

SHORT TERM (24 HOURS) AND LONG TERM (1 YEAR) ASSESSMENTS OF RELIABILITY IN OLDER ADULTS: CAN ONE REPLACE THE OTHER?

T. Abe, S.J. Dankel, S.L. Buckner, M.B. Jessee, K.T. Mattocks, J.G. Mouser, Z.W. Bell, J.P. Loenneke

Abstract: There may be some individuals who do not adapt favorably to an exercise stimulus. This is most commonly determined by assessing the error of the measurement across two separate testing sessions separated by a short period of time. It has been recommended that this error be assessed over the same time frame as the intervention. We examined the 24-h test-retest reliability ($n=18$, aged 42 to 64 years) of forearm muscle thickness, handgrip strength, and “muscle quality” and compared that to the reliability observed when visits are separated by 1-year ($n=80$, aged 60 to 79 years). The measurement errors were greater in all measured variables following test-retest separated by 1-year than the test-retest separated by 24-hours. Our findings suggest that a time-matched control group is likely important to fully capture the error of the tester as well as the error associated with random biological variability within a timed intervention.

Key words: Long-term reliability, muscle size, muscle function, older adults, strength, ultrasound.

Introduction

Cross-sectional and longitudinal studies have reported that values of handgrip strength appear to decrease gradually with increasing age in both men and women, although the age at which handgrip strength starts to decline differs among those studies (1, 2). Similar declines are observed with muscle mass, and the ratio of muscle strength/size is often used as an index of muscle quality (3).

Resistance exercise is commonly recommended for older individuals in an effort to mitigate these proposed declines in muscle size and strength (4). Moreover, long term (~ 1 year) resistance training is capable of producing favorable changes in muscle size and strength (5). Although this is a well-accepted finding, recent reports suggest that there may be some individuals within these interventions who do not actually respond favorably to the exercise stimulus (6, 7). Knowing if these differential responders exist may lead to better exercise prescriptions and more favorable clinical end points. One way these individuals may be accounted for is by determining the magnitude of the measurement error (i.e. tester error in combination with biological variability). This variability

is most commonly assessed using two separate testing sessions separated by 24-48 hours but sometimes can be separated by up to 1-2 weeks (8-10). In addition, this short-term test-retest reliability would ideally be completed on a population similar to that which is going to undergo the long-term training study.

As noted in several recent papers (11, 12), short-term reliability may not be appropriate and, it is suggested, this is best assessed using control groups of similar duration as the actual intervention. Given the added burden this creates for the researcher by including a separate comparative arm that is not actually receiving an intervention, it is important to determine whether or not there are differences between short-term and long-term tests of reliability on biomarkers of muscle mass and function. Thus, the aim of this study was to examine the 24 hours test-retest reliability (Experiment 1) of forearm muscle size, handgrip strength, and an index of muscle quality in older adults and compare that to the reliability observed when visits are separated by 1 year (Experiment 2).

Methods

For Experiment 1, eighteen apparently healthy adults (men = 9, women = 9) between the ages of 42 and 64 [mean of 54 (SD 6)] years were measured twice with 24 hours between measurements. For Experiment 2, eighty healthy older adults (men = 34, women = 46) between the ages of 60 and 79 years [mean of 72 (SD 3)] were

Department of Health, Exercise Science, & Recreation Management, Kevser Ermin Applied Physiology Laboratory, The University of Mississippi, University, MS 38677, USA.

Corresponding Author: Takashi Abe, PhD Department of Health, Exercise Science, & Recreation Management, The University of Mississippi, 224 Turner Center, University, MS 38677, USA, Phone: +1 (662) 915-5844, Email: t12abe@gmail.com

Received April 6, 2018

Accepted for publication April 26, 2018

measured twice with 1 year between measurements. Participants had no orthopedic abnormalities (e.g. surgery or trauma) in their upper and lower extremities. All participants performed structured regular exercise (mainly walking and/or golf, three to five times per week) for at least 2 years. All participants signed a written informed consent to participate in the study, which was approved by the Ethics Committee of the University. The same investigator completed all ultrasound and strength measurements of the Experiments 1 and 2 (i.e. intra-observer reliability).

Participants were instructed to refrain from any vigorous physical activity for 24 h prior to the testing. Body mass and standing height were measured to the nearest 0.1 kg and 0.1 cm, respectively, by using an electronic weight scale and a stadiometer. During each visit, ultrasound images were taken from the anterior forearm for quantification of forearm muscle thickness. Muscle thickness was measured using B-mode ultrasound (Aloka SSD-500, Tokyo, Japan) on the right side of the anterior forearm at 30% of the distance from the styloid process of the ulna to the head of the radius. The measurements were made while subjects stood with the elbow extended and the forearm supinated. A linear transducer with a 7.5-MHz scanning head was coated with water-soluble transmission gel to provide acoustic coupling and reduce pressure by the scanning head to achieve a clear image. The scanning transducer was placed on the skin surface of the measurement site using the minimum pressure required, and cross sections of each muscle were imaged. Three images were printed (Toshiba Super Sonoprinter TP-8010, Tokyo, Japan). Muscle thickness was measured as the distance between the subcutaneous adipose tissue-muscle interface and muscle-bone interface of the ulna (MT-ulna), as described

previously (13), and the average of the three was used for data analysis.

Maximum voluntary handgrip strength was measured using a calibrated Smedley (TKK-5401 Grip-D, Takei Scientific Instruments, Tokyo, Japan) hand dynamometer. All participants were right handed and were instructed to: 1) maintain an upright standing position; 2) keep their arms at their side; and 3) hold the dynamometer in the right hand with the elbow extended downward without squeezing. Participants were allowed to perform one test trial followed by two maximum trials with a 1-minute rest period between attempts. The highest value was used for analysis.

Muscle quality in the forearm was defined as a ratio of handgrip strength to forearm muscle thickness (MT-ulna) (3). The MT-ulna includes two major flexor muscles and there is a strong correlation between MT-ulna and MRI-measured forearm flexor muscle cross-sectional area (14).

Data are presented as mean and standard deviation (SD). The mean and SD of the difference between Visit 1 and Visit 2 (SDdifference) was calculated for body mass, MT-ulna, handgrip strength, and the ratio of muscle strength to size. The minimal difference was formulated as follows: SDdifference \times 1.96. Pearson product correlations were performed to determine the associations between Visit 1 and Visit 2. The technical error of measurement (TEM) and the coefficient of reliability (R) were calculated (15). The coefficient of variation was also calculated as the SDdifference divided by the mean of the Visit 1 and Visit 2. Statistical significance was set at $P \leq 0.05$.

Results

For Experiment 1, the correlation coefficients between testing visits were 0.999, 0.984, 0.995, and 0.973 for

Table 1
Short-term (24 h) and long-term (1 yr) test-retest reliability of ultrasound measured forearm muscle thickness (MT-Ulna), handgrip strength, and forearm muscle quality (fMQ) in older adults

	24-h Test-Retest (n=18)			
	Visit 1	Visit 2	Difference	MD
Age (yr)	54 (6)			
Height (cm)	164.4 (9.0)			
Body mass (kg)	61.8 (12.4)	61.6 (12.4)	0.22 (0.57)	1.1
MT-Ulna (cm)	3.4 (0.4)	3.4 (0.5)	0.02 (0.04)	0.07
Handgrip (kg)	40.1 (10.9)	38.7 (11.4)	1.45 (2.03)	3.97
fMQ (kg/cm)	11.5 (2.0)	11.1 (2.2)	0.34 (0.55)	1.07
	One Year Test-Retest (n=80)			
	Visit 1	Visit 2	Difference	MD
Age (yr)	72 (3.8)			
Height (cm)	155.9 (8.0)			
Body mass (kg)	58.2 (9.2)	57.8 (8.9)	0.44 (2.9)	5.6
MT-Ulna (cm)	3.3 (0.3)	3.4 (0.3)	-0.09 (0.19)	0.37
Handgrip (kg)	31.8 (7.9)	29.5 (6.8)	2.3 (3.3)	6.4
fMQ (kg/cm)	9.4 (1.7)	8.5 (1.4)	0.91 (1.06)	2.07

MD, minimal difference

body mass, handgrip strength, MT-ulna, and the ratio of muscle strength to size ($p < 0.05$). The coefficients of variation were 0.4%, 3.9%, 0.8%, and 3.2% for body mass, handgrip strength, MT-ulna, and the ratio of muscle strength to size, respectively. For Experiment 2, the correlation coefficients between testing visits were 0.948, 0.908, 0.865, and 0.797 for body mass, handgrip strength, MT-ulna, and the ratio of muscle strength to size ($p < 0.05$). The coefficients of variation were 2.0%, 6.8%, 3.3%, and 8.5% for body mass, handgrip strength, MT-ulna, and the ratio of muscle strength to size, respectively. The minimal differences were greater in all measured variables following test-retest separated by one year than the test-retest separated by 24 hours (Table 1). The relative TEM values for handgrip strength, MT-ulna and the ratio of strength to size in Experiment 1 were acceptable (4.4%, 1.1%, and 4.0%, respectively) and all the R values were above 0.95 (0.98, 0.99, and 0.96 respectively). For Experiment 2, however, the relative TEM values were higher (9.4%, 4.5%, and 10.9%, respectively) and the R values were lower (0.85, 0.83, and 0.63 respectively) compared with the results of Experiment 1.

Discussion

The present investigation found large differences between short term (Experiment 1) and long term intra-rater test-retest reliability (Experiment 2) evaluated by the correlation coefficient, coefficient of variation, minimal difference, relative TEM, and coefficient of reliability. Many studies assess long term (i.e. months) changes in variables yet rely on a short test-retest (i.e. a few days) period for informing them on the error needed to surpass (i.e. minimal difference) in order to determine “real” or meaningful changes (8-10). In this study, the minimal difference between testing visits in forearm muscle thickness, handgrip strength and the ratio of strength to size was higher in the one year test-retest compared with the short-term test-retest. Similar results were observed in the TEM and coefficient of reliability. Our findings are of particular importance given the recent attention paid to responders (favorable and adverse) and non-responders to exercise (6, 9, 16).

Some of these studies have relied on short-term test retest reliability to inform them of the variability needed to surpass in order to classify someone into a particular category (6, 9). This is often completed on the same group of people included in the intervention or on a previous sample of individuals who are similar to those included in the present study. If short-term assessments could indeed assess a similar amount of variability as that observed over the course of a longer time frame, then it would allow for individuals to be placed into an actual exercise intervention group rather than a time matched non-exercise control group. This would help remove a recruiting burden on the investigator by not

having to recruit a control group and it would also remove the potential ethical dilemma of withholding perceived “treatment” (e.g. exercise). However, our findings suggest that a time-matched control group is likely important to fully capture the error of the tester as well as the error associated with random biological variability within a given time frame. Though it would have been preferable to perform the short-term test-retest on the same individuals as one year sample, we do not feel this potential limitation meaningfully impacts the interpretation.

Funding: This study was supported in part by the Japanese Society of Wellness and Preventive Medicine funded research.

Conflicts of interest: The authors declare that they have no conflict of interests relevant to the content of this study.

Acknowledgements: Our appreciation is extended to the volunteers who participated in this study.

References

- Lauretani F, Russo CR, Bandinelli S, Bartali B, Cavazzini C, Di Iorio A, et al. Age-associated changes in skeletal muscles and their effect on mobility: an operational diagnosis of sarcopenia. *J Appl Physiol* 2003; 95(5): 1851–1860.
- Rantanen T, Masaki K, Foley D, Izmirlian G, White L, Guralnik JM. Grip strength changes over 27 yr in Japanese-American men. *J Appl Physiol* 1998; 85(6): 2047–2053.
- Abe T, Thiebaud RS, Loenneke JP. Age-related change in handgrip strength in men and women: is muscle quality a contributing factor? *Age (Dordr)* 2016; 38(1), 28.
- Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee IM, et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. *Med Sci Sports Exerc* 2011; 43(7): 1334–1359.
- Pyka G, Lindenberger E, Charette S, Marcus R. Muscle strength and fiber adaptations to a year-long resistance training program in elderly men and women. *J Gerontol* 1994; 49(1): M22–M27.
- Bouchard C, Blair SN, Church TS, Earnest CP, Hagberg JM, Hakkinen K, et al. Adverse metabolic response to regular exercise: is it a rare or common occurrence? *PLoS One* 2012; 7(5): e37887.
- Loenneke JP, Fahs CA, Abe T, Rossow LM, Ozaki H, Pujol TJ, et al. Hypertension risk: Exercise is medicine* for most but not all. *Clin Physiol Funct Imaging* 2014; 34(1): 77–81.
- DeFreitas JM, Beck TW, Stock MS, Dillon MA, Kasishke II PR. An examination of the time course of training-induced skeletal muscle hypertrophy. *Eur J Appl Physiol* 2011; 111(11): 2785–2790.
- Barbalho MSM, Gentil P, Izquierdo M, Fisher J, Steele J, Raiol RA. There are no no-responders to low or high resistance training volume among older women. *Exp Gerontol* 2017; 99: 18–26.
- Loenneke JP, Rossow LM, Fahs CA, Thiebaud RS, Mouser GJ, Bemben MG. Time-course of muscle growth, and its relationship with muscle strength in both young and older women. *Geriatr Gerontol Int* 2017; 17(11): 2000–2007.
- Atkinson G, Batterham AM. True and false interindividual differences in the physiological response to an intervention. *Exp Physiol* 2015; 100(6): 577–588.
- Hopkins WG. Individual response made easy. *J Appl Physiol* 2015; 118(12): 1444–1446.
- Abe T, Thiebaud RS, Loenneke JP, Ogawa M, Mitsukawa N. Association between forearm muscle thickness and age-related loss of skeletal muscle mass, handgrip and knee extension strength and walking performance in old men and women: a pilot study. *Ultrasound Med Biol* 2014; 40(9): 2069–2075.
- Abe T, Nakatani M, Loenneke JP. Relationship between ultrasound muscle thickness and MRI-measured muscle cross-sectional area in the forearm: a pilot study. *Clin Physiol Funct Imaging* 2017 Aug 7. doi:10.1111/cpf.12462
- Perini TA, de Oliveira GL, Ornallas JS, de Oliveira FP. Technical error of measurement in anthropometry. *Rev Bras Med Esporte* 2005; 11(1): 86–90.
- Churchward-Venne TA, Tieland M, Verdijk LB, Leenders M, Dirks ML, de Groot LC, et al. There are no nonresponders to resistance-type exercise training in older men and women. *J Am Med Dir Assoc* 2015; 16(5): 400–411.