The reliability of the 1RM strength test for untrained middle-aged individuals

Itamar Levinger\textsuperscript{a,*}, Craig Goodman\textsuperscript{a}, David L. Hare\textsuperscript{b}, George Jerums\textsuperscript{c}, Deidre Toia\textsuperscript{d}, Steve Seliga\textsuperscript{a}

\textsuperscript{a}Centre for Ageing, Rehabilitation, Exercise and Sport, School of Human Movement, Recreation and Performance, Victoria University, Australia
\textsuperscript{b}Department of Cardiology and University of Melbourne, Austin Health, Australia
\textsuperscript{c}Department of Endocrinology and University of Melbourne, Austin Health, Australia
\textsuperscript{d}Department of Cardiology, Austin Health, Australia

Received 18 April 2007; received in revised form 2 October 2007; accepted 18 October 2007

**Summary** The one-repetition maximum (1RM) test is considered the gold standard for assessing muscle strength in non-laboratory situations. Since most previous 1RM reliability studies have been conducted with experienced young participants, it is unclear if acceptable test–retest reliability exists for untrained middle-aged individuals. This study examined the reliability of the 1RM strength test of untrained middle-aged individuals. Fifty-three untrained males ($n = 25$) and females ($n = 28$) aged $51.2 \pm 0.9$ years participated in the study. Participants undertook the first 1RM test (T1) 4–8 days after a familiarisation session with the same exercises. 1RM was assessed for seven different exercises. Four to eight days after T1, participants underwent another identical 1RM test (T2). Ten weeks later, 27 participants underwent a third test (T3). Intraclass correlation coefficients (ICC), typical error as a coefficient of variation (TEcv), retest correlation, repeated measures ANOVA, Bland–Altman plots, and estimation of 95% confidence limits were used to assess reliability. A high ICC (ICC $> 0.99$) and high correlation ($r > 0.9$) were found for all exercises. TEcv ranged from 2.2 to 10.1%. No significant change was found for six of the seven exercises between T1 and T2. Leg press was slightly higher at T2 compared to T1 ($1.6 \pm 0.6\%$, $p = 0.02$). No significant change was found between T2 and T3 for any exercise. 1RM is a reliable method of evaluating the maximal strength in untrained middle-aged individuals. It appears that 1RM-testing protocols that include one familiarisation session and one testing session are sufficient for assessing maximal strength in this population.

© 2007 Sports Medicine Australia. Published by Elsevier Ltd. All rights reserved.

**Introduction** Resistance training has been of great interest to exercise scientists and health professionals...
who study both healthy individuals, and those who suffer from chronic conditions including cardiovascular diseases, diabetes mellitus, and overweight/obese individuals. Since an increase in muscle strength is the most common and important benefit of resistance training, an accurate determination of muscle strength to properly evaluate the efficacy of the training is essential. Laboratory-based methods for evaluating maximal muscle strength include the use of isometric dynamometers and isokinetic dynamometers. These two methods, however, usually require sophisticated laboratory equipment and personnel trained in their use. Furthermore, these tests are not very specific for the types of movement patterns commonly used in typical fitness regimes. In contrast, the one-repetition maximum (1RM) method, defined as the maximal weight that can be lifted once with correct lifting technique, is comparatively simple and requires relatively inexpensive non-laboratory equipment. Moreover, because the 1RM test can be performed using the same patterns as those undertaken by the exercising individuals during their normal training, it is increasingly gaining acceptance as the gold standard for assessing muscle strength. Furthermore, previous studies have reported that the 1RM method to assess muscle strength is safe for healthy adults and also for patients with cardiovascular disease. The test—retest reliability of the 1RM demonstrates high intraclass correlation coefficients (ICC). However, as most 1RM reliability studies have been conducted with experienced healthy young participants (age 18–30 years), it is unclear whether this test—retest reliability is applicable to untrained middle-aged individuals who are increasingly the subject of exercise intervention studies. It has been suggested that the reliability of strength tests in older populations may be lower due to decreased muscle strength and joint stability. To date, there is a lack of data on the test—retest reliability of 1RM tests performed by untrained middle-aged individuals for a range of different resistance exercises. Most studies that have examined the test—retest reliability of maximal strength in middle-aged and older populations have used isokinetic dynamometers and not isoinertial-based (gym) machines.

A familiarisation process prior to 1RM strength testing is essential for ensuring reliable test results and minimize learning effect or systematic bias. Furthermore, it has been shown that without a familiarisation process prior to strength testing, there is a significant increase in the expression of muscle strength between two consecutive strength tests performed a few days apart. Some investigators have suggested that older individuals should undergo between 8–9 sessions of 1RM testing in order to increase the consistency of the 1RM measurements. Multiple familiarisation sessions, however, may not be practical for training studies that examine strength changes for several different resistance exercises as well as numerous other functional parameters. This would unnecessarily increase the time requirement of each participant and extend the duration of the study, both of which might increase the likelihood of participant dropout. As such, the purpose of this study was to examine the test—retest reliability of the 1RM strength test of untrained middle-aged individuals following one familiarisation session.

Methods

Fifty-three untrained males (n = 28) aged 51.2 ± 0.9 years volunteered to participate in the study (participants’ characteristics are shown in Table 1). Participants were included, whether or not they had cardiovascular risk factors such as overweight/obesity, hypertension, dyslipidemia and hyperglycaemia. Participants were on a range of medications including beta-blockers (two participants), calcium channel blockers (two), ACE inhibitors (four), diuretics (one), statins (two), metformin (one), and hormone replacement therapy (six). They were only included if they had not been involved in regular aerobic physical activity in the previous 6 months or resistance training in the previous 5 years or more. Participants were excluded from the study if they were involved in vigorous regular exercise activity for more than 90 min a week or had documented heart disease such as chronic heart failure or coronary artery disease. Each participant received an explanation about the nature of the study and after a medical clearance signed an informed consent document. The study protocol was approved by the Human Research Ethics Committees of both Victoria University and Austin Health.

Four to eight days prior to the first one-repetition maximum test (T1), participants performed a familiarisation session with the resistance training equipment (Life Fitness and CalGym, Caloundra, QLD, Australia). During the familiarisation session, correct lifting and breathing technique were taught and practiced using submaximal and near-maximal loads. 1RM was defined as the heaviest weight a participant could lift once with a proper lifting technique, without compensatory movements. 1RM strength was assessed for seven different exercises.
comprising in order, chest press, leg press, lateral pull-down, triceps pushdown, knee extension, seated row and biceps curl. Exercises involving large muscle groups were performed first followed by those involving small muscle groups. In order to facilitate the recovery and reduce the effect of fatigue, exercises were alternated between the upper and lower body. The tests commenced after a light warm-up (3-min walking at self-selected speeds on a treadmill). The maximal strength test protocol included one set of 10 repetitions at a relatively light load that served as a specific warm-up, followed by a gradual increase in load until 1RM was achieved. The rate of the gradual increase in load was dependent on the participant’s self-perceived capacity, and it ranged from 1 to 10 kg for biceps curl and triceps pushdown, up to 100 kg for leg press and between 1 and 20 kg for the rest of the exercises. The 1RM was achieved within 3–6 attempts. The rest period between attempts was 1 min, and between each specific exercise, volunteers recovered for 2 min. Four to eight days after the T1, participants underwent another 1RM test (T2) and, for a subgroup of 27 participants, again after 10 weeks (T3). T2 and T3 test protocols and the test conditions were identical to test T1, and the same investigators conducted the testing on each occasion. The 27 participants from the subgroup were asked not to change their activity level and this was monitored.

Reliability in this study was analysed according to the recommendation of Hopkins and Atkinson and Nevil. Intraclass correlation coefficients were used as they are a common measurement of reliability that enables comparisons with other studies. The ICC method was based on a repeat measurement of maximal strength (single value) (2, 1) and the same investigator conducted the tests. Retest correlation was measured by Pearson correlation coefficient. Within-subject variation was reported as typical error, and was expressed as a coefficient of variation (TEcv). This allows for comparisons between different studies, and is usually less affected by machine or operator error. Also, we report the lower and upper confidence limits (95%) of the above measurements. Changes in the means between tests, which examines random change and systematic bias, were assessed by repeated-measures ANOVA. Log transformation (multiplied by 100 to maintain precision) was performed prior to the data analyses as data were not normally distributed. Data distribution was examined by normal probability plots and the Kolmogorov–Smirnov test because heteroscedastic errors are the norm in studies in the area of exercise science. Log transformation was performed using a spreadsheet designed for this purpose. Bland–Altman plots were used to display visual representations of the errors against true values by plotting the difference between T2 and T1 against the mean of T2 and T1. Data were analysed using SPSS (Version 15). Data are presented as means ± standard error of measurements (SEM), unless otherwise stated. All statistical analyses were conducted at the 95% level of significance.

Results

No significant injuries occurred during the study, except for mild muscle soreness that is common with unaccustomed exercise for untrained individuals. Some participants reported this mild soreness 1–3 days after some of the tests.

Raw data of T1 and T2 for all seven exercises (53 participants) are presented in Table 2. Also reported are the ICC and the change in mean, TEcv (%) and the correlation between the two tests. A high ICC (>0.97) and high correlation (r > 0.9) were found for all exercises and for the total strength (sum of all seven exercises). TEcv (%) indicated typical errors range of 2.2–7.2%. In addition, no significant systematic bias was found between total weight lifted at T1 and T2, nor for six of the individual exercises including chest press, triceps pushdown, seated row, leg extension, biceps curl and lateral pull (Table 2). There was a significant

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Participants’ characteristics (mean ± S.E.M.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All participants</td>
</tr>
<tr>
<td>Gender (M/F)</td>
<td>25/28</td>
</tr>
<tr>
<td>Age (years)</td>
<td>51.2 ± 0.9 (40–69)</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>168.5 ± 1.3 (152.0–186.0)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>78.6 ± 2.3 (40–113.6)</td>
</tr>
<tr>
<td>Waist (cm)</td>
<td>91.9 ± 1.9 (59.1–120.5)</td>
</tr>
<tr>
<td>BMI (kg m⁻²)</td>
<td>27.5 ± 0.7 (16.7–39.22)</td>
</tr>
</tbody>
</table>

Note: BMI, body mass index.
The reliability of the 1RM strength test for untrained middle-aged individuals

Table 2  Muscle strength and the reliability for the 53 participants who underwent T1 and T2

<table>
<thead>
<tr>
<th>Exercise</th>
<th>T1 (kg ± S.E.)</th>
<th>T2 (kg ± S.E.)</th>
<th>ICC</th>
<th>Reliability after log-transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Δ mean (%) (95% CI)</td>
</tr>
<tr>
<td>Chest P</td>
<td>52.5 ± 3.5</td>
<td>52.5 ± 3.6</td>
<td>0.99</td>
<td>−0.7 (−3.1 to 1.7)</td>
</tr>
<tr>
<td>Leg P</td>
<td>195 ± 11.2</td>
<td>198.8 ± 11.4</td>
<td>0.99</td>
<td>1.6 (0.3 to 2.9)</td>
</tr>
<tr>
<td>Lat P</td>
<td>50.5 ± 2.3</td>
<td>50.0 ± 2.3</td>
<td>0.99</td>
<td>−1.1 (−2.3 to 0.2)</td>
</tr>
<tr>
<td>Tri P</td>
<td>27.4 ± 1.4</td>
<td>27.7 ± 1.4</td>
<td>0.98</td>
<td>0.9 (−1.1 to 2.9)</td>
</tr>
<tr>
<td>Knee E</td>
<td>46.9 ± 2.1</td>
<td>48.0 ± 2.2</td>
<td>0.97</td>
<td>1.7 (−0.7 to 4.2)</td>
</tr>
<tr>
<td>Seated R</td>
<td>50.8 ± 2.4</td>
<td>50.3 ± 2.4</td>
<td>0.99</td>
<td>−1.1 (−2.4 to 0.2)</td>
</tr>
<tr>
<td>Biceps C</td>
<td>18.8 ± 1.1</td>
<td>19.1 ± 1.2</td>
<td>0.98</td>
<td>1.1 (−1.6 to 3.9)</td>
</tr>
<tr>
<td>All</td>
<td>434.3 ± 21.0</td>
<td>438.2 ± 21.5</td>
<td>0.99</td>
<td>0.7 (−0.2 to 1.6)</td>
</tr>
</tbody>
</table>

Note: ICC, intraclass correlation coefficients; TE, typical error; CV, coefficient of variation; Chest P, chest press; Leg P, leg press; Lat P, lateral pull-down; Tri P, triceps push down; Knee E, knee extension; Seated R, seated row; Biceps C, biceps curl; Total, total weight lifted.

* p < 0.05.

Figure 1  (A) Bland—Altman plot for all 53 participants. (B) Bland—Altman plot for 27 subgroup of participants that also were tests at T3. Solid line represents mean difference between two tests. Dashed lines represent Δ difference between two tests ± 2 S.D. For both panels data point represents comparison for the total strength (sum of all seven exercises).

but small difference in leg press 1RM with T2 being slightly higher than T1 (1.6 ± 0.6%, p = 0.02). The Bland—Altman plot within cases for T1 and T2 for total strength demonstrated also that most results for individual participants are within 2 standard deviation (Fig. 1A.).

The reliability data of the 27 individuals in the subgroup who performed T1, T2 and T3 are presented in Table 3. ICC was >0.96 between T2 and T1 and T3 and T2. TECv (%) range was 1.9—10.1%. In addition, the retest correlation was >0.9. No significant systematic bias was found between any of the exercises. Most data points on the Bland—Altman plot for total strength were within 2 standard deviation for T2 versus T3 (Fig. 1B).

Discussion

Accurate evaluations of muscle strength are important to prescribe safe and effective resistance training intensities and to evaluate the efficacy of training. The main finding of the current study is that the 1RM test, after one familiarisation session, using standard resistance training equipment, is a reliable and simple tool for assessing maximal strength for untrained middle-aged individuals across a wide array of resistance exercises.

To our knowledge, this study is the first to examine the reliability of the 1RM test in a wide range of isoinertial resistance exercises, compared to only one or two exercises in previous studies. The present study had a relatively large sample size (n = 53) compared to other studies that examined the reliability of the 1RM test.²¹,²³,³² The present study demonstrated high ICC across all exercises examined with inexperienced middle-aged individuals. A high ICC was observed both between T1 and T2 and T2 and T3. These results are similar to the reported test—retest reliability of the 1RM test for experienced resistance-trained young individuals. The ICC of 1RM testing is specific to the exercise and ranges between 0.64 and 0.99 for exercises such as leg press,²⁰,²¹,³² bench press,²⁰,²¹,³³ lateral pull-down,²² and leg extension.²² However, it has been suggested that ICC may overestimate
### Table 3

<table>
<thead>
<tr>
<th>Exercise</th>
<th>T1 (kg ± S.E.)</th>
<th>T2 (kg ± S.E.)</th>
<th>T3 (kg ± S.E.)</th>
<th>ICC (T3-T2)</th>
<th>Reliability after log-transformation (T3 vs. T2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>∆ mean (%) (95% CI)</td>
<td>TE as a CV (%) (95% CI)</td>
<td>Rel. test r</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chest P</td>
<td>53.9 ± 5.6</td>
<td>209.7 ± 16.6</td>
<td>55.0 ± 5.9</td>
<td>0.972</td>
<td>−1.6 (−6.7 to 3.9)</td>
</tr>
<tr>
<td>Leg P</td>
<td>201.1 ± 15.5</td>
<td>203.4 ± 15.8</td>
<td>209.7 ± 16.6</td>
<td>0.993</td>
<td>1.0 (−1.1 to 3.2)</td>
</tr>
<tr>
<td>Lat P</td>
<td>52.5 ± 3.6</td>
<td>209.7 ± 16.6</td>
<td>55.0 ± 5.9</td>
<td>0.995</td>
<td>−0.3 (−1.6 to 1.0)</td>
</tr>
<tr>
<td>Tri P</td>
<td>28.0 ± 2.1</td>
<td>28.4 ± 2.2</td>
<td>28.7 ± 2.2</td>
<td>0.997</td>
<td>−0.8 (−2.6 to 1.0)</td>
</tr>
<tr>
<td>Knee E</td>
<td>46.8 ± 3.1</td>
<td>48.6 ± 3.2</td>
<td>47.3 ± 3.5</td>
<td>0.990</td>
<td>−3.6 (−6.7 to 0.7)</td>
</tr>
<tr>
<td>Seated R</td>
<td>53.0 ± 3.8</td>
<td>52.4 ± 3.8</td>
<td>52.7 ± 4.0</td>
<td>0.995</td>
<td>−0.5 (−2.0 to 1.1)</td>
</tr>
<tr>
<td>Biceps C</td>
<td>19.4 ± 1.7</td>
<td>19.6 ± 1.7</td>
<td>19.3 ± 1.7</td>
<td>0.997</td>
<td>−4.2 (−3.2 to 0.1)</td>
</tr>
<tr>
<td>Total</td>
<td>442.1 ± 31.9</td>
<td>446.2 ± 32.4</td>
<td>458.0 ± 32.7</td>
<td>0.964</td>
<td>48.8 (31.2 to 66.4)</td>
</tr>
</tbody>
</table>

Note: ICC, intraclass correlation coefficient; TE, typical error; CV, coefficient of variation; Chest P, chest press; Leg P, leg press; Lat P, lateral pulldown; Tri P, triceps pushdown; Knee E, knee extension; Seated R, seated row; Biceps C, biceps curl; Total, total weight lifted. Data were analysed after log transformation.

Typical error (within-subject variation) is an important measure of reliability in the field of exercise and sports science, and it reflects the change in scores between tests. Hopkins (2000) recommended reporting of the typical error as coefficient of variation (TEcv) as it allows direct comparisons between different studies, and it is also less affected by different machines, analysers or diverse cohorts of participants. The lower the TEcv value, the higher the reliability (less variation). It is difficult to assess if the TEcv in this study indicates high, good or low reliability, as we found only one other study that reported this value for maximal strength testing. Symons et al. examined the reliability of concentric, isometric and eccentric isokinetic knee extensor strength in older women. They reported that the TEcv ranged from 7 to 11% for the knee extensors. The authors concluded that the interpretation depends on the precision required but that this range was adequate for assessing muscle strength before training. In the current study, we found lower values between T1 and T2 (2.2–6.5%) and between T2 and T3 (2.4–10.1%), compared to Symons et al. The difference in the TEcv range may be due to the lack of a familiarisation session in their protocol testing.

Changes in the mean (systematic bias) also indicate consistency between tests. It has been assumed that an increase in muscle strength between two tests can be the result of a learning effect and as such, several familiarisation trials may be necessary before applying an exercise training intervention. In the current study, there were no significant changes in the maximal weight lifted from T1 to T2 for six of seven exercises. The only difference was observed in leg press exercise with the T2 being slightly higher than T1. However, this difference between leg press strength at T1 and T2 was small (representing a difference of only 1.9% or ~3 kg) compared to the expected change in strength following resistance training of least 15%. Further evidence suggesting that minimal or no learning effects took place in the present study may be seen for the participants that performed...
a third 1RM test (T3). No significant improvement was observed for all exercises from T2 to T3 (see Table 3).

Our study also demonstrates that similar to healthy young individuals,32,33 one familiarisation session may be sufficient before assessing maximal strength in an untrained middle-aged population. Previously we have shown that one familiarisation session is adequate for assessing maximal strength (isokinetic dynamometer) in elderly clinical populations.25 However, Ploutz-Snyder and Giamis23 reported that older individuals should undergo between eight and nine sessions of 1RM testing (isokinetic dynamometer) in order to increase the consistency of the 1RM measurement. It is important to note that the Ploutz-Snyder and Giamis23 protocol included a 1RM test every 48 h over 2—3 weeks. This type of familiarisation protocol may actually induce a training effect due to neural adaptations.9 This may be valid for investigators who wish to examine increases in muscle strength following resistance training due solely to muscle hypertrophy. However, to our knowledge, most trials with untrained or clinical populations use resistance training as a mode of clinical or therapeutic intervention.37 As such, changes in motor coordination (i.e. inter-muscular co-ordination) and other neural adaptations should be considered as part of the benefits of the training.38,39 In other words, if the aim of the study is to examine the early neural adaptations as a result of resistance training, then muscle strength should be assessed before chronic exposure to resistance exercise that occurs with long familiarisation processes.

Furthermore, in training studies with clinical populations, strength testing is usually not the only outcome measure with other physiological, biochemical, functional and psychological parameters also being evaluated. Due to the complexity of these types of studies, a long familiarisation process may lead to an increase in the dropout rates of participants, particularly those with low or marginal levels of motivation40 who would benefit the most from clinical intervention studies.

Practical implications

- The 1RM method is a reliable and simple method to evaluate maximal strength in wide array of resistance exercises in untrained middle-aged individuals.
- 1RM-testing protocols that include one familiarisation session and one testing session separated by 4—8 days are sufficient for assessing maximal strength in this untrained population.

References


