Clinical Rehabilitation 2007; **21:** 258–265

Muscle strength testing with one repetition maximum in the arm/shoulder for people aged 75+ – test–retest reliability

E Rydwik, **C Karlsson** Department of Geriatric Medicine, Research and Development Unit, Jakobsberg Hospital, Karolinska Institutet, Järfälla, **K Frändin** Department of Neurobiology, Care Sciences and Society, Division of Physiotherapy, Huddinge and **G Akner** Department of Geriatric Medicine, Research and Development Unit, Jakobsberg Hospital, Karolinska Institutet, Järfälla, Sweden

Received 4th July 2006; returned for revisions 6th August 2006; revised manuscript accepted 25th August 2006.

Objective: To investigate the reliability of a muscle strength test of the arm/shoulder in elderly people, aged 75 and older, and to compare subjects with and without previous muscle strength training experience.
Design: Reliability study – test–retest.
Setting: Research centre for the elderly.
Main measures: One repetition maximum (1 RM) was measured using an arm/shoulder strength-training device (Pull Down, Norway). Two measurements were conducted, approximately one week apart.
Results: Forty people were included in the study and 34 completed both sessions. Eleven participants had previous muscle strength training experience on the indicated device. There was a high correlation between the test sessions, $r = 0.97$ for both groups. The analysis of 95% limits of agreement for the mean difference was $-4.3/+6.9$ kg for the group without and $-3.0/+6.4$ kg for the group with previous experience, respectively.
Conclusion: One repetition maximum evaluated by the Pull Down device seems to be a reliable and safe method for dosing and evaluating a muscle strength training programme for elderly people. The observed variation of approximately $-4/+7$ kg cannot be interpreted as an effect of muscle training, but is more likely an effect of learning, fluctuations in daily condition and/or motivation.

Introduction

Muscle loss occurs at a rate of 5% per decade starting in the fourth decade and in people 80 years and older muscle mass has on average declined 50% when compared with young controls.[1] However, decreases in voluntary strength do not become apparent before the age of 60.[2] These changes can lead to a decline in functional performance and reduced activities of daily living (ADL) and constitute a major component in the development of frailty.[3–5] It has been shown that high-intensive progressive muscle strength training has a favourable effect on muscle strength and functional performance for both healthy and frail elderly people.[6,7]

To determine the necessary intensity level for planning a progressive muscle strength training programme, different types of measurements are

Address for correspondence: Elisabeth Rydwik, Department of Geriatric Medicine, Research and Development Unit, Jakobsberg Hospital, Karolinska Institutet, 177 31 Järfälla, Sweden. e-mail: elisabeth.rydwik@sll.se

10.1177/0269215506072088

important. Muscle strength is defined as a maximum isometric contraction[8] or as the maximal force that a muscle or muscle group can generate.[9] Dynamic muscle strength can be measured by one repetition maximum (1 RM), defined as the maximum weight a person can lift only once in a complete range of motion.[10]

One repetition maximum has been shown to be reliable for experienced weight lifters in the upper and lower extremities.[11] Elderly people appear to need 2–3 test sessions to achieve a reliable test.[12–14] Muscle groups tested in these studies were mainly the leg press, bench press[12,13] and in one study knee extension.[14]

We hypothesized that elderly subjects with previous experience of muscle strength training would be more reliable in a test–retest investigation. The aim of this study was to investigate the reliability of test–retest of 1 RM of the arm/shoulder in elderly people, aged 75 and older, and to compare subjects with or without previous muscle strength training experience.

Material and methods

All subjects were recruited from a database of 1700 subjects that previously had shown interest in participating in research concerning nutrition and physical activity, aged 75 years and older living in the community of Solna, Sweden. Two groups were recruited: one group with no previous muscle strength training experience (group 1) and one group who had already experienced muscle strength training (group 2). The latter group had previously participated in a physical training programme for frail elderly (manuscript in preparation) and were recruited to this study only to allow comparison between subjects with and without previous muscle strength training experience, and there was no intervention during this test–retest study. Figure 1 shows a flowchart summarizing the inclusion process, including drop-out.

Exclusion criteria for both groups were age under 75, stroke during the last two years, myocardial infarction during the last six months, congestive heart failure above class 2 according to the New York Heart Association classification, and less than 7 points on the 9-item Mini-Mental State Examination short form.[15]

The study was approved by the Research Ethics Committee at Karolinska Institutet.

Examination of baseline characteristics

Physical activity level was estimated with a six-graded activity scale including household activities.[16,17] Personal ADL was estimated with the Swedish Functional Independence Measure (FIM) form.[18,19] Instrumental ADL was estimated using the Instrumental Activity Measure (IAM), developed as a supplement to FIM.[20,21]

Body weight and height were measured by standard procedures and the body mass index (BMI) was calculated by dividing the body weight (kg) by height^2 (m).

Procedure

The study was performed in our research centre for elderly people in Solna, a suburb to Stockholm, Sweden. One repetition maximum was measured using a muscle strength training device for the arm/shoulder (Pull Down, Norway) (Figure 2a,b). The 1 RM measurement was in accordance with the recommendations of the American College of Sports Medicine.[22]

The test leader (CK) conducted the measurements on each subject in two different test sessions, approximately one week apart. To avoid fluctuations in daily condition, the measurements were performed at about the same time of day. Several muscles and muscle groups were involved in the movement, for example the biceps brachii, pectoralis major and latissimus dorsi.

In the Pull Down device, the peak load is provided at the end of the range of motion, 180° shoulder flexion. The subjects started the measurement at zero degrees shoulder flexion (starting position), and were instructed to resist the movement eccentrically against the load up to 180° shoulder flexion, if possible, and then to perform a concentric movement back to starting position (Figure 2c). The concentric phase was the actual phase tested.

A measuring tape was used to ensure that full range of motion for each subject was achieved. The height of the chair was fixed. If the subjects did not reach the floor well enough to ensure stability, a footstool was used.

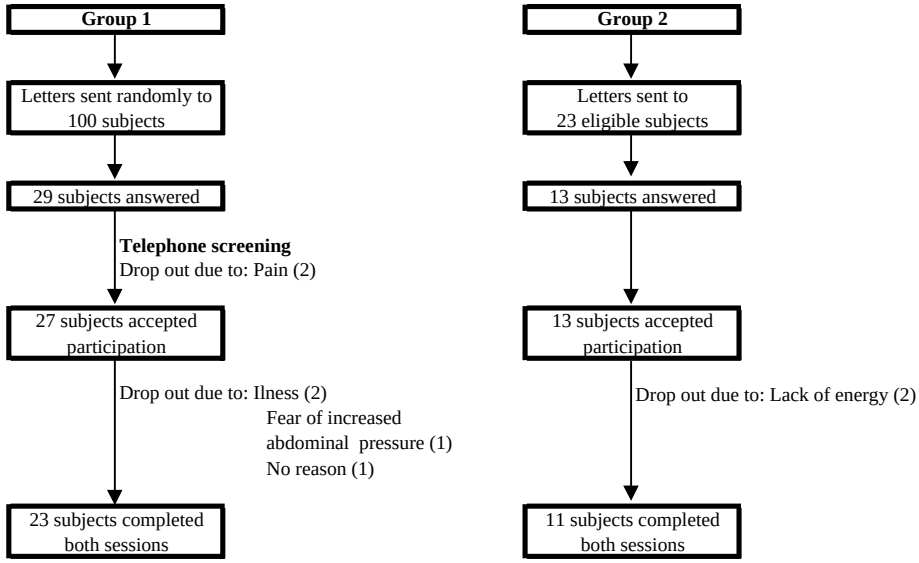

Figure 1 Flow of recruitment, inclusion and reasons for drop-out during the study.

Before the measurement began, the subjects were instructed to warm up for about 5 min, including walking in combination with movements of the arms and legs. For subjects with walking disabilities, these exercises took place in a sitting position.

After warming up, the subjects were familiarized with the device by performing 10 repetitions at the lowest load.

The test leader then estimated the starting load according to gender and BMI and if the subject could perform two repetitions, the load was increased until the subject could only perform one repetition in the individual full range of motion without compensatory behaviour. Resting periods between attempts were similar to those suggested by Phillips *et al.*,[13] where subjects were allowed to

Figure 2 (a) Starting position. (b) Full range of motion and peak of load. (c) End position.

rest for 1 min if scoring the effort ≤ 3 and 2 min if scoring the effort > 3 on the Borg CR-10 scale.[23]

Before the second session, the subjects were asked if they had experienced any injuries or muscle soreness after the first session, and if so how it had affected their daily life.

Statistical analyses

Pearson correlation analyses were conducted, as well as Bland and Altman's 95% limits of agreement[24–26] in JMP 5.0 (SAS Institute, USA) and Excel 2000 (Microsoft, USA). The 95% limits of agreement were introduced by Bland and Altman as an alternative and complement to the correlation coefficient for method comparison studies. Two methods may be highly correlated, yielding a high value for the correlation coefficient, although the agreement is low. These analyses were conducted for each group as well as for the total group.

To analyse the variance between groups in 1 RM, an *F*-test was conducted in SAS version 8.2 (SAS Institute, USA). The analyses of baseline characteristics were conducted in JMP 5.0 (SAS Institute, USA) using Student's *t*-test for continuous data with normal distribution and the Wilcoxon/Kruskal–Wallis test for ordinal data and continuous data with skewed distribution.

Continuous data are described with means and standard deviations (SD) and/or a confidence interval of 95% (CI 95%) and ordinal data with medians and 1st and 3rd quartiles.

Results

The baseline characteristics are shown in Table 1. Women in group 1 were significantly younger, had a higher BMI, were more physically active and less dependent in ADL compared with the women in group 2. Men in group 1 were significantly more physically active and less dependent in ADL compared with the men in group 2.

The differences at baseline were expected due to the inclusion criteria of group 2.

Test–retest

The subjects in both groups performed an average of six attempts (range 4–9) at each test session. There was a high correlation between the measurements, $r = 0.97$ for both groups respectively, and for the combined group as well (Figure 3). The 95% limits of agreement for the mean difference were $-4.3/+6.9$ kg for group 1 and $-3/+6.4$ kg for group 2 (Figures 4 and 5). In the combined group the 95% limits of agreement for the mean difference were $-2.6/+5.6$ kg.

The systematic bias was, on average, 1.3 with a 95% confidence interval [0.09–2.5] for group 1, 1.7 [0.2–3.3] for group 2 and 1.4 [0.5–2.4] for the total group, possibly indicating a small, but statistically significant, learning effect for both groups.

The *F*-test showed no significant differences between the groups.

Three subjects in group 1 and two in group 2 reported muscle soreness after the first test session, but not to the extent that it affected their daily lives. No injuries were reported.

Discussion

The study showed a high reliability of 1 RM for elderly people aged 75 and older using the Pull Down device. The reliability was equally high in

Table 1 Baseline characteristics

	Group 1 ($n = 23$)		Group 2 ($n = 11$)	
	Women ($n = 10$)	Men ($n = 13$)	Women ($n = 5$)	Men ($n = 6$)
Age, years (mean)	80.3 (3.1)	81.5 (3.1)	85.4 (1.7)*	83 (4.7)
BMI, kg/m^2 (mean)	26.5 (4.7)	24.4 (3)	21.8 (3.1)*	21.8 (4.6)
Physical activity, points (median)	4 (3–4)	4 (4–4)	2 (2–3)*	3 (2–4)‡
Personal ADL, points (median)	89 (88–91)	91 (91–91)	79 (78–85)*	85 (81–90)‡
Instrumental ADL, points (median)	55 (54–56)	54 (50–56)	23 (21–29)*	45 (25–48)‡

Values are mean (SD) or median (q1–q3).
*$P < 0.05$ between women in group 1 and 2, ‡ $P < 0.05$ between men in group 1 and 2.

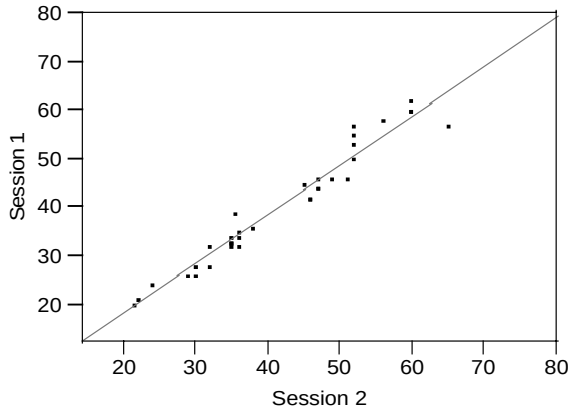

Figure 3 Results of sessions 1 and 2 in kg for the combined group, with line of equality. $r = 0.97$, $P < 0.0001$.

Clinical messages

- Using one repetition maximum for evaluating muscle strength in the arm/shoulder is feasible and safe for healthy and frail elderly people.
- One repetition maximum is highly reliable using the Pull Down device. A variation of −4/+7 kg cannot be considered a change in muscle strength, but rather an effect of learning, fluctuating daily condition and motivation.

both groups, despite the fact that there were significant differences between the groups at baseline. Furthermore, there was no significant difference in 1 RM between the groups indicating that previous muscle strength training experience does not affect the results; however this interpretation must be tested in a larger study.

The analysis of 95% limits of agreement showed that when 1 RM was performed in a second session, about 95% of the repeated measurements fell within −4/+7 kg of the first session. The limits of agreement were asymmetrically distributed around zero, due to a mean systematic bias of 1.3 and 1.7 kg for group 1 and group 2, respectively. This finding may indicate a learning effect or differences in motivation and daily condition rather than a true change in muscle strength.

The observed difference of −4/+7 kg may seem large, but one must consider the many muscles and muscle groups involved in the movement. When measuring only one muscle group, for example elbow flexors, a difference of −4/+7 kg would be much too high to consider the method reliable.

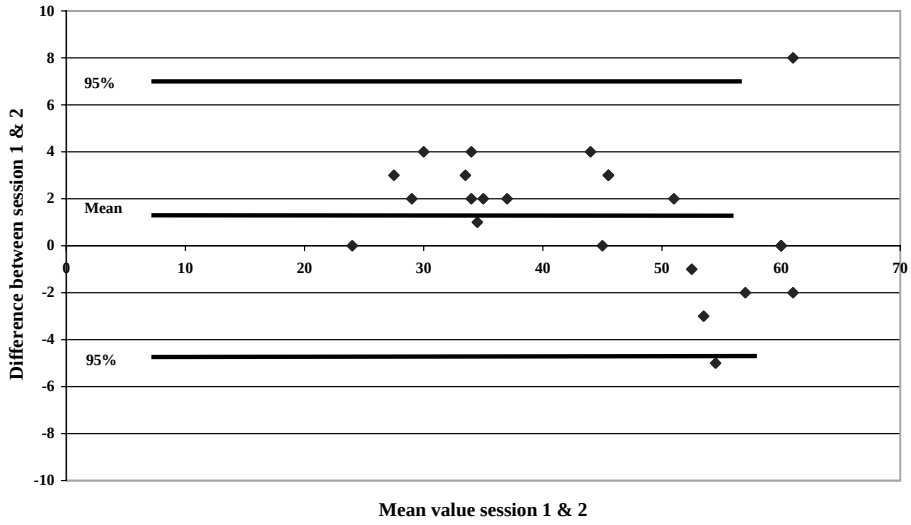

Figure 4 Bland and Altman 95% limits of agreement analysis (−4.3/+6.9 kg) for subjects in group 1.

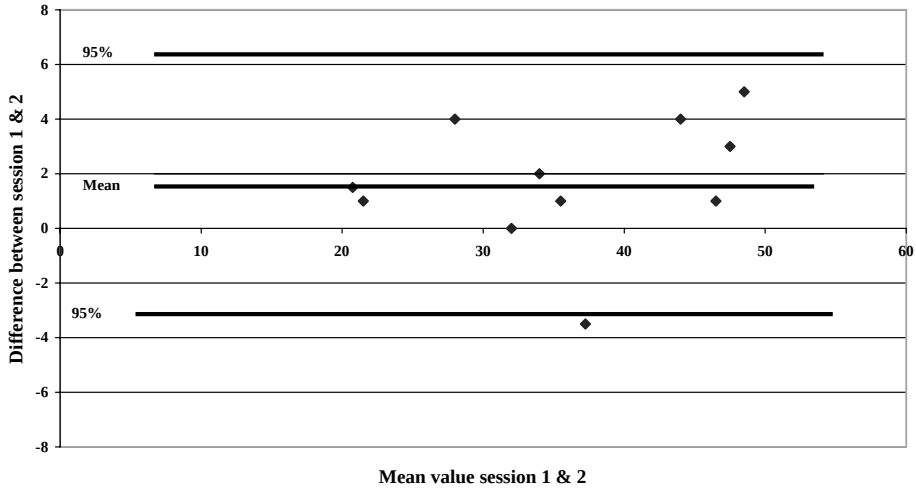

Figure 5 Bland and Altman 95% limits of agreement analysis (−3/+6.4 kg) for subjects in group 2.

Only 5 of 34 (15%) subjects reported muscle soreness after the first session and none of them reported that it had affected their daily living. No subject reported an injury. These observations suggest that it is safe to use the presented device to conduct 1 RM measurements in both healthy and frail elderly subjects. Previous studies reported injuries in a range of 2.4–19%.[27,28] One study reported muscle soreness in 70% of the subjects, however, without affecting daily living.[28] Our clinical experience in elderly care has shown that professional staff are often reluctant to motivate elderly people to push their limits, despite evidence of effect of high-intensive muscle strength training. Since our results show that 1 RM in the arm/shoulder is safe to conduct we believe that this study will contribute to a necessary change in attitudes.

One limitation of the study is that the subjects performed an average of six attempts before reaching 1 RM, indicating a possible risk of exhaustion before reaching the maximum weight. This possibility constitutes an intrinsic problem when measuring 1 RM. To circumvent exhaustion and decrease the risk of injuries, The American College of Sports Medicine suggests that after warm-up the initial load should amount to 60–80% of the predicted 1 RM and reaching 1 RM within 3–5 attempts.[22] This study fulfilled these requirements only for some of the subjects, mainly because it is more difficult to estimate the starting load for this group of elderly people.

Another limitation is the small sample size in group 2. Altman[29] argues that there should be at least 50 subjects in an analysis of 95% limits of agreement; otherwise there might be a risk that the ranges vary too much.

When implementing assessment methods in clinical settings, it is important to strive for simplicity, minimal device cost and a non-time consuming procedure. The 1 RM procedure presented here meets all of these requirements. The test can be conducted on existing equipment and is simple, highly reliable and safe for elderly people. Since progressive muscle strength training has been shown to be effective for elderly people, the importance of dosing and monitoring a muscle strength training programme is vital for high-quality management, while at the same time considering the effects of learning, fluctuating daily condition and motivation.

Acknowledgement

We would like to thank Göran Humble at Kebo Care for lending us the Pull Down device and Anna Törner for statistical advice.

Competing interests

There is no conflict of interest for any of the authors. Göran Humble, Kebo Care lent us the

device free of charge and was not involved in the design of the study.

Authors contribution

The first, third and fourth (ER, KF, GA) writers were involved in the design of the study. The second author (CK) was involved in the assessments and the analyses of the results, she also wrote the first draft in Swedish. ER wrote the article; KF and GA supervised and commented on the writing and analysis of the manuscript.

References

1 Porter MM, Vandervoort AA, Lexell J. Aging of human muscle: structure, function and adaptability. *Scand J Med Sci Sports* 1995; **5**: 129–42.
2 Morley JE, Baumgartner RN, Roubenoff R, Mayer J, Sreekumaran Nair K. Sarcopenia. From the Chicago meetings. *J Lab Clin Med* 2001; **137**: 231–43.
3 Fried LP, Tangen CM, Walston J *et al.* Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci* 2001; **56**: M146–56.
4 Kinney JM. Nutritional frailty, sarcopenia and falls in the elderly. *Curr Opin Clin Nutr Metab Care* 2004; **7**: 15–20.
5 Keysor JJ. Does late-life physical activity or exercise prevent or minimize disablement? *Am J Prev Med* 2003; **25** (suppl): 129–36.
6 Fiatarone Singh MA. Exercise to prevent and treat functional disability. *Clin Geriatr Med* 2002; **18**: 431–62.
7 Latham NK Bennett DA, Stretton CM, Andersson CS. Systematic review of progressive resistance strength training in older adults. *J Gerontol* 2004; **59A**: 48–61.
8 Enoka RM. *Neuromechanical basis of kinesiology*, second edition. Human Kinetics, 1994.
9 Wilmore HJ, Costill LD. *Physiology of sport and exercise*, second edition. Human Kinetics, 1991: 84.
10 Abernethy P, Wilson G, Logan P. Strength and power assessment, issues, controversies and challenges. *Sports Med* 1995; **19**: 401–17.
11 Sale DG. Testing strength and power. In MacDougall JD, Wenger HA, Green HJ eds. *Physiological testing of the high-performance athlete*. Human Kinetics, 1991: 21–103.
12 Rikli RE, Jones CJ, Beam WC, Duncan SJ, Lamar B. Testing versus training effects on 1 RM strength assessment in older adults. *Med Sci Sports Exerc* 1996; **28**: S153 abstract.
13 Phillips WT, Batterham AM, Julie E, Valenzuela JE, Burkett LN. Reliability of maximal strength testing in older adults. *Arch Phys Med Rehabil* 2004; **85**: 329–34.
14 Ploutz-Snyder LL, Giamis EL. Orientation and familiarization to 1 RM strength testing in old and young women. *J Strength Cond Res* 2001; **15**: 519–23.
15 Callahan CM, Unverzagt FW, Hui SL, Perkins AJ, Hendrie HC. Six-item screener to identify cognitive impairment among potential subjects for clinical research. *Med Care* 2002; **40**: 771–81.
16 Mattiasson-Nilo I, Sonn K, Johannesson K, Gosman-Hedström G, Person G.B, Grimby G. Domestic activities and walking in elderly: Evaluation from a 30-hour heart rate recording. *Aging* 1990; **2**: 191–98.
17 Frändin K, Grimby G. Assessment of physical activity, fitness and performance in 76-year-olds. *Scand J Med Sci Sports* 1994; **4**: 41–46.
18 Kidd D, Stewart G, Baldry J *et al.* The Functional Independence Measure: a comparative validity and reliability study. *Disabil Rehabil* 1995; **1**: 10–14.
19 Grimby G, Gudjonsson G, Rodhe M, Stibrant Sunnerhagen K, Sundh V, Östensson M–L. The functional independence measure in Sweden: Experience for outcome measurement in rehabilitation medicine. *Scand J Rehabil Med* 1996; **28**: 51–62.
20 Daving Y, Andrén E, Grimby G. Inter-rater agreement using the Instrumental activity measure. *Scand J Occup Ther* 2000; **7**: 33–38.
21 Grimby G, Andrén E, Daving Y, Wright B. Dependence and perceived difficulty in daily activities in community-living stroke survivors 2 years after stroke. *Stroke* 1998; **29**: 1843–49.
22 American College of Sports Medicine (ACSM). *ACSM's guidelines for exercise testing and prescription*. Williams and Wilkins, 1993.
23 Borg GA. Psychophysical bases of perceived exertion. *Med Sci Sports Exerc* 1982; **14**: 377–81.
24 Rousson V, Gasser T, Saifert B. Assessing intrarater, interrater and test-retest reliability of continuous measurements. *Statist Med* 2002; **21**: 3431–46.
25 Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999; **8**: 135–60.
26 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **i**: 307–10.
27 Pollock ML, Carroll JF, Graves JE *et al.* Injuries and adherence to walk/jog and resistance training

programs in the elderly. *Med Sci Sports Exerc* 1999; **23**: 1194–200.
28 Shaw CE, McCully KK, Posner JD. Injuries during the one repetition maximum assessment in the elderly. *J Cardiopulm Rehabil* 1995; **15**: 283–87.
29 Altman DG. Some common problems in medical research. In Altman DG ed. *Practical statistics for medical research*. Chapman and Hall, 1991: 398–403.