Reliability of Maximal Strength Testing in Older Adults

Wayne T. Phillips, PhD, FACSM, Alan M. Batterham, PhD, FACSM, Julie E. Valenzuela, MS, Lee N. Burkett, PhD

ABSTRACT. Phillips WT, Batterham AM, Valenzuela JE, Burkett LN. Reliability of maximal strength testing in older adults. Arch Phys Med Rehabil 2004;85:329-34.

Objectives: To determine (1) the reliability of a maximal strength test (1 repetition maximum) [1-RM] in older adults and (2) the impact of differing periods of familiarization.

Design: Within-subject, repeated trials of maximal strength.

Setting: Community-based senior center.

Participants: Forty-seven independently living men (n=16) and women (n=31), with a mean age of 75.4±4.7 years.

Interventions: None.

Main Outcome Measures: Systematic error (shift in mean) and random error (% coefficient of variation [%CV]) was assessed between consecutive pairs of 1-RM trials.

Results: For the bench press, systematic error was virtually eliminated for men between trials 2 and 3 (0.7%; 95% confidence interval [CI], –2.7% to 4.3%). The CV was stable (4.7%–7.3%) across all trials in both genders. For the leg press, a significant but clinically small systematic error (3.6%, $P<.05$; 95% CI, 0.8–6.6) was evident for women between trials 2 to 3. The CV was reduced across trial pairs by 3.3% for men and 0.9% for women. Three versus 6 or more sessions of familiarization produced small clinical differences in systematic error (≤4.1%) and CV (≤0.2%) between trials 2 and 3 for both lifts.

Conclusions: Reliability is an indispensable requirement for valid test outcomes. Our results show that, in this group of older adults, 3 familiarization sessions and 2 to 3 test trials produced highly reliable 1-RM measures. Additional periods of familiarization added little to test reliability. Effective reliability testing for 1-RM is a practical and attainable goal for outcomes based practitioners.

Key Words: Functional assessment; Measurement error; Elderly; Rehabilitation.

A GROWING BODY OF RESEARCH has consistently reported the importance of strength training in a variety of health-, functional-, and/or rehabilitation-related functional tasks.[1-5] The validity of such results, however, depends on accurate and appropriate assessment of reliability of the functional task under consideration and the strength measure being used. The value and importance of reliable assessment methods in geriatrics has been stressed for more than a decade.[6,7] However, although the literature contains numerous reports of reliability testing for functional tasks[8,9] and for isokinetic strength testing,[10-13] reliability is rarely reported for maximal isotonic strength testing (1 repetition maximum [1-RM]) in older adults.[14-16] Where reliability has been otherwise reported, it has typically been expressed solely as a correlation, usually a Pearson correlation coefficient or an intraclass correlation coefficient (ICC). The sole use of such statistics, however, is inappropriate and insufficient to accurately assess reliability[17] and has recently been the subject of discussion in this journal.[18] The Pearson correlation coefficient is a bivariate statistic and so is inappropriate for assessing reliability of a univariate repeated-measures test such as 1-RM.[19] The ICC, a univariate measure, is also inappropriate for confidently assessing reliability because, like the Pearson correlation coefficient, it is strongly influenced by the distribution of the data.[17,20] A wide between-subject variation, as recently illustrated by Keller et al[11] for isokinetic testing, will therefore lead to a high ICC regardless of the actual agreement between the variables under consideration. In addition, although the effects of lack of familiarization on reliability are well accepted,[19,21] few studies report pretesting familiarization sessions, and no study has previously reported the potential effect of differing periods of familiarization on the reliability of 1-RM testing.

Recent authoritative reviews and texts[19-23] have expressed concern at the lack of attention paid to reliability in the exercise science literature and have stated that, in order for reliability to be confidently assessed, both random and systematic sources of error should be appropriately measured. Random (within-subjects or typical) error includes biologic or psychologic variations within subjects and variance due to test equipment and apparatus or the personnel administering the test. Systematic error is a nonrandom difference in mean scores across trials and is generally considered to be due to learning effects secondary to a lack of familiarization. From a methodologic perspective, addressing systematic error may be even more relevant in studies involving older adults who typically report little familiarity with strengthening activities[24] and are thus more likely to exhibit systematic variability in baseline maximal strength testing responses.

This study assessed both systematic and random (typical) error components of reliability during repeated trials of 1-RM strength testing in a population of older adults and compared reliability assessments between participants who completed 3 sessions of familiarization with those who completed 6 or more sessions of familiarization. This is the first study of which we are aware that has (1) simultaneously and appropriately assessed both systematic and random error during 1-RM testing in older adults and (2) compared the reliability of differing lengths of familiarization.

METHODS

Participants

Forty-seven independently living men (n=16) and women (n=31) (mean age, 75.4±4.7y; range, 60–91y) were recruited from a local community senior center. Before entry into the study, all subjects were required to complete an extensive demographic and lifestyle questionnaire and to obtain physician consent for participation. Exclusion criteria included any

From the Department of Exercise and Wellness, Arizona State University East, Mesa, AZ (Phillips, Valenzuela, Burkett); and Department of Sport and Exercise Science, University of Bath, Bath, UK (Batterham).

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the author(s) or upon any organization with which the author(s) is/are associated.

Reprint requests to Wayne T. Phillips, PhD, FACSM, Dept of Exercise and Wellness, 7001 E Williams Field Rd, Arizona State University East, Mesa, AZ 85212-0180, e-mail: *wphillips@asu.edu*.

0003-9993/04/8502-8121$30.00/0
doi:10.1016/j.apmr.2003.05.010

Table 1: Subject and Familiarization Data

Subjects	Age ± SD (y)	Weight ± SD (lb)	3 Familiarization Sessions	6+Familiarization Sessions	Total N
Men	75.8±7.5	186.6±22.3	8	8	16
Women	75.2±17.3	157.0±40.1	12	19	31
Total	75.4±14.7	167.0±37.6	20	27	47

musculoskeletal, visual, orthopedic, or neuromuscular deficits that would preclude exercise and/or interfere with correct exercise technique; cognitive deficits as measured by the Folstein Mini-Mental State Examination[25]; and physician recommendation for nonparticipation. All subjects provided written consent for the study, which was approved by the institutional review board.

Procedures

All subjects attended Monday, Wednesday, and Friday for familiarization and testing sessions. Most subjects were new to resistance training (RT); however, some had attended a low-to-moderate intensity RT class for 6 to 30 weeks before the commencement of this study. Others, because of an extended delay in obtaining physician consent, were allowed to continue with their familiarization sessions before actual testing to maintain interest and adherence to the study. All subjects were initially naive to 1-RM testing and completed at least 3 sessions of familiarization before testing. During the familiarization period, subjects became acquainted with the test administrators, the exercise equipment, and exercise techniques. The importance of safe and effective lifting was continuously emphasized. We paid particular attention to low back positioning for the bench press in view of the increased risk for low back pain caused by excessive arching of the back during strenuous lifting efforts. Grip width was selected for each participant at a distance that required forearms to be vertical and elbows directly under the bar, thus providing a biomechanically efficient starting point for the lift.

Following extensive explanation and demonstration of correct technique, participants first performed each lift at a low intensity, with emphasis being placed on correct breathing and body mechanics. Over the course of the familiarization period, they progressed to multiple repetitions and sets of low-to-moderate intensity. Reliability testing was applied to a representative upper-body (supine chest press) and lower-body (seated leg press) lift. On test days, the test administrator again explained the test to the subject and demonstrated the actual lifts. To minimize potential measurement error, all testing was carried out at the same time of day, with the same testers, in the same facility, and with the same protocol. Immediately before testing, subjects performed a 5-minute general warm-up on a stationary recumbent cycle, followed by a more specific warm-up of the actual lift to be performed, using 5 to 10 repetitions of a light load as determined by the test administrator. All testing was conducted according to a standardized procedure[26] (appendix 1). Three separate testing sessions for each lift were conducted on nonconsecutive days. Both test lifts were performed during the same session in random order. All lifts were conducted on a Paramount plate-loaded multistation unit.[a]

Data Analysis

All analyses were conducted using SPSS, version 10,[b] statistical analysis software for Windows. Following the recommendations of Hopkins,[23] separate analyses were performed on consecutive pairs of trials (trials 1 and 2, 2 and 3) for both male and female samples. This permits an exploration of learning effects on the random error that may be obscured if all 3 trials were subjected at once to a repeated-measures analysis of variance (ANOVA). Pairwise plots (not illustrated) of each participant's difference scores (trial 2 minus trial 1, trial 3 minus trial 2) plotted against the mean for the 2 trials concerned suggested the presence of heteroscedasticity in the data, that is, differences between trials increased in proportion to the absolute amount of weight lifted in the 1-RM. This violated the assumption of homoscedasticity, or constant error variance. In addition, Shapiro-Wilks tests of pairwise difference scores revealed violations of the normal distribution assumption. Therefore, a normalizing and variance stabilizing natural logarithm transformation was applied to all data points.

Systematic bias (change in the mean) was calculated from a paired *t* test, with 95% confidence intervals (CIs) representative of the likely range for the "true" mean difference in the population. Mean differences between consecutive pairs of trials were expressed as a percentage change in mean. This was calculated via the back-transformation of the mean pairwise differences between the 2 trials according to the formula: (percentage change in the mean$=100[e^x-1]$), where x is the mean difference between trials for the log-transformed variables. The typical error or noise (random within-subject variation) from trial to trial was computed from the standard deviation (SD) of the trial to trial differences (SD_{DIFFS}). The random error is equal to the SD_{DIFFS} divided by $\sqrt{2}$ (equivalent to the root mean squares error [RMSE] from an ANOVA analysis). A percentage coefficient of variation (CV) can then be derived using the formula $\%CV=100(e^s-1)$, where s is the random error.[23] In addition to the trial-to-trial comparisons, the 3 trials in each exercise test were treated with a repeated-measures ANOVA. This enabled the calculation of the "average" random error across the 3 trials using the same formula as above, with random error defined as the square root of the RMSE term from the ANOVA variance partitioning of the log-transformed data input.

We also conducted a similar group analysis according to those who had received 3 sessions of familiarization and those who had received 6 or more sessions of familiarization and compared the relative clinical effect of these 2 familiarization periods on systematic and random error.

RESULTS

Fifty-three subjects were recruited to participate in this study (35 women, 18 men); however, 4 women and 2 men were unable to complete all test sessions for reasons not associated with the study and so were excluded from the analysis. Forty-seven subjects completed 3 test sessions each for the bench press and leg press. Twenty-one subjects completed 3 familiarization sessions while the remaining 26 completed between 6 and 30 sessions. No injuries were elicited during the testing process. Subject and familiarization data appear in table 1.

Table 2: Descriptive Data and Reliability Analyses for Consecutive Trials: Trial 2 Versus Trial 1

Test	n	Trial 1 Mean ± SD (lb)	Trial 2 Mean ± SD (lb)	% Shift in Mean (95% CI)	%CV (95% CI)
Bench press					
Women	31	62.6±16.5	63.3±16.8	1.1 (−1.5 to 3.8)	5.2 (4.3–7.3)
Men	16	102.4±30.3	107.3±31.1	5.0 (1.1–9.4)*	5.4 (4.0–8.7)
Leg press					
Women	31	215.4±47.8	216.0±51.1	−0.1 (−3.2 to 3.1)	6.3 (5.1–8.7)
Men	16	270.7±52.9	269.4±49.8	−0.4 (−5.1 to 4.6)	6.7 (5.1–10.9)

*Significant difference (P=.018) between trial means.

Reliability Results by Gender

Bench press. Men (table 2) demonstrated a significant demonstrated a significant systematic bias between trials 1 and 2, which was not apparent by trial 3 (table 3). Random error (%CV) was essentially stable across pairs of trials at approximately 5% of the mean. Random error for all 3 trials taken as a unit was also approximately 5% of the mean (table 4).

Women (tables 2, 3) demonstrated negligible systematic bias (≈1–2lb [≈.45–.9kg]) across all pairs of trials. Random error (%CV) was similar to that of the male group. Trials 1–2 (table 2) and for all trials taken as a unit (table 4).

Leg press. Men (tables 2, 3) demonstrated negligible systematic bias and a tendency for mean random error to decrease across all pairs of trials. The random error for all 3 trials taken as a unit was approximately 4% of the mean (table 4).

Women (tables 2, 3) demonstrated negligible systematic bias between trials 1 and 2, but a statistically significant, although small, change in the mean occurred between trials 2 and 3 (3.6%). The random error across consecutive pairs of trials was relatively stable (≈5%–6%). The random error for all 3 trials taken as a unit was approximately 6% (table 4).

There is little clinical difference between male and female reliability data. In table 2, the largest differences occurred in systematic error between the male and female bench press (3.9%); all other systematic mean differences were less than 2%. Random error was also similar for both the male and female bench press and leg press. Table 3 similarly shows small differences in systematic error (<2%) between men and women. Differences in random error also did not exceed 2.5%.

Reliability Results by Length of Familiarization Period

For the bench press, differences in systematic error between 3 and 6 or more familiarization periods were 4.1% between trials 1 and 2 (4.8% vs 0.7%), with the larger shift in the mean occurring with 3 familiarization periods (table 5). Differences were also 4.1% between trials 2 and 3 (–0.6% vs 3.5%), although interestingly the larger shift in mean now occurred with 6 or more familiarization periods. Random error differences did not exceed 1.1% for both pairs of trials.

For the leg press, the shift in mean was 2.9% for 3 familiarization sessions between trials 1 and 2, versus a similar shift (−2.4%) for 6 or more familiarization sessions but in the opposite direction (table 5). For trials 2 to 3, differences in systematic error between familiarization periods were virtually eliminated (3.2% for 3 sessions vs 3.0% for 6+ sessions). Differences in random error between the varying lengths of familiarization in the leg press were small and were reduced with the third test trial (2.6%–0.2%).

DISCUSSION

The identification of valid relationships between strength (1-RM) and health-related outcomes in older adults depends on the confidence one can place in 1-RM test findings and/or the identification of potential limitations to the test's reliability. Previous studies[15] have reported that a single day of familiarization is insufficient to elicit stable and precise measures of 1-RM. Other studies and authoritative reviews[11,18,20,23] have also reported that methods of assessing reliability have often been inappropriate or insufficient. In assessing both systematic and random error aspects of reliability, we found that, for both men and women, 3 familiarization sessions and 2 to 3 1-RM tests were sufficient to confidently determine the 1-RM in this group of older adults. For some lifts, a third 1-RM test trial added to the stability and/or precision of the test.

Bench Press

In our study, male test performance on the bench press demonstrated a small but significant systematic bias between the first 2 trials of 1-RM performance. The 95% CI indicated that the true systematic bias in the population might be as high as 9.4% of the mean. Hence, a learning effect across the first 2 trials was evident despite the prior familiarization sessions. There was no systematic bias, however, between trials 2 and 3 (table 3), indicating that the addition of a third trial was necessary to eliminate the learning effect and to determine a stable baseline. In contrast to this, women demonstrated no systematic bias across all pairs of trials (tables 2, 3), indicating that the prior familiarization sessions were sufficient to elimi-

Table 3: Descriptive Data and Reliability Analyses for Consecutive Trials: Trial 3 Versus Trial 2

Test	n	Trial 2 Mean ± SD (lb)	Trial 3 Mean ± SD (lb)	% Shift in Mean (95% CI)	%CV (95% CI)
Bench press					
Women	31	63.3±16.8	64.4±15.4	2.3 (–1.4 to 6.1)	7.3 (6.0–10.2)
Men	16	107.3±31.1	108.4±31.7	0.7 (–2.7 to 4.3)	4.7 (3.5–7.5)
Leg press					
Women	31	216.0±51.1	223.9±52.8	3.6 (0.8–6.6)*	5.6 (4.6–7.8)
Men	16	269.4±49.8	275.5±53.8	2.0 (–0.5 to 4.7)	3.4 (2.6—5.5)

*Significant difference (P=.016) between trial means.

Table 4: 1-RM ANOVA Results for Consecutive 1-RM Trials

Test	F	*P*	Overall %CV	95% CI
Bench press				
Women	$F_{2,60}=2.39$	.10	6.2	5.3–7.6
Men	$F_{2,30}=6.65$	.004*	4.9	3.9–6.5
Leg press				
Women	$F_{2,60}=3.26$	.04†	6.2	5.3–7.6
Men	$F_{2,30}=0.58$	.56	4.3	3.2–6.8

*Significant trial 1 to trial 2 (table 2).
†Significant trial 2 to trial 3 (table 3).

nate learning effects, and the addition of a third trial was not necessary to establish a stable baseline 1-RM. For men, the addition of a third trial reduced random error (increased precision), whereas in women random error increased by 2.1%.

Leg Press

No systematic bias was evident for men in our study across all 3 trials (tables 2, 3) and the mean random error decreased (increased precision) across consecutive pairs of trials. Hence, although a third trial appeared unnecessary with respect to systematic bias in this lift, where appropriate, it may be a useful strategy to reduce noise in the measurement and thus increase the ability to detect smaller changes in strength in a person after an intervention. The random error for all 3 trials taken as a unit was 4.3% of the mean (table 4).

In women (tables 2, 3), there was no systematic bias evident between trials 1 and 2. The comparison of trials 2 and 3, however, revealed a statistically significant, although small, shift in the mean 1-RM of 3.6% or approximately 8lb (3.6kg). This may indicate a delayed learning effect subsequent to increased confidence and/or more specific familiarization with exerting maximal effort in trials 1 and 2. The random error across consecutive pairs of trials for this lift was relatively stable (≈5%–6%) with a value of around 13 to 14lb (5.85–6.3kg; 6.2% of the grand mean) for all 3 trials taken as a unit (table 4). Although a shift in the mean of this magnitude may be small enough to be acceptable in some clinical or training settings, additional tests appear to be necessary if further reductions in systematic and random error were deemed appropriate or desirable. The magnitude of our random error scores were generally lower than those reported by others for isotonic testing.[16,27] However, a comparison of our data with previous work is problematic owing to lack of standardization in the methods used to assess reliability. Salem et al[16] estimated CVs for repeat 1-RM trials by calculating SDs for each person and then averaging these CVs to derive a CV for all participants. In this approach, however, the CV may be contaminated by shifts in the mean. Our approach, in contrast, calculates and presents the CV separately from the shift in the mean. Our measure of random error is thus uncontaminated and represents a more appropriate indicator of the precision of strength measurements.

Awareness of the importance of systematic and random error and the use of appropriate statistical techniques to accurately identify the magnitude (or limitations) of such measures is crucial to establishing a valid single measurement of strength. If a stable and valid baseline is not reliably established and reported, erroneous conclusions may be drawn from the data. Rikli et al[15] have elegantly illustrated the potential impact on systematic error when omitting familiarization and baseline reliability testing of 1-RM. Before implementing a strength-training program in a group of older adults, they reported significant differences between the means of the first 2 baseline 1-RM trials (large systematic error) but no differences between trials 2 and 3. The magnitude of this difference was such that had the first trial been taken as the "true" 1-RM, the apparent posttraining strength gains in this study would have been overestimated by 30% to 50%. Of relevance to our study, the single session of familiarization reported by Rikli was not sufficient to remove the large systematic error between the first 2 trials. More than a single period of familiarization appears to be necessary to counteract the effects of systematic error in 1-RM test measures in this population, and it would also be important to know the minimum or optimum number of familiarization sessions necessary to obtain acceptable reliability for 1-RM testing.

Familiarization Periods

Ours is the first study of which we are aware that has examined and reported both systematic and random error aspects of reliability and has compared the effect of differing lengths of familiarization on these measures. Table 5 displays reliability data by period of familiarization. Data for men and women were pooled because gender-specific analyses of familiarization sessions produced no meaningful differences in systematic or random error. Examination of table 5 shows little clinical difference between systematic and random error measures for 3 sessions or 6 or more sessions of familiarization. For both sets of lifts therefore we feel that such differences are not likely to be clinically meaningful in an applied setting and that the additional familiarization sessions add little or nothing to either the stability or the precision of the measure. Indeed, in some instances, longer familiarization periods paradoxically resulted in larger shifts in mean and larger random error measures than the shorter familiarization periods.

The data from this study suggest that, with our protocol, clinicians, researchers, or practitioners can be confident of detecting worthwhile increases in strength exceeding 5% to 6%. For example, if a training intervention were applied to a

Table 5: Descriptive Data and Reliability by Number of Familiarization Sessions

Familiarization Sessions	% Shift in Mean Trial 2 − Trial 1 (95% CI)	Random Error %CV (95% CI)	% Shift in Mean Trial 3 − Trial 2 (95% CI)	Random Error %CV (95% CI)
Bench press				
3 (n=20)	4.8 (1.8–8.1)	4.6 (3.6–7.0)	–0.6 (–4.6 to 3.6)	6.5 (5.0–9.9)
6+ (n=27)	0.7 (–2.3 to 3.9)	5.7 (4.6–8.1)	3.5 (0.1–7.1)	6.3 (5.1–9.0)
Leg press				
3 (n=20)	2.9 (0.0–6.0)	4.5 (3.5–6.7)	3.2 (0.0–6.6)	4.9 (3.8–7.4)
6+ (n=27)	–2.4 (–6.0 to 1.4)	7.1 (5.7–10.2)	3.0 (0.2–5.9)	5.1 (4.1–7.2)

NOTE. Data for men and women were pooled because individual analysis of male and female reliability data produced no meaningful differences in systematic or random error.

man from the sample in the current study, an improvement in bench press 1-RM of at least 6lb (2.7kg) would be required to have reasonable confidence in the precision of the change score. If strength gains of less than this were considered clinically meaningful (in the specific context in which the results were being used), then either more familiarization, more trials, or a more precise strength test would be needed to confidently detect the change.

CONCLUSIONS

Although the heterogeneous nature of the older adult population prevents extrapolation of our results to the general population, this study has highlighted the important contribution of familiarization and repeated trials to the reduction of systematic bias and random error in 1-RM testing for older adults. Following a familiarization period of 3 sessions before baseline testing, systematic bias did not exceed 3.5% within 3 test trials. Additional familiarization sessions did not improve reliability. The significant shift in the mean of 3.6% for female leg press at trial 3 was not considered clinically meaningful. The precision of our measure, represented as random error or %CV, was approximately 5% to 6% for both exercises in both genders. We believe this is an acceptable level of precision for most purposes, including the single measurement of strength to relate to other health, physical function, and/or rehabilitation outcomes, as well as the tracking of strength gains in individuals exposed to a training intervention.

We recommend that researchers, clinicians, and practitioners using 1-RM tests in older adults naive to resistance exercise, conduct and report reliability tests that address both systematic and random error. Three sessions of familiarization together with 2 to 3 1-RM test sessions appear to be sufficient to confidently establish reliability. We have shown that such procedures will reduce systematic bias and random error, maximizing the precision of measurements and the detection of clinically significant health or rehabilitation-related effects or relationships.

APPENDIX 1: SAMPLE SCRIPT AND PROTOCOL TO DETERMINE 1-RM BENCH PRESS

(A 1-RM is defined as the maximum amount of weight able to be lifted with good form)

NOTE. This protocol will be explained and demonstrated to participants during the familiarization period and immediately before to testing.

1. Explain which muscle groups the lift primarily affects. "The Bench Press primarily affects the muscles of the chest and the backs of the upper arm." NOTE. Tester will point to muscle locations as part of explanation.
2. Demonstrate lift with accompanying verbal explanation.
3. Position participant in basic lifting position. "Lie on the bench with your head toward the bar." "Place your feet on either side of the bench or on top of the bench, whichever is the most comfortable." NOTE. Tester will check for excessive arching of back and adjust feet and body position accordingly.
4. Explain and/or demonstrate correct grip. "Grasp the bar firmly at the positions indicated." NOTE. Hands are positioned so that a right angle is obtained at the elbow, forearms are vertical, and elbows are directly beneath bar.
5. Ensure body position is correct. "Make sure your body is in the center of the bench and your back is flat." NOTE. Tester will check for excessive arching of back, by sliding hand under lumbar spine, and adjust feet/body position accordingly.
6. Ensure bar is positioned in the correct starting position. "Make sure bar is directly over the mid-chest area and elbows are vertical." NOTE. Body position will be monitored and recorded by tester.
7. Adjust starting height of bar (where appropriate). "For bench press, elbow joint should be at 90°." NOTE. Bar position will be noted and recorded by tester.
8. Remind about correct breathing technique. "Take a breath in to prepare for the lift, and breathe out as you push the bar steadily upward." "Blow the bar up." "Breathe in as you lower the bar steadily."
9. Perform several lifts at low or zero resistance to reestablish familiarity with movement and correct lifting technique. Encourage and monitor technique at all times.
10. Set initial resistance at a level slightly above that of the warm-up resistance (ie, 1 or 2 blocks, 5–15lb [2.25–6.75kg]). This will vary between participants according to their perceived or observed effort during the warm-up.
11. Perform 1 lift with good technique.
12. Ask performer to rate how hard they perceived the lift to be on a rating of perceived exertion (RPE) scale of 6 (very, very easy) to 20 (most I could possibly do). NOTE. Tester will also monitor difficulty of lift by observing the speed and effort at which it is performed by the participant.
13. Rest 1 minute for RPE scores below 12, rest 2 minutes for RPE scores above 12.
14. Add 5 to 10lb (2.25–4.5kg) depending on the RPE.
15. Repeat process to momentary muscular failure (ie, they cannot continue) or to volitional fatigue (ie, they do not wish to continue).
16. Record maximum weight lifted. NOTE. When failure occurs, it may be appropriate in certain cases to remove some of the added weight and attempt another maximum effort at a slightly lower resistance. NOTE. If tester considers that a maximal effort was not attempted for any reason, this should be recorded in the Comments section of the data sheet. NOTE. Test should be repeated on at least 1 more occasion (separated by a minimum of 48h) to ensure reliability. NOTE. A similar protocol will be adopted for all lifts. Performers will be constantly monitored for feedback on any pain and/or discomfort experienced during the testing period. Correct breathing and lifting techniques will also be constantly emphasized during each exercise.

References

1. Hurley BF, Hagberg JM. Optimizing health in older persons: aerobic or strength training. Exerc Sport Sci Rev 1998;26:61-89.
2. Phillips WT, Haskell WL. "Muscular fitness"—easing the burden of disability in elderly adults. J Aging Phys Activity 1995;3:261-89.
3. Johnston R, Wolinsky FD. The structure of health status among older adults: disease, disability, functional limitations and perceived health. J Health Soc Behav 1993;34:105-21.
4. Fried LP, Guralnik JM. Disability in older adults: evidence regarding significance, etiology and risk. J Am Geriatr Soc 1997;45:92-101.
5. Guralnik JM, LaCroix AZ, Abbott RD, et al. Maintaining mobility in late life. I. Demographic characteristics and chronic conditions. Am J Epidemiol 1993;137:845-57.
6. Comprehensive Geriatric Assessment American Geriatrics Society (AGS). Comprehensive geriatric assessment position statement. 1993. Available at: http://www.americangeriatrics.org/products/positionpapers/cga.shtml. Accessed June 16, 2003.
7. Reuben DB, Solomon DH. Assessment in geriatrics. Of caveats and names. J Am Geriatr Soc 1989;37:370-2.
8. Jette AM, Jette DU, Ng D, Plotkin DJ, Bach MA. Are performance-based measures sufficiently reliable for use in multicenter

trials? Musculoskeletal Impairment (MSI) Study Group. J Gerontol A Biol Sci Med Sci 1999;54:M3-6.
9. Rikli RE, Jones CJ. Development and validation of a functional fitness test for community living older adults. J Aging Phys Activity 1999;7:129-61.
10. Ford-Smith CD, Wyman JF, Elswick RK, Fernandez T. Reliability of stationary dynamometer muscle strength testing in community-dwelling older adults. Arch Phys Med Rehabil 2001;82:1128-32.
11. Keller A, Hellesnes J, Broz JI. Reliability of the isokinetic trunk extensor test, Biering-Sorensen test and Astrand bicycle test: assessment of intraclass correlation coefficient and critical difference in patients with chronic low back pain and healthy individuals. Spine 2001;26:771-7.
12. Madsen OR. Trunk extensor and flexor strength measured by the Cybex 6000 dynamometer. Assessment of short-term and long-term reproducibility of several strength variables. Spine 1996;21: 2770-6.
13. Morris-Chatta R, Buchner DM, deLateur BJ, Cress ME. Isokinetic testing of ankle strength in older adults: assessment of inter-rater reliability and stability of strength over 6 months. Arch Phys Med Rehabil 1994;75:1213-6.
14. Landsdorfer JE, Phillips WT, Burkett LN. Reliability of maximal strength testing in older adults [abstract]. Med Sci Sports Exerc 2000;32(Suppl):S243.
15. Rikli RE, Jones CJ, Beam WC, Duncan SJ, Lamar B. Testing versus training effects on 1RM strength assessment in older adults [abstract]. Med Sci Sports Exerc 1996;28(Suppl):S153.
16. Salem GJ, Wang MY, Sigward S. Measuring lower extremity strength in older adults: the stability of isokinetic versus 1RM measures. J Aging Phys Activity 2002;10:489-503.
17. Bland JM, Altman DG. Statistical method for assessing agreement between two methods of clinical measurement. Lancet 1986; 1(8476):307-10.
18. Madsen OR. Reliability of muscle strength testing quantified by the intraclass correlation coefficient [letter]. Arch Phys Med Rehabil 2002;83:582.
19. Thomas JR, Nelson JK. Research methods in physical activity. 3rd ed. Champaign: Human Kinetics; 1996.
20. Batterham AM, George KP. Reliability in evidence-based clinical practice: a primer for allied health professionals. Phys Ther Sport 2000;1:54-61.
21. Baumgartner TA, Jackson AS. Measurement for evaluation in physical education and exercise science. 5th ed. Dubuque (IA): Brown & Benchmark; 1995.
22. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med 1998;26:217-38.
23. Hopkins WG. Measures of reliability in sports medicine and science. Sports Med 2000;1:1-15.
24. US Department of Health and Human Services. Physical activity and health: a report of the Surgeon General. Atlanta: Dept of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Diseases Prevention and Health Promotion; 1996.
25. Folstein MF, Folstein SE, McHugh PR. “Mini-mental state.” A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12:189-98.
26. Phillips WT. Strength assessment and training in older adults. Paper presented at: The International Conference on Aging and Physical Activity; 1997 Sept 18-21; Austin (TX).
27. Fiatarone MA, Marks EC, Ryan ND, Meredith CN, Lipsitz LA, Evans WJ. High intensity strength training in nonagenarians: effects on skeletal muscle. JAMA 1990;263:3029-34.

Suppliers

a. Paramount Fitness Corp, 6450 E Bandini Blvd, Los Angeles, CA 90040.
b. SPSS Inc, 233 S Wacker Dr, 11th Fl, Chicago, IL 60606.