many authors (see, for example, 2, 4, 16). Many of their discussions are concerned with the inappropriate use of Pearson's r as a measure of agreement. If pairs of measurements are plotted, and if these measurements are in close agreement, then the points on the graph are often located close to a straight line that passes through the origin and has a slope of 1. The frequently used ICC assesses the proximity of the points to this particular line, whereas Pearson's r assesses the proximity to *any* straight line. There are many circumstances where the values of ICC and Pearson's r are close: in this study, the values of ICC and Pearson's r differ by at most 4% (cf. Tables III–V). From the algebraic relationship between ICC and Pearson's r—equation [3]—if n is large, if the variance of the first measurement is not too different from that for the second, and if the difference between mean squares JMS and EMS is reasonably small, then ICC and Pearson's r will be close. Importantly, ICC can also be used to measure the agreement between several repeated measurements, whereas Pearson's r only allows for comparisons of two measurements. On the other hand, ICC cannot be applied to assess agreement between measurements made in different units, nor when different properties are being measured even if the same unit is used. When assessing test–retest reliability in which the same measurements are made of the same property in the same units on two or more occasions, ICC is the preferred measure.

To interpret reliability measures, Fleiss (7) recommends that ICC values above 0.75 represent excellent reliability while values between 0.4 and 0.75 represent fair to good reliability. However, the decision whether a method is sufficiently reliable depends upon the specific application: for example, an ICC of 0.80 may be acceptable if the method is going to detect large differences as a result of an intervention, whereas a higher ICC may be required to detect very small differences.

To address the inherent variability within the method, measurement errors need to be calculated. The most common measurement errors reported are the standard error of the measurement, SEM, the method error, ME, and the coefficient of variation, CV. The values of SEM and ME in Tables III–V are close, which is not surprising as they essentially express the same property (see Methods; Statistical analyses). Since SEM is derived from the ICC, it can be used to assess the error when several measurements are made, whereas ME can only be used if two measurements are made. Both SEM and ME can be used to determine sample and effect sizes in power analysis and to calculate prediction intervals, allowing a clinician to determine if measurements from a patient, as part of a treatment or therapy, represent a real change or if the measurements are within the range of the error of the method. The coefficient of variation, CV, is independent of the units of measurement, and is useful as a descriptive tool (2).

When two repeated measurements are considered, the use of graphs is advocated, in which the ordinate is the difference between measurements and the abscissa is the mean of the two measurements (4). If there is a systematic variation—for example, the second measurement is more often larger than the first—then the points are not distributed equally about the zero line. From these graphs possible outliers—measurements that are wrong or are substantially different from the other measurements—can be easily identified. The possibility that the differences between measurements increase as the mean value of the measurement increases (or decreases) can also be addressed.

Reliability of peak torque, work and torque-time

Reliability for concentric ankle dorsiflexion peak torque at all five angular velocities measured on the Biodex dynamometer was excellent according to the recommendations of Fleiss (7) (ICC 0.78 to 0.93; Table III) but tended to be lower for 120°/s and 150°/s. Similar results have been reported on other isokinetic dynamometers (1, 8, 13, 22). The only previous reliability study using the Biodex dynamometer (19) showed considerably lower reliability. Wennerberg (19) suggested that the reasons for this low reliability could have been the inconsistent positioning and the very short time between test occasions. We addressed both of these factors in the design of the test protocol, which most likely explains the much higher reliability in the present study.

The SEM and ME for peak torque showed no consistent pattern of change with increasing angular velocity, although they were slightly higher at the two fastest velocities (Table III). Since the absolute values for peak torque decreased with increasing velocity (cf. Table II, Fig. 2), and since SEM and ME represent absolute values, the CV doubled between 30°/s and 150°/s (from 6.4 to 13.5%) (Table III). In addition, the graphs showed a wider range of peak torque difference for the three higher angular velocities (Fig. 3C–E). Thus, even though ICCs for peak torque at all five angular velocities were considered excellent, the higher measurement errors at faster velocities have to be considered when the sample or effect sizes are determined.

Reliability for work was very similar to that for peak torque with reliability coefficients being excellent (ICC 0.76 to 0.88, Table IV). Because the actual values for work were smaller than for peak torque, the values of SEM and ME were lower, but the value for CV was slightly larger. Only one study has previously evaluated the reliability for ankle dorsiflexion work (30°/s and 120°/s), and the reliability coefficients and absolute values were similar to ours (22).

Work is often advocated in clinically oriented situations, for example after injuries, as it is considered to add information about performance of the tested muscle or muscle group (12). As in a previous study of ankle dorsiflexion strength (22), we also found a very high correlation between peak torque and work for all five angular velocities (cf. Table VI). This indicates a close relationship between peak torque and work, and in healthy young adults the addition of work to the set of measurements appears to add no information.

As a representation of time-critical neuromuscular performance, the Biodex test report includes an index of the rate of tension development: the torque at 200 ms. In the Biodex manual it is stated: “The time of 0.20 seconds is pre-selected because it has been documented that upon heelstrike it takes the leg extensors 0.20 seconds to develop enough force to support the body in normal ambulation.” However, it was found that dorsiflexion peak torque for several subjects was reached before 0.20 seconds at most velocities. Therefore, we decided to reduce the time point at which the torque was recorded. For each velocity a unique time was selected based on the occurrence of peak torque, and within the restrictions of the analysis software. The torque which occurred at this specific time was defined as ‘torque–time’.

Reliability for torque–time was excellent for 30°/s and 60°/s (ICC of 0.84 and 0.89, respectively), and fair/good for the higher velocities (ICC 0.61 to 0.74) (Table V). The CVs for 120°/s and 150°/s were greater than 19% (Table V), and the graphs showed a similar pattern as for peak torque (Fig. 4A–E). Reasons for the lower reliability at the higher velocities could include the relatively low sampling frequency (100 Hz) and/or the acceleration phase of the movement, known to occur for all isokinetic dynamometers (5).

The absolute values of torque-time, like peak torque and work, declined with increasing angular velocity. Unlike peak torque, 82% of the maximum torque–time values were obtained from the first contraction. The relationship between peak torque and torque–time was high (Pearson’s r 0.74 to 0.94) in these subjects. This is not unexpected as peak torque and torque–time are obtained from the same measurements. As the peak torque and torque–time values were seldom obtained from the same contractions, the high correlation also indicates that the two parameters are related physiologically. Further studies are needed to understand how, for example, the fibre type composition, metabolism and muscle cross-sectional area contribute to peak torque and torque–time.

In conclusion, we have determined that concentric ankle dorsiflexion peak torque, work and torque-time in young healthy men and women can be reliably examined using the Biodex isokinetic dynamometer. The inclusion of torque-time could add to the clinical evaluation of time-criticial neuromuscular performance. For young healthy individuals, measurements of work appear to add no further information. For the statistical analyses of test–retest reliability, it is recommended that the ICC and assessments of measurement errors (SEM, ME or CV) are included, and that the data are presented graphically to indicate any systematic variations.

ACKNOWLEDGEMENT

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, Loo and Hans Osterman Foundation, Gun and Bertil Stohne 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. Andersen, H.: Reliability of isokinetic measurements of ankle dorsal and plantar flexors in normal subjects and in patients with peripheral neuropathy. Arch Phys Med Rehabil *77:* 265–268, 1996.
2. Armitage, P. & Berry, G.: Statistical methods in medical research. Blackwell Scientific Publications, Oxford, 1994.
3. Baumgartner, T. A. & Jackson, A. S.: Measurement for evaluation in physical education and exercise science. Wm. C. Brown Publishers, Dubuque, 1991.
4. Bland, J. M. & Altman, D. G.: Statistical methods for assessing agreement between two methods of clinical measurement. Lancet *8476:* 307–310, 1986.
5. Dvir, Z.: Isokinetics: Muscle testing, interpretation, and clinical applications. Churchill Livingstone, New York, 1995.
6. Figoni, S. F. & Morris, A. F.: Effects of knowledge of results on reciprocal isokinetic strength and fatigue. J Orthop Sports Phys Ther *6:* 190–197, 1984.
7. Fleiss, J. L.: The design and analysis of clinical experiments. John Wiley & Sons, New York, 1986.
8. Karnofel, H., Wilkinson, K. & Lentell, G.: Reliability of isokinetic muscle testing at the ankle. J Orthop Sports Phys Ther *11:* 150–154, 1989.
9. Keating, J. L. & Matyas, T. A.: The influence of subject and test design on dynamometric measurements of extremity muscles. Phys Ther *76:* 866–889, 1996.
10. Morris-Chatta, R., Buchner, D. M., de Lateur, B. J., Cress, M. E. & Wagner, E. H.: Isokinetic testing of ankle strength in older adults: assessment of inter-rater reliability and stability of strength over six months. Arch Phys Med Rehabil *75:* 1213–1216, 1994.
11. Peacock, B., Westers, T., Walsh, S. & Nicholson, K.: Feedback and maximum voluntary contraction. Ergonomics *24:* 223–228, 1981.
12. Perrin, D. H.: Isokinetic exercise and assessment. Human kinetics. Champaign, 1993.
13. Porter, M. M., Vandervoort, A. A. & Kramer, J. F.: A method of measuring standing isokinetic plantar and dorsiflexion peak torques. Med Sci Sports Exerc *28:* 516–522, 1996.
14. Roebroeck, M. E., Harlaar, J. & Lankhorst, G. J.: The application of generalizability theory to reliability assessment: an illustration using isometric force measurements. Phys Ther *73:* 386–400, 1993.
15. Rothstein, J. M.: Measurement and clinical practise: theory and application. *In* Measurement in physical therapy (ed. J. M. Rothstein). Churchill Livingstone, New York, 1985.
16. Shrout, P. E. & Fleiss, J. L.: Intraclass correlations: uses in assessing rater reliability. Psychol Bull *86:* 420–428, 1979.
17. Stratford, P. W., Bruulsema, A., Maxwell, B., Black, T. & Harding, B.: The effect of inter-trial rest interval on the assessment of isokinetic thigh muscle torque. J Orthop Sports Phys Ther *11:* 362–366, 1990.
18. Timm, K. E., Gennrich, P., Burns, R. & Fyke, D.: The mechanical and physiological performance reliability of selected isokinetic dynamometers. Isokin Ex Sci *2:* 182–190, 1992.
19. Wennerberg, D.: Reliability of an isokinetic dorsiflexion and plantar flexion apparatus. Am J Sports Med *19:* 519–522, 1991.
20. Wilhite, M. R., Cohen, E. R. & Wilhite, S. C.: Reliability of concentric and eccentric measurements of quadriceps performance using the Kin-Com dynamometer: the effect of testing order for three different speeds. J Orthop Sports Phys Ther *15:* 175–182, 1992.
21. Wolfson, L., Judge, J., Whipple, R. & King, M.: Strength is a major factor in balance, gait, and the occurrence of falls. J Gerontol *50A (Special Issue):* 64–67, 1995.
22. Woodson, C., Bandy, W. D., Curis, D. & Baldwin, D.: Relationship of isokinetic peak torque with work and power for ankle plantar flexion and dorsiflexion. J Orthop Sports Phys Ther *22:* 113–115, 1995.
23. Öberg, B., Bergman, T. & Tropp, H.: Testing of isokinetic muscle strength in the ankle. Med Sci Sport Exc *19:* 318–322, 1987.

Accepted January 8, 1999

Address for offprints:

Anna Maria Holmbäck, PT, MSc
Department of Physical Therapy
Lund University
Box 5134
SE-220 05 Lund
Sweden