

Dewhurst, Susan and Bampouras, Theodoros (2014) Intraday reliability and sensitivity of four functional ability tests in older women. *American Journal of Physical Medicine and Rehabilitation*, 93 (8). pp. 703-707.

Downloaded from: <http://insight.cumbria.ac.uk/id/eprint/1649/>

Usage of any items from the University of Cumbria's institutional repository 'Insight' must conform to the following fair usage guidelines.

Any item and its associated metadata held in the University of Cumbria's institutional repository Insight (unless stated otherwise on the metadata record) may be copied, displayed or performed, and stored in line with the JISC fair dealing guidelines (available [here](#)) for educational and not-for-profit activities

provided that

- the authors, title and full bibliographic details of the item are cited clearly when any part of the work is referred to verbally or in the written form
 - a hyperlink/URL to the original Insight record of that item is included in any citations of the work
- the content is not changed in any way
- all files required for usage of the item are kept together with the main item file.

You may not

- sell any part of an item
- refer to any part of an item without citation
- amend any item or contextualise it in a way that will impugn the creator's reputation
- remove or alter the copyright statement on an item.

The full policy can be found [here](#).

Alternatively contact the University of Cumbria Repository Editor by emailing insight@cumbria.ac.uk.

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:

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

²University 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

26 interventions, functional assessments were developed to replicate everyday tasks^{5,6} and, hence, evaluate
27 the impact of the intervention on real-life situations.

28 Functional ability tests frequently used in older populations to assess their functional performance
29 include the 6m maximum walking speed⁶, the timed 8-foot up-and-go, the chair sit-and-reach and the
30 back scratch flexibility tests.⁵ Some of the advantage of these tests is the relevance to ‘real-life’
31 movements, low demand on resources, low equipment cost involved and the quick results they provide.⁵
32 The applicability of such measurements aside, any assessment tool must be valid and reliable to avoid
33 erroneous conclusions on the effectiveness of a particular intervention. Many of these tests have
34 previously been shown to be valid and reliable when compared to more advanced measures.⁵ Good test-
35 retest (inter-day) reliability of functional tests is important in enabling accurate evaluation of intervention
36 programmes (e.g. Capodaglio et al.¹, Carvalho et al.², Hallage et al.⁴, Thomas et al.⁶). Indeed, intraclass
37 correlation coefficient for test-retest reliability of these measures was high (i.e. > 0.90), indicating
38 stability of the test scores over time.

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

48 Although reliability of a test is useful knowledge for the test itself, it provides little information
49 and assistance to practitioners to make judgments from a single individual’s result. Every test
50 performance includes an inherent random variation as a result of biological variability in the execution of
51 movement.⁸ Quantification of this random variation, can be made with the use of typical error (TE⁸),

52 which provides a readily available score that indicates the magnitude of the random variation form
53 measurement to measurement. Such a value allows the practitioner to determine whether the inevitable
54 variability between two trials is within an acceptable range (i.e. equal to or below the TE score).
55 Differences between two trials above the TE indicate that other factors are impacting on the result (e.g.
56 lack of clarity in instruction, poor measurement technique, incorrect execution of task), and the test needs
57 to be redone.

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

63

64 **METHODS**

65 **Participants**

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

72

73 **Functional Ability Tests**

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

78

79 ***6m Maximum Walking Speed (SPEED)***

80 SPEED evaluates neuromuscular function and has been found to improve with increased
81 strength¹¹ and body weight unloading speed⁶ in older adults. To assess maximal walking speed,
82 participants started from a static standing position and walked as fast as they could to the end of a 9m
83 course⁶. Visible markers were placed at the start, 6m and 9m. Time taken from start to 6m was recorded
84 using a stop watch (Seiko, Seiko S-Yard Co Ltd, Tokyo, Japan) and speed was calculated.

85

86 ***Timed 8-Foot Up-and Go (TUG)***

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

93

94 ***Chair Sit-and-Reach (CSR)***

95 The CSR test is a widely used test of back and hamstring flexibility⁵. To assess back and
96 hamstring flexibility, while sitting on a chair with the legs stretched out in front, the participant was asked
97 to reach down towards their toes. Participants were asked to maintain their foot at 90° of dorsiflexion with
98 their toes relaxed in natural position. The distance between the extended fingers and the tip of the toes
99 was measured. Left side (CSR_left), right side (CSR_right) and the average of the two (CSR) was used
100 for further analysis.

101

102 ***Back Scratch (BS)***

103 The BS is a widely used test assessing upper body flexibility⁵. To assess shoulder range of
104 motion, the participant had one hand reaching down over the shoulder and the other one up the middle of
105 their back. The distance between the extended fingers of the two hands was measured. Data was analyzed
106 as left side (BS_left), right side (BS_right), depending on which hand was reaching down over the
107 shoulder, and the average of the two (BS).

108

109 **Data Analysis**

110 Heteroscedasticity of data was checked by examining the uniformity of the scatter when change
111 scores were plotted against the mean scores. As heteroscedasticity was absent raw scores were used for
112 further analysis. Reliability and sensitivity were calculated as suggested by Hopkins⁸. Intraclass
113 correlation coefficient (ICC, calculated as $1 - \text{typical error}^2 / \text{mean between-subject standard deviation}$
114 between trials) and typical error (TE, calculated as standard deviation of the change scores between trials
115 / square root of 2) were calculated between trials (ie trial 2 v trial 1, trial 3 v trial 2)⁸. ICC provided an
116 indication of agreement between trials¹¹ while TE an indication of the error expected from measurement
117 to measurement⁸. Descriptive data are given as mean (SD).

118

119 **RESULTS**

120 Descriptive statistics for all tests for all sessions can be found in Table 1. All tests produced high
121 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
122 between trials.

123

INSERT TABLE 1 AND 2 HERE

124

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

128

INSERT TABLE 3 HERE

129

130 **DISCUSSION**

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

136 ICC has been widely used and suggested for reliability studies¹², however, its interpretation can
137 be challenging, as various ICC thresholds have been used. For example, Fleiss¹³ suggested an ICC > 0.75
138 as 'excellent', while Nunnally and Bernstein¹⁴ stated that an ICC > 0.8 results from small measurement
139 error. The ICC scores for all the tests in the present study confirmed the high reliability of the functional
140 ability tests used as all ICC were above 0.8 suggesting high agreement between the measurements. ICC
141 for inter-day reliability of these tests was provided as part of the tests development by Rikli and Jones⁵
142 and ranged 0.90 – 0.96. The ICCs in the present study add to the high inter-day reliability of these
143 functional tests, as they indicate high intra-day reliability too.

144 Notwithstanding the importance of validity and reliability in measurement, the sensitivity of a
145 measure is an important factor^{8,15} which is less widely reported. Although there is no uniformly accepted
146 measure of sensitivity¹⁵, the use of TE is suggested as the TE is easily interpreted and can be readily used
147 to assess accuracy of the measurement.⁸ As the TE indicates the error expected from repeating a test in
148 raw units, it can be used as a threshold for its consistency. When using one of the functional ability tests
149 described above, practitioners should record two performances. If the difference between the two scores is
150 below the TE, they can be confident that this is the true score of the individual and no subsequent trial is
151 required.

152 Of interest from the measurement of flexibility was the TE values for both the CSR and BS tests
153 being different from left to right side (refer to Table 3). This, in addition to the high intra-participant
154 variability, suggests that the use of an average value of the left and right, as typically reported for both

155 tests^{2,4,5}, should be revisited. This average value may mask side differences that are important to identify.
156 For example, the BS test involves a combination of shoulder movements (abduction, adduction) as well
157 rotation (internal, external). These movements allow everyday tasks to be completed (e.g. combing hair,
158 putting on clothes)⁵ and therefore, it is of importance to know whether both sides are equally capable to
159 achieve those aims. Similarly, any loss of ankle mobility on one ankle might impact on the CSR score,
160 offering erroneous results on ‘flexibility’ of back and hamstrings. Unlike the recommendation by Rikli
161 and Jones⁵ to present the average of the left and right for the CSR and the BS tests, we suggest that each
162 arm movement is examined separately to enable examination of flexibility differences as well as
163 application of a more reflective TE.

164

165 **CONCLUSIONS**

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

175

176 **REFERENCES**

- 177 1. Capodaglio P, Capodaglio EM, Ferri A, et al. Muscle function and functional ability improves more
178 in community-dwelling older women with a mixed-strength training programme. *Age Ageing*.
179 2005;34:141-147.

- 180 2. Carvalho MJ, Marques E, Mota, J. Training and detraining effects on functional fitness after a
181 multicomponent training in older women. *Gerontology*. 2009;55:41-48.
- 182 3. de Vreede PL, Samson MM, van Meeteren NL, et al. Functional-task exercise versus resistance
183 strength exercise to improve daily function in older women: a randomized, controlled trial. *J Am*
184 *Geriatr Soc*. 2005;53: 2-10.
- 185 4. Hallage T, Krause MP, Haile L, et al. The effects of 12 weeks of step aerobics training on functional
186 fitness of elderly women. *J Strength Cond Res*. 2010;24:2261-2266.
- 187 5. Rikli RE, Jones CJ. Development and validation of criterion-referenced clinically relevant fitness
188 standards for maintaining physical independence in later years. *Gerontologist*.
189 2012;doi:10.1093/geront/gns071.
- 190 6. Thomas EE, De Vito G, Macaluso A. Speed training with body weight unloading improves walking
191 energy cost and maximal speed in 75- to 85-year-old healthy women. *J Appl Physiol*.
192 2007;103:1598-1603.
- 193 7. Butler AA, Menant JC, Tiedemann AC, Lord SR. Age and gender differences in seven tests of
194 functional mobility. *J Neuroeng Rehabil*. 2009;6:31-39.
- 195 8. Hopkins WG. A new view of statistics. *Internet Society for Sport Science*.
196 2000;http://sportsoci.org/resource/stats.
- 197 9. Skelton DA, Greig CA, Davies JM, et al. Power and related functional ability of healthy people aged
198 65-89 years. *Age Ageing*. 1994;23:371-377.
- 199 10. Greig CA, Young A, Skelton DA, et al. Exercise studies with elderly volunteers. *Age Ageing*.
200 1994;23:185-189.
- 201 11. Schlicht J, Camaione DN, Owen SV. Effect of intense strength training on standing balance, walking
202 speed, and sit-to-stand performance in older adults. *J Gerontol A Biol Sci Med Sci*. 2001;56:M281-6.
- 203 12. Riddle DL, Finucane SD, Rothstein JM, Walker ML. Intrasession and intersession reliability of
204 hand-held dynamometer measurements taken on brain-damaged patients. *Phys Ther*. 1989;69:182-
205 194.

- 206 13. Fleiss JL. *Design and analysis of clinical experiments*. New York, NY: John Wiley & Sons; 1986.
- 207 14. Nunnally J, Bernstein I. *Psychometric Theory* (3rd ed). New York, NY: McGraw Hill; 1994.
- 208 15. Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical
209 therapy. *Phys Ther*. 2006;86:735-743.
- 210