Published in final edited form as:

J Am Geriatr Soc. 2008 November ; 56(11): 2118–2123. doi:10.1111/j.1532-5415.2008.01953.x.

Tests of Muscle Strength and Physical Function: Reliability and Discrimination of Performance in Younger and Older Men and Older Men with Mobility Limitations

Nathan K. LeBrasseur, PT, PhD^{*,†}, Shalender Bhasin, MD, Renee Miciek, MS, and Thomas W. Storer, PhD

Section of Endocrinology, Diabetes & Nutrition Department of Medicine Boston University Medical Center Boston, MA 02118

Abstract

Objectives—To compare the reliability of muscle strength and physical function measures in younger and older men.

Design—Cross-sectional.

Setting—Academic research center.

Participants—Thirty younger men, 31 older men and 39 older men with mobility limitations.

***Correspondence:** Nathan K. LeBrasseur, PT, PhD Section of Endocrinology, Diabetes & Nutrition Boston University Medical Center 670 Albany Street, 2nd Floor Boston, MA 02118 Phone: 617.414.1031 Fax: 617.638.8217 nlebrass@bu.edu. **Alternate:** Shalender Bhasin, MD Section of Endocrinology, Diabetes & Nutrition Boston University Medical Center 670 Albany Street, 2nd Floor Boston, MA 02118 Phone: 617.414.2951 Fax: 617.638.8217 bhasin@bu.edu.

†Current Address: Pfizer Global Research & Development Cardiovascular & Metabolic Disease Eastern Point Road MS 8220-3001 Groton, CT 06340 Phone: 860.686.3165 Fax: 860.686.0565 nathan.lebrasseur@pfizer.com

Author Contributions Nathan K. LeBrasseur: Study concept and design, data acquisition, analysis and interpretation, manuscript preparation.

Shalender Bhasin: Study concept and design, data analysis and interpretation, manuscript preparation

Renee Miciek: Data acquisition, analysis and interpretation, manuscript preparation

Thomas W. Storer: Study concept and design, data analysis and interpretation

Sponsor's Role The sponsor's had no role in the design, methods, subject recruitment, data collection, analysis or preparation of the manuscript.

Elements of Financial/Personal Conflicts	*Author 1 LeBrasseur		Author 2 Bhasin		Author 3 Miciek		Author 4 Storer	
	Yes	No	Yes	No	Yes	No	Yes	No
Employment or Affiliation		X		X		X		X
Grants/Funds		X		X		X		X
Honoraria		X		X		X		X
Speaker Forum		X		X		X		X
Consultant		X		X		X		X
Stocks		X		X		X		X
Royalties		X		X		X		X
Expert Testimony		X		X		X		X
Board Member		X		X		X		X
Patents		X		X		X		X
Personal Relationship		X		X		X		X

Measurements—Test-retest measures of 1-repetition maximum (1RM), unloaded and loaded 50m walk and stair climb, and a lift and lower task. Reliability was assessed by intra-class correlation (ICC) analysis and the Bland Altman (BA) method.

Results—Leg and chest press 1RM measures identified significant differences between the groups, exhibited excellent test-retest reliability in younger men, older men and older men with mobility limitations (ICCs = 0.946–0.994) and minimal bias between trial 1 and 2 (BA = improvement of 21.1 and 1.1N for leg and chest press, respectively). Test-retest measures of the time to walk 50m and climb 12 steps also demonstrated excellent agreement (ICCs = 0.980–0.988 and 0.942–0.992, respectively) and minimal bias (BA = 0.755–1.007 and 0.141–0.361 sec faster, respectively). When a subject repeated these measures carrying a modest load ICCs remained > 0.940, bias was similar and the tests better discriminated between the groups. The lift and lower measure demonstrated excellent agreement (ICCs = 0.925–0.947), minimal bias (1.4–2.9 more shelves) and revealed significant differences between groups.

Conclusion—Measures of muscle strength and physical function can be performed in younger men, older men and older men with mobility limitations with high reliability. In future clinical trials, more challenging measures of performance may better discriminate amongst higher functioning study participants.

Keywords

Muscle strength; physical function; aging; sarcopenia; anabolic therapies

INTRODUCTION

The loss of muscle strength and limitations in physical function consequent to advancing age represent a far-reaching public health burden in the United States. Impairments in strength and limitations in physical function are predictive of falls^{1, 2}, disability³, hospitalization⁴, quality of life⁵ and even mortality^{6–8}. The aging of the population and socioeconomic costs of these sequelae have fueled a dramatic surge in the effort to develop anabolic therapies (e.g., selective androgen receptor modulators, growth hormone secretagogues and myostatin antagonists) that not only increase muscle mass, but improve muscle performance and physical function. Thus, proving the efficacy of future anabolic, function promoting therapies will be dependent on the selection of appropriate outcome measures that demonstrate high reliability and the ability to capture a change in performance in the chosen target population.

The objective of this study was to determine the test-retest reliability of common measures of muscle performance and physical function in younger men, older men and older men with mild to moderate limitations in physical functioning. We hypothesized that test-retest reliability of standardized measures would be higher in younger than in older men, and higher in older men than in older men with mobility limitations. We also hypothesized that when common measures of physical function were made more challenging they would better discriminate between the cohorts being studied. The data presented in this report are intended to aid in the design of future clinical trials.

METHODS

Subjects

Inclusion criteria for young healthy men included age of 18–50 years. Inclusion criteria for older men and older men with mobility limitations included age of ≥ 60 years, community-dwelling, baseline testosterone levels < 400 ng/dl. In addition, inclusion criteria for older men with mobility limitations included self-reported difficulty in walking two blocks on a

level surface or in climbing a flight of stairs and mild to moderate limitations in physical mobility as defined by a score of 4 to 9 on the Short Physical Performance Battery (SPPB)⁹. Subjects aged 45 years or more completed a supervised graded exercise test prior to enrollment. For all studies, exclusion criteria included myocardial infarction or fracture within the past 6 months and other orthopedic, cardiac (e.g., symptomatic coronary artery disease or uncontrolled hypertension), cognitive or neurological impairments that would prohibit participation. All subjects provided written informed consent. All study procedures were in accordance with institutional (Boston University Medical Center) guidelines.

Muscle Strength

Maximal voluntary strength of the lower extremities and the upper body was quantified by measuring the one repetition maximum (1RM) for the leg press and chest press, respectively, with Keiser A420 pneumatic resistance machines and integrated software (Keiser Sport, Fresno, CA). For the recumbent leg press, subjects started in a position of approximately 90 degrees of knee flexion and extended their legs to a position approximating 0 degrees. A standardized protocol was executed that consisted of warm up repetitions and a sequence of progressively increased resistances approaching the subjects' 1RM that were each separated by fixed rest periods. The resistance was then adjusted and 1RM attempts were performed, each separated by a set rest period, until the 1RM was determined.

For the seated chest press, subjects were positioned to align the machine's handles with the midline of the sternum in the horizontal plane and the anterior aspect of the chest in the frontal plane. The subjects' knees were positioned at 90 degrees of flexion and their feet firmly planted on the floor or an adjustable foot stool. The 1RM was determined using the protocol described above for the leg press. For 1RM measures, the first and second trials were separated by at least 2 but not more than 7 days.

50 Meter Walk

The time to walk 50m was measured using a switch mat and infrared timing system (Lafayette Instrument Company, Lafayette, IN) and recorded to the nearest 0.001 second. Subjects were instructed to walk as fast as possible (running was prohibited) and allowed to use assistive devices (e.g., canes and walkers) as needed. The first and second trials were separated by a rest period of 1.5 minutes.

Loaded 50 Meter Walk

The procedures to measure the time to walk 50m as fast as possible were repeated having the subjects carry two canvas tote bags equally weighted with standard Olympic weight plates whose sum was equivalent to 25 percent (young men and older men) or 20 percent (older men with mobility limitations) of their body weight. Subjects who were unable to perform this measure (e.g., due to impaired balance or inability to carry bags while using an assistive device) were excluded from the analyses.

Stair Climb

The time to ascend a single flight of stairs consisting of 12 steps (step height = 16 cm) was assessed using a switch mat timing system (Lafayette Instrument Company, Lafayette, IN) and time was recorded to the nearest 0.001 second. Subjects were instructed to climb the stairs as fast as possible while touching every step and allowed to use the handrail only if needed. The first and second trials were separated by a rest period of 1.5 minutes.

Loaded Stair Climb

The protocol to measure the time to ascend 12 steps as fast as possible were repeated having the subjects carry two equally weighted canvas tote bags (as described and used for the loaded 50m walk). Subjects who were unable to perform this task (primarily due to safety concerns or need to use a handrail) were excluded from the analyses.

Lift and Lower

As a measure of upper body function, subjects were instructed to lift a weighted basket (equivalent to 15 percent of body weight) from a shelf positioned at standard desk height (78.5 cm) and place it on a shelf positioned at their respective shoulder height, then to a shelf positioned at their respective head height and then to lower it back down in the reverse sequence. Subjects repeated this sequence as many times as possible in 1 minute and the number of shelves completed was recorded. The first and second trials were separated by a 2 minute rest period.

Data Analysis

Descriptive statistics were calculated for all subjects and are reported as means $\pm$ standard error. One-way analysis of variance (ANOVA) was used to compare means across groups. When ANOVAs generated statistically significant results, Tukey-Kramer post test analyses were performed to compare group means. The reliability of the measures was determined by examining the agreement between trial 1 and trial 2 using two methods: intra-class correlation (ICC) analysis and the Bland Altman method. For ICCs, a two-way mixed-model for repeated measures was used. In the Bland-Altman method, the difference between trial 1 and trial 2 (trial 2 subtracted from trial 1) was plotted against the mean of the two measurements. The mean difference between trials is reported as bias. The bias $\pm$ the coefficient of reliability (the standard deviation of the bias multiplied by 1.96) is reported as the limits of agreement. Descriptive statistics, ANOVAs, and Bland Altman analyses were performed using GraphPad Prism Statistical Software 4.03 for Windows. ICCs were calculated using SPSS 15.0.0 for Windows. Statistical significance was defined *a priori* as $p < 0.05$.

RESULTS

Subjects

Thirty younger men (38.7 ± 1.4 yrs, BMI = 26.3 ± 0.7), thirty-one older men (68.4 ± 1.4 yrs, BMI = 28.2 ± 0.7) and thirty-nine older men with mobility limitations (73.3 ± 0.8 yrs, BMI = 30.0 ± 0.8) who met the inclusion criteria were included in the analyses. One subject in the older men group and two subjects in the older men with mobility limitations group did not perform the loaded lower body or upper body physical function measures and were excluded from their respective analyses.

The Reliability of Measures of Muscle Strength in Younger and Older Men and Older Men with Mobility Limitations

Younger men exerted 559N (25%) and 1035N (58%) more force than older men and older men with mobility limitations in the leg press, respectively (both $p < 0.001$) (Figure 1A); and 1RM measures were 476N (27%) greater in older men compared to older men with mobility limitations ($p < 0.001$). Significant differences were also observed in the chest press between younger men (752 ± 36 N), older men (502 ± 20 N) and older men with mobility limitations (383 ± 19 N) ($p < 0.001$ for comparisons to younger men, and $p < 0.05$ between the 2 older cohorts).

Across all subjects, the ICCs for leg press and chest press 1RM test-retest measures were 0.990 (95% confidence interval (CI): 0.985–0.993) and 0.996 (95% CI: 0.994–0.997), respectively. In the Bland-Altman analysis, the mean difference (bias) between trial 1 and trial 2 for the leg press was an increase of 1.1% or 21.1N (95% CI: –47.5 to 5.32N) (Figure 1A) and for chest press was an increase of 0.4% or 1.12N (95% CI: –7.14 to 4.89N). Despite significant differences in mean values, the reliability, bias and limits of agreement (generated from the coefficients of reliability) of the 1RM measures in the 3 groups examined were highly similar and are detailed in Tables 1 and 2.

The Reliability of Measures of Lower Body Function in Younger and Older Men and Older Men with Mobility Limitations

Older men with mobility limitations required 14.84 more seconds to walk 50m as fast as possible compared to younger men ($p < 0.001$), and 10.49 more seconds than older men ($p < 0.001$) (Figure 1B). A statistically significant difference was not noted between younger and older men. Increasing the intensity of the task by having subjects carry loaded tote bags resulted in greater differences between younger men and older men with mobility limitations (21.53 ± 0.85 vs. 36.47 ± 1.67 sec, respectively) and resulted in a significant difference in the time (5.53 sec slower) for older men to walk 50m relative to younger men ($p < 0.05$). A similar trend was observed when comparing group means for the time to climb a standard flight of stairs. As illustrated in Figure 1C, there was not a significant difference in unloaded stair climbing time between younger men and older men. In the loaded trials, however, older men and older men with mobility limitations required 1.59 ($p < 0.05$) and 4.45 ($p < 0.001$) more seconds than younger men (3.55 ± 1.67 sec), respectively, to complete the task.

Across groups, test re-test measures of the time to walk 50m without and with a load demonstrated excellent reliability (ICC = 0.992 (95% CI: 0.988–0.995) and 0.993 (95% CI: 0.989–0.996), respectively). The Bland Altman analysis revealed a small bias, or improvement, between trial 1 and trial 2 of 0.915 sec (95% CI: 0.516 to 1.340 sec) for unloaded (Figure 1C) and 0.684 sec (95% CI: 0.359 to 1.009 sec) for loaded 50m walk measures. Examination of the three groups independently revealed that test re-test reliability was excellent in the unloaded as well as the loaded walking tests (Table 1), and $< 5\%$ bias (an improvement of 1.007 sec) was noted from trials 1 to 2 (Table 2). Similarly, the stair climb measures exhibited excellent agreement between the two trials. The ICCs were 0.990 (95% CI: 0.985–0.994) and 0.986 (95% CI: 0.978–0.991) for unloaded and loaded test re-test measures, respectively. The bias between trial 1 and trial 2 was an improvement of 0.226 sec (95% CI: 0.110 to 0.342 sec) for the unloaded stair climb (Figure 1D), and 0.275 sec (95% CI: 0.116 to 0.434 sec) for the loaded trials. Both groups of older men exhibited excellent reliability in unloaded and loaded stair climb measures (all ICCs > 0.940) (Table 1). Younger men also demonstrated excellent agreement in loaded trials, however, good agreement was observed between the two unloaded stair climb trials (ICC = 0.791). This appeared to result from the awkward nature of the task for the younger men who customarily run up a flight of stairs without touching every step as required in the protocol. Bias and limits of agreement for the stair climb measures are reported in Table 2.

The Reliability of a Measure of Upper Body Function in Young Men, Older Men and Older Men with Mobility Limitations

Comparisons of the means revealed older men and older men with mobility limitations completed 19% ($p < 0.05$) and 47% ($p < 0.001$) fewer shelf transfers than younger men, respectively; and older men with mobility limitations completed 34% fewer shelves than older men ($p < 0.001$) (Figure 4B).

When all subjects were combined, the test-retest reliability of the lift and lower measure was excellent (ICC = 0.972 (95% CI: 0.955–0.983). The bias between trial 1 and trial 2 was an improved score of 2.30 shelves (95% CI: –3.20 to –1.41 shelves) (Figure 1E). As detailed in Table 1, agreement was excellent within each of the 3 groups examined (all ICCs > 0.925) and bias was similar between groups (Table 2).

DISCUSSION

In this study we examined the reliability of key measures of muscle performance and physical function in younger men, older men and older men with mild mobility limitations. Despite differences in mean performance in the 3 cohorts examined, the described protocols and instrumentation for determining the 1RM and physical function demonstrated good to excellent test-retest reliability and minimal bias in each group. Importantly, when the selected measures of physical function were made more challenging we observed that they better discriminated between the two higher functioning groups and continued to demonstrate a high degree of reliability.

In the present study, the 1RM revealed significant differences in strength between younger men and older men and between the two groups of older men distinguished by their performance in a composite measure of mobility; namely, the SPPB. In addition, we demonstrate that the administration of a standardized 1RM testing protocol designed to provide adequate warm-up to optimize performance, prevent injury and minimize fatigue has excellent test-retest reliability for both the leg press and the chest press; even in older men with mobility limitations. The Bland Altman analyses further suggest that factors (e.g., familiarization or fatigue) commonly affecting performance outcomes are largely addressed with this 1RM testing protocol and independent of age and mobility status as the difference between trials is impressively small in each of the cohorts examined (e.g., an improvement of 15.71 or 40.66N in the leg press is roughly equivalent to 1.6 or 4.1kg) and subsequently, neither statistically nor clinically significant. Incorporation of highly reliable strength measures in future clinical trials (of either higher functioning or mobility limited older participants) may minimize subject burden by eliminating the need for repeated measures.

In older individuals, measures of physical function have emerged as the principal outcomes to characterize muscle health because of their strong predictive ability for hospitalization⁴, disability³ and even mortality^{6, 7}. Walking distance in a defined time period (e.g., the 6-minute walk) and the time to walk a predetermined distance (e.g., 4m, 8m or 400m) are the most studied outcomes of lower body function in epidemiological and intervention studies. The low ceiling of many of these measures is widely recognized. As an illustration, the time to walk an intermediate distance of 50m failed to distinguish older men from younger men despite significant differences in lower extremity muscle strength between these two groups.

In recognition of this issue, recent reports have supported the use of more challenging metrics of physical performance in higher functioning community dwelling older individuals, such as the timed 400m walk^{10–12}. In this study, we observed that adding a modest load to the 50m walk and stair climb as well as administering a challenging measure of upper body function, revealed differences between higher functioning younger men and older men without mobility limitations. These data support the notion that more arduous tasks display higher ceilings and may better discriminate among higher functioning participants in future clinical trials. We posit that these laboratory-based measures of physical function also have clear relevance to common and necessary activities of daily living (e.g., carrying groceries and performing a variety of household chores) and maintaining independence; similar to walking more moderate distances such as 400m¹³. The responsiveness of these metrics to changes in muscle mass or strength, however, remains to

be determined in efficacy trials of anabolic, function promoting therapies. It should be noted that 2 subjects that required walkers for ambulation and one subject with severe osteoarthritis in his hands were not able to perform the loaded trials as described. Thus, these more rigorous measures may not be beneficial in trials targeting more impaired populations.

The repeated measures of lower and upper body physical function demonstrated excellent agreement and reliability in the statistical methods employed. In the majority of measures (largely independent of group), however, the bias and its 95% CI suggests a consistent, albeit small, improvement in trial 2. We conclude that until minimal clinically important differences are clearly established for such metrics, performing two trials to capture a subjects' best performance will improve the quality of the captured data and its interpretation.

It should be noted that this relatively small study used the baseline data collected from subjects participating in 3 separate clinical trials. The 2 older cohorts were recruited to participate in studies investigating testosterone therapy in men with low total testosterone levels (< 400 ng/dl). This may influence the generalizability of the findings and qualify the comparisons to eugonadal younger men. In addition, all measures were performed by 2 unblinded examiners. While extreme steps were taken to standardize the testing protocols and prevent human measurement errors and testing bias, this should be disclosed.

In summary, well-designed protocols employing appropriate instrumentation that allow precise measurements of performance can quantify muscle strength and physical function in younger men, older men and older men with mobility limitations with similar high degrees of reliability. More challenging measures of lower and upper body physical function also demonstrate excellent reliability and provide better discrimination among higher functioning individuals. Future studies are warranted to examine the association of these described measures of physical function with muscle mass and strength and their responsiveness to function promoting anabolic therapies.

Acknowledgments

NKL was supported by NIH/NIA grant K01AG031154. This work was supported by NIH grants U01AG01436906 and R01HD047722 and a grant from Solvay Pharmaceuticals to SB. The authors would like to thank the study coordinators Newsha Lajevardi, Pierre Daou, Connie Dzekov and Emma Pinjic for their efforts in subject recruitment.

Funding Sources: NKL was supported by NIH grant K01AG031154. This work was supported by NIH grants U01AG01436906 and R01HD047722 and a grant from Solvay Pharmaceuticals to SB.

REFERENCES

- [1]. Tinetti ME, Speechley M, Ginter SF. Risk factors for falls among elderly persons living in the community. *N Engl J Med* 1988;319:1701–1707. [PubMed: 3205267]
- [2]. Low Choy NL, Brauer SG, Nitz JC. Age-related changes in strength and somatosensation during midlife: rationale for targeted preventive intervention programs. *Ann N Y Acad Sci* 2007;1114:180–193. [PubMed: 17934051]
- [3]. Guralnik JM, Ferrucci L, Simonsick EM, et al. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. *N Engl J Med* 1995;332:556–561. [PubMed: 7838189]
- [4]. Studenski S, Perera S, Wallace D, et al. Physical performance measures in the clinical setting. *J Am Geriatr Soc* 2003;51:314–322. [PubMed: 12588574]
- [5]. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *Jama* 1995;273:59–65. [PubMed: 7996652]

- [6]. Hardy SE, Perera S, Roumani YF, et al. Improvement in usual gait speed predicts better survival in older adults. *J Am Geriatr Soc* 2007;55:1727–1734. [PubMed: 17916121]
- [7]. Newman AB, Simonsick EM, Naydeck BL, et al. Association of long-distance corridor walk performance with mortality, cardiovascular disease, mobility limitation, and disability. *Jama* 2006;295:2018–2026. [PubMed: 16670410]
- [8]. Takata Y, Ansai T, Akifusa S, et al. Physical fitness and 4-year mortality in an 80-year-old population. *J Gerontol A Biol Sci Med Sci* 2007;62:851–858. [PubMed: 17702876]
- [9]. Guralnik JM, Simonsick EM, Ferrucci L, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol* 1994;49:M85–94. [PubMed: 8126356]
- [10]. Sayers SP, Guralnik JM, Newman AB, et al. Concordance and discordance between two measures of lower extremity function: 400 meter self-paced walk and SPPB. *Aging Clin Exp Res* 2006;18:100–106. [PubMed: 16702778]
- [11]. Chang M, Cohen-Mansfield J, Ferrucci L, et al. Incidence of loss of ability to walk 400 meters in a functionally limited older population. *J Am Geriatr Soc* 2004;52:2094–2098. [PubMed: 15571549]
- [12]. Pahor M, Blair SN, Espeland M, et al. Effects of a physical activity intervention on measures of physical performance: Results of the lifestyle interventions and independence for Elders Pilot (LIFE-P) study. *J Gerontol A Biol Sci Med Sci* 2006;61:1157–1165. [PubMed: 17167156]
- [13]. Hadley EC. Testing interventions to preserve walking ability: progress against disability, one step at a time. *J Gerontol A Biol Sci Med Sci* 2007;62:834–836. [PubMed: 17702873]

Figure 1. Differences in and the Reliability of Measures of Strength and Physical Function in Younger Men, Older Men and Older Men with Mobility Limitations*

(*Mean maximum voluntary strength for the leg press was determined by the 1RM measure (A) and revealed significant differences between groups. The time to walk 50m (B) and the time to climb a standard flight of stairs as fast as possible (C) were measured using an electronic timing system and demonstrated significant differences in younger and older men compared to older men with mobility limitations (+ML). Upper body function was determined by counting the number of shelves placed at desk, shoulder and top-of-head height that a subject could lift and lower a weighted basket onto in 60 sec and was significantly different between groups (D). Group means (+ SE) were derived from the subjects' best performance in two separate trials. * $p < 0.05$ and ‡ $p < 0.001$. Bland-Altman analyses for all subjects (younger men = closed triangles, older men = open squares, and older men with mobility limitations = open circles) are presented in the right hand column for the chosen measures. The mean difference, or bias (trial 1 - trial 2), is indicated by the solid horizontal line and the limits of agreement (difference $\pm$ coefficient of reliability) are demarcated by the dashed horizontal lines. For the 1RM (A) and lift and lower measure (B), the negative mean difference reflects that subjects consistently generated greater force and completed more shelves, respectively, in trial 2. For the 50m walk and stair climb, the positive mean difference indicates that subjects consistently completed the tasks in less time in trial 2.)

Table 1

Intra-class Correlation Coefficients (ICCs) of Performance Measures *.

	Younger Men	Older Men	Older Men + ML [†]
Leg Press 1RM	0.984 (0.969–0.992)	0.946 (0.879–0.976)	0.983 (0.964–0.992)
Chest Press 1RM	0.994 (0.989–0.997)	0.983 (0.960–0.993)	0.984 (0.964–0.993)
50 m Walk	0.987 (0.967–0.995)	0.980 (0.954–0.991)	0.988 (0.972–0.994)
Loaded 50 m Walk	0.977 (0.941–0.991)	0.988 (0.972–0.994)	0.991 (0.980–0.996)
Stair Climb	0.791 (0.486–0.915)	0.942 (0.872–0.973)	0.992 (0.983–0.996)
Loaded Stair Climb	0.940 (0.847–0.976)	0.976 (0.941–0.990)	0.978 (0.952–0.990)
Lift and Lower	0.925 (0.775–0.975)	0.962 (0.917–0.983)	0.947 (0.875–0.978)

* Test-retest measures of muscle strength and physical function were used for determination of ICCs. An ICC ≥ 0.80 represents good agreement and an ICC ≥ 0.900 is considered to reflect excellent agreement between trials.

[†] Mobility Limitations (ML)

Table 2

Bias and Limits of Agreement of Performance Measures^{*}.

	Younger Men			Older Men			Older Men + ML [†]		
	Bias	LOA		Bias	LOA		Bias	LOA	
Leg Press 1RM (N)	-15.71 (-57.73–26.32)	-283.34–251.93		-40.66 (-97.7–16.37)	-371.42–236.10		-25.04 (-63.66–13.58)	-231.42–181.33	
Chest Press 1RM (N)	-1.49 (-12.34–9.36)	-68.87–65.90		-5.95 (-15.08–3.19)	-48.34–36.45		3.71 (-6.54–13.96)	-47.08–54.49	
50 m Walk (sec)	0.935 (0.509–1.361)	-0.798–2.667		0.755 (0.140–1.371)	-2.166–3.677		1.007 (0.254–1.761)	-3.546–5.561	
Loaded 50 m Walk (sec)	0.472 (-0.071–1.016)	-1.804–2.748		0.331 (-0.221–0.884)	-2.292–2.955		1.186 (0.601–1.772)	-1.655–4.028	
Stair Climb (sec)	0.175 (-0.250–0.375)	-0.688–1.038		0.361 (0.170–0.553)	-0.589–1.311		0.141 (-0.068–0.349)	-0.972–1.254	
Loaded Stair Climb (sec)	0.205 (0.095–0.315)	-0.254–0.664		0.207 (0.071–0.343)	-0.482–0.900		0.390 (-0.033–0.812)	-1.702–2.481	
Lift and Lower (shelves)	-2.53 (-4.10 to -0.960)	-8.91–3.85		-2.93 (-4.31 to -1.54)	-9.78–3.93		-1.39 (-3.24–0.46)	-9.79–7.00	

^{*} Bland Altman mean differences (reported as bias (95% CI)) were calculated by subtracting the outcome of trial 1 to determine the reliability of the strength (1RM) and physical function measures. For strength measures, the generation of greater force in trial 2 would result in a negative bias, while for physical function outcomes, a faster performance in trial 2 would result in a positive bias. The inclusion of zero in the 95% CI indicates that no significant systematic difference was present (e.g., leg press 1RM for younger men). In contrast, for measures such as the 50m walk, the second trial is consistently better than the first and results in a positive bias and confidence interval that does not include zero. The limits of agreement (LOA) are generated by calculating the bias $\pm$ coefficient of reliability of the measures.

[†] Mobility Limitations (ML)