Intra-Day Reliability and Sensitivity of Four Functional Ability Tests in Older Females

Authors:

Susan Dewhurst, PhD[1,2]; Theodoros M. Bampouras, MSc[1*]

Affiliations:

[1]University of Cumbria, Faculty of Health and Wellbeing, Sport and Physical Activity, Active Ageing Research Group, Lancaster, UK

[2]University of Strathclyde, Strathclyde Institute of Pharmacy and Biomedical Sciences, Glasgow, UK

Correspondence:

Theodoros M. Bampouras

Faculty of Health and Wellbeing, Sport and Physical Activity, Active Ageing Research Group University of Cumbria, Bowerham Road, Lancaster, LA1 3JD

Phone: +44 1524 590837

Email: theodoros.bampouras@cumbria.ac.uk.

Disclosures:

Financial disclosure statements have been obtained, and no conflicts of interest have been reported by the authors or by any individuals in control of the content of this article.

Intra-Day Reliability and Sensitivity of Four Functional Ability Tests in Older Females

ABSTRACT

Functional tests are commonly used to evaluate functional ability of older individuals, however, intra-day reliability and sensitivity are required to enable informed decisions on whether repeated trials are necessary and to ensure that the values obtained from a single session are a patient's true score. The study aimed to investigate the intra-day reliability and sensitivity of commonly used functional tests in older individuals. Seventy one healthy older females (71.7 (7.3) years, 64.8 (10.2) kg, 1.58 (0.07) m) performed the 6m maximum walking speed, timed 8-foot up-and-go, chair sit-and-reach, and back scratch tests three times in one single session, with one minute between trials. Reliability was examined using intraclass correlation coefficient (ICC) while sensitivity using typical error (TE) between all trials. All tests were highly reliable (ICC range 0.89-0.99), indicating no need for a familiarization trial. TE between trials 2-1 were 0.06 ms^{-1}, 0.42 s, 1.13 cm, 0.92 cm for the 6m maximum walking speed, timed 8-foot up-and-go, chair sit-and-reach, and back scratch tests, respectively. Practitioners should perform two tests to examine whether the difference between them is less than the TE reported here. These results should help practitioners ensure that scores obtained from an individual from these functional tests are a true reflection of their functional ability rather than measurement error.

Key Words: Functional Ability Tests, Neuromuscular Performance, Test-Retest, Typical Error

INTRODUCTION

The ability to successfully perform everyday tasks (e.g. rising out of a chair, successfully crossing the road, avoiding raised objects) allows for independent ageing. Various exercise interventions have been used in older populations in an attempt to maintain functional ability.[1-4] To quantify the success of such interventions, functional assessments were developed to replicate everyday tasks[5,6] and, hence, evaluate the impact of the intervention on real-life situations.

Functional ability tests frequently used in older populations to assess their functional performance include the 6m maximum walking speed[6], the timed 8-foot up-and-go, the chair sit-and-reach and the back scratch flexibility tests.[5] Some of the advantage of these tests is the relevance to ‘real-life’ movements, low demand on resources, low equipment cost involved and the quick results they provide.[5] The applicability of such measurements aside, any assessment tool must be valid and reliable to avoid erroneous conclusions on the effectiveness of a particular intervention. Many of these tests have previously been shown to be valid and reliable when compared to more advanced measures.[5] Good test-retest (inter-day) reliability of functional tests is important in enabling accurate evaluation of intervention programmes (e.g. Capodaglio et al.[1], Carvalho et al.[2], Hallage et al.[4], Thomas et al.[6]). Indeed, intraclass correlation coefficient for test-retest reliability of these measures was high (i.e. > 0.90), indicating stability of the test scores over time.

However, when individuals are going through a functional ability screening process, their performance is assessed in a single session. It is therefore critical that the assessor is confident that that the score they measure is a true reflection of the individual’s ability on a given test. Similarly, it is common in a research setting that these tests are administered to compare performances between groups at a single point in time (e.g. Butler et al.[7]) and therefore for accurate comparisons to be made, the researcher must be confident in the scores recorded for each group. As reliability is an indication of measurement error (and, thus, high reliability suggests low measurement error), it is important to know the same-day test-retest (intra-day) reliability of these tests, to make informed decisions about their use and ensure appropriate amount of trials is given before a score is recorded.

Although reliability of a test is useful knowledge for the test itself, it provides little information and assistance to practitioners to make judgments from a single individual’s result. Every test performance includes an inherent random variation as a result of biological variability in the execution of movement.[8] Quantification of this random variation, can be made with the use of typical error (TE[8]), which provides a readily available score that indicates the magnitude of the random variation form measurement to measurement. Such a value allows the practitioner to determine whether the inevitable variability between two trials is within an acceptable range (i.e. equal to or below the TE score). Differences between two trials above the TE indicate that other factors are impacting on the result (e.g. lack of clarity in instruction, poor measurement technique, incorrect execution of task), and the test needs to be redone.

Given the importance of reliable intra-day testing and the need for a threshold that will guide and inform practitioner's decision on the correct number of test repetitions while avoiding unnecessary repetitions, the aim of the present study was to assess the intra-day reliability and sensitivity of these commonly used functional tests in older females. As females deteriorate faster than males in functional ability, they are in need of accurate screening processes to allow effective monitoring.[9]

METHODS

Participants

Seventy one healthy, physically active older females (mean (SD): age 71.7 (7.3) [range: 60-84 years] years, body mass 64.8 (10.2) kg, stature 1.58 (0.07) m) participated in the study after giving written informed consent. Participants had no known neuromuscular disorders and were considered medically stable, according to the criteria described by Greig et al.[10]. Ethical approval was obtained from the Ethics committee of the University of Strathclyde and all procedures followed were in accordance with the Helsinki Declaration of 1975.

Functional Ability Tests

Participants performed three trials of functional ability tests used to assess a number of parameters important in everyday living tasks.[5,6] All tests were done in one single session with one minute rest given between trials. Tests were performed in a randomized order and no prior familiarization was given for any test.

6m Maximum Walking Speed (SPEED)

SPEED evaluates neuromuscular function and has been found to improve with increased strength[11] and body weight unloading speed[6] in older adults. To assess maximal walking speed, participants started from a static standing position and walked as fast as they could to the end of a 9m course[6]. Visible markers were placed at the start, 6m and 9m.Time taken from start to 6m was recorded using a stop watch (Seiko, Seiko S-Yard Co Ltd, Tokyo, Japan) and speed was calculated.

Timed 8-Foot Up-and Go (TUG)

TUG poses various stresses to the neuromuscular system by a range of challenges including generation of leg force to lift the individual off the chair without using their arms and assume a balanced upright position, walk at high speed, change direction and return at high speed while turning to resume a seated position. To assess the integration of these parameters (power, speed, agility and dynamic balance)[5], time taken to raise from a seated position, walk 2.44m (8 feet), turn and return to the seated position, was recorded (Seiko, Seiko S-Yard Co Ltd, Tokyo, Japan).

Chair Sit-and-Reach (CSR)

The CSR test is a widely used test of back and hamstring flexibility[5]. To assess back and hamstring flexibility, while sitting on a chair with the legs stretched out in front, the participant was asked to reach down towards their toes. Participants were asked to maintain their foot at 90˚ of dorsiflexion with their toes relaxed in natural position. The distance between the extended fingers and the tip of the toes was measured. Left side (CSR_left), right side (CSR_right) and the average of the two (CSR) was used for further analysis.

Back Scratch (BS)

The BS is a widely used test assessing upper body flexibility[5]. To assess shoulder range of motion, the participant had one hand reaching down over the shoulder and the other one up the middle of their back. The distance between the extended fingers of the two hands was measured. Data was analyzed as left side (BS_left), right side (BS_right), depending on which hand was reaching down over the shoulder, and the average of the two (BS).

Data Analysis

Heteroscedasticity of data was checked by examining the uniformity of the scatter when change scores were plotted against the mean scores. As heteroscedasticity was absent raw scores were used for further analysis. Reliability and sensitivity were calculated as suggested by Hopkins[8]. Intraclass correlation coefficient (ICC, calculated as 1-typical error^2 / mean between-subject standard deviation between trials) and typical error (TE, calculated as standard deviation of the change scores between trials / square root of 2) were calculated between trials (ie trial 2 v trial 1, trial 3 v trial 2)[8]. ICC provided an indication of agreement between trials[11] while TE an indication of the error expected from measurement to measurement[8]. Descriptive data are given as mean (SD).

RESULTS

Descriptive statistics for all tests for all sessions can be found in Table 1. All tests produced high ICC (range: trial 2 v trial v 1, 0.89 – 0.98; trial 3 v trial 2, 0.90 – 0.99; Table 2), indicating high reliability between trials.

INSERT TABLE 1 AND 2 HERE

TE values for all tests can be found in Table 3. All tests produced low TE values, with almost all variables (SPEED, TUG, SR_right, SR, BS_right, BS) demonstrating a reduced TE between trial 3 v 2 compared to between trial 2 v 1 (Table 3). Hence, further reference to and suggestions about TE will be from trial 3 v 2.

INSERT TABLE 3 HERE

DISCUSSION

The present findings indicate that the tests have high intra-day reliability and sensitivity, as suggested by the very similar ICC and TE scores between trials 2-1 and 3-2. Therefore, we posit that firstly, a familiarization trial is not necessary and secondly, that practitioners should initially conduct two trials. If the difference between trial 1 and trial 2 is smaller than the TE reported here, then the practitioner can be confident that this is the patient's true score.

ICC has been widely used and suggested for reliability studies[12], however, its interpretation can be challenging, as various ICC thresholds have been used. For example, Fleiss[13] suggested an ICC > 0.75 as 'excellent', while Nunnally and Bernstein[14] stated that an ICC > 0.8 results from small measurement error. The ICC scores for all the tests in the present study confirmed the high reliability of the functional ability tests used as all ICC were above 0.8 suggesting high agreement between the measurements. ICC for inter-day reliability of these tests was provided as part of the tests development by Rikli and Jones[5] and ranged 0.90 – 0.96. The ICCs in the present study add to the high inter-day reliability of these functional tests, as they indicate high intra-day reliability too.

Notwithstanding the importance of validity and reliability in measurement, the sensitivity of a measure is an important factor[8,15] which is less widely reported. Although there is no uniformly accepted measure of sensitivity[15], the use of TE is suggested as the TE is easily interpreted and can be readily used to assess accuracy of the measurement.[8] As the TE indicates the error expected from repeating a test in raw units, it can be used as a threshold for its consistency. When using one of the functional ability tests described above, practitioners should record two performances. If the difference between the two scores is below the TE, they can be confident that this is the true score of the individual and no subsequent trial is required.

Of interest from the measurement of flexibility was the TE values for both the CSR and BS tests being different from left to right side (refer to Table 3). This, in addition to the high intra-participant variability, suggests that the use of an average value of the left and right, as typically reported for both tests[2,4,5], should be revisited. This average value may mask side differences that are important to identify. For example, the BS test involves a combination of shoulder movements (abduction, adduction) as well rotation (internal, external). These movements allow everyday tasks to be completed (e.g. combing hair, putting on clothes)[5] and therefore, it is of importance to know whether both sides are equally capable to achieve those aims. Similarly, any loss of ankle mobility on one ankle might impact on the CSR score, offering erroneous results on 'flexibility' of back and hamstrings. Unlike the recommendation by Rikli and Jones[5] to present the average of the left and right for the CSR and the BS tests, we suggest that each arm movement is examined separately to enable examination of flexibility differences as well as application of a more reflective TE.

CONCLUSIONS

The functional ability tests examined in the present study are highly reliable when performed within a short period of time and can reflect the individual's real score. In addition, assessment of an individual's performance during a functional ability screening can be easily achieved by immediate comparison of their values to the TE provided here. As the ICC and TE scores between trial 2-1 are similar to scores between trial 3-2, this suggests that no familiarization trial is needed for these tests. Two trials should be performed to allow the practitioner to assess whether the difference obtained is less than the TE reported here, meaning the practitioner can be confident that it is a true score. Future studies should consider the use of separate left and right side flexibility measures, as well as separate left and right chair sit-and-reach flexibility measures, in order to examine side to side flexibility differences.

REFERENCES

1. Capodaglio P, Capodaglio EM, Ferri A, et al. Muscle function and functional ability improves more in community-dwelling older women with a mixed-strength training programme. *Age Ageing*. 2005;34:141-147.

2. Carvalho MJ, Marques E, Mota, J. Training and detraining effects on functional fitness after a multicomponent training in older women. *Gerontology*. 2009;55:41-48.
3. de Vreede PL, Samson MM, van Meeteren NL, et al. Functional-task exercise versus resistance strength exercise to improve daily function in older women: a randomized, controlled trial. *J Am Geriatr Soc*. 2005;53: 2-10.
4. Hallage T, Krause MP, Haile L, et al. The effects of 12 weeks of step aerobics training on functional fitness of elderly women. *J Strength Cond Res*. 2010;24:2261-2266.
5. Rikli RE, Jones CJ. Development and validation of criterion-referenced clinically relevant fitness standards for maintaining physical independence in later years. *Gerontologist*. 2012;doi:10.1093/geront/gns071.
6. Thomas EE, De Vito G, Macaluso A. Speed training with body weight unloading improves walking energy cost and maximal speed in 75- to 85-year-old healthy women. *J Appl Physiol*. 2007;103:1598-1603.
7. Butler AA, Menant JC, Tiedemann AC, Lord SR. Age and gender differences in seven tests of functional mobility. *J Neuroeng Rehabil*. 2009;6:31-39.
8. Hopkins WG. A new view of statistics. *Internet Society for Sport Science*. 2000;http://sportsci.org/resource/stats.
9. Skelton DA, Greig CA, Davies JM, et al. Power and related functional ability of healthy people aged 65-89 years. *Age Ageing*. 1994;23:371-377.
10. Greig CA, Young A, Skelton DA, et al. Exercise studies with elderly volunteers. *Age Ageing*. 1994;23:185-189.
11. Schlicht J, Camaione DN, Owen SV. Effect of intense strength training on standing balance, walking speed, and sit-to-stand performance in older adults. *J Gerontol A Biol Sci Med Sci*.2001;56:M281-6.
12. Riddle DL, Finucane SD, Rothstein JM, Walker ML. Intrasession and intersession reliability of hand-held dynamometer measurements taken on brain-damaged patients. *Phys Ther*. 1989;69:182-194.

13. Fleiss JL. *Design and analysis of clinical experiments*. New York, NY: John Wiley & Sons; 1986.

14. Nunnally J, Bernstein I. *Psychometric Theory* (3rd ed). New York, NY: McGraw Hill; 1994.

15. Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. *Phys Ther*. 2006;86:735-743.