J Rehabil Med 2002; 34: 91–95

TEST-RETEST RELIABILITY IN ISOKINETIC MUSCLE STRENGTH MEASUREMENTS OF THE SHOULDER

Jetty van Meeteren,[1] Marij E. Roebroeck[2] and Henk J. Stam[1,2]

From the [1]Department of Rehabilitation Medicine, University Hospital Rotterdam, and [2]Institute of Rehabilitation Medicine, Erasmus University Rotterdam, The Netherlands

Test-retest reliability is important for long-term follow-up; however, data on the reliability of isokinetic dynamometry of the shoulder are scarce. Twenty subjects (50% male) were measured; 10 with asymmetrical use of the arms (mean age 27 years) and 10 used their arms symmetrically (mean age 32 years). A Biodex® dynamometer (Multi joint system 2) was used. Abduction/adduction and external/internal rotation were measured following a standardized protocol. Performed scheme: two sessions with a 2-week interval, all measurements were done with 60°/second (5 repetitions) and respectively 120°/second and 180°/second (10 repetitions). Differences in the mean peak torques, split for muscle group and gender were significant. Intraclass correlation coefficients ranged from 0.69 to 0.92. This implies good to excellent reliability in research on groups. To determine test-retest reliability of two consecutive individual measurements smallest detectable differences (SDD) were computed and ranged from 21% to 43%. It is questionable whether the SDDs are small enough to detect real changes in muscle strength.

Key words: isokinetic dynamometry, shoulder, reliability.

J Rehabil Med 2002; 34: 91–95

Correspondence address: Jetty van Meeteren, Department of Rehabilitation Medicine, University Hospital Rotterdam, PO Box 2040, NL-3000 CA Rotterdam, The Netherlands. E-mail: vanmeeteren@revd.azr.nl

Accepted October 22, 2001

INTRODUCTION

Isokinetic dynamometry of the shoulder has been increasingly used in clinical practice since 1980 (1). The technique can be used to evaluate both the function of a joint and the effectiveness of therapy, because objective parameters (e.g. muscle strength and range of motion) can be measured. Much research on isokinetic dynamometry has been focused on the lower extremity, especially the knee joint. Isokinetic dynamometry of the knee is known to be reliable with a good instruction, standardization of the test procedure and familiarization with the equipment (1, 2). Although several studies on isokinetic dynamometry of the shoulder are available, these do not address test-retest reliability of the measurements (2–5). We found only one study on this topic, addressing both intra- and inter-day variability of isokinetic shoulder abduction and adduction measurements (6). Their results indicated excellent reliability between days, with Pearson's correlation coefficients varying between 0.87–0.97 for abduction and 0.95–0.99 for adduction measurements.

An important prerequisite for correct interpretation of measurement results is that they are reliable. Reliability depends on consistency of measurement results and, thus, to a relative absence of measurement errors. Reliability over two sessions, i.e. test-retest reliability, is necessary, because long-term follow-up is of clinical importance. When the test-retest reliability is good, unilateral comparison over a period of time is possible.

Several factors can influence the reliability of isokinetic measurements of the shoulder joint. First, in isokinetic dynamometry the axis of the dynamometer has to be lined out to the axis of the joint. The axis of the dynamometer has a fixed position, but there is no consensus about the localization of the functional joint axis of the shoulder. The glenohumeral joint has an extensive range of motion in several planes and the axis of the glenohumeral joint moves about 8 cm in flexion/extension and abduction/adduction movements (5). The influence of this phenomenon on the reliability of the measurement results is unknown.

Second, isokinetic measurements of the shoulder joint are done in several different positions, e.g. sitting, standing, lying supine and with different angles of abduction and flexion of the shoulder (7, 8); however, it is not known which is the best or the most reliable position. The rotation movement is often done with 90° abduction or 90° flexion of the shoulder but with this method the subacromial structures are very vulnerable. Soderberg & Blaschak (7) studied 6 different positions for the rotational movement and reported significant differences in the maximal torque measured. The highest peak torque is seen in the neutral position, i.e. sitting with no abduction and flexion of the shoulder. The influence of the used position on reliability of the measurements is not known; most studies have used only one position. There is no consensus about the position to be used in the abduction and adduction movement.

Third, the choice of the preset angular velocity in isokinetic measurements of the shoulder joint is arbitrary rather than scientifically based. Low and high angular velocities are often used; the assumption is that a low angular velocity relates to maximal voluntary contraction and a high angular velocity relates to muscle coordination which is important in functional activities. The motivation for the used angular velocities (often 60°/second, 180°/second and sometimes 300°/second) is not given. For sportsmen using their arms, high angular velocities (>180°/second) are often used.

 ISSN 1650–1977

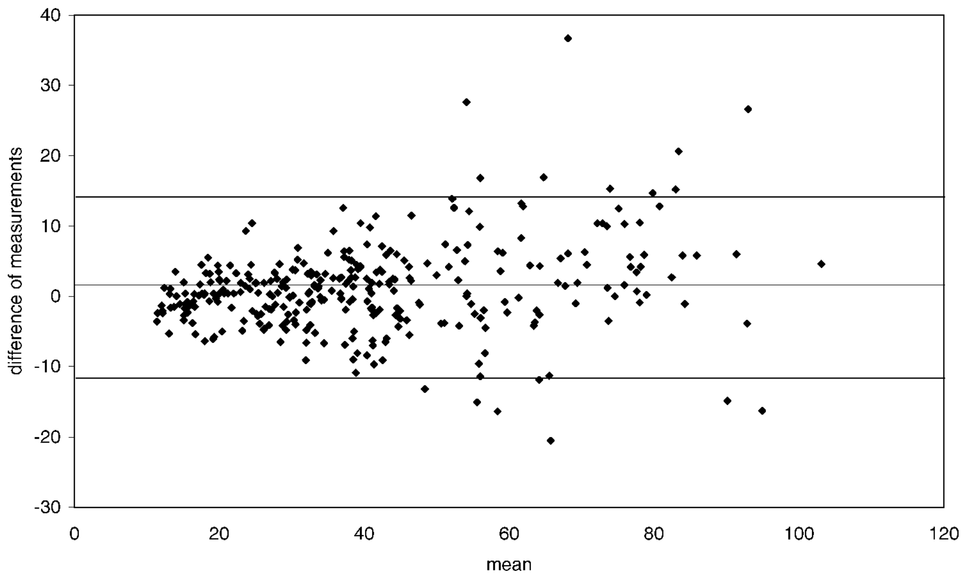

Fig. 1. Scatter plot of the mean maximal peak torque of the two sessions versus the difference between the maximal peak torques. Reference lines of the mean of the differences and ±2 S.D.

The aim of the present study was to determine the test-retest reliability of torques measurements of the shoulder joint in healthy subjects using a Biodex® isokinetic dynamometer, Multi joint system 2.

METHOD

Subjects

In 20 subjects (healthy men and women of varying ages without pain and diseases) isokinetic measurements of the shoulders were done with a Biodex®, Multi joint system 2. Ten of the subjects were active sportspersons with asymmetrical use of the arms (referred to here as the asymmetrical group) and 10 were either not sportspersons or sportspersons with symmetrical use of the arms (referred to as the symmetrical group). In both groups 50% was male. In the asymmetrical group 8 subjects were right-handed and 2 had no clear dominance; their mean age was 27 (SD 9.6) (range 22–54) years, mean body mass 70 (range 61–92) kg and mean height 177 (range 163–187) cm. In the symmetrical group 8 subjects were right-handed and 2 left-handed; their mean age was 32 (SD 12.7) (range 21–57) years, mean body mass 74 (range 62–97) kg and mean height 178 (range 152–196) cm. In the symmetrical group two examiners performed the measurements, whereas in the asymmetrical group only one examiner made measurements. All subjects were informed about the study and all gave informed consent. Approval for the study was obtained from the Ethics Committee of the University Hospital Rotterdam and Erasmus University Rotterdam.

Test protocol

All measurements were done according to a standardized protocol: measurements were done in a sitting position, subjects were strapped down with two bands across the chest, one across the pelvis and one across the contralateral leg, a footrest was used, gravity correction was used. The following muscle groups were measured: abductors, adductors, external and internal rotators. Assumptions for the abduction/adduction movement: chair was rotated 45° with the leg of the T-frame, back of the chair 45°, powerhead rotated 30° with the back of the chair, powerhead was overturned 45°, axis of the dynamometer was placed in the middle of the circle of movement during the abduction/adduction movement, this was done visually. Assumptions for the external/internal rotation: chair was rotated 90° with the leg of the T-frame, back of the chair was nearly vertical, powerhead parallel to the chair, powerhead was overturned 67.5°, axis of the dynamometer was placed in the longitudinal axis of the humerus through the olecranon.

Two sessions with a 2-week interval were performed. As reported in other studies (2, 9–11) low and high angular velocities were used: for abduction/adduction 60°/second and 120°/second and for external/internal rotation 60°/second and 180°/second. In a pilot study we found that preset angular velocities higher than 120°/second in abduction/adduction movement and higher than 180°/second in external/internal rotation could not be exceeded by healthy subjects who were all active sportsmen (unpublished data). At the low angular velocity 5 repetitions were made and at the high angular velocity 10 repetitions were made. The maximal peak torque of these repetitions was determined, because this is reported to be the most used parameter in isokinetic dynamometry. Both shoulders were measured; the side where the measurements started was determined by randomization. Preceding the measurements there was a warming-up period in which the movement was done three times sub maximal. The rest period between the two angular velocities was 60 seconds. There was no verbal or visual feedback.

Statistical analysis

The scatter plot of the mean versus the difference of the peak torques of two sessions showed a greater difference at higher mean peak torques values (Fig. 1). Because of the proportional difference a logarithmic transformation (12) of the raw data was done (Fig. 2).

An ANOVA was performed using the transformed data. Analysis of the different subgroups, i.e. split for gender, sport, muscle group, side and angular velocity showed that the differences in error variance for sport, side and angular velocity were negligible. So, in further analysis only the two subgroups muscle group and gender were discerned. A general linear model of ANOVA was used with subject and session as random factors. The estimated variance components were determined: between-subject (var Subject) and within-subject (var Session and var Subject*Session). Analysis was performed using SPSS.

From the estimated variance components the intraclass correlation coefficient (ICC) was determined; the ICC is the ratio of variance of interest (between-subject variance) over variance of interest and error variance (between-subject plus within-subject variance) (13).

The standard error of measurement (S.E.M.) was calculated with the estimated variance components; the S.E.M. is the square root of the error variance (14, 15). From the S.E.M. the 95% confidence interval ($\pm 1.96 \times$ S.E.M.) and the smallest detectable difference (SDD) between two measurements ($1.96 \times \sqrt{2} \times$ S.E.M.) were determined. The latter index is practical in individual muscle strength measurements. Only differences between two measurements that exceed the SDD represent a real (non-error) change in peak moment. Antilogs of the S.E.M., 95% confidence interval and SDD give the proportional indexes of measurement errors (in %).

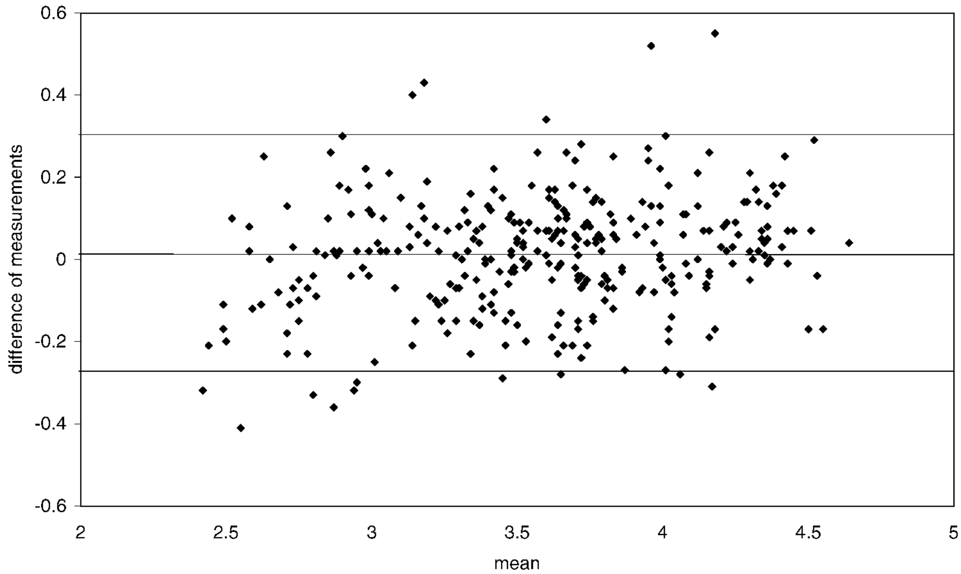

Fig. 2. Scatter plot after logarithmic transformation of the data.

RESULTS

Table I gives the mean peak torque (Nm) and standard deviation (S.D.) of two sessions for each muscle group, split for gender ($n = 10$). Analysis of the non-logarithmic data (ANOVA; F-value) shows that all differences between muscle groups and gender are statistically significant ($p < 0.05$).

Table II gives the estimated variance components and the intraclass correlation coefficient (ICC). The between-subject variation is larger than the within-subject variation. The contribution of the factor session to the within-subject variance is smaller than that of the interaction component, i.e. interaction between subject and session. The ICC ranges from 0.69 to 0.92.

Table III gives the proportional indexes of the measurement error (percentages). In women, the (S.E.M.) ranges from 8% to 14% and in men from 7% to 12%. The SDD in women ranges from 25% (abduction) to 43% (adduction) and in men from 21% (adduction) to 37% (abduction).

DISCUSSION

Few data are available on the test-rest reliability of isokinetic dynamometry of the shoulder joint. Most studies on isokinetic dynamometry have used a Cybex II®. In the present study a Biodex® was used. Compared with the Cybex II®, the Biodex® dynamometer has more possibilities including a greater range of angular velocities, greater maximal torque and the possibility to measure the active range of motion. Comparison of our data with that of other studies (1, 8–10) showed that the peak torques of external and internal rotation are similar: the range of other studies being respectively 9.5–35.3 and 23.1–62.4 Nm (1). In our study peak torques of abduction are somewhat higher and of adduction lower than in other studies which report abduction to range from 26.6 to 56.6 and adduction from 31.0 to 108.5 Nm (1). A possible explanation for this difference is that in our study gravity correction was made. As our measurements were done in a sitting position, muscle strength values may differ from those in studies which made no gravity correction.

Test-retest or inter-session reliability is important for the correct interpretation of measurement results in a clinical setting. Good or excellent test-retest reliability means that measurement results of two different sessions (when no differences in muscle strength are expected) are the same. In order to judge development or progression of a disease or effectiveness of a therapy, long-term follow-up of patients is necessary, thus measurements must be reliable between different sessions.

From the ANOVA it is clear that variance components refer to subject, session and the interaction between subject and session. Inter-session or test-retest reliability is not influenced by the variance component subject (i.e. inter-individual differences in muscle strength), but by the variance components session and the interaction of subject and session. Both the variance com–ponents session and the interaction between session and subject contribute to the error variance, with a relatively large contribution of the interaction component. Interaction of subject and session implies that some subjects achieved larger moments in the second session, whereas others achieved better results in the

Table I. *Data (mean and standard deviation; S.D.) of peak torque for each muscle group for males (M) and females (F) of the two sessions*

	Gender	Mean (S.D.) (Nm)	
		Session 1	Session 2
Abduction	F	36.4 (7.9)	36.9 (8.2)
	M	63.2 (15.5)	63.1 (18.7)
Adduction	F	37.6 (6.9)	41.1 (10.8)
	M	66.7 (14.1)	70.5 (15.8)
External rotation	F	17.0 (3.6)	15.9 (3.3)
	M	29.5 (5.8)	29.3 (7.1)
Internal rotation	F	24.9 (6.9)	26.1 (7.0)
	M	48.5 (12.4)	50.0 (14.1)

Table II. *Test-retest reliability results by muscle groups and gender: the estimated variance components and the intraclass correlation coefficient (ICC), determined with logarithmic transformed data*

		Between-person (var Su)	Within-person (var Se)	(var Su*Se)	ICC
Abduction	F	0.044	0.0001	0.007	0.86
	M	0.073	0.0001	0.013	0.85
Adduction	F	0.038	0.006	0.011	0.69
	M	0.049	0.0009	0.004	0.91
External rotation	F	0.040	0.002	0.012	0.74
	M	0.055	0.000	0.008	0.87
Internal rotation	F	0.062	0.0003	0.014	0.81
	M	0.072	0.0004	0.006	0.92

Var Su = varSubject, var Se = varSession, var Su*Se = varSubject*Session.
ICC = var Su/(var Su + var Se + var Su*Se).

first session. In this phenomenon, both learning effects and some demotivation (which can differ between subjects) may coincide.

Although there is a displacement of the glenohumeral joint axis in abduction/adduction movements (5), it can be concluded from this study that this does not influence the reliability. In contrast to the abduction/adduction movement, there is no displacement of the joint axis in the external/internal rotational movement. In the present study there is no difference in reliability between the abduction/adduction and the external/internal rotational movements. In this study one position was used for the measurements, in both movements a sitting position; therefore, it is not possible to determine the influence on the reliability. The used position is practical and the subacromial structures are better protected than with 90° abduction of the shoulder. Furthermore, we found no influence of the angular velocity on the reliability. All measurements were done according to a standardized protocol which is important for reliability. In clinical use, in most cases, different therapists perform the measurements. In the present study the inter-tester reliability was not determined, but there was no difference in reliability between the symmetrical (two examiners) and asymmetrical group (one examiner).

We determined the intraclass correlation coefficient (ICC) for the different muscle groups in both males and females. In the present study the ICC ranged from 0.69 to 0.92. This implies good to excellent reliability (16) of the measurements when used in research on groups of patients. The ICC of adduction in women was the lowest. For this muscle group in women the between-subject variance (varSu) is relatively small; which may be a reason for the moderate ICC. Comparison with other studies is not possible. One study (6) determined reliability at group level using regression analysis and Pearson correlation coefficients, which are not comparable with ICCs.

The ICC is a proportional index of reliability in which the error variance is weighted against the between-subject variance. For clinical use a proportional index of reliability is not informative, but rather indices of absolute reliability focusing on the error variance (such as the S.E.M., 95%-CI and SDD) which can be interpreted for two consecutive measurements in individuals. As said before, data were logarithmically transformed before ANOVA was done (12), for clinical use, however, a non-logarithmic index is necessary. The antilog of the difference between two values on a log scale is a dimensionless ratio. In this study the percentage resulting from the antilog ratio is used, implying that only differences of at least 21–43% of the first measurement should be interpreted as a real change in muscle strength.

Thus, we conclude that for isokinetic dynamometry of the shoulder the test-retest reliability of measurement results of individual subjects, expressed as the SDD is less satisfactory than for groups of subjects as was concluded from the ICCs. There are no studies of isokinetic dynamometry of the shoulder with which to compare. If we compare the SDDs of the present study with estimated from data on knee flexion and extension of Harding et al. (17) and of Stratford (18) we see higher SDDs in our study. The SDDs of these studies (17, 18) ranged from 12% to 16%. In order to improve the reliability in shoulder measurements additional measures can be considered, for example to perform an extra session. By using means over two sessions, the within-subject variance can be divided by 2 (13), resulting in a SDD ranging from 15% to 29%.

Test-retest reliability of isokinetic dynamometry of the shoulder is better for groups of subjects than for measurements for application in individuals. In individual measurements the SDD, found in this study, can be used. It is questionable whether the SDD is small enough to be sufficiently sensitive to detect

Table III. *The proportional indexes of measurement error by muscle group and gender*

		S.E.M. (%)	95% CI (%)	SDD (%)
Abduction	F	8	±17	25
	M	12	±25	37
Adduction	F	14	±29	43
	M	7	±15	21
External rotation	F	13	±26	39
	M	9	±19	28
Internal rotation	F	13	±26	39
	M	8	±17	25

S.E.M. = standard error of measurement; 95% CI = 95% confidence interval; SDD = smallest detectable difference.

clinically relevant change in patients, because no data about this subject are known. This needs further research. Another study could focus on bilateral comparison, i.e. a comparison between affected and non-affected joints can be used. If there is no significant difference between the dominant and non-dominant shoulder in healthy subjects, in patients the non-affected shoulder can be used as a reference.

REFERENCES

1. Dvir Z. Isokinetics: Muscle testing, interpretation and clinical application. First Edition. Longman Group Limited, Churchill Livingstone, 1995.
2. Frontera WR, Hughes VA, Dallal GE, Evans WJ. Reliability of isokinetic muscle strength testing in 45- to 78-year-old men and women. Arch Phys Med Rehabil 1993; 74: 1181–1185.
3. Bohannon RW. Measurement, nature, and implications of skeletal muscle strength in patients with neurological disorders. Clin Biomech 1995; 10: 283–292.
4. Nitschke JE. Reliability of isokinetic torque measurements: A review of the literature. Austr Phys 1992; 38: 125–134.
5. Walmsley RP. Movement of the axis of rotation of the glenohumeral joint while working on the Cybex II dynamometer. Part I. Flexion/extension and part II. Abduction/adduction. Isokin Exerc Sci 1993; 3: 16–26.
6. Magnusson SP, Gleim GW, Nicholas JA. Subject variability of shoulder abduction strength testing. Am J Sports Med 1990; 18: 349–353.
7. Soderberg GJ, Blaschak MJ. Shoulder internal and external rotation peak torque production through a velocity spectrum in differing positions. J Orthop Sports Phys Ther 1987; 518–524.
8. Reid DC, Oedekoven G, Kramer JF, Saboe LA. Isokinetic muscle strength parameters for shoulder movements. Clin Biomech 1989; 4: 97–104.
9. Connelly Maddux RE, Kibler WB, Uhl T. Isokinetic peak torque and work values for the shoulder. J Orthop Sports Phys Ther 1989; 264–269.
10. Ivey FM, Calhoun JH, Rusche K, Bierschenk J. Isokinetic testing of shoulder strength: normal values. Arch Phys Med Rehabil 1985; 66: 384–386.
11. Shklar A, Dvir Z. Isokinetic strength relationships in shoulder muscles. Clin Biomech 1995; 10: 369–373.
12. Bland JM, Altman DG. Statistical methods for assessing agreements between two methods of clinical measurement. Lancet 1986; 8: 307–310.
13. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. 2nd edition. Oxford: Oxford Medical Publications, 1996.
14. Roebroeck ME, Harlaar J, Lankhorst GJ. The application of generalizability theory to reliability assessment: an illustration using isometric force measurements. Phys Ther 1993; 73: 386–401.
15. Shavelson RJ, Webb NM, Rowley GL. Generalizability theory. Am Psych 1989; 44: 922–932.
16. Fleiss JL. The design and analysis of clinical experiments. New York: John Wiley; 1986.
17. Harding B, Black T, Bruulsema A, Maxwell B, Stratford P. 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: 218–224.
18. Stratford P. Reliability of a peak knee extensor and flexor torque protocol: a study of post ACL reconstructed knees. Phys Canada 1991; 43: 27–30.