Reliability of Maximal Voluntary Muscle Strength and Power Testing in Older Men

E. Todd Schroeder, Ying Wang, Carmen Castaneda-Sceppa, Gregory Cloutier, Alberto F. Vallejo, Miwa Kawakubo, Nicole E. Jensky, Susan Coomber, Stanley P. Azen, and Fred R. Sattler

Departments of Biokinesiology and Physical Therapy, Preventive Medicine, and Medicine, Keck School of Medicine, University of Southern California, Los Angeles.

Nutrition, Exercise Physiology and Sarcopenia Laboratory, Jean Mayer USDA Human Nutrition Research Center on Aging at Tufts University, Boston, Massachusetts.

Background. Maximal voluntary muscle strength (MVMS) and leg power are important measures of physical function in older adults. We hypothesized that performing these measures twice within 7–10 days would demonstrate a >5% increase due to learning and familiarization of the testing procedures.

Methods. Data were collected from three studies in older adult men (60–87 years) and were divided into two cohorts defined by study site and type of exercise equipment. MVMS was assessed in 116 participants using the one-repetition maximum method at two separate study visits for the chest press, latissimus pull-down, leg press, leg flexion, and leg extension exercises along with unilateral leg extension power.

Results. Test–retest scores were not different and did not exceed 0.8 ± 9.0% in Cohort 1 or 2.3 ± 9.8% in Cohort 2, except for leg extension, which improved by 6.6 ± 14.4% (p < .009) and 3.4 ± 6.8% (p < .016), respectively. Repeat tests were closely correlated with initial tests (all p < .001). Pearson correlation coefficients ranged from 0.74 for leg extension power to 0.96 for leg press. Coefficients of variation were <10% (4.2%–9.0%) for all exercises except for leg extension power, which was 15.5%.

Conclusions. Our findings demonstrated that test–retest measures of MVMS and power in older adult men do not differ by more than 2.3% except for leg extension, and have relatively low coefficients of variation using data collected from three studies. Moreover, these findings were similar between two study sites using different equipment, which further supports the reliability of MVMS and power testing in older adult men.

MEASURES of maximal voluntary skeletal muscle strength and power are essential indicators of physical function in older adults (1–3) and are important for quantifying outcomes of study interventions. Few studies obtain multiple assessments of maximal voluntary muscle strength (MVMS) and power prior to interventions, which may result in an overestimation of the success of the intervention (e.g., resistance training regimens, anabolic drug therapies) if there is learning or familiarization with the testing procedures. This is of particular importance in older adults in whom improvements in strength may be small but have functional significance.

In a previous study (4), we reported average strength gains of approximately 11% following 12 weeks of treatment with an anabolic androgen in older adults. Performing two baseline strength assessments allowed us to confidently accept any changes in strength as a direct result of the study intervention rather than attribute those changes to learning and familiarization which may account for 5%–10% of the gains (5,6). Indeed, evidence suggests that in untrained older adults, a single testing session will result in neural facilitation and motor-learning adaptations that enhance strength (5,7). Therefore, subsequent testing sessions may demonstrate improvements in strength and power due to these adaptations. Apprehension about performing maximal strength and power exercises in a laboratory environment may limit performance during the initial evaluation session in those individuals not accustomed to such testing (8,9). Such psychological factors may contribute to submaximal performance on the initial evaluation followed by improvements in subsequent testing sessions.

The test–retest reliability of measures of maximal skeletal muscle strength in older adults have primarily been limited to isokinetic and isometric protocols (6,10,11) or handheld dynamometers (12). Similar studies have been conducted in young adults with the majority of investigations assessing reliability using isokinetic dynamometers (13). One-repetition maximum (1-RM) testing has been validated in older adult populations (14–19), is relatively safe (20), and is preferred to multiple repetition-maximum testing (e.g., 5-RM or 10-RM), which requires estimations to obtain maximal strength values and often is limited by muscular fatigue. However, few studies (5,16,19,21) have reported test–retest reliability for MVMS using the 1-RM method in older adults, and those few were limited to lower body exercises.

Regardless, performing multiple pre-intervention assessments of MVMS and power to familiarize participants with testing procedures may be impractical in some older persons with limited access to travel and resources (e.g., facilities or training personnel). Thus, it is important to validate the efficacy of performing repeated measures of 1-RM strength of both upper and lower body muscle groups and leg
extension power prior to study interventions. We, therefore, hypothesized that performing these tests twice within 7–10 days would demonstrate a greater than 5% increase due to learning and familiarization. Additionally, we sought to compare the test–retest reliability for the same measures conducted at a different university participating in our multicenter study.

METHODS

Study Design

The data reported here are from three studies in older adult men (n = 116). The first study (Study 1, n = 16) was an investigator-initiated, dose-ranging, double-blind, placebo-controlled trial to determine the dosing effect of an oral androgen on muscle and metabolism (4). The second study (Study 2, n = 32) was an investigator-initiated, double-blind, placebo-controlled investigation to determine the magnitude and durability of effects of an anabolic androgen (18,22). The third study (Study 3, n = 42) was a multicenter, investigator-initiated, double-blind, placebo-controlled investigation to determine the effects of testosterone and growth hormone on muscle and metabolism. All studies were performed at the University of Southern California (USC) National Center for Research Resources (NCRR)–funded General Clinical Research Center and the Clinical Exercise Research Center in the Department of Biokinesiology & Physical Therapy. Study 3 was also performed at the Nutrition, Exercise Physiology and Sarcopenia Laboratory in the Jean Mayer USDA Human Nutrition Research Center on Aging (HNRCA), Tufts University (n = 26). The studies and informed consents were approved and annually reviewed by the Institutional Review Boards of both the Los Angeles County-USC Medical Center and the Tufts–New England Medical Center.

Study Population

Men 60–87 years of age were recruited from the Los Angeles communities surrounding the USC Health Sciences Campus and the greater Boston area. In brief, to be eligible for these studies, participants must not have participated in regular resistance training, physical activity (with the exception of a walking program), or competitive sports for the previous 6 months. Participants had to have a body mass index (BMI) ≤ 35 kg/m², repeated resting blood pressure < 180/95 mmHg, prostate specific antigen (PSA) ≤ 4.1 ng/mL, serum hematocrit ≤ 50%, alanine aminotransferase less than three times the upper limit of normal, and serum creatinine < 2 mg/dL. Participants with untreated endocrine abnormalities (e.g., diabetes, hypothyroidism), active inflammatory conditions, or cardiac problems (heart failure, myocardial infarction, or angina) in the proceeding 3 months were excluded. A maximal cycle ergometer exercise test with a ramped protocol, 12-lead electrocardiogram, and blood pressure monitoring to achieve a heart rate ≥ 85% of age-predicted maximum was administered prior to resistance exercise testing to identify participants at possible risk for exercise-induced cardiac ischemia, abnormalities in heart rhythm, or abnormal blood pressure response during exercise testing.

Evaluation of Muscle Strength

MVMS was assessed using the 1-RM method (23) at the test and retest visits 7–10 days apart. The 1-RM was defined as the greatest resistance that could be moved through a defined range of motion using proper technique. Prior to strength testing, participants warmed up on a cycle ergometer or by walking for 5 minutes. At the first testing session, participants were instructed how to perform the selected exercises and were allowed to practice the exercises with minimal resistance. After the participants were comfortable with the equipment, they performed five warm-up repetitions at low intensity (estimated to be ≤ 50% of participant’s 1-RM based on his perceived exertion during familiarization) and 3–5 repetitions at higher intensity (estimated to be ≤ 75% of his 1-RM) for each exercise. The resistance was increased with each subsequent attempt, and the magnitude of increase estimated so that the participant failed at his respective maximum in fewer than eight attempts (to avoid fatigue). Ninety seconds of rest was given between each repetition attempt. The total test time to complete the 1-RM assessments for all exercises was 45–60 minutes. These procedures were standardized for all studies and at both study sites.

At USC, the 1-RM was determined for the bilateral leg press, leg extension, leg flexion, latissimus pull-down (lat pull), and chest press exercises on Keiser A-300 pneumatic equipment (Keiser Corp., Fresno, CA). The leg press and chest press machines displayed force as Newtons. Because Newtons cannot be accurately converted to kilograms on these machines, strength data are reported as Newtons for the two exercises at USC. Additionally, the leg flexion and extension exercises were not performed in Study 1. At the HNRCA site, all exercises (bilateral leg press, leg extension, leg flexion, lat pull, and chest press) were performed on selectorized weight-stack resistance exercise machines (Cybex VR2; Cybex International Inc., Medway, MA), on which the measures of force are reported in kilograms. The greatest 1-RM measured for each exercise during the two strength testing sessions was used as the value for MVMS. The same investigators performed all tests for each participant at their respective study sites.

Evaluation of Muscle Power

Unilateral leg extension power (W) using the Bassey Power Rig (University of Nottingham, Nottingham, U.K.) was only determined at USC for Studies 2 and 3, and has been described elsewhere (24). In brief, participants are seated and place the right foot on the foot pedal with the arms folded and trunk slightly forward. The participants are instructed to perform the leg extension procedure exerting as much force as possible and as fast as possible. At least 10–12 trials were performed until a plateau was reached and the highest score achieved was recorded for leg power.

Body Composition by Dual-Energy X-Ray Absorptiometry

Whole-body dual-energy x-ray absorptiometry (DEXA) scans (Hologic QDR-4500, version 7.2 software; Waltham, MA) were performed to quantify total and regional lean
RELIABILITY OF STRENGTH TESTING IN OLDER MEN

Table 1. Demographic and Body Composition Characteristics of Study Cohorts

<table>
<thead>
<tr>
<th>Variable</th>
<th>Cohort 1 (N = 90)</th>
<th>Cohort 2 (N = 26)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y</td>
<td>72 ± 5</td>
<td>69 ± 3</td>
</tr>
<tr>
<td>Height, m</td>
<td>1.74 ± 0.07</td>
<td>1.76 ± 0.08</td>
</tr>
<tr>
<td>Weight, kg</td>
<td>82.5 ± 11.9</td>
<td>84.4 ± 10.7</td>
</tr>
<tr>
<td>Body mass index, kg/m²</td>
<td>27.2 ± 3.5</td>
<td>27.3 ± 3.3</td>
</tr>
<tr>
<td>Total lean body mass, kg</td>
<td>60.8 ± 10.9</td>
<td>59.8 ± 6.5</td>
</tr>
<tr>
<td>% Fat</td>
<td>26.4 ± 5.7</td>
<td>26.2 ± 4.6</td>
</tr>
</tbody>
</table>

**Notes:** Cohort 1: University of Southern California (USC) using Keiser resistance machines; Cohort 2: Human Nutrition Research Center on Aging (HNRCA) using Cybex selectorized weight-stack resistance machines. SD = standard deviation.

Body mass and fat mass. One blinded, experienced technician analyzed all scans for both sites. The coefficient of variation (CV) for repeated measures was < 1% for lean and fat mass.

**Statistical Considerations**

Statistical analyses were performed by site (USC, Cohort 1; HNRCA, Cohort 2). Analyses for Cohort 1 (n = 90) included data from Studies 1, 2, and 3, and analyses for Cohort 2 (n = 26) included data from Study 3. For each cohort, summary statistics were calculated for demographic characteristics (age, height, weight, BMI) as well as body composition variables from the DEXA scan (total lean body mass and % fat). Pearson correlation coefficients were used to determine the association between the test and retest values for the strength variables. In addition, we calculated the absolute change (retest–test) and the relative change ((retest–test)/test) × 100, and tested for their significance using a Student paired t test. Random effect analysis of variance (ANOVA) was conducted to contrast between-and within-person variability to calculate CVs using the following formula: CV = 100 × (within-person standard deviation [SD]/within-person mean). For these analyses, the test variables were log transformed. Bland–Altman plots were used to describe the mean change for each exercise in both cohorts. All results are reported at the .05 level of significance (two-sided), and all analyses were performed using SAS 9.0 (Cary, NC).

**RESULTS**

Table 1 summarizes the demographic as well as body composition characteristics of the two cohorts that were similar except for age. Data from a total of 116 participants were analyzed, 90 from Cohort 1 and 26 from Cohort 2. In Cohort 1, the following number of participants did not perform the retest for these exercises: 2 for the leg press, chest press, and lat pull; 4 for the leg flexion; 5 for the leg extension; and 1 for the leg extension power test. In Cohort 2, 1 participant did not perform the retest for the chest press and 2 participants did not perform the lat pull exercises. Reasons for not completing these tests included sore joints (n = 11), fatigue (n = 2), or missed appointments (n = 2).

The MVMS and power tests are reported in Table 2. We observed consistent retest patterns across the tests for both study cohorts. The consistency across the two repeated tests is confirmed by the small (< 10%) CV from all exercises (range from 4.2% to 9.0%) (Table 2) except for the Bassey leg extension power, which had a CV of 15.5%. In addition, there was a strong linear association between the test and retest as confirmed by the large Pearson correlation coefficients (Cohort 1: range from 0.74 for the Bassey power test to 0.94 for the chest press; Cohort 2: range from 0.91 for the chest press to 0.98 for the leg extension, all p values < .0001).

Table 3 summarizes the test–retest differences for MVMS and power. For Cohort 1, the absolute mean change was 0, except for leg press (5 ± 146 Newtons), leg extension (3 ± 7 kg), and the Bassey power test (–3.8 ± 39.8 W). No significant difference was detected for most strength tests (p > .05) except for the leg extension test where an average improvement of 3 ± 7 kg was found to be statistically significant (p = .012). Similar results were observed for the relative change, where there was no significant change (p > .05) except for the leg extension where we observed a 6.6 ± 14.4% improvement comparing retest to test values (p = .009). We observed consistent results for Cohort 2 with

Table 2. Absolute Values and Pearson Correlation Coefficients for Maximal Muscle Strength and Power

<table>
<thead>
<tr>
<th>Test</th>
<th>Cohort 1</th>
<th>Cohort 2</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strength Test</td>
<td>N</td>
<td>Test</td>
<td>Retest</td>
</tr>
<tr>
<td>Leg press, Newtons</td>
<td>88</td>
<td>1367 ± 328</td>
<td>1372 ± 328</td>
</tr>
<tr>
<td>Chest press, Newtons</td>
<td>88</td>
<td>224 ± 50</td>
<td>224 ± 50</td>
</tr>
<tr>
<td>Leg flexion, kg</td>
<td>70</td>
<td>66 ± 13</td>
<td>66 ± 12</td>
</tr>
<tr>
<td>Lat pull, kg</td>
<td>88</td>
<td>55 ± 12</td>
<td>55 ± 12</td>
</tr>
<tr>
<td>Leg extension, kg</td>
<td>37</td>
<td>63 ± 15</td>
<td>66 ± 14</td>
</tr>
<tr>
<td>Bassey power, W</td>
<td>73</td>
<td>185.5 ± 53.4</td>
<td>181.6 ± 57.0</td>
</tr>
</tbody>
</table>

**Notes:** Cohort 1: University of Southern California (USC) using Keiser resistance machines; Cohort 2: Human Nutrition Research Center on Aging (HNRCA) using Cybex selectorized weight-stack resistance machines. CV = coefficient of variation; lat pull = latissimus pull-down.

*CV = 100 × (within-person standard deviation [SD]/within-person mean).

1All values significant, p < .001.

2Mean ± SD.
no significant change ($p > .10$) in absolute values for the leg press ($3 \pm 8$ kg), chest press ($-1 \pm 5$ kg), leg flexion ($1 \pm 5$ kg), and lat pull ($0 \pm 6$ kg). The leg extension exercise showed a significant $3 \pm 4$ kg ($p = .006$) absolute and $3.4 \pm 6.8\%$ ($p = .016$) relative increase from test to retest.

Bland–Altman plots are presented in Figures 1 and 2 with the $x$ axis representing the average of the test–retest scores and the $y$ axis representing the difference of the test–retest scores for each exercise. The plots from Cohort 1 (Figure 1) show a mean change of 0 for the leg press, chest press, leg flexion, and lat pull exercises. The leg extension exercise shows a mean change $> 0$ and more variability from test to retest. Additionally, the Bassey leg extension power demonstrated a small decrease in mean change with large variability among participants from test to retest. The plots from Cohort 2 (Figure 2) show a mean change near 0 for the leg press, chest press, leg flexion, and lat pull exercises, with the leg extension exercise showing the greatest mean change and variability from test to retest.

### Discussion

Our findings demonstrate that performing two tests prior to study interventions to account for familiarization and learning may not be necessary in older adult men for most strength and power tests. We hypothesized that performing one test followed by a second test 7–10 days later would result in a greater than 5% increase in MVMS and power measurements in our two cohorts. We theorized that the improvement in strength and power in a subsequent testing session would be the result of familiarization and learning of the testing procedures. Indeed, reports by other investigators (6,21,25) have demonstrated improvements after repeat measurements without intervention in older adults, and have attributed those improvements to motor learning adaptations. Contrary to these previous reports and rejecting our hypothesis, there were nonsignificant improvements of less than 2.3% in test–retest scores for all but the leg extension exercise in our two cohorts. In fact, the other five exercises performed differed by less than 0.7% in the test–retest data analyzed for Cohort 1, and had CVs less than 8% (except for leg extension power, which demonstrated the highest CV of 15.5%). Similarly, except for leg extension, the maximum difference for the four exercises in Cohort 2 was only 2.3% with a maximum CV of 7.8%. The higher CV of 15.5% for leg power may relate to the fact that this test is technique dependent; we speculate that it is somewhat more difficult to learn.

A second important finding is that the results are data compiled from three different studies in older adults conducted at one study site and corroborated by data collected in older adults from a second study site using different equipment. The likelihood of demonstrating test–retest reliability for strength and power testing should be greater when data from multiple studies and study sites are included in the analyses. The reliability between test–retest measures of strength and power is further supported by the fact that data collection methods differed (e.g., type of exercise machines, testers).

It is noteworthy that the leg extension exercise was the only measurement to show significant improvement in test–retest scores in both cohorts. Cohort 1 demonstrated a $6.6 \pm 14.4\%$ ($p = .009$) increase from test to retest, with Cohort 2 showing a smaller but also significant increase from test to retest of $3.4 \pm 6.8\%$ ($p = .016$). A possible explanation for this finding in Cohort 1 may be the small sample size of 37 compared to the 70 or 88 participants analyzed for the other exercises. However, Cohort 2 demonstrated the same finding with a sample size of 26 for the leg extension exercise and showed nonsignificant changes in four other exercises with a sample size ranging from 24 to 26. Thus, this finding is difficult to explain and may be inherent to measurement of strength for the leg extension exercise. Salem and colleagues (5) reported a significant 2.9% reduction in 1-RM strength for leg extension from Test 1 to Test 2 in persons aged...
51–78 years. Therefore, multiple baseline evaluations of MVMS for the leg extension exercise should be performed as there appear to be inconsistent findings across studies.

Previous reports on the reliability of strength testing using isokinetic dynamometers have demonstrated the need for multiple session testing (11), and have reported larger CVs compared to 1-RM testing (13,21). Findings from Ordway and colleagues (13) demonstrated large CVs using isokinetic dynamometry even with a familiarization session. Although the importance of a familiarization session on strength
measurements has been reported (26), we did not perform a separate familiarization session for 1-RM strength testing, and we report CVs that are less than the average (8%-10%) reported in other studies for isokinetic and isometric testing in older adults (11).

Although our findings may be of value to geriatric researchers and clinicians, there are several limitations to consider. First, our results relate only to relatively healthy, community-dwelling older men as tested in these studies. Performing 1-RM testing in younger populations or older
persons with known heart or lung disease, hypertension, or other chronic conditions may not produce the same results, although maximal strength testing appears to be safe and effective even in nonagenarians (15). Second, we did not evaluate test–retest reliability of maximal strength and power in older women; therefore, we cannot conclude that there would have been similar findings in women. Lastly, although our sample size was larger than those used in previous studies, our findings should be corroborated with larger studies using several hundred participants.

Summary

Our findings demonstrated that test–retest measures of MVMS and power in older men did not differ by more than 2.5% and have relatively low CVs, except for the leg extension using data collected from three studies. Our findings suggest that it may not be necessary to perform multiple baseline tests to eliminate or minimize potential test–retest improvement as a result of familiarization and learning of the testing procedures for most exercises. However, the significant change in the leg extension exercise and the large variability in the Bassey leg extension power test suggest that both may need to be repeated if selected. Moreover, these findings were similar between two study sites using different equipment, which further supports the reliability of MVMS and power testing in older men. Lastly, these findings may be of value to geriatric researchers studying sarcopenia as MVMS and power are important measures related to physical function in older adults.

Acknowledgments

This work was supported in part by National Institutes of Health (NIH) Grant R01 AG18169, National Center for Research Resources General Clinical Research Center (NCRR GCRC) Grant MO1 RR00043, by a grant-in-aid from Savient Pharmaceuticals and Solvay Pharmaceuticals, by the U.S. Department of Agriculture Cooperative Agreement 58-1950-9-001, and by NIH GCRC Grant MO1 RR000054.

We gratefully acknowledge the participants who committed substantial time and efforts to our studies.

Correspondence

Address correspondence to E. Todd Schroeder, PhD, University of Southern California, Departments of Medicine and Biokinesiology & Physical Therapy, 1540 East Alcazar St., CHP-155, Los Angeles, CA 90089. E-mail: eschroed@usc.edu

References