Published in final edited form as:

Isokinet Exerc Sci. 2011 ; 19(2): . doi:10.3233/IES-2011-0400.

Relative reliability of three objective tests of limb muscle strength

Richard W. Bohannon^{a,*}, Deborah J. Bubela^a, Susan R. Magasi^b, and Richard C Gershon^b

^aProgram in Physical Therapy, Department of Kinesiology, University of Connecticut, Storrs, CT, USA

^bDepartment of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

Abstract

If measures of muscle strength are to be broadly applied, they should be objective, portable, quick, and reliable. Through this component of the NIH Toolbox study we sought to compare the test-retest reliability of 3 tests of muscle strength that are objective, portable, and quick: the five-repetition sit-to-stand test (FRSTST), hand-grip dynamometry (HGD), and belt-stabilized hand-held dynamometry (BSHHD) of knee extension. Three sets of each test were performed- 1 warm-up and 2 maximal. Measures from the maximal tests obtained 4 to 10 days apart were compared. Reliability was described using descriptive statistics, intraclass correlation coefficients (ICCs) and 4 measures of response stability: standard error of measurement (SEM), method error (ME), coefficient of variation of SEM (SEM_{CV}), and coefficient of variation of variation of ME (ME_{CV}). The ICCs of all tests were good (> 0.853). Measures of response stability showed less variability between test and retest for FRSTST and HGD than for BSHHD. In conclusions all 3 tests demonstrated good test-retest reliability. However, greater differences would need to be observed between tests sessions to conclude that a real change had occurred in measures obtained by BSHHD.

Keywords

Muscle strength; measurement; reliability; response stability

1. Introduction

The implications of muscle strength for mobility and other important outcomes is well established [2, 3]. It is appropriate, therefore, that muscle strength be assessed during the examination of both apparently healthy individuals and patients with known limitations [9]. Although many methods are available for measuring muscle strength, those that are objective, quick, and portable are particularly compelling. Included among such methods are hand-grip dynamometry (HGD), hand-held dynamometry (HHD), and sit-to-stand testing. The reliability of these tests has been described extensively using reliability coefficients (eg, intraclass correlation coefficients) [1,4,10,12,14–18]. However, other indicators of reliability such as response stability (measurement variability) have received little attention. Information on response stability is particularly important if such tests are to be interpreted in regard to those obtained at an earlier time [13]. Specifically, information regarding

response stability allows a tester to determine if a retest measure is truly different from the earlier measure rather than within the bounds of normal variability associated with repeated testing. The purpose of this study, therefore, was to describe the relative reliability of HGD, belt-stabilized HHD (BSHHD), and five-repetition sit-to-stand testing (FRSTST) using both reliability coefficients and indicators of response stability.

2. Method

This study was part of the validation component of the NIH Toolbox study, an investigation designed to select a short but inclusive battery of portable, low-cost and lay-administered tests of cognitive, motor, sensory, and emotional health and function for use in future cohort studies [8]. Data for this study were gathered for the motor domain of the NIH Toolbox by 7 trained testers at 2 participating sites (University of Connecticut and Rehabilitation Institute of Chicago). The institutional review boards at both sites approved the study.

2.1. Participants

All participants provided written informed consent before testing. Inclusion required that participants were fluent in English, were able to walk without an assistive device, were between 14 and 85 years of age, and had no heart, vascular, lung, or bone/joint problems precluding their standing from a chair or climbing steps. Of 184 motor domain participants, 28 returned for a second visit within 4 to 10 days of an initial visit to establish reliability.

2.2. Procedures

Basic demographic (age, gender) and anthropometric (height, weight [body mass index]) data were first obtained. Thereafter, bilateral HGD, bilateral BSHHD, and FRSTST were conducted in random order. For each participant a warm-up trial was performed followed by 2 maximal efforts. HGD was conducted in accordance with the protocol recommended by the American Society of Hand Therapy [6]. Specifically, participants were tested while they were seated, their arms were against their sides, their elbows were flexed 90 degrees and the Jamar dynamometer was in the second handle position. BSHHD was conducted while subjects were seated on an elevated commercially available chair with foam padding and stabilizing straps for their proximal thighs and waist. Their knees were at about 90 degrees of flexion (i.e., their legs were vertical). A dynamometer stabilizing belt passed around a bar secured behind the back legs of the chair and over a calibrated MicroFET HHD that was placed against the anterior legs of participants just proximal to the malleoli (Fig. 1). Participants were asked to take a second or two to come to maximal effort and to then continue trying to straighten their knee as hard as possible until the tester asked them to stop (about 4 seconds later). The FRSTST required participants to stand up from, and sit down on, a 43 cm high armless chair 5 times as quickly as possible. Participants folded their arms across their chests and were instructed to stand-up completely and make firm contact when sitting. Timing began on the command “go” and ceased when the participants sat after the fifth stand-up.

2.3. Data analysis

For all analyses only the 2 maximal efforts were employed. Specifically, the first, best and mean of the 2 efforts were used. Descriptive statistics (mean, standard deviation [SD]) were calculated for each, as were paired t-tests, to determine if performance differences existed between the sessions. Intraclass correlation coefficients (ICCs) were used to characterize reliability for men and women participants (separately and combined). The standard error of measurement ($SEM [SD_{(set\ of\ scores)} \sqrt{1-ICC}]$), method error ($ME [SD_{(difference\ scores)} / \sqrt{2}]$), coefficient of variation of SEM ($CV_{SEM} [(SEM/mean_{(overall)}) \times 100]$), and coefficient of variation of method error ($CV_{ME} [(ME/mean_{(overall)}) \times 100]$) were used to describe response

stability [13] for men and women combined. The Statistical Package for Social Sciences (version 18.0) was used for all analyses.

3. Results

Of the 28 participants half were of each gender. Their ages ranged from 15 to 85 years (mean 45.7 ± 23.5). Table 1 summarizes the results relevant to reliability for men and women participants combined. The FRSTST and BSHHD measurements all differed (improved) significantly between sessions whereas the HGD measures did not. Regardless of the procedure, the magnitudes of ICCs for men (0.870–0.968) and women (0.680–0.983) were mostly good and did not differ significantly. Consequently only the ICCs for all 28 men and women participants (0.853–0.975) are reported in Table 1. The magnitudes of these ICCs were consistently lower for BSHHD than for FRSTST and HGD; however, the confidence intervals overlap. The SEMs and MEs for FRSTST cannot be compared to those of the dynamometer measures as the units are different. The SEMs and MEs for the dynamometer measures can be compared, particularly as dynamometer measured hand grip and knee extension forces were not particularly disparate. The SEMs and MEs for BSHHD were, on average, slightly more than twice those for HGD. The CV_{SEM} and CV_{ME} , which can be compared across all tests, show that percentage differences for FRSTST and HGD (7.1%) are less than those for BSHHD (9.7%).

4. Discussion

Consistent with previous studies, the reliability coefficients of the present investigation support the reliability of the 3 measures used to quantify muscle strength [1, 4,10,12,14–18]. The present study goes further, however, in describing the response stability of the 3 measures investigated. Information regarding response stability might be used in following individuals over time to ascertain if their natural course or response to therapy surpasses normal test-retest variability. The response stability of the tests was not equivalent. On the basis of all indicators of response stability the FRSTST and HGD demonstrated far less variability between sessions than BSHHD. Consequently, the difference between a retest value and an initial test value required to conclude that a real change had taken place would be greater for BSHHD. This along with expense and ease of performance may favor the FRSTST and HGD over BSHHD. Regardless, knee extension strength is a measure of proven importance to function [5,11] and the reliability of BSHHD certainly supports its use. Ongoing NIH Toolbox work will explore alternative portable methods of measuring isometric knee extension force which may further improve response stability.

This study has several limitations. First, it involved community-dwelling individuals without notable activity limitations. While it is important to know the reliability of strength measures obtained from such individuals, the reliability found may not generalize to some populations. Second, the sample was relatively small. This reality notwithstanding, the sample size exceeds the 15 or 20 indicated by Fleiss [7] to be sufficient to provide an estimate of reliability. Moreover the confidence intervals for the ICCs take sample size into account.

We conclude on the basis of their ICCs, the FRSTST, HGD, and BSHHD can be considered to demonstrate good test-retest reliability. However, measures of response stability favor the FRSTST and HGD over BSHHD for repeated testing.

Acknowledgments

This study is funded with Federal funds from the Blueprint for Neuroscience Research, National Institutes of Health under Contract HHS-N-260-2006-00007-C. The content presented herein does not necessarily represent the official

views of the National Institutes of Health or the National Institute of Aging. Dr Bohannon is a consultant with Hoggan Health Industries, the manufacturer of the hand-held dynamometer used in belt-stabilized testing. We acknowledge the role of Ying-Chih Wang in managing the project at the Rehabilitation Institute of Chicago.

References

1. Arnold CM, Warkentin KD, Chilibeck PD, Magnus CRA. The reliability and validity of handheld dynamometry for the measurement of lower-extremity muscle strength in older adults. *Journal of Strength and Conditioning Research*. 2010; 24:815–824. [PubMed: 19661831]
2. Bohannon RW. Nature, implications, and measurement of limb muscle strength in patients with orthopaedic or neurologic disorders. *Physical Therapy Practice*. 1992; 2:22–31.
3. Bohannon RW. Hand-grip dynamometry predicts future outcomes. *Journal of Geriatric Physical Therapy*. 2008; 31:1–10.
4. Bohannon RW, Schaubert KL. Test-retest reliability of grip-strength measures obtained over a 12-week interval from community-dwelling elders. *Journal of Hand Therapy*. 2005; 18:426–428. [PubMed: 16271690]
5. Eriksrud O, Bohannon RW. Relationship of knee extension force to independence in sit-to-stand performance in patients receiving acute rehabilitation. *Physical Therapy*. 2003; 83:544–551. [PubMed: 12775200]
6. Fess, EE. Grip strength. In: Casanova, JS., editor. *Clinical Assessment Recommendations*. 2. Chicago, Ill: American Society of Hand Therapists; 1992. p. 41-45.
7. Fleiss, JL. *The Design and Analysis of Experiments*. New York: John Wiley and Sons, Inc; 1999. p. 8
8. Gershon RC, Cella D, Fox NA, Havlik RJ, Hendrie HC, Wagster MV. Assessment of neurological and behavioral function: the NIH Toolbox. *Lancet Neurology*. 2010; 9:138–139. [PubMed: 20129161]
9. *Guide to Physical Therapist Practice*. *Physical Therapy* (2010). 81:80–81.
10. Hamilton A, Balnave R, Adams R. Grip strength testing reliability. *Journal of Hand Therapy*. 1994; 7:163–170. [PubMed: 7951708]
11. Hortobágyi T, Mizelle C, Bean S, DeVita P. Old adults perform activities of daily living near their maximal capacities. *Journals of Gerontology: Medical Science*. 2003; 58A:453–460.
12. Kelin BM, McKeon PO, Gontkof LM, Hertel J. Hand-held dynamometry: reliability of lower extremity muscle testing in healthy, physically active, young adults. *Journal of Sport Rehabilitation*. 2008; 17:160–170. [PubMed: 18515915]
13. Lexell JE, Downham DY. How to assess the reliability of measures in rehabilitation. *American Journal of Physical Medicine and Rehabilitation*. 2005; 84:719–723. [PubMed: 16141752]
14. Molenaar HM, Zuidam JM, Selles RW, Stam HJ, Hovius SE. Age-specific reliability of two grip-strength dynamometers when used by children. *Journal of Bone and Joint Surgery [American]*. 2008; 90:1053–1059.
15. Mong Y, Teo TW, Ng SS. 5-repetition sit-to-stand test in subjects with chronic stroke: reliability and validity. *Archives of Physical Med and Rehabilitation*. 2010; 91:407–413.
16. O'Shea SD, Taylor NF, Paratz JD. Measuring muscle strength for people with chronic obstructive pulmonary disease: retest reliability of hand-held dynamometry. *Archives of Physical Medicine and Rehabilitation*. 2007; 88:32–36. [PubMed: 17207672]
17. Schaubert KL, Bohannon RW. Reliability and validity of three strength measures obtained from community-dwelling elderly persons. *Journal of Strength and Conditioning Research*. 2005; 19:717–720. [PubMed: 16095431]
18. Schaubert KL, Bohannon RW. Reliability of the sit-to-stand test over dispersed test sessions. *Isokinetics and Exercise Science*. 2005; 13:119–122.

Fig. 1.
Belt-stabilized hand-held dynamometer for measuring knee extension force.

Table 1
Statistics summarizing the reliability of toolbox strength measures obtained from 28 participants

Test*	Measure	Session 1 mean (SD)	Session 2 mean (SD)	Inter-session difference mean (SD)	T (p)	ICC (95% CI)	SEM	CV _{SEM} (%)	ME	CV _{ME} (%)
FRSTST (sec)	First	8.1 (2.3)	7.8 (2.3)	0.3 (0.7)	2.48 (0.020)	0.954 (0.903–0.978)	0.5	6.2	0.5	6.2
	Best	7.7 (2.1)	7.4 (2.3)	0.3 (0.6)	2.39 (0.024)	0.963 (0.922–0.983)	0.4	5.3	0.4	5.6
	Mean	8.0 (2.2)	7.6 (2.3)	0.3 (0.5)	3.61 (0.001)	0.975 (0.946–0.988)	0.3	4.5	0.3	4.5
HGD (N): left	First	345.2 (97.4)	342.5 (97.9)	2.7 (34.7)	0.40 (0.690)	0.938 (0.870–0.971)	24.2	7.1	24.5	7.1
	Best	351.8 (96.1)	357.6 (100.1)	–5.8 (29.4)	–1.08 (0.293)	0.955 (0.906–0.979)	20.4	5.7	20.6	5.9
	Mean	340.3 (94.3)	344.7 (97.0)	–4.4 (23.6)	–0.99 (0.331)	0.965 (0.926–0.983)	17.7	5.2	16.7	4.9
HGD (N): right	First	366.5 (101.9)	364.7 (97.4)	1.3 (35.1)	0.23 (0.823)	0.937 (0.870–0.971)	25.6	7.0	24.9	6.8
	Best	375.0 (97.9)	380.7 (106.3)	–5.3 (34.7)	–0.84 (0.408)	0.942 (0.879–0.973)	23.6	6.2	24.6	6.5
	Mean	367.4 (98.7)	369.2 (101.0)	–2.2 (32.5)	–0.34 (0.733)	0.948 (0.890–0.975)	22.5	6.1	22.9	6.2
BSHHD (N): left	First	377.6 (140.1)	415.0 (122.8)	–37.4 (71.2)	–2.78 (0.010)	0.853 (0.708–0.929)	53.7	13.6	50.3	12.7
	Best	399.4 (145.0)	429.2 (121.9)	–29.8 (60.0)	–2.61 (0.015)	0.899 (0.795–0.952)	46.1	11.1	42.5	10.3
	Mean	386.1 (142.8)	417.7 (118.3)	–31.6 (64.3)	–2.59 (0.015)	0.879 (0.756–0.942)	49.7	12.4	45.6	11.3
BSHHD (N): right	First	384.3 (149.5)	425.7 (146.8)	–41.8 (66.3)	–3.33 (0.003)	0.900 (0.796–0.953)	47.3	11.7	46.9	11.6
	Best	408.8 (158.8)	443.5 (146.3)	–34.7 (64.9)	–2.83 (0.009)	0.910 (0.813–0.957)	47.6	11.2	45.9	10.8
	Mean	393.6 (153.5)	424.8 (140.1)	–30.7 (56.0)	–2.91 (0.007)	0.928 (0.850–0.966)	41.2	10.1	39.6	9.7

* FRSTST = five repetition sit-to-stand testing; HGD = hand-grip dynamometry; BSHHD = belt-stabilized hand-held dynamometry.