Isokinetics and Exercise Science 22 (2014) 17–26
DOI 10.3233/IES-130522
IOS Press

Test-retest relative and absolute reliability of knee extensor strength measures and minimal detectable change

Asunción Ferri-Morales[a,b,*], Luis M. Alegre[b], Angel Basco[a] and Xavier Aguado[b]
[a]*Department of Nursing and Physiotherapy, University of Castilla-La Mancha, Toledo, Spain*
[b]*Biomechanics of Human Movement and Sport, Research, University of Castilla-La Mancha, Toledo, Spain*

Received 10 September 2013
Accepted 1 January 2013

Abstract.
BACKGROUND: Few studies have assessed the test-retest reliability of the isokinetic strength at 90°/s and the isometric parameters for knee extensors.
OBJECTIVES: To assess the reliability of isokinetic and isometric parameters of knee extensors, and to determine the smallest real difference between the test-retest measures.
METHODS: Knee extensor strength was measured twice, 4 to 5 days apart, using a Biodex dynamometer in forty subjects (12 men, 28 women). The protocols consisted of 5 concentric contractions at 90°/s and 5 isometric contractions.
RESULTS: For women, test-retest reliability was very high for isometric and isokinetic peak-torque ($\text{ICC}_{c,1} > 0.9$), moderate for work ($\text{ICC}_{c,1} = 0.82$) and questionable for angle of peak torque (Angle-PT), mean Angle-PT and rate of torque development ($\text{RTD}_{0-100\ \text{ms}}$) (ICC < 0.67). The measurement errors were small for all variables, the SEM% ranged between 3.1% to 7.4%, and SRD% from 8.6% to 19.9%, except for $\text{RTD}_{0-100\ \text{ms}}$ that was the most variable (SEM% $=$ 16.4%; SRD $=$ 47.5%). No differences were found in Angle-PT and $\text{RTD}_{0-100\ \text{ms}}$ with regard to sex.
CONCLUSIONS: The isokinetic values obtained at 90°/s and isometric peak-torque for knee extensors are highly repeatable with a standardized test protocol and the SRD values can be used to detect real changes. Alternative strategies of analysis should be developed to improve reliability of Angle-PT and $\text{RTD}_{0-100\ \text{ms}}$. Angle-PT and $\text{RTD}_{0-100\ \text{ms}}$ do not appear affected by gender

Keywords: Rehabilitation, reproducibility of results, dynamometry, torque

1. Introduction

In both rehabilitation and sports medicine, accurate and reliable measurements of muscle function are required to assess the impact of therapeutic interventions as well as the effects of physical training. Muscle strength in both health and disease can be assessed isokinetically and isometrically using dynamometers. To be clinically meaningful, the assessment procedure must be reliable and sensitive enough to determine whether a finding indicates impairment or improvement and to evaluate the outcomes of therapeutic intervention. The values must be defined to provide guidance in deciding whether an observed change upon reassessment is within the limits of assessment error or whether there has been a true change.

For a long time, correlation coefficients, such as the intraclass correlation (ICC), were almost exclusively reported as evidence of measurement variability in the test-retest reliability of isokinetic knee strength studies. However, ICC is a relative index of reliability that

*Corresponding author: Asunción Ferri Morales, Departamento de Enfermería y Fisioterapia, Avda Carlos III s/n 45071 Toledo, Spain. Tel.: +34 925 268 800 ext 5820; Fax: +34 925 268 811; E-mail: asuncion.ferri@uclm.es.

ISSN 0959-3020/14/$27.50

provides information about a measure's ability to discriminate among subjects but neither expresses measurement error in a clinically meaningful terms [1], nor does it reveal the retest variability in the units of measurement that the clinician must consider when interpreting the significance of a retest score based on an earlier test score. Examples of absolute indices that capture this notion of within-subject variability are the standard error of measurement (SEM) for a single test and the smallest real difference (SRD) for the interpretation of change [2]. The SRD is an index that reports the minimum change that can be interpreted as a true difference and has been defined by Beckerman [2], as the amount of change required to be 95% confident that an observed change in scores reflects real change in the variable. Both the SEM and SRD are readily interpretable because they are expressed in the units of the measurements used in their calculation.

Keating and Matyas's reviews [3,4] evidenced that subject-related factors such as age, weight, sex, and athletic condition should be considered in reliability studies of torque measurements obtained with a dynamometer. Research shows that forces decline with increasing age, that heavier subjects tend to produce higher strength values than lighter subjects, and forces generated by males generally exceed those generated by females when participants are grouped by age and athletic background.

Literature has reported that strength variables are usually heteroscedastic [5,6], thus, if the men are stronger than women, the amount of error observed for the men could be larger than that observed in the women and it may therefore be inappropriate to pool data across gender when establishing estimates of test-retest measurement reliability, because the errors could be underestimated for males and overestimated for females [3].

Previous reports have suggested "high" to "very high" reliability for the isokinetic variables for knee extensors at different angular velocities [7–9], but very few studies have evaluated this reliability at 90°/s in healthy young subjects using the Biodex dynamometer. Montgomery [8], reported that the reliability was generally higher at slower angular velocities and higher for knee extension than flexion. Given that reliability varies with angular velocity tested, we chose to assess the test-retest reliability of different parameters at 90°/s. This assessment would be useful in studies where the force-velocity relationship needs to be evaluated. Similarly, the absolute reliability of isometric knee extensor contractions has been poorly documented, as has the reliability of other parameters, such as the angle of peak-torque (Angle-PT) and the rate of torque development (RTD). The RTD is an indirect measure that may be associated with the ability of the neuromuscular system to generate force or torque rapidly and can be evaluated using the isometric mode [10]. The RTD measures the slope of the torque-time curve. This parameter has been associated with the rapid postural adjustments required and utilized to recover stability from a loss of balance and the rapid force increases during a change of direction task in sports [11]. The torque-angle curve displayed during an isokinetic test is influenced by the muscle force and the length of the moment arm change throughout the joint movement [12]. The Angle-PT has been utilized as an external reference of the length-tension relation of the muscles that cross a given joint and is sensitive to shifts in this relationship [13,14]. Furthermore, the changes in this variable are meaningful when we compare muscle groups with and without muscle strain injuries [14] or analyze the effects of different protocols on muscle damage [15].

Therefore, the main aims of this study were to assess in young women the test-retest reliability of several strength parameters of knee extensor muscles under isokinetic (90°/s) and isometric tests and to define the limits for the smallest change that indicate a real change in all of them. A secondary objective was to determine whether there are differences between men and women in the Angle-PT and the $\mathrm{RTD}_{0-100\ \mathrm{ms}}$.

2. Materials and methods

2.1. Subjects

Forty university students, 12 men and 28 women (mean (sd) age, 20.2 (3.3) years), participated voluntarily in this study. The participants' characteristics are presented in Table 1. The exclusion criteria were as follows: having a neurologic, cardiovascular, musculoskeletal, or systemic disease; performing sports or physical activity more than twice a week; and performing lower body strength training during the study period. Before testing, we obtained informed consent from all of the subjects. The study was approved by the local Institutional Review Board and was performed in accordance with Helsinki Declaration of 1964, revised in 2008.

2.2. Instrumentation

We used the Biodex System 3 Pro isokinetic dy-

Table 1
Study population characteristics

	Measure	Women $n = 28$	Men $n = 12$
Age (y)	Mean (SD)	19.5 (2.2)	22.1 (4.6)
	Median (range)	18.5 (18 to 26)	21.0 (18 to 31)
Heigth (m)	Mean (SD)	1.65 (0.05)	1.77 (0.05)
	Median (range)	1.65 (1.51 to 1.76)	1.79 (1.69 to 1.86)
Body mass (kg)	Mean (SD)	56.5 (8.6)	77.3 (12.2)
	Median (range)	53.7 (41 to 75)	76.1 (60 to 100)
BMI* (kg/m^2)	Mean (SD)	20.7 (2.6)	24.5 (3.6)
	Median (range)	20.3 (17.30 to 27.9)	24.0 (20.1 to 29.7)

*BMI: Body mass index.

namometer (Biodex Medical, Shirley, NY, USA) and the Biodex Advantage Software program, v. 4.0. We collected the Biodex data at a frequency of 100 Hz. The dynamometer was calibrated weekly throughout the study in accordance with the manufacturer's specifications. Studies have documented that the mechanical reliability of this dynamometer is excellent [16].

2.3. *Experimental protocol*

The subjects were scheduled to participate in three sessions (4 to 5 days apart), at the same time of day. All of the tests were performed on the right leg. The first session was designed to familiarize subjects with the dynamometer and the testing protocol. This session simulated the experimental test sessions. Weight, height, and data related to the testing position were recorded to serve as a reference for each individual subject in successive sessions. In the second and third sessions, the isokinetic and isometric strength tests were performed.

The subjects sat on the adjustable chair of the dynamometer with the seat tilted back to an angle of 85° relative to the horizontal level; the subjects were stabilized by thigh, pelvic, and torso straps. The lateral femoral epicondyle was carefully aligned with the center of the dynamometer shaft to establish the axis of rotation around the knee joint. The length arm of the dynamometer was adjusted to ensure that the ankle strap rested comfortable 2 cm to 3 cm above the lateral and medial malleoli. We recorded all of the chair settings to ensure setup reproducibility on subsequent visits. For the isokinetic test, the participant's range of motion was 90° (from flexion to 0° of full extension). Each subject was required to fold their arms across their chest. For the isometric test, the arm of the dynamometer was blocked at 60° of knee flexion, which has been demonstrated to be the angle of maximal isometric force generation [17]. Gravity correction was performed according to the manufacturer's specifications.

Before testing, all of the subjects completed the standardized dynamic warm-up, which consisted of cycling for 5 minutes at 90 W and approximately 60 rpm and a specific warm-up before each test, with 3 series of concentric contractions for knee extensors. The first familiarization of three consecutive submaximal and slightly stronger contractions was followed by 30 s of rest; the second of two submaximal, stronger repetitions was followed by 60 s of rest; and the third of two maximal contractions was followed by 90 s of rest. In the specific warm-up for the isometric test, the participants performed 3 submaximal contractions at 60° knee flexion. The test was started after these warm-up contractions.

The isokinetic test consisted of five consecutive maximal voluntary concentric contractions from 90° flexion to full knee extension at a velocity of 90°/s with a 5-s rest between contractions during which the lower limb was passively returned to the starting position.

The isometric test consisted of five consecutive, maximal, voluntary 4-s contractions with a 15-s rest between contractions. The rest between strength tests was 2 minutes.

The testing procedures were administered to all of the subjects by the same investigator, who provided standard verbal encouragement (to push as hard and as fast as possible) in addition to the visual feedback from the Biodex computer screen to facilitate maximal effort contractions [3,18,19].

2.4. *Outcome measures*

Variable definitions are presented in Table 2. For the isokinetic test, the primary outcome measures were the highest peak torque isokinetic concentric (Peak-Torque$_{\mathrm{CON}}$) and Work. The secondary outcome measures were angle of highest peak torque (Angle-PT) and mean angle of peak toque (Mean Angle-PT). For the isometric test, the primary outcome measures were

Table 2
Variable definitions

Variables (units)	Description
Isokinetic test	
Peak-Torque$_{\mathrm{CON}}$ (N·m)	The highest peak torque recorded from the five isokinetic concentric repetitions at 90°/s for each session.
Work (J)	The product of peak torque by the angular distance in each repetition. We considered the greatest mechanical work recorded from the five isokinetic repetitions, in each isokinetic test.
Angle-PT (°)	The angle of knee flexion at which the highest peak torque occurred, in each isokinetic test.
Mean Angle-PT (°)	The mean of the angles of knee flexion at which the highest peak torque occurred from the five repetitions, in each isokinetic test.
Isometric test	
Peak-Torque$_{\mathrm{ISOM}}$ (N·m)	The highest peak torque recorded from the five isometric repetitions, in each session.
RTD$_{0-100\ \mathrm{ms}}$ (N·m/s)	The greatest torque increase in 100 ms recorded from the five isometric repetitions, in each isometric test.

the highest peak torque isometric (Peak-Torque$_{\mathrm{ISOM}}$) and the secondary outcome measure was the rate of torque development in 100 ms (RTD$_{0-100\ \mathrm{ms}}$).

2.5. *Statistical analyses*

All of the variables, including differences in test scores (test 2 – test 1) were tested for normality using the Shapiro-Wilk's test. For non-normally distributed variables, we used nonparametric tests. The means and standard deviations, as well as the medians and ranges for each variable, were calculated on days 1 and 2 to describe the characteristics of the test scores of the participants. Paired t-tests or Wilcoxon tests, depending on the data distribution, were used to test for significant differences between the test and retest scores. Independent-samples t-test was used to test the mean differences between test 1 and test 2 by sex. Effect size (Cohen's d) was calculated to examine the practical significance of the differences between tests when there were significant differences in the t-test. Cohen suggested ES of 0.2, 0.5 and 0.8 as small, moderate and large effects [20].

Heteroscedasticity was determined by a significant positive correlation between the absolute differences (scores of test 2 minus test 1) and the individual mean of test scores in the two testing sessions [21].

To provide a visual analysis of the differences between test 1 and test 2 measurements, Bland-Altman plots with limits of agreement (LOA) were generated [22]. In these graphs, the difference between test sessions (test 2 minus test 1) is plotted against the mean of the two test sessions for each subject (Fig. 1).

Several statistical methods were used to determine the test-retest relative and absolute reproducibility [23]. The relative index of reliability used was the ICC derived from the analysis of variance for a two-way mixed-effects model with a 95% confidence interval (CI) was used to evaluate the relative reliability of the measures between days. The type of ICC used was ICC$_{(C,k)}$, specifically, ICC$_{(C,1)}$ for single measurements (e.g., Peak-Torque, Angle-PT) and ICC$_{(C,5)}$ for Mean Angle-PT because it was based on the mean of 5 repetitions (the subscript indicates the following: C, consistency; and k, the number of repetitions included in the mean) [24]. As a general rule, we considered an ICC over 0.90 as high, between 0.80 and 0.0.89 as moderate and below 0.80 as questionable [25].

The absolute test-retest reliability was evaluated using the following measures: the change in the mean between the test sessions ($\overline{d}$), together with the 95% CI for $\overline{d}$ and the SEM, that is, 1 standard deviation of the distribution of error associated with a single test score [4]. It was calculated as the square root of the mean square error term from the ANOVA [1,26]. The SEM was used to calculate the SRD. We define the SRD as the 95% CI of error associated with repeated measurements and was determined by the formula: SRD $= 2.05 \times\sqrt{2}\times$ SEM; where, $\sqrt{2}\times$ SEM is the error associated with the repeated measurements (because there is error associated with both the first and second measurement) and 2.05 value is the t statistic associated with a 95% CI and 27 degrees of freedom ($t_{0.025,27}$). Only differences between two measurements that exceed the SRD represent a real (non-error) change in scores for a subject. According to Hopkins [6], when there are two levels of trials, the SEM can alternatively be calculated as the SD of the difference scores (SD$_{\mathrm{diff}}$) divided by $\sqrt{2}$ (SEM $=$ SD$_{\mathrm{diff}}/\sqrt{2}$) and thus the SRD as the 95% CI derived from SD$_{\mathrm{diff}}$ (SRD $= t_{\alpha/2,df}\times$ SD$_{\mathrm{diff}}$). An "error band" around the mean difference of the two measurements ($\overline{d}\pm$ SRD) provides the LOA for 95% CI [27].

The SEM and the SRD are expressed in the actual units of measurement. Both values may be expressed in relative terms, as a percentage relative to the mean for all observations [28] (SEM% $=$ (SEM/mean) $\times$ 100; SRD% $=$ (SRD/mean) $\times$ 100, where the mean is the

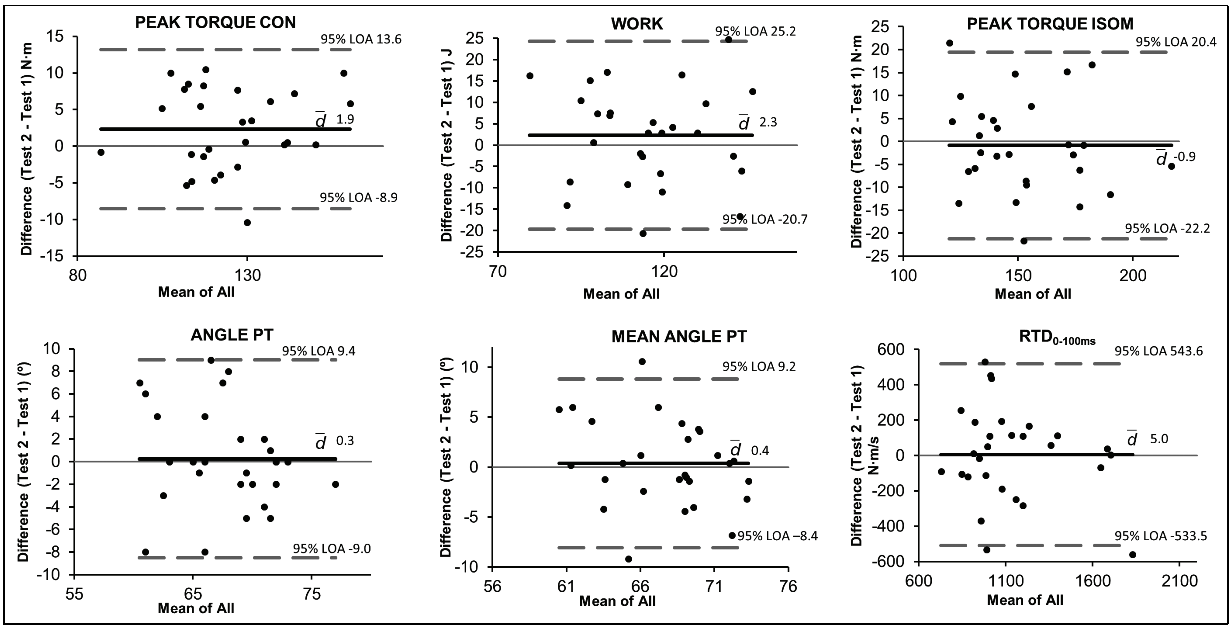

Fig. 1. *Bland-Altman plots*. The difference between the two measurements per subject is plotted against the mean of the two measurements. We can appreciate the lines of the upper and lower limits of agreement (LOA) for 95% CI, the line of the mean of the difference (bias) and the line of perfect agreement (line crossing zero). CON, isokinetic concentric; PT, peak torque; ISOM, isometric; RTD, rate of torque development.

mean of all of the data from both test occasions). Both SEM% and SRD% are independent of the units of measurements.

We conducted the statistical analyses using SPSS for Windows, v. 19.0 (SPSS Inc., Chicago, IL, USA). For Bland-Altman plots, we used Analyse-it for Microsoft Excel, v. 3.0 (Analyse-it Software, Ltd., Leeds, UK). An alpha level of 0.05 was set for determining the significance of all of the test results.

3. Results

The mean values and SDs from tests 1 and 2 for the isokinetic and isometric strength variables are presented in Table 3. For women, there were significant differences for Peak-Torque$_{\mathrm{CON}}$ between test 1 and test 2 ($t = -2.23, p = 0.03$,) with systematically higher scores for the second test. However, ES of this difference was small (Cohen's $d = 0.14$).

There were significant differences between men and women in all the variables analyzed ($p < 0.0001$) except for the Angle-PT, Mean Angle-PT and RTD$_{0-100\ \mathrm{ms}}$ ($p > 0.05$).

There were no positive relationships between the absolute magnitude of the variable and the associated error in any of the variables tested, therefore all distributions were homoscedastic. The Pearson's correlation coefficients for these relationships are shown in Table 3.

Table 4 reports the relative and absolute reliability data for all of the variables for women. All peak torque and work measures had an ICC$_{(C,1)}$ greater than 0.8 (from 0.82 for Work to 0.95 for Peak-Torque$_{\mathrm{CON}}$). However, ICC$_{(C,5)}$ for Mean Angle-PT and ICC$_{(C,1)}$ for Angle-PT and RTD$_{0-100\ \mathrm{ms}}$ were lower than 0.67. The reliability for Mean Angle-PT appeared greater than the reliability of the Angle-PT (ICC, 0.67 and 0.50, respectively).

The SEM for all variables, except for RTD, were low, with SEM% less than 7.5%. The isometric RTD$_{0-100\ \mathrm{ms}}$ demonstrated the highest levels of variability (SEM $=$ 185.8 N·m/s, SEM% $=$ 16.4%). The Peak-Torque$_{\mathrm{CON}}$ showed lower error in absolute and percentage value than the Peak-Torque$_{\mathrm{ISOM}}$ (SEM, 3.9 N·m, SEM% 3.1% and SEM, 7.3 N·m, SEM% 4.8%, respectively).

The minimum change that can be considered real change was 11.3 N·m for Peak-Torque$_{\mathrm{CON}}$, 21.3 N.m for Peak-Torque$_{\mathrm{ISOM}}$ and around 9° for Angle-PT and Mean-Angle PT. The 'random error bands' that represent the smallest detectable change for a single subject (95% CI for SRD $=$ LOA) are shown in Bland-Altman plots (Fig. 1).

Table 3
Means and standard deviations for isometric and isokinetic tests

Variables	Women ($n = 28$)					Men ($n = 12$)				
	Test 1	Test 2	p (t test)	$\bar{d}$	r (p)	Test 1	Test 2	p (t test)	$\bar{d}$	r (p)
Isokinetic results										
Peak-Torque$_{\mathrm{CON}}$ (N·m)	123.7 (16.5)	126.1 (17.1)	0.03	2.3 (5.5)	−0.05 (0.98)	205.3 (26.9)†	209.7 (31.7)	0.21	4.4 (12.8)	0.17 (0.58)
WORK (J)	114.0 (18.9)	116.3 (18.4)	0.29	2.3 (11.2)	−0.05 (0.98)	180.7 (25.4)	184.0 (24.9)	0.38	3.3 (13.3)	−0.01 (0.97)
Angle-PT (°)	67.7 (5.1)	67.9 (4.3)	0.82	0.2 (4.5)	−0.37 (0.06)	68.0 (4.2)	69.8 (2.9)	0.12	1.8 (4.1)	−0.39 (0.21)
Mean angle-PT (°)	67.5 (4.8)	67.9 (3.8)†	0.65	0.4 (4.3)	−0.24 (0.22)	67.8 (4.7)	69.7 (2.7)	0.11	2.0 (4.0)	−0.45 (0.14)
Isometric results										
Peak-Torque$_{\mathrm{ISOM}}$ (N·m)	152.9 (25.5)	152.0 (23.8)	0.66	−0.9 (10.4)	0.004 (0.98)	222.9 (25.7)	224.0 (35)	1.0	1.1 (13.7)	0.01 (0.97)
RTD$_{0-100\ \mathrm{ms}}$ (N·m/s)	1131.1 (338.2)†	1136.1 (288.0)	0.92	5.0 (262.7)	−0.05 (0.81)	1328.5 (282.1)	1221.4 (304.2)	0.14	−107.1 (252.8)	−0.29 (0.37)

CON, isokinetic concentric; PT, peak torque; ISOM, isometric; RTD$_{0-100\ \mathrm{ms}}$, rate of torque development from 0 to 100 ms; data Test 1 and data Test 2 are mean (standard deviations) of raw values; $\bar{d}$, mean of the difference (test 2 – test 1) and standard deviations of the mean of the differences in brackets; r, Pearson's correlation coefficient between the absolute difference and the individual mean; p, significant level. If $r > 0$ and $p < 0.05$, data are heteroscedastic. In this sample data were homoscedastic; †Not normally distributed ($p < 0.05$), using Shapiro-Wilk test.

Table 4
Relative and absolute reliability scores for isokinetic and isometric tests in women ($n = 28$)

	Relative reliability	Absolute reliability					
	ICC[a] (95% CI)	$\bar{d}$ (95% CI) units	SEM units	SEM%	SRD (95% CI) units	SRD%	LOA (95% CI)
Peak-Torque$_{\mathrm{CON}}$(N·m)	0.95 (0.89 to 0.97)	2.3 (0.19 to 4.4)*	3.9	3.1	11.3	9.0	2.3 ± 11.3 (−8.9 to 13.6)
Work (J)	0.82 (0.65 to 0.91)	2.3 (−2.1 to 6.6)	7.9	7.4	23.0	19.9	2.3 ± 23.0 (−20.7 to 25.2)
Peak-Torque$_{\mathrm{ISOM}}$(N·m)	0.91 (0.82 to 0.96)	−0.9 (−4.9 to 3.1)	7.3	4.8	21.3	14.0	−0.9 ± 21.3 (−22.2 to 20.4)
Angle-PT (°)	0.56 (0.25 to 0.77)	0.3 (−1.5 to 1.9)	3.2	4.7	9.2	13.6	0.3 ± 9.2 (−9.0 to 9.4)
Mean angle-PT (°)	0.67 (0.28 to 0.85)	0.4 (−1.3 to 2.0)	3.1	4.5	8.8	13.0	0.4 ± 8.8 (−8.4 to 9.2)
RTD$_{0-100\ \mathrm{ms}}$(N·m/s)	0.65 (0.37 to 0.82)	5.0 (−96.8 to 106.9)	185.8	16.4	538.5	47.5	5.0 ± 5.38 (−533.5 to 543.6)

ICC, intraclass correlation coefficient; [a]ICC$_{(C,1)}$ for all peak torque, Work, Angle-PT and RTD; ICC$_{(C,5)}$ for Mean Angle-PT (average of five repetitions). CI, confidence interval; $\bar{d}$, systematic bias, mean of the difference between the two tests (test 2 minus test 1); SEM, standard error of measurement; SEM%, coefficient of variation of standard error of measurement; SRD, smallest real difference; SRD%, coefficient of variation of smallest real difference; LOA, Limits of agreement; *significant difference between the 2 sessions (95% CI does not include zero).

4. Discussion

Our findings demonstrate that the dynamometer Biodex System 3-Pro and the protocol used for the measurement of knee extensors strength showed good relative and absolute reliability in the main variables studied (Peak-Torque$_{\mathrm{CON}}$, Work, and Peak-Torque$_{\mathrm{ISOM}}$) in healthy young women. The ICC$_{(C,1)}$ values were greater than 0.8, and the SEM% were less than 7.5% (Table 4). We also found poor relative reliability values for the Angle-PT (ICC$_{(C,1)}$ = 0.56) and mean Angle PT (ICC$_{(C,5)}$ = 0.67). However, the absolute reliability values of these variables were acceptable, with an SEM of approximately 3°. The parameter that showed the greatest variability was the RTD$_{0-100\ \mathrm{ms}}$.

As men are on average 40% stronger than women, and in strength variables, such as peak torque, the magnitude of error usually increases when the variable value does so, data analysis was separately conducted by gender. The pooled group analysis would have given underestimated reliability values in men and overestimated in women, compared to those obtained in a separate analysis. As an example with our data for the Peak-Torque$_{\mathrm{CON}}$, SEM in the women's group was 3.9 N·m, 9 N·m in the men's group and 5.8 N·m in the pooled group. Thus, the reliability analysis was performed only on the women group because the sample size of men (n = 12) was considered insufficient for to obtain acceptable conclusions by statistical insensitivity [3].

4.1. Relative reliability

Relative reliability was quantified using ICC. The ICC represents the degree to which the individuals preserve their position in the group relative to others in a sample with repeated measures [5,6]. It is based on the ratio between the true score variance (between subjects variability) and the total variance (between subjects variability plus error variance) [26] The magnitude of the ICC is dependent on the variability in the data as well as the true score variance. Thus, a larger variation between the individuals' scores will result in a higher ICC, even if the variation within individuals is the same [29].

The ICC model used in the current study (ICC$_{C,k}$) is the same as the model ICC$_{(3,K)}$ used by Shrout and Fleiss [29]. These models only consider random error [26]. A value of 1 represents perfect agreement, whereas a value of 0 represents no agreement. In the current study, the ICC$_{(C,1)}$ values ranging from 0.82 for Work to 0.91 and 0.95 for Peak-Torque$_{\mathrm{ISOM}}$ and Peak-Torque$_{\mathrm{CON}}$, respectively (Table 4), demonstrated good relative reliability, which indicated that these measures generally exhibited consistency for repeated measurements. The comparisons between the ICCs found in this study and the ICCs obtained in other studies may be confounded by differences in the raw sample scores. The need for caution in when comparing ICC values has been widely documented [5,6,26]. Harding [7] demonstrated high reliability for Peak-Torque$_{\mathrm{CON}}$ (ICC, 0.98) in a population and protocol similar to ours (young women, mean age 21.6 (1.6) years; weight 55.6 (6.6) Kg; mean of Peak-Torque$_{\mathrm{CON}}$ 135.18 (24.86) N·m), but their sample size was smaller than ours (n = 14). ICC for Peak-torque ISOM in the literature ranged from 0.86 to 0.97 and were comparable to those obtained in this study [19,30].

The ICC$_{(C,k)}$ values for the Angle-PT, the Mean Angle-PT and the RTD$_{0-100\ \mathrm{ms}}$ were below 0.67, and the lower CI limits were as low as 0.25, 0.28 and 0.37, respectively. Measuring these parameters from single or the mean of several repetitions has been reported to give low relative reliability [7,13,31,32]. Therefore, alternative approaches have been proposed in the literature, for example, polynomial fitting [14] from torque-angle curves or averaging the torque values at given angles [13]. The main reason why the ICC$_{(C,k)}$ is lower for Angle-PT and Mean Angle-PT is that the range of the data is much narrower than in the other variables (see Fig. 1 for the range). It is known that the ICC can provide misleading results if the sample is homogeneous, giving low values for ICC even if the differences between subjects' scores across test conditions are small [26,29]. Rogosa et al. [33] explained that "the difference score can be an accurate and useful measure of individual change even in situations where the reliability is low" (p. 341). Accordingly, even with low ICCs, if the SRD is acceptable and if typical changes in scores exceed the SRD, the test scores may still have useful clinical applicability. Whether the observed errors are acceptable depends on the intended application of the measurements, in particular whether subjects might be expected to demonstrate changes in scores across time that typically exceed the observed SRDs.

4.2. Absolute indices of reliability

Absolute indicators of reliability describe the degree to which an individual's observed score will vary with repeated measurements [6]. They are expressed in units of measurement, so they are easy to understand. Any observed test score from a subject is composed of two elements: the true score and a margin of error. Any observed score contains errors. The smaller the error, the more reliable the measure. To find the magnitude of the true score, we need to determine the magnitude of the error.

If there has been no intervention between the test and the retest, the change between test 1 and test 2 scores should be considered the result of measurement error and is attributed to random and systematic error [6,19]. In the current study, when analyzing changes in the mean ($\overline{d}$, Table 4), the positive values for strength parameters indicate a tendency to a better performance in the second test. Thus, with a single familiarization session, the learning effects were small, in spite of the significant differences for Peak-Torque$_{\mathrm{CON}}$. In order to assess the practical significance of this difference, we used an effect size index indicating a very small magnitude of the observed effect (Cohen's $d = 0.14$). In a clinical context, it is most likely impractical to perform more than a single familiarization session.

The SEM refers to the random variation in a participant's score associated with a single measurement (within-subject variability) [6]. In our study, to reduce the SEM, protocol-related errors were kept to a minimum by following standardized procedures (only one experienced examiner, warm-up, stabilization-seating, and alignment of the dynamometer and lever). For a given group of subjects, the SEM is linked to the mean of the measurements, and, in the strength variables, appears to increase when changes in the mean scores increase. Thus, information about the mean of the measurements from which SEMs are determined is required for comparison of SEMs associated with different test procedures [4]. SEM% is useful for comparing reliability among different measures and across different studies [6], and also when the error that a subject demonstrates in repeated test scores is related to the magnitude of their score (heteroscedasticity), because if the error is positively correlated with score magnitude, the application of SEMs obtained from the low-scoring subjects to the high-scoring subjects would be erroneous. If the errors are not correlated to individual mean scores (homoscedasticity), an absolute error that makes no assumptions based on score magnitude would then be preferable [4,21].

For Isokinetic knee extension strength tested at 90°/s and 0°/s, Symons [19] reported SEM% values of 9.2% for Work versus to 7.4% for Peak-Torque$_{\mathrm{CON}}$ and 8.2% for Peak-Torque$_{\mathrm{ISOM}}$ in older women. These percentages of error are greater than those reported in the present study in these variables, and were assessed in samples with lower mean scores than the ours. These findings suggests that the reliability of the strength tests in the knee extensors is greater in young than in older women. Our SEM value for the Peak-Torque$_{\mathrm{CON}}$ was lower than that reported by Harding [7] in young women (3.9 N·m versus 5.5 N.m). However, for the Peak-Torque$_{\mathrm{ISOM},}$ Molczyk [30] reported higher SEM values than those found in the present study (SEM, 3.4 N·m versus 7.3 N·m) with their isometric peak torque mean values slightly lower than ours (Peak-Torque$_{\mathrm{ISOM}}$, 145 N·m versus 152 N·m).

The SEM can be used to define the amount by which a person's score needs to change, to be sure that the change in the measure can be considered real and not just due to the noise in the measurement (random error). The SRD provides the theoretical lower boundary of detectable change. The SRD around the mean difference provides the limits of agreement (LOA). The LOA include the systematic and random error and they represent the 95% probability range for the difference between a subject's measurements from these two test occasions [22]. After an intervention, if the difference in a score is outside this reference range, it could be considered to represent a real change.

Based on the SRD in this study (Table 4), we are 95% confident that a change higher than 11.3 N·m for Peak-Torque$_{\mathrm{CON}}$ of the knee extensors is necessary to indicate that there has been a real change in performance. The changes in this variable reported in the literature after resistance training programs are typically well above this value [34–36]; For Peak-Torque$_{\mathrm{ISOM}}$ and for work a real improved performance would require an increase of at least 21.3 N·m or 23 J at 95% CI, respectively. As for Peak-Torque$_{\mathrm{CON}}$, these values are well below the changes expected after an intervention; thus, the results of SRD provided in the present study indicate clinical applications for the measurements obtained under our test procedures.

For both Angle-PT and Mean Angle-PT the SEM was approximately 3° within an actual range of approximately 20° (from 57° to 78°). These values show fairly homogeneous (low variability) and good absolute indexes of reliability in these two variables, in contrast to the lower ICC$_{(C,k)}$ values obtained (discussed above). With such a restricted range, even a very small within subject variability in measures of Angle-PT or Mean Angle-PT across sessions can substantially decrease the ICC values obtained. Harding [7] found

SEM values similar to those reported here (3.3° for Angle-PT at 60°/s) in young women. On the other hand, the 95% CI for SRD was 9.2° and 8.8° for Angle PT and Mean-Angle PT. Thus, e.g. if the Angle-PT was 68° in the first test, it is likely that the second score might be as low as 58.8° (68° − 9.2°) or as high as 77.2° (68°+ 9.2°) With these degrees of variation, the use of this protocol to assess Angle-PT only would make sense for interventions that lead to changes of at least 9° in this parameter. The protocol, which is utilized for the Angle-PT analysis, has been used in previous research [7,31,37] and was proposed by the manufacturer. Alternative approaches, such as the use of polynomials to analyze several repetitions [12], or the assessment of the fluctuations in the torque angle relationship in discrete intervals [13], have been proposed. Further research is warranted to confirm the reliability of these approaches.

The greatest variability was found in the isometric $RTD_{0-100\ ms}$; the SEM was 185.8 N·m (SEM% = 16.4%), which is not surprising because it is a parameter that usually shows greater variability than the dynamic and isometric peak torque [38,39]. The SRD scores obtained in our study suggest that intervention studies performed with these protocols, in populations similar to ours, would need to find large differences (> 538 N· m/s, 47.5% related to the mean; SRD and %SRD, respectively) to be 95% confident that a real change has occurred. Although in published results for RTD measures large differences in the isometric knee extension test after strength training were reported, with a 42.6% mean improvement [40], the wide range of error obtained in this study for RTD leads us to question its usefulness in clinical practice. In measurements with such large variability, the 95% confidence limits for the SEM might be too stringent for a decision limit. We could calculate narrower confidence intervals (e.g., the %SRD might be 39.4% or 30.1% for a 90% or 80% CI, respectively).

The assessment of these SRD values is important to clinicians, so that they can decide with 95% confidence whether the intervention has produced a real change in a patient's score and to enable evidence-based clinical decisions for adapting interventions to improve the quality of a rehabilitation or training program.

4.3. Study limitations

These results can only be applied to untrained healthy young subjects or to the uninjured leg of a clinical patient. For example, the variability is likely to be greater for older subjects [41] or those with musculoskeletal [3] or neurological [42] injuries compared with the subjects of this study.

Comparing the results of our study with the published data will only be appropriate when similar procedures are used [3]. Therefore, the comparisons made were limited, given the difficulty of finding comparable studies to ours in terms of the study population, measurement protocols, type of dynamometer used, contraction velocity used, and measurement reliability calculations applied.

5. Conclusions

The current study shows a standardized and reproducible description of a protocol that can be implemented for the isokinetic assessment of knee extensor strength at 90°/s. The Peak-Torque_{CON} and $\text{Peak-Torque}_{ISOM}$ for knee extensors have high test-retest reliability and the SRD values provided have fair clinical applicability. On the other hand, the RTD_{0-100} reliability is of questionable utility and the SRD values indicate no clinical applicability. Relative and absolute reliability were slightly greater for Mean Angle-PT than for Angle-PT. The SRD values associated with these two variables have clinical application only when large changes are expected. For improved clinical utility of the Angle-PT measure, given its variability when traditional analyzing procedures are applied, it would be necessary to devise and validate strategies to reduce the unpredictable error in these measurements. The Angle-PT, Mean Angle-PT and $RTD_{0-100\ ms}$ variables do not appear to be affected by gender.

References

[1] Stratford PW, Goldsmith CH. Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. Phys Ther. 1997;77(7):745-50.

[2] Beckerman H, Roebroeck ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek AL. Smallest real difference, a link between reproducibility and responsiveness. Qual Life Res. 2001;10(7):571-8.

[3] Keating JL, Matyas TA. The influence of subject and test design on dynamometric measurements of extremity muscles. Phys Ther. 1996;76(8):866-89.

[4] Keating JL, Matyas TA. Unpredictable error in dynamometry measurements: a quantitative analysis of the literature. Isokinetics and Exercise Science 7. 1998;7:107-21.

[5] Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med. 1998;26(4):217-38.

[6] Hopkins WG. Measures of reliability in sports medicine and science. Sports Med. 2000;30(1):1-15.
[7] Harding B, Black T, Bruulsema A, Maxwell B, Stratford PW. Reliability of a reciprocal test protocol performed on the kinetic communicator: an isokinetic test of knee extensor and flexor strength. J Orthop Sports Phys Ther. 1988;10(6):218-23.
[8] Montgomery LC, Douglass LW, Deuster PA. Reliability of an isokinetic test of muscle strength and endurance. J Orthop Sports Phys Ther. 1989;10(8):315-22.
[9] Tredinnick TJ, Duncan PW. Reliability of measurements of concentric and eccentric isokinetic loading. Phys Ther. 1988;68:656-9.
[10] Webber SC, Porter MM. Reliability of ankle isometric, isotonic, and isokinetic strength and power testing in older women. Phys Ther. 2010;90(8):1165-75.
[11] Blazevich AJ, Cannavan D, Horne S, Coleman DR, Aagaard P. Changes in muscle force-length properties affect the early rise of force in vivo. Muscle Nerve. 2009;39(4):512-20.
[12] Brughelli M, Cronin J. Altering the length-tension relationship with eccentric exercise: implications for performance and injury. Sports Med. 2007;37(9):807-26.
[13] Blazevich AJ, Cannavan D, Coleman DR, Horne S. Influence of concentric and eccentric resistance training on architectural adaptation in human quadriceps muscles. J Appl Physiol. 2007;103(5):1565-75.
[14] Brughelli M, Nosaka K, Cronin J. Application of eccentric exercise on an Australian Rules football player with recurrent hamstring injuries. Phys Ther Sport. 2009;10(2):75-80.
[15] Brockett CL, Morgan DL, Proske U. Human hamstring muscles adapt to eccentric exercise by changing optimum length. Med Sci Sports Exerc. 2001;33(5):783-90.
[16] Drouin JM, Valovich-mcLeod TC, Shultz SJ, Gansneder BM, Perrin DH. Reliability and validity of the Biodex system 3 pro isokinetic dynamometer velocity, torque and position measurements. Eur J Appl Physiol. 2004;91(1):22-9.
[17] Thorstensson A, Grimby G, Karlsson J. Force-velocity relations and fiber composition in human knee extensor muscles. J Appl Physiol. 1976;40(1):12-6.
[18] O'Sullivan A, O'Sullivan K. The effect of combined visual feedback and verbal encouragement on isokinetic concentric performance in healthy females. Isokinetics and Exercise Science. 2008;16:47-53.
[19] Symons TB, Vandervoort AA, Rice CL, Overend TJ, Marsh GD. Reliability of isokinetic and isometric knee-extensor force in older women. J Aging Phys Act. 2004;12(4):525-37.
[20] Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale (NJ): Lawrence Erlbaum Associates. 1988.
[21] Dvir Z. How much is necessary to indicate a real improvement in muscle function? A review of modern methods of reproducibility analysis. Isokinetics and Exercise Science. 2003;11:49-52.
[22] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-10.
[23] Holmbäck AM, Lexell JE. Reproducibility of isokinetic ankle dorsiflexor strength and fatigue measurements in healthy older subjects. Isokinetics and Exercise Science. 2007;15:263-70.
[24] McGraw KO, Wong SP. Forming Inferences About Some Intraclass Correlation Coefficients. Psychological Methods. 1996;1(1):30-46.
[25] Vincent WJ. Statistics in Kinesiology. 3rd Ed. Human Kinetics, Champaign, IL. 2005:196 pp.
[26] Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005;19(1):231-40.
[27] Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8(2):135-60.
[28] Lexell JE, Downham DY. How to assess the reliability of measurements in rehabilitation. Am J Phys Med Rehabil. 2005;84(9):719-23.
[29] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin. 1979;86(2):420-8.
[30] Molczyk L, Thigpen LK, Eickhoff J, Goldgar D, Gallagher JC. Reliability of Testing the Knee Extensors and Flexors in Healthy Adult Women Using a Cybex II Isokinetic Dynamometer. J Orthop Sports Phys Ther. 1991;14(1):37-41.
[31] Phillips BA, Lo SK, Mastaglia FL. Isokinetic and isometric torque values using a Kin-Com dynamometer in normal subjects aged 20 to 69 years. Isokinetics and Exercise Science. 2000;8:147-59.
[32] Caruso JF, Brown LE, Tufano JJ. The reproducibility of isokinetic dynamometry data. Isokinetics and Exercise Science. 2012;20:239-53.
[33] Rogosa DR, Willett JB. Demonstrating the reliability of the difference score in the measurement of change. Journal of Educational Measurement. 1983;20(4):335-43.
[34] Blazevich AJ, Gill ND, Deans N, Zhou S. Lack of human muscle architectural adaptation after short-term strength training. Muscle Nerve. 2007;35(1):78-86.
[35] Suetta C, Andersen JL, Dalgas U, Berget J, Koskinen S, Aagaard P, et al. Resistance training induces qualitative changes in muscle morphology, muscle architecture, and muscle function in elderly postoperative patients. J Appl Physiol. 2008;105(1):180-6.
[36] Suetta C, Hvid LG, Justesen L, Christensen U, Neergaard K, Simonsen L, et al. Effects of aging on human skeletal muscle after immobilization and retraining. J Appl Physiol. 2009;107(4):1172-80.
[37] Wilhite MR, Cohen ER, Wilhite SC. 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. 1992;15(4):175-82.
[38] de Ruiter CJ, Kooistra RD, Paalman MI, de Haan A. Initial phase of maximal voluntary and electrically stimulated knee extension torque development at different knee angles. J Appl Physiol. 2004;97(5):1693-701.
[39] Sleivert GG, Wenger HA. Reliability of measuring isometric and isokinetic peak torque, rate of torque development, integrated electromyography, and tibial nerve conduction velocity. Arch Phys Med Rehabil. 1994;75(12):1315-21.
[40] Blazevich A. Are training velocity and movement pattern important determinants of muscular rate of force development enhancement? Eur J Appl Physiol. 2012;112(10):3689-91.
[41] Ordway NR, Hand N, Briggs G, Ploutz-Snyder LL. Reliability of knee and ankle strength measures in an older adult population. J Strength Cond Res. 2006;20(1):82-7.
[42] Flansbjer UB, Holmback AM, Downham D, Lexell J. What change in isokinetic knee muscle strength can be detected in men and women with hemiparesis after stroke? Clin Rehabil. 2005;19(5):514-22.