J. Phys. Ther. Sci.
27: 967–971, 2015

Technical Note

Reliability of measurements of knee extensor muscle strength using a pull-type hand-held dynamometer

Terumi Suzuki[1)]

[1)] Department of Physical Therapy, University of Tokyo Health Science: 4-11 Ochiai, Tama city, Tokyo 206-0033, Japan

Abstract. [Purpose] The purpose of this study was to examine the intra-examiner and inter-examiner reliabilities of measurements of knee extensor muscle strength using a pull-type hand-held dynamometer (HHD). [Subjects] Fifty-four healthy adults (35 males; average age, 23 years) participated in this study. [Methods] Knee extensor muscle strength of each leg was measured three times using the HHD. To examine the intra- and inter-examiner reliabilities, measurements were performed by two examiners, a physical therapist and a physical therapy student. [Results] The intra-examiner reliabilities, ICC (1, 1) and ICC (1, 3) ranged from 0.94–0.99. The inter-examiner reliabilities, ICC (2, 1) and ICC (3, 1) ranged from 0.90–0.92 for the right leg, and 0.88−0.90 for the left leg. Neither constant nor proportional errors were found by Bland-Altman analysis. [Conclusion] Intra-examiner and inter-examiner reliabilities were acceptable, indicating that muscle strength can be measured with the pull-type HHD without dependence on skill of measurement. Pain was not caused by measurements with the pull-type HHD.
Key words: Inter-examiner reliability, Intra-examiner reliability, Hand-held dynamometer

(This article was submitted May 29, 2014, and was accepted Oct. 1, 2014)

INTRODUCTION

Measurement of muscle strength in rehabilitation is important for the evaluation of treatment outcomes. Simple manual muscle testing can be performed, or a measuring device can be used for quantitative measurement of muscle strength. In particular, a hand-held dynamometer (HHD) is a relatively inexpensive and very portable instrument, and has been used in a variety of situations for measuring muscle strength.

Previous reports have demonstrated the high reliability of measuring muscle strength with an HHD[1, 2)]. However, it has also been reported that the results of measuring muscle strength with an HHD are influenced by the position of the testee's lower extremity, fixation of the device, and the skill of the examiner[3–7)]. For improving reliability, a strap or a fixing frame has been used to make fixation holding the measuring device by hand[8)], and both methods provide high reliability. Also, using an HHD attached with a belt has been reported to be high reliable, overcoming the problem of examiners' skills[9–12)].

In terms of measurement modes, it has been reported that the break test requires greater strength of the examiner to resist the strength of the examinee than the make test[13–15)]. Furthermore, the reliability of the results of the break test has been reported to be low[16)].

When measuring muscle strength with an HHD attached by a belt, muscle strength is measured by pushing a part of the body against the sensor pad. In measurement of isometric knee extension strength, a sensor pad is attached to the front of the distal lower extremity to measure strength. Some patients complain of pain and discomfort at the distal lower extremity when this method is used, and the pain becomes a limiting factor, preventing the exertion of full muscle strength, particularly in cases of high muscle strength values.

When the thickness and shape of the cushion of the sensor pad are changed for the measurement of isometric knee extension strength using an HHD attached by a belt, values of muscle strength are reproducible, but muscle strength values are strongly influenced by the pain experienced at the time of measurement. Therefore, to improve the appropriateness of measured values of muscle strength, it may be necessary to use a pad of sufficient thickness with an appropriate curve to eliminate pain.

In rehabilitation, the muscle strength of elderly individuals and subjects with reduced muscle mass is often measured. Compared with healthy adults, soft tissue at the front of the distal lower extremity is often thin, and, accordingly, the measurement procedure can be painful. Therefore, in measurement of isometric knee extension strength using an HHD with attachment of a fixation belt, pain at the front of the distal lower extremity remains a problem.

To overcome this problem, Mobie, a new device for measuring muscle strength using a pull sensor, rather than

Corresponding author. Terumi Suzuki (E-mail: t-suzuki@u-ths.ac.jp)

pushing a limb against the sensor, was developed.

The intra-class correlation coefficient (ICC) is used as a statistical analysis method of reliability. However, in cases of data with large individual differences and measurements with marked variability, inter-examiner differences and errors become relatively small and the coefficient becomes large. Recently, the reliability of an HHD was examined using standard error of measurement (SEM) and minimal detectable change (MDC) in addition to ICC[17–22].

There are random errors and systematic errors, and the former can be reduced by increasing the number of subjects and repetition of measurements, whereas the latter is an error in which measurements deviate with a specific bias, and it cannot be minimized by increasing the number of subjects or measurements. Therefore, it is necessary to extract systematic errors.

Systematic errors are proportional and constant. Proportional errors increase or decrease with the true value with a specifi bias, whereas constant errors have aspecific bias irrespective of the true value. Systematic errors can be identified using Bland-Altman analysis[23], and using the degree of “variability” between two measurements, it is possible to determine the reliability of the measurement and its appropriateness for clinical application.

In this study intra-examiner reliability and inter-examiner reliability of an HHD with a pull sensor was examined using ICC and Bland-Altman analysis.

SUBJECTS AND METHODS

The subjects were 54 healthy adults (35 males and 19 females; mean age, 23.2±6.8 years; height, 166.3±8.4 cm; weight, 60.0±12.6 kg) (Table 1). The bilateral lower extremities (108 legs) of all subjects were examined. The 35 subjects (21 males and 14 females) in their 20s were examined the most. The aims and contents of the study were explained to the subjects verbally and with a written document before measurement. All participants provided their written informed consent. Permission to conduct this study was provided by the Human Research Ethics Committee of Tokyo college of allied medicine physical therapy.

A manual muscle strength monitor (Mobie; Sakai Medical Co., Ltd.) was used. Mobie is an HHD with a thin sensor. Unlike conventional HHDs that measure force by pushing the body against the sensor, this device senses muscle strength by pulling a distortion gauge, and muscle strength can be measured with a fixation belt attached.

To examine inter-examiner differences in the Mobie HHD measurements, measurements were performed by two examiners: one female physical therapist (PT) (Examiner A) with 21 years of clinical experience and skill in handling HHDs, and one male third-grade student on a four-year physical therapy course (Examiner B). He practiced the method of the measurement that had defended notes the day before the measurements began. On the first day, Examiner A and Examiner B measured one leg of each subject, and the next day they measured the other leg. Measurements of each leg were repeated three times. A rest of one minute was provided between measurements. For all measurements, isometric knee extension was performed at maximum effort for approximately 3 seconds.

Table 1. The characteristics of the subjects (n=54)

Age (yrs)	23.2 ± 6.8
Height (cm)	166.3 ± 8.4
Weight (kg)	60.0 ± 12.6
BMI (kg/m^2)	21.6 ± 4.1

Data are means ± SD

For the measurements, the subjects sat on the edge of a bed with their feet not touching the floor, and their arms crossed in front of the body. Subjects were asked; to maintain the trunk in the upright position; to keep the buttocks on the bed; and to exert muscle power to extend the knee joint without a swing movement.

Prior to measurement, an appropriate posture for measurement was explored. Since the muscle to be measured is the femoral quadriceps, the occurrence of compensatory movement due to exertion of muscle power in isometric knee extension was examined in two postures: supporting both arms at the back of the bed with the pelvic tilted backward and the trunk in extension, similar to the posture for manual muscle testing; and with both arms crossed in front of the trunk.

Elevation of the pelvis was more likely to appear in the posture similar to that for manual muscle testing than in the posture with both arms crossed in front of the trunk, and it was necessary for the examiner to fix both sides of the pelvis manually in addition to controlling the device to block this compensaory movement. Therefore, considering rapid and simple measurement with blocking of compensaory movement in clinical practice, the posture with both arms crossed in front of the trunk was selected for this study.

Calibration was always performed prior to beginning the three measurements. The pad of the measuring device was attached around the distal lower extremity and controlled with a belt fastened to the examination pole so that the lower extremity was perpendicular to the floor. The sensor was set parallel to the floor with the fixation belt, and positioned midway between the lower extremity and the examination bed. Since this measuring device senses tension, a hand of the examiner supported to the sensor and to prevent gravity-derived downward tension during calibration and measurement. In addition, to avoid popliteal pain by compression measurement, a bath towel was spread over the site. Subjects were always encouraged to exert maximum muscle power.

ICC was calculated based on the muscle strengths measured by Examiners A and B, and intra-examiner reliability and inter-examiner reliability were examined using Bland-Altman analysis. When no systematic error was recognized, the MDC of the 95% confidence interval (MDC_{95}) was calculated. MDC_{95} and SEM were calculated according to the following formula.

$$MDC_{95} = SEM \times 1.96 \times \sqrt{2}\ .$$

SEM = standard deviation of difference between two measurements (S) / $\sqrt{2}$.

Inter-examiner reliability was examined with ICC (1,1) and (1,3) over the three measurements. The first and second measured values were examined bilaterally using Bland-Altman analysis. For inter-examiner reliability, ICC (2,1) and ICC (3,1) were calculated to examine both the random effect and the fixation effect, respectively. In both intra-examiner reliability and inter-examiner reliability, systematic errors and the limit of agreement (LOA), the permissible range of errors between two measurements, were calculated. LOA was calculated according to the following formula.

$[(d - 1.96s) + tSE_{LOA}] \sim [(d + 1.96s) - tSE_{LOA}]$

*d: average of difference between two

measurements $SE_{LOA} = \sqrt{3S^2 / n}$

t: t value for the degree of freedom n − 1

For data analysis, IBM SPSS Statistics 21 was used.

RESULTS

Table 2 shows the averages of muscle strength as measured by Examiner A and Examiner B. The average right muscle strength was 45.6 ± 9.0 kgf in males and 28.2 ± 3.8 kgf in females by Examiner A, and 45.9 ± 8.4 kgf in males and 27.9 ± 4.9 kgf in females by Examiner B. With regard to intra-examiner reliability, both ICC (1,1) and ICC (1,3) were 0.9 or higher (Table 3). Bland-Altman analysis indicated there were no constant or proportional errors by Examiner A or Examiner B (Table 3).

The LOA of the right extremity was −3.97 to 3.17 by Examiner A and −5.95 to 5.35 by Examiner B, while the LOA of the left extremity was −6.50 to 4.50 by Examiner A and −4.86 to 5.06 by Examiner B. Since no systematic error was found, MDC_{95} of the right and left extremities were 4.7 kgf and 7.3 kgf, respectively, by Examiner A, and 7.5 kgf and 6.7 kgf, respectively, by Examiner B. These results indicate that when muscle strength was measured twice in the same subject, MDC_{95} within 4.7 kgf for the right extremity by Examiner A was considered a measurement error and MDC_{95} more than 4.7 kgf was judged as true change caused by intervention for the subject.

In terms of inter-examiner reliability, both ICC (2,1) and ICC (3,1) in the first through the third measurements were 0.9 or higher for the right extremity and 0.8 or higher for the left extremity (Table 4). In the first through the third measurements, neither constant nor proportional errors were identified by Bland-Altman analysis (Table 4). No subject complained of pain due to the pad attached around the lower extremity while muscle power was being exerted.

DISCUSSION

To investigate intra-examiner and inter-examiner reliability of isometric muscle strength as measured using the Mobie HHD with a pull sensor, measurement was repeated three times by a PT with clinical experience and a PT student without clinical experience. Previously, we reported that for subjects in their 20s[24]), which accounted for the largest proportion of subjects, isometric knee extension strength of males and females was 60.4 ± 8.1 kgf and 37.1 ± 8.9 kgf, respectively. Compared with these results, the muscle strength of the subjects in this study was slightly weaker. Therefore, compensaory movement was unlikely to have occurrde and both intra-examiner variability and inter-examiner variability in measurements were smaller, and systematic errors were unlikely to have occurred.

With regard to intra-examiner reliability, ICC (1,1) was

Table 2. The average values of the measured strength

	A examiner (kgf)	B examiner (kgf)
Right (average 3 times)	39.4 ± 10.6	39.5 ± 11.2
Left (average 3 times)	39.4 ± 10.6	39.2 ± 9.9
Right		
1 test	39.0 ± 10.7	39.1 ± 11.6
2 test	39.5 ± 10.4	39.3 ± 11.1
3 test	39.9 ± 10.8	40.2 ± 11.1
Left		
1 test	38.5 ± 10.6	39.3 ± 9.6
2 test	39.5 ± 10.5	39.1 ± 10.3
3 test	40.2 ± 10.7	39.3 ± 10.1

Data are means ± SD

Table 3. The intra-examiner reliabilities of the measured strength

				Bland - Altman analysis			
				Adding error		Proportional error	
	ICC (1,1)	ICC (1,3)	LOA	95% confidence interval	Presence or absence	The slope of the regression line	Presence or absence
examiner A Right	0.96	0.99	−3.97 to 3.17	−1.05 to 0.26	absence	0.14	absence
examiner A Left	0.94	0.98	−6.50 to 4.5	−2.01 to 0.01	absence	0.06	absence
examiner A Right	0.95	0.98	−5.95 to 5.35	−1.34 to 0.74	absence	0.16	absence
examiner A Left	0.94	0.98	−4.86 to 5.06	−1.13 to 0.75	absence	−0.21	absence

Table 4. The intra-examiner reliabilities of the measured strength

				Bland - Altman analysis			
				Adding error		Proportional error	
	ICC (2,1)	ICC (3,1)	LOA	95% confidence interval	Presence or absence	The slope of the regression line	Presence or absence
A Right 1 test	0.92	0.92	−9.02 to 14.50	−1.74 to 1.52	absence	−0.17	absence
2 test	0.92	0.92	−8.53 to 14.21	−1.48 to 1.68	absence	−0.13	absence
3 test	0.90	0.90	−10.26 to 16.0	−2.13 to 1.53	absence	−0.04	absence
A Left 1 test	0.89	0.90	−10.17 to 14.53	−2.52 to 0.92	absence	0.17	absence
2 test	0.90	0.90	−8.97 to 15.73	−1.13 to 0.75	absence	0.02	absence
3 test	0.88	0.88	−9.31 to 17.34	−1.06 to 2.66	absence	0.1	absence

0.94 to 0.96, and the reliability was judged as "excellent". No systematic error was found, indicating that there is no problem with intra-examiner reliability. When ICC (1,1) and ICC (1,3) were compared, ICC (1,3) was slightly higher showing that reliability was higher, as expected, when the same examiner carried out the same measurement more than once.

With regard to inter-examiner reliability, both ICC (2,1) and ICC (3,1) were 0.88 to 0.92, and the reliability was considered "good" to "excellent". For both legs, ICC (2,1) and ICC (3,1) were slightly reduced as measurement was repeated from the first through the third time, and the variability was attributed to fatigue. However, since ICC ranged between 0.88 and 0.92 and no constant or proportional errors were found, there was no problem inter-examiner reliability.

From the study of Examiner A with clinical experience and skill in handling HHDs, and those by Examiner B, a PT student with no experience of measuring muscle strength with HHD, the examiner can carry out measurement even if he/she has no experience on the operation of the device as long as he/she gives the subjects the notification at measurement and measures according to the predetermined procedure.

Although pain remains a problem in muscle strength measurement with an HHD and a fixation belt, the device used in this study caused no pain and had no influence on the muscle power exerted. Therefore, when knee extension strength is measured using an HHD with a pull sensor, the measurement is accurate and, since the subjects experienced no pain while exerting muscle power, measurement is possible without inducing pain in cases of thin soft tissue at the front of distal lower extremities in frail elderly people. Moreover, quantitative muscle strength measurement is expected to be possible in a variety of subjects.

In this study, the subjects were healthy adults, but compensatory movement is likely during the exertion of muscle power by subjects with high muscle power levels such as athletes. However, compensatory movement is unlikely to occur in the elderly. Whether knee joint extension without a swing movement is possible or not influences the measurement errors of the measuring device. It will be necessary to further investigate the utility of measurements with a pull-type HHD using subjects with different muscle power levels, elderly people, and patients.

REFERENCES

1) Bohannon RW: Reference values for extremity muscle strength obtained by hand-held dynamometry from adults aged 20 to 79 years. Arch Phys Med Rehabil, 1997, 78: 26–32. [Medline] [CrossRef]
2) Bohannon RW: Test-retest reliability of hand-held dynamometry during a single session of strength assessment. Phys Ther, 1986, 66: 206–209. [Medline]
3) Kelln BM, McKeon PO, Gontkof LM, et al.: Hand-held dynamometry: reliability of lower extremity muscle testing in healthy, physically active, young adults. J Sport Rehabil, 2008, 17: 160–170. [Medline]
4) Torborg K, Bandholm T, Schick M, et al.: Hip strength assessment using Handheld dynamometry is subject to intertester bias when tester are of different sex and strength. Scand J Med Sci Sports, 2011, 10: 1–7. [Medline]
5) Mulroy SJ, Lassen KD, Chambers SH, et al.: The ability of male and female clinicians to effectively test knee extension strength using manual muscle testing. J Orthop Sports Phys Ther, 1997, 26: 192–199. [Medline] [CrossRef]
6) Deones VL, Wiley SC, Worrell T: Assessment of quadriceps muscle performance by a hand-held dynamometer and an isokinetic dynamometer. J Orthop Sports Phys Ther, 1994, 20: 296–301. [Medline] [CrossRef]
7) Wikholm JB, Bohannon RW: Hand-held dynamometer measurements: tester strength makes a difference. J Orthop Sports Phys Ther, 1991, 13: 191–198. [Medline] [CrossRef]
8) Lu TW, Hsu HC, Chang LY, et al.: Enhancing the examiner's resisting force improves the reliability of manual muscle strength measurements: comparison of a new device with hand-held dynamometry. J Rehabil Med, 2007, 39: 679–684. [Medline] [CrossRef]
9) Brinkmann JR: Comparison of a hand-held and fixed dynamometer in measuring strength of patients with neuromuscular disease. J Orthop Sports Phys Ther, 1994, 19: 100–104. [Medline] [CrossRef]
10) Gagnon D, Nadeau S, Gravel D, et al.: Reliability and validity of static knee strength measurements obtained with a chair-fixed dynamometer in subjects with hip or knee arthroplasty. Arch Phys Med Rehabil, 2005, 86: 1998–2008. [Medline] [CrossRef]
11) Thorborg K, Bandholm T, Hölmich P: Hip- and knee-strength assessments using a hand-held dynamometer with external belt-fixation are inter-tester reliable. Knee Surg Sports Traumatol Arthrosc, 2013, 21: 550–555. [Medline] [CrossRef]
12) Kramer JF, Vaz MD, Vandervoort AA: Reliability of isometric hip abductor torques during examiner- and belt-resisted tests. J Gerontol, 1991, 46: M47–M51. [Medline] [CrossRef]
13) Bohannon RW: Make tests and break tests of elbow flexor muscle strength. Phys Ther, 1988, 68: 193–194. [Medline]
14) van der Ploeg RJ, Oosterhuis HJ: The "make/break test" as a diagnostic tool in functional weakness. J Neurol Neurosurg Psychiatry, 1991, 54:

248–251. [Medline] [CrossRef]
15) Stratford PW, Balsor BE: A comparison of make and break tests using a hand-held dynamometer and the Kin-Com. J Orthop Sports Phys Ther, 1994, 19: 28–32. [Medline] [CrossRef]
16) Agre JC, Magness JL, Hull SZ, et al.: Strength testing with a portable dynamometer: reliability for upper and lower extremities. Arch Phys Med Rehabil, 1987, 68: 454–458. [Medline]
17) Knols RH, Aufdemkampe G, de Bruin ED, et al.: Hand-held dynamometry in patients with haematological malignancies: measurement error in the clinical assessment of knee extension strength. BMC Musculoskelet Disord, 2009, 10: 31. [Medline] [CrossRef]
18) Koblbauer IF, Lambrecht Y, van der Hulst ML, et al.: Reliability of maximal isometric knee strength testing with modified hand-held dynamometry in patients awaiting total knee arthroplasty: useful in research and individual patient settings? A reliability study. BMC Musculoskelet Disord, 2011, 12: 249. [Medline] [CrossRef]
19) Roebroeck ME, Harlaar J, Lankhorst GJ: Reliability assessment of isometric knee extension measurements with a computer-assisted hand-held dynamometer. Arch Phys Med Rehabil, 1998, 79: 442–448. [Medline] [CrossRef]
20) Kropmans TJ, Dijkstra PU, Stegenga B, et al.: Smallest detectable difference in outcome variables related to painful restriction of the temporomandibular joint. J Dent Res, 1999, 78: 784–789. [Medline] [CrossRef]
21) Ottenbacher KJ, Johnson MB, Hojem M: The significance of clinical change and clinical change of significance: issues and methods. Am J Occup Ther, 1988, 42: 156–163. [Medline] [CrossRef]
22) Roebroeck ME, Harlaar J, Lankhorst GJ: The application of generalizability theory to reliability assessment: an illustration using isometric force measurements. Phys Ther, 1993, 73: 386–395, discussion 396–401. [Medline]
23) Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. Lancet, 1986, 1: 307–310. [Medline] [CrossRef]
24) Hirasawa Y, Hasegawa T, Matushita K, et al.: Health adult's isometric knee extension strength. PT J, 2004, 38: 330–332 (in Japanese).